Relational Skills and Intelligence:

Developing a Functional Account of Intellectual Performance and its Enhancement

Dylan Colbert, BA (Hons.), MSc.

Thesis submitted to the Department of Psychology, Faculty of Science and Engineering, in fulfilment of the requirements for the degree of Doctor of Philosophy, National University of Ireland Maynooth.

February 2019

Research Supervisor: Dr. Bryan Roche

Head of Department: Prof. Andrew Coogan

Table of Contents

Acknowledgements

While only one name will appear on the spine of this thesis, this was the furthest thing from a one-man effort. As such, I would like to express the deepest, sincerest and most enduring gratitude to the following people, who have each contributed massively in making this possible.

Firstly, to my supervisor Bryan Roche – I am just one more of the many individuals who are eternally indebted to your many virtues as a mentor and guide. Your level of dedication, zeal and integrity is without equal, and you have been a genuine inspiration for me in my fledgling years as an academic. However, beyond all of this, I have learned just as much from you as a human being – and you truly are one of the finest I have had the pleasure of encountering.

To all in the Department, thank you for giving me the most wonderful of academic homes over the past 8 years. In particular, I would like to thank Derek, Ann, Victoria and Deirdre for looking out for me as you have since the earliest days of this journey. To Fiona Lyddy, thank you for giving me a start, and helping me sneak my way into academia. I would also like to thank all of the postgrads, past and present, who were always on hand to distract me from what I was actually supposed to be doing. Finally, I would like to thank Siobhain O'Doherty, who was a large part of my early days in the Department.

To Sarah Cassidy, it feels like decades ago that this all started – thank you for your time, knowledge and constant support. You have been an invaluable resource, and I am so lucky to have you in my corner. To my fellow Road Warrior, Steven Gannon, thank you for all of your hard work, and all of those horrible early-morning trips across the country!

To Aoife Malone, Seafra Barrett and Luke Tuohy, each of you have made a huge contribution to this thesis, and as such, I will be forever grateful for your diligence and effort.

To everyone at Drimnagh Castle CBS, in particular, Eugene Duffy, the belief you showed in me and my ideas means the absolute world. Without the opportunities you so graciously afforded me, so much of this work simply would not have been possible.

To everyone at Summerhill College, the commitment you have displayed to this research stream has been incredible. In particular, Declan Brennan, you are a pure force of nature in serving as an advocate for our work, and a real driving force behind the progress we are making.

I am extremely fortunate to be surrounded by the most incredible friends. Thank you for helping me maintain my sanity over all these years. In particular, Jack and Julian, I'm not sure you will ever understand how important the odd weekend of being a dinosaur has been.

To Cian, thanks for all the 'za, burritos and 'cold ones' watching the big game.

As I have mentioned, only one name will appear on the spine of this thesis, but realistically, virtually all of the hard yards have been run by others who selflessly worked themselves to the bone to give me this opportunity. This Doctorate was not earned by administering tests and crunching stats, it was earned by sweeping floors, laying bricks and lifting crates.

To my Grandad Colbert, thank you for teaching me never to allow anyone or anything to force me into a backward step.

To my Grandad Salmon, thank you for showing me the importance of duty and service to those you hold dear.

To my Nanny Colbert, thank you for exemplifying grit, resilience and endurance. You have been an inspiration.

To my Nanny Salmon, thank you for being the embodiment of kindness, decency and honesty. I have learned more from you than you know.

To my parents, if I even came close to expressing how fortunate and grateful I am for having you, I would probably double the length of this thesis. You have dedicated the last 30 years of your lives to giving me every single advantage that I could imagine. Things like this are made easy when I can count on the unwavering belief, constant support and teak-tough grounding you have provided for me. This thesis is not mine, it is yours. And it's almost time to call yourself M.O.D, ma.

To Holly, not one of the 100,000 words contained in this thesis was written without you in my mind. It's been a long road, and I cannot thank you enough for making it such a wonderful journey. You inspire me to be the best I can be. I love you to the moon and back.

Abstract

Recent research has implicated the potential utility of reconceptualising general intelligence as representing proficiency in a behavioural skillset known as relational responding. Indeed, a growing literature base proposes that many of the competencies that are traditionally conceived to comprise ‘intelligence’ can actually be understood from this more functional perspective. In addition, as these relational skills are inherently malleable and open to amelioration, a number of analyses have suggested that intellectual function can be improved by training and targeting these skills. In light of this emerging research stream, the current thesis entailed two primary aims: 1) to assess the efficacy of relational skills training in improving intellectual and academic performance and 2) to further investigate the relationship between the wider range of relational frames and intellectual function as a potential means of developing a functional alternative to traditional IQ assessments based on behaviour-analytic principles.

In Experiment 1, the efficacy of the SMART program in significantly improving relational responding proficiency was confirmed, using a large sample of Irish secondary school students. Experiment 2 extended upon this finding by analysing the utility of this program in improving intellectual performance using a single-blind randomised controlled trial, reporting significant gains in Full-Scale, Verbal and Performance IQ.

As pre-intervention levels of relational ability were found to be an important determinant of post-intervention outcomes, Experiment 3 endeavoured to further investigate this pattern by administering SMART to the youngest, normally-developing sample to date using a crossover design. Statistical analyses revealed the apparent delimiting impact of low levels of intellectual ability at baseline, with only a small

proportion of the sample completing the training program within a 4-month period. In light of this finding, Experiment 4 represented the first analysis of the SMART: Remedial system, a training protocol specifically designed to establish the arbitrarily-applicable relational skills deemed prerequisite for the main SMART program, as a means of allowing younger children, and those with lower levels of relational skill/intellectual performance to access the benefits the SMART program may provide. Results indicated that such skills were successfully established in a sample of children presenting with additional educational needs and below-average IQ.

Due to the recurrent finding that SMART is an efficacious means of improving intellectual function, Experiment 5 assessed the impact of this training on academic performance in a large sample of secondary school students. SMART was found to produce significant improvements on the Irish Department of Education's academic aptitude assessment of choice, the Drumcondra Reasoning Tests.

Experiments 6 & 7 aimed to elucidate the relationship between specific frames of relational responding and intellectual skills by administering two relational skills assessments alongside gold-standard metrics of intellectual performance. Such analyses identified the relational frames of coordination, opposition and comparison as being most closely associated to intellectual function. In addition, such analyses provide important insights into the role of analogical and deictic relational responding in intellectual performance.

The results of the current thesis combine to suggest that relational skills interventions may facilitate potentially life-changing improvements in both intellectual and academic performance at a level of magnitude and consistency that has not been replicated by other 'cognitive enhancement' protocols. In addition, the insights gleaned from the current set of analyses add further weight to the suggestion that intelligence may be reconsidered as a clearly-defined, functionally-understood, and malleable behavioural repertoire, rather than an invariant, trait-based, mentalistic construct.

List of Tables

List of Figures

List of Appendices

Chapter 1:

Literature Review

1.1 Introduction

Despite its far-reaching and empirically-validated contribution to a wide range of domains, including education (Cooper, 1982; Sulzer-Azaroff & Gillat, 1990), economics (Clarke, 2003; DiClemente & Hantula, 2003; Foxall, 1944, 2001, 2015, Foxall, Roma, Reed, DiGennaro Reed, & Hursh, 2016; Furreboe & Sandaker, 2017; Yousafzai, Foxall, & Pallister, 2010), healthcare (Compas, Keefe, Haaga, Leitenberg, & Williams, 1998; Glanz, Rimer, & Viswanath, 2008; Marteau, Dieppe, Foy, Kinmonth, & Schneiderman, 2006; Trask et al., 2002) and psychotherapy (Butler, Chapman, Forman, & Beck, 2006; Cuijpers et al., 2013; Van Etten & Taylor, 1998; Virués-Ortega, 2010), there is a prevalent perception that the utility of the behaviour analytic approach does not extend to the study of intelligence (Abramson, 2013; Block, 1981; Putnam, 1983; Schlinger, 2003).

Indeed, since Spearman's conceptualisation of a single general factor of intelligence (termed 'g'; 1904), the focus of intelligence research has displayed a clear shift toward a more essentialist and mentalistic approach. Critically, however, due to the nature of *g*, the ability to measure this variable has yet to be clearly established, with much of its supposed measurement thus far being indirect, through the use of Intelligence Quotient (IQ) tests (Richardson, 2002; Richardson & Norgate, 2015). As *g* is viewed as the general, overarching intellectual capacity which expresses itself through various specific mental abilities (e.g. verbal fluency, mathematical computation, memory), IQ test items are believed to vary in the extent to which they 'tap' or measure *g* (Gottfredson, 1998; Jensen, 1998; Spearman, 1904). A 'perfect' measure of *g*, however, is yet to be developed, with Raven's Matrices (Raven & Court, 2000) traditionally being conceived as the closest current approximation (Jensen, 1998; Spearman, 1946; Thorndike, Hagen, & Sattler, 1986). However, as acknowledged by Neisser et al. (1996), the inaccessibility of *g* has resulted in various interpretations of what this factor represents, ranging from mental energy (Spearman, 1927), abstract reasoning ability (Gustafsson, 1984) and processing speed (Hale & Jansen, 1994; Myerson, Hale, Zheng, Jenkins, & Widaman, 2003; Reed & Jensen, 1992; Sheppard & Vernon, 2008). More recently, *g* has been argued to be equivalent to working memory (Engle, Laughlin, Tuholski, & Conway, 1999; Kyllonen & Christal, 1990; Unsworth & Engle, 2005).

In light of the general lack of consensus on its definition, intelligence therefore has typically been operationalised in terms of what IQ tests measure (Boring, 1923; Richardson & Norgate, 2015; van der Maas, Kan, & Borsboom, 2014), insofar as IQ tests are said to merely "define the theory of intelligence that the test is intended to measure" (Naglieri, 2008, p. 68). In essence, through assessing an individual's performance on a range of mental tasks, IQ tests attempt to reduce a wide-ranging spectrum of intellectual behaviours into a unitary, quantitative factor (Cassidy, Roche, & O'Hora, 2010). IQ test scores are argued to reflect a stable, invariant and non-malleable trait (Jensen, 1980; Juliano, Haddad, & Carroll, 1988; Locurto, 1991; Ramsden et al., 2011; Reynolds, Gutkin, Dappen, & Wright, 1979; Spearman, 1927). However, empirical research has increasingly suggested that IQ test scores may not be as immutable as once assumed, with the Flynn effect identifying substantial rises in IQ test performance throughout the 20th and 21st century (Flynn, 1984, 1998, 2007). Interestingly, there is emerging evidence to propose that the Flynn effect may have stalled or even reversed in recent times (Dutton & Lynn, 2014, 2015; Dutton, van der Linden, & Lynn, 2016; Pietschnig & Voracek, 2015; Shayer & Ginsburg, 2009; Shayer, Ginsburg, & Coe, 2007; Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 2005; Woodley & Meisenberg, 2013). Nevertheless, the very fact that population IQ scores undergo such non-linear fluctuations across time has been attributed to a wide variety of environmental (Ceci, 1991; Dickens & Flynn, 2001; Flynn, 2007; Lynn, 1990, 2009), social (Blair, Gamson, Thorne, & Baker, 2005; Brand, 1987; Ceci, 1991) and genetic factors (Jensen, 1998; Mingroni, 2004, 2007), and suggests that the stability espoused by trait-based theories of intelligence may have been exaggerated. Indeed, there is now accumulating evidence within the literature which argues that intellectual ability is in fact, a pliable concept which is influenced by the environment (see Dickens & Flynn, 2001; Schlinger, 2003; Sternberg, 2008). In fact, emerging evidence from cognitive (Basak, Boot, Voss, & Kramer, 2008; Buschkuehl et al., 2008; Harrison et al., 2013; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011), educational (Brinch & Galloway, 2012; Campbell, Ramey, Pungello, Sparling, & Miller-Johnson, 2002; Ceci, 1991; Jencks, 1972) and neuroscientific investigations (Mackey, Miller Singley, & Bunge, 2013; Mackey, Hill, Stone, & Bunge, 2011; Mackey, Miller Singley, Wendelken, & Bunge, 2015) propose that intelligence may be improved through environmental interventions.

Some of the most noteworthy research thus far carried out on the malleability of intelligence has emerged from a behaviour-analytic paradigm with a number of investigations (e.g. Cassidy, Roche, Colbert, Stewart, & Grey, 2016; Cassidy, Roche, & Hayes, 2011; Hayes & Stewart, 2016) asserting that intellectual performance can be improved by systematically training a key behavioural repertoire known as relational responding or relational skill. The theoretical basis for such training protocols is based on a behaviour-analytic account of language and cognition known as Relational Frame Theory (Hayes, Barnes-Holmes, & Roche, 2001) which proposes that much of the skills/competencies/abilities that we conceive as constituting ‘intelligence’ can be understood as relational responding, defined as the act of responding to one stimulus in accordance to its contextually-controlled relation to another stimulus. Indeed, a small number of correlational studies have been published and appear to provide an empirical basis to this proposed relationship between intellectual performance and relational responding (Colbert, Dobutowitsch, Roche, & Brophy, 2017; O'Hora, Pelaez, & Barnes-Holmes, 2005; O'Hora et al., 2008; O'Toole & Barnes-Holmes, 2009). Given that relational responding is conceptualised as a generalised operant (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000) and is therefore inherently accessible to manipulation and improvement, a number of interventions have been developed in an attempt to improve the fluency of these relational skills as a means of improving intellectual function. In particular, the SMART program has thus far displayed considerable, but tentative, efficacy in this regard (see Amd & Roche, 2018; Cassidy et al., 2016; Cassidy et al., 2011; Hayes & Stewart, 2016). The current thesis aims to extend upon these findings by further investigating the efficacy of relational skills training programs as a means of improving intellectual performance through a series of investigations across age and ability ranges. In addition, as a small number of studies have indicated a close relationship between intellectual ability and relational responding, a second aim of the current research is to further elucidate the nature of this relationship, with a view of potentially developing a measure of relational responding that may serve as a functional alternative to traditional IQ assessments.

1.2 Early Approaches to Intelligence

Perhaps the most prominent conceptualisation of intelligence is rooted in Spearman's general factor theory (1904), which posits that there is a latent factor or faculty ("g") that influences performance on all intelligence measures, and which stays stable throughout a person's lifetime. In a precursor to modern factor analysis, Spearman's theory was based on the positive intercorrelations (the 'positive manifold') found between an individual's level of performance on tasks which assess a range of intellectual skills (i.e. that those who display above-average performance on one task, tend to display above-average performance on other tasks). To explain this relationship, Spearman (1904; 1927) argued for the existence of a single factor representing "mental energy" labelled as 'g', which comprised three subcomponents: the apprehension of experience, the eduction of relations and the eduction of correlates. Such was the influence of Spearman's single factor theory, Herrnstein & Murray (1994) proposed that this reorientation of intellectual performance "shaped both the development and much of the methodological controversy about mental tests ever since" (p.2).

At this point, it is illuminating to acknowledge the juxtaposition of the views of arguably the two pre-eminent intelligence researchers at the turn of the last century, the aforementioned Charles Spearman and Alfred Binet, who is credited, alongside his frequent collaborator Théodore Simon, as the developer of the first practical measurement of intellectual performance, publishing it the year after Spearman's seminal work (Binet & Simon, 1905). Indeed, it was Binet's early attempt to quantify individual differences in intellectual performance which served as the foundations upon which today's IQ tests are built, none more so than the widely-administered, gold-standard Stanford-Binet assessment which is currently in its 5th edition (Roid, 2003). Following the passing of a law that designated that primary education was compulsory for all children between the ages of 6 and 13 in France (see Prost, 1989), the issue of how to educate those at the lower end of the intellectual spectrum became a pressing social issue (Nicolas, Andrieu, Croizet, Sanitioso, & Burman, 2013). Therefore, as head of the *Société Libre pour l'Etude Psychologique de l'Enfant* [SLEPE; Free Society for the Psychological Study of the Child], Binet was mandated to construct an assessment of intellectual ability that would facilitate the identification of children who required additional educational support (Binet & Simon, 1905d; 1916).

Unlike some of his contemporaries, most notably influential French neurologist Desire-Magloire Bourneville (see Bourneville, 1895, 1898, 1899; Gateaux-Mennecier, 1989, 2002), who argued for the exclusion of children with additional educational needs from mainstream schooling, Binet deemed all levels of intellectual deficit as treatable, and maintained that it was not a question of "if" such children may be treated, but "how"(Binet, 1905a):

> A few modern philosophers seem to lend their moral support to these deplorable verdicts when they assert that an individual's intelligence is a fixed quantity which cannot be increased. We must protest and react against this brutal pessimism…With practice, training, and above all method, we manage to increase our attention, our judgement, and literally to become more intelligent than we were before. (p. 301)

As outlined by Schlinger (2003), Binet had avoided the use and definition of the nounified 'intelligence' in his work, as this may have directed him towards identifying the referent of this term. Instead, Binet utilised an adjectival form, 'intellectual skill', as this firmly tied his studies in functional, behavioural accounts of this repertoire, rather than trait-based, essentialist conceptualisations which, theoretically-speaking, would appear to preclude the possibility of intervention. Through the use of their new testing battery, Binet & Simon were able to identify which students required additional support and provide what Binet termed 'mental orthopaedics' to address areas of difficulty (Siemsen et al., 2017). Following the widespread administration of these mental orthopaedic interventions, Binet (1909) proposed that these training programs had resulted in not only increased knowledge, but increased intelligence on the part of his students, many of whom now performed at an intellectual level that allowed them to join mainstream schooling.

Following the construction of the early Binet-Simon scales, however, two important developments fostered a reconceptualisation of not only what these tests of intellectual performance measure, but what intelligence itself constitutes. For one, as a result of the apparent utility of the Binet-Simon scales, this testing battery was adopted by Louis Terman, alongside his colleagues at Stanford University, for use in the United

States (Terman, 1916). Most pertinent to the current discussion is Terman's most seminal contribution to the field of intelligence: the calculation of an Intelligence Quotient (IQ) score. Whereas Binet and Simon used their testing battery to identify an individual's 'mental age' based on what type of tasks they were capable of completing, Terman, influenced by the work of William Stern, was the first to use an IQ score, based on the ratio between an individual's chronological age and mental age, multiplied by 100 to remove fractions (i.e., $IQ = \left(\frac{Mental\ Age}{Chronological\ age}\right) x\ 100$). While Binet did not live long enough to see this application of his testing battery, Simon strongly opposed the use of this statistic viewing it as a misuse of their instrument and a oversimplified and reductionist description of intellectual performance, terming this development as a 'betrayal' (Wolf, 1973, p. 203).

Several researchers have proposed that it was this development, the computation of an IQ score, alongside Spearman's proposed general factor theory of intelligence that led to the reification of intelligence (e.g., Fancher, 1985; Gould, 1981; Schlinger, 2003). The logical error of reification refers to the process by which the presence of an abstraction, such as a verbal label (in this case *g* or general intelligence) is erroneously used as a basis to infer the physical existence of its referent (Bell, Staines, & Michell, 2001). Indeed, the positive manifold analysed by Spearman does not offer any insight in itself into the nature of intelligence *per se*, but rather, at best, serves as support for the assertion that the measures of 'intelligence' studied by Spearman may measure the same capacity or proficiency. This proposal, however, does not in itself extend to an explanation of what that capacity represents (Gottfredson, 1998). Indeed, while Spearman proposed that this general factor reflected general intelligence, a number of other possible explanations have been offered for this shared variance, perhaps most notably processing speed (Hale & Jansen, 1994; Jensen, 1998; Myerson et al., 2003; Sheppard & Vernon, 2008; Vernon & Jensen, 1984; Zheng, Myerson, & Hale, 2000) and working memory (Engle et al., 1999; Kyllonen & Christal, 1990; Unsworth & Engle, 2005).

Given that the pursuit of an objective measurement of intelligence was still in its formative stages at the time of Spearman's 1904 publication, it would appear premature to propose that the relatively rudimentary intelligence assessments at Spearman's disposal could facilitate such a grandiose interpretation of the positive intercorrelations found between scores derived from these assessments. Indeed, the 'intellectual' measures utilised in his 1904 experiments were entirely inappropriate, consisting of teacher ratings of 'cleverness' and student ratings of peers' 'common sense' provided by the two oldest members of the sample for Spearman's first experiment, and school grades (Classics, English, French & Mathematics) for his second.

Viewed from a different perspective, and moving beyond Spearman's data, the recurrent finding that individuals tend to perform to a similar level on a range of measures constructed to assess a specific capacity may represent a general theoretical convergence on what tests of that capacity should constitute, rather than providing an empirical basis for unitary nature of that same capacity. As such, according to the linguistic conventions advocated by Maccorquodale & Meehl (1948), the general factor theorised by Spearman satisfies the criteria for classification as a hypothetical construct, as it refers to an unobservable process or entity that cannot be completely reduced to empirical terms.

As Schlinger, (2003) points out, committing the logical error of reification almost inevitably entails a further logical error: circular reasoning. The endurance of this error can be summed up by Boring's (1923) famous statement that "intelligence is what intelligence tests measure". Circular reasoning is the logical error of proposing that the construct used to explain a given effect or relationship (the *explanans*) is equivalent to the effect/relationship that it is proposed to explain (the *explanandum*;

Boag, 2011). Indeed, in the case of the general factor of intelligence, the positive manifold was explained as this general factor, while at the same time, the general factor was explained by the positive manifold. In this way, Spearman merely described (rather than explained) this finding by applying a verbal label to it based on an inappropriate testing battery that was insufficient in supporting the application of such a label (Howe, 1990). At worst, this has been termed "linguistic sleight-of-hand" (Matthews, Zeidner, & Roberts, 2002, p. 88), a term which perhaps unfairly infers intentional deception on Spearman's part. At best, while referring to this common factor offers clear pragmatic utility, this reorientation has been viewed as entirely circular, and thus offered no further explanatory power in expounding upon the nature of intelligence.

This issue of circularity is widespread in psychology (see Hahn, 2011), and while it may seem innocuous, the proliferation of psychological constructs can eventually result in a loss of awareness of the mere descriptive function a label may serve, and a reorientation of the label such that it may be considered to explain the phenomenon it refers to. While this may appear to be an issue of semantics, its implications are far-reaching. As expressed by Skinner, (1953), the practice of applying descriptive labels to constructs is "dangerous because it suggests that we have found the cause and therefore look no further" (p. 31). The reification of such labels into constructs can directly influence the course of a given research stream and subsequent theory creation (Maccorquodale & Meehl, 1948). It, therefore, comes as no surprise that Spearman's general factor theory exerted a profound influence on subsequent investigations into intelligence to an extent such that it may have disparaged attempts to provide more functional accounts by propagating an essentialist conceptualisation of this 'capacity'.

Spearman's general factor theory essentially reduced intellectual performance from the functional account advocated by Binet & Simon, into an essentialist, single-factor theory reflecting a hypothetical construct rather than a repertoire of observable skills. However, as noted by (Howe, 1990), in a criticism of the general factor theory of intelligence:

> The absence of logical grounds for assuming that intelligence must have a conceptual status other than that of a descriptive or labelling construct does not justify our ruling out the possibility that there might still exist a quality of intelligence which can help to account for people's abilities. (p. 491)

In light of this point, the purpose of the current thesis is not to suggest that intercorrelations between performance on various intellectual tasks do not exist, nor is it contested that there may be underlying skills that may exert a central, foundational influence in contributing to proficiency across the spectrum of intellectual tasks. The purpose of the current discussion is to probe as to whether this development (i.e. the reconceptualisation of intelligence as a statistical abstraction reflecting a single factor) represented a progression in the accuracy and validity of intelligence theory and in addition, whether the zeitgeist of a single factor theory offered pragmatic utility regarding the manner in which we conceptualise intellectual performance. As expressed by Skinner (1938) who, despite marked theoretical reservations regarding the use of hypothetical constructs, acknowledged the practical utility of usage of the construct of intelligence:

> The existence of a popular term does create some presumption in favor of the existence of a corresponding experimentally real concept, but this does not free us from the necessity of defining the class and demonstrating the reality if the term is to be used for scientific purposes. (p. 42)

Therefore, the current thesis aims to highlight certain inconsistencies, misconceptions and/or errors, from both a historical and empirical perspective that may challenge the acceptance of essentialist, trait-based and/or unitary conceptualisations of intelligence and highlight the potential utility and appropriateness of functional accounts of intellectual performance. Furthermore, it is hoped that such a discussion may highlight the potential contribution of behaviour analysis regarding investigations into the nature and measurement of intelligence.

1.3 Behaviour Analysis and Intelligence

While behaviour analysis has displayed a degree of ambivalence toward intelligence as a research item, one of the most vocal proponents of the utility of a behaviour-analytic approach to this field, Arthur Staats, has called for the need to 'behaviourize' psychology (Staats, 1996) due to his assertion that intelligence research (and psychological research in general) lacks analyses of phenomena in terms of behaviour and therefore lacks explanatory power. Furthermore, without such behavioural accounts, Staats posits that it is difficult to explain how such phenomena are established and may be subsequently modified. Staats (1968) emphasised the central role of learning in intelligent behaviour, defining intelligence as "a wide sample of (learned) basic behavioural skills the child (or adult) has acquired which are important to the acquisition of further skilled behaviours" (p. 389).

As such, an individual's level of intellectual performance at any given time was viewed as their current position in a cumulative-hierarchical learning process based on their experiences up to that point. In this way, Staats proposes that complex cognition and intellectual skills can be viewed as basic behavioural repertoires (BRRs), which reflect "actual stimulus-response constellations that have to be stipulated" (Staats, 1996, p.193), rather than underlying mentalistic faculties. Staats (1996) argues that the topography of IQ test items, and their perceived dissimilarity to practical real-world skills and academic tasks, may contribute to such assessments being perceived inaccurately as metrics of a 'deeper' mental potential or ability. However, the nature of these hidden faculties is very much demystified once these test items are analysed from a functional perspective (see Section 1.4.3).

Staats (1996) outlines a number of such BRRs that contribute to intellectual performance, such as Verbal-Motor (the ability to comprehend language and use verbal information to regulate motor activity), Verbal-Image (the elicitation of sensory responses as a result of verbal information), Verbal-Emotional (the elicitation of emotional responses as a result of verbal information), Verbal-Labelling (ability to respond verbally to external stimuli) and Verbal-Expression (written expression of language). Staats proposed that in many cases, traditional IQ assessments specifically assess these repertoires. For example, tests of vocabulary and block design, hallmarks of multiple gold-standard IQ assessments, can be considered measures of the Verbal-Labelling and Verbal-Motor repertoire, respectively. However, Staats argues that most intelligence tests provide a metric of the generalised effect of sophisticated BRRs by 'sampling' the extent of a child's learning experiences, the repertoires this learning has produced, and the level of learning the child can demonstrate in novel situations. By doing this, it becomes clear that the utility of IQ tests is not in assessing competency in completing its entailed test items specifically, but rather in assessing BRRs that facilitate learning and thus influence performance on such test items (Staats, 1971; Staats & Burns, 1981). Staats & Burns (1981) empirically investigated this generalised effect and demonstrated that training in specific BRRs, Verbal-Labelling and Verbal-Expression resulted in improved performance on two WPPSI subtests developed to assess different domains of intellectual performance: Mazes and Geometric Designs. In this way, Staats posited that it was an 'underlying' proficiency in these malleable, observable BRRs that explained the general factor of intelligence, rather than hidden, essentialistic capacities of the individual.

While Staats invokes BRRs as a means of defining intelligence without appeal to a hypothetical construct, it has been levelled that despite this, Staats has failed to avoid another of the logical errors that behaviourists traditionally accuse mentalistic accounts of committing: circular reasoning. To borrow the example outlined by Holth (2003), if we were to explain a child's failure to complete a specific IQ test item as being due to the absence of a given BRR, does the observational basis for this inference differ in any substantial way to that of proposing that the child simply was unable to complete the tasks, without suggesting the absence of a BRR as being the cause? And if so, does the invocation of BRRs as an explanation provide any additional explanatory power? As such, even when defining phenomena in operational terms as Staats attempted, the issue of circularity is difficult to avoid. As expounded by Holth (2003), due to this circularity (and the entailed failure to identify the dependent and independent variables in this relationship), it is impossible to ascertain the nature of the relationship between BRRs and intelligence in terms of cause and effect (i.e. does intelligence lead to the establishment of BRRs, or vice versa?).

The issue of discriminating the relationship between behaviours deemed to constitute or contribute to intelligence (such as Staats' BRRs) and intellectual performance itself (if these two phenomena can indeed be disentangled), highlights the centrality of the behaviour-analytic aim for both prediction and control in theorising upon such issues. Indeed, explanation, in behaviour-analytic terms, refers to "prediction and control (of a given phenomenon) with adequate scope of precision" (Hayes & Brownstein, 2016, p. 179).

As explained by Hayes and Brownstein (2016), the existence of behaviour-behaviour relations, as specified by various accounts of intelligence based purely on correlational analyses (e.g. Ackerman, Beier, & Boyle, 2005; Colom, Abad, Quiroga, Shih, & Flores-Mendoza, 2008), clearly do not provide the causal accounts required for prediction and control. Indeed, as discussed by Hayes and Brownstein, while this is readily acknowledged when the related behaviours operate in similar domains (e.g. poker playing and Monopoly playing), this point may be overlooked when dealing with topographically- and functionally-distinct behaviours. For example, Hayes and Brownstein propose that while a positive correlation between competency in playing Monopoly and playing poker would rarely be mistaken to infer a cause-effect relationship, a correlation between aggression (i.e. aggressive behaviour) and poker skill may be (i.e. we may infer that poker skill may be explained at least partially by lower levels of aggression). As such, in providing a behavioural account on intelligence, it is important to identify the independent and dependent variables in such behaviour-behaviour relations in order to avoid simply repackaging the fallacies of circular reasoning and category error so often derided of mentalistic accounts and hypothetical constructs.

In order to do this, it is essential to investigate such behaviour-behaviour relations by confirming that variance in the proposed dependent variable can be attributed to variance in the proposed independent variable, and not extraneous influences (Hersen & Barow, 1976; Johnston & Pennypacker, 1980; Sidman, 1960). Importantly, this analysis necessitates reliable assessment and operational definition of each variable in order to ensure the validity of any inferences drawn (Hobbs, Moguin, Tyroler, & Lahey, 1980; Peterson, Homer, & Wonderlich, 1982). Regarding the relationship between the construct of intelligence and IQ test performance, while the dependent variable (IQ test performance) satisfies the above criteria, this may not be the case for the independent variable in this relationship, i.e. ‘intelligence’. Therefore, in order to provide a valid behavioural account of intelligence, it is necessary to provide an operationally-defined behavioural repertoire which can be accurately measured (dependent variable), and that can be empirically-validated as the cause or determinant of performance on tasks traditionally and commonly perceived to assess intellectual skill or ability (independent variable). One such account will be subsequently discussed in Section 1.4.

1.3.1 Applied Behaviour Analysis and Intellectual Performance

Despite behaviour analysis' relative disinterest in intelligence as a major topic of investigation, an extensive research literature proposes that applied behaviour-analytic (ABA) interventions may facilitate the amelioration of a wide range of behaviours, skills and abilities, including those considered to constitute intellectual performance. Much of the more recent work in ABA and Early Intensive Behavioural Intervention (EIBI) has stemmed from Lovaas' (1987) pioneering study which found that systematic one-on-one treatment based on operant conditioning principles cannot only catalyse improvements in adaptive and verbal behaviour, but also produce significant increases in intellectual ability. While it must be noted that attempts to replicate the results of Lovaas' seminal investigation have had mixed success (Eikeseth, Smith, Jahr, & Eldevik, 2002b; Jacobson, Mulick, & Green, 1998; Sallows & Graupner, 2005; Sheinkopf & Siegel, 1998; Smith, Groen, & Wynn, 2000), a number of meta-analyses concerning ABA and EIBI interventions have proposed that such protocols appear to be efficacious in improving performance on various psychometric measures of intelligence for individuals diagnosed with autism spectrum disorders (Eikeseth et al., 2002). In an analysis of 11 ABA interventions aimed at improving intellectual performance, Peters-Scheffer, Didden, Korzilius, & Sturmey (2011) reports mean score increases of 12 and 11.1 standardised points on measures of Full-Scale and non-verbal IQ respectively, with a Cohen's *d* statistic of 2 indicating a very large effect size. Howlin, Goode, Hutton, & Rutter (2004) similarly report IQ increases of between 8 and 31 points across 11 EIBI interventions. In a more recent analysis of 22 ABA studies assessing the impact of EIBI interventions on IQ, Makrygianni, Gena, Katoudi, & Galanis (2018) reported mean score increases of 14.3 and 10.9 standardised points on verbal and non-verbal IQ tests.

In relation to the current thesis, there are two major implications of this well-validated effect of ABA in improving performance on psychometric measures of intelligence. For one, if such interventions can bring about improvement in intellectual performance by targeting and shaping demonstrable and malleable behaviours, it follows that it may, therefore, be possible to reconceptualise intelligence by proposing a functional account based on observable skills, rather than underlying, unobservable, mentalistic traits. In addition, these post-intervention gains would appear to complement reports emanating from various other disciplines, such as cognitive psychology (e.g. Basak et al., 2008; Jaeggi et al., 2008, 2011; Stephenson & Halpern, 2013), educational psychology (Brinch & Galloway, 2012; Campbell et al., 2002; Ceci, 1991; Jencks, 1972) and neuroscience (Mackey et al., 2011, 2013, 2015), in challenging the notion that intellectual ability represents an invariant, inaccessible capacity that is not amenable to modification or improvement. Such themes will be further explored in subsequent sections.

1.4 Relational Frame Theory

As discussed previously, from a behaviour-analytic perspective (De Houwer, Barnes-Holmes, & Moors, 2013), trait definitions of intelligence are considered to fall victim to the errors of reification and circular reasoning (Gottfredson, 1998; Howe, 1990; Schlinger, 2003), and therefore are wholly incongruent with the behaviourist tradition (Schlinger, 2003; Skinner, 1974). Thus, behaviour analysts embrace a more functional account of "intelligence", in which the term merely refers to a measurable quality of a set of actions, which are intricately linked to their context and are thus amenable to experimental manipulation (Cassidy et al., 2011; Hayes & Stewart, 2016; Schlinger, 2003). In effect, intellectual abilities are viewed as malleable, with IQ tests functioning solely to provide an index of the fluency of the skills involved. According to this view, the stability of IQ scores across time does not imply the existence of an underlying trait, but merely reflects stability in the learning environment and the unfolding of intellectual development at a typical rate.

Recent developments within the field of behavior analysis, most notably under the rubric of Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001), have begun to progressively explore the utility of a more functional approach of conceiving intellectual behaviour (Dymond & Roche, 2013; O'Hora et al., 2005; O'Toole, Barnes-Holmes, Murphy, O'Connor, & Barnes-Holmes, 2009). In particular, RFT highlights the *derived* and *generative* nature of human language and cognition (O'Toole et al., 2009) and proposes that much of what we consider intellectual ability can be reconsidered as a form of behaviour known as relational responding (Cassidy et al., 2010; Hayes et al., 2001). Relational responding is defined as the act of responding to one stimulus in accordance to its relation to another stimulus and comes in two general forms, nonarbitrarily- and arbitrarily-applicable, based on the stimuli being responded to, and the nature of relations between these stimuli.

Non-arbitrary relational responding is a more basic form of responding which is governed by the physical properties of stimuli (e.g., size, shape, colour). For example, when asked to select the "longest" stick amongst an array, an individual is engaging in non-arbitrary relational responding, as their response is governed by the physical form of the stimuli being related (i.e. sticks of varying length) and their respective relation to each other (i.e. response requires identifying which stick is longest). Indeed, a wide range of species have exhibited this form of relational responding, notably various species of primate, including chimpanzees (Beran & Washburn, 2002; Dugdale & Lowe, 2000; Finch, 1942; Flemming & Kennedy, 2011; Hashiya & Kojima, 2001; Haun & Call, 2009; Hopkins & Washburn, 2002; Hribar, Haun, & Call, 2011; Nissen, Riesen,

& Nowlis, 1938; Parr, Winslow, Hopkins, & De Waal, 2000), gorilla (Haun & Call, 2009; Vonk, 2003), orangutan (Haun & Call, 2009; Hribar et al., 2011; Vonk, 2003), baboons (Zurcher et al., 2010), capuchin monkeys (D'Amato & Colombo, 1985; D'Amato & O'Neill, 1971; D'Amato & Worsham, 1972, 1974; Etkin & D'Amato, 1969; Truppa et al., 2010; Truppa, Mortari, Garofoli, Privitera, & Visalberghi, 2011; Worsham & D'Amato, 1973) and rhesus macaques (Davachi & Goldman-Rakic, 2001; Harmon, Strong, & Pasnak, 1982; Hopkins & Washburn, 2002; Mello, 1971; Parr et al., 2000; Sliwa, Duhamel, Pascalis, & Wirth, 2011; Zimmerberg, Glick, & Jarvik, 1971). Non-primate mammals and bird species have also demonstrated such a capacity in experimental settings, including, amongst others, dolphins (Herman & Gordon, 1974; Herman, Hovancik, Gory, & Bradshaw, 1989; Herman & Thompson, 1982; Kilian, Von Fersen, & Güntürkün, 2005; Roitblat, Penner, & Nachtigall, 1990), rats (Dunnett, Martel, & Iversen, 1990; Hampson, Jarrard, & Deadwyler, 1999; Porritt & Poling, 2008; Stanhope, McLenachan, & Dourish, 1995), dogs (Kuśmierek & Kowalska, 2002), crows (Goto & Watanabe, 2009, 2012; Koehler, 1950; Moll & Nieder, 2015; Smirnova, Lazareva, & Zorina, 2000, 2003; Smirnova, Zorina, Obozova, & Wasserman, 2015; Zorina & Smirnova, 1996) and pigeons (Bodily, Katz, & Wright, 2008; Cumming & Berryman, 1961; Grant, 1976; Lind, Enquist, & Ghirlanda, 2015; Skov-Rackette, Miller, & Shettleworth, 2006; Wright, Cook, Rivera, Sands, & Delius, 1988; Wright & Delius, 1994, 2005).

Due to the relative complexity of our socio-verbal environment however, humans show an apparently unique capacity for arbitrarily-applicable relational responding (AARR). AARR is a specific form of responding whereby the responding is not governed by the formal physical properties of the stimuli involved, but rather on contextual cues, known as relational frames, which specify a particular relationship between these stimuli (Steele & Hayes, 1991). As the centrality of the physical relatum (i.e. topography) is diminished for this form of relational responding, AARR can be applied to a wide range of concepts and contexts and is thus considered a generalised, functionally-defined operant (Barnes-Holmes, Barnes-Holmes, & Cullinan, 2000).

A number of relational frames have been identified, such as coordination (e.g., A is the same as B; Hayes et al., 2001), distinction (e.g., A is different to B; Roche & Barnes, 1997), opposition (e.g., A is opposite to B; Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004), comparison (e.g., A is greater than/less than B; Dymond & Barnes, 1995; Roche, Barnes-Holmes, Barnes-Holmes, Stewart, & O'Hora, 2002), temporality (e.g., A is before/after B; O'Hora, Roche, Barnes-Holmes, & Smeets, 2002), analogy (e.g., A is to B as C is to D; Stewart, Barnes-Holmes, Roche, & Smeets, 2002), hierarchy (e.g., A subsumes/belongs to B; Griffee & Dougher, 2002) and deixis (e.g., A is here and B is there; McHugh, Barnes-Holmes, & Barnes-Holmes, 2004). Furthermore, a number of subcomponent frames have been identified, including spatial relations (a subset of comparison relations, e.g. A is left of B; May, Stewart, Baez, Freegard, & Dymond, 2017), containment relations (a subset of hierarchical relations e.g., A is inside B; Slattery, Stewart, & O'Hora, 2011), and metaphor (a subset analogical relations, e.g. A to B is like C to D; Stewart & Barnes-Holmes, 2001).

Derived relational responding occurs once the establishment of relational network based on previously established relations is sufficient to allow the inference of novel, untrained relations within that network. Historically, the traditional behavioural perspective focuses primarily on directly established contingencies between stimuli, but derived relational responding, on the other hand, is not dependent on direct contingency respondent, operant or generalisation process (Blackledge, 2003; Hayes et al., 2001). For example, having learned hundreds of examples that some relatum A is *greater than* some relatum B, a verbally-able human is capable of responding appropriately when told that an abstract relatum X (such as an abstract character) is greater than another abstract relatum Y (e.g., another abstract character of the same size). Furthermore, if an individual is explicitly taught that relatum A is greater than relatum B, and relatum B is the same as relatum C, it is possible to derive the relation between relata A and C (i.e. A is more than C) based on their respective relation to relatum B, even though this relation has never been trained directly. In this way, if an individual is trained that Brian is older than Theresa and that Theresa is older than Cian, he/she can thus derive additional, unspecified relations between these three relata based on the aforementioned relational premises. In this example, the individual can infer than Brian is therefore older than Cian (and that Cian is younger than Brian), based on each relatum's specified relationship to Theresa.

The derivation of novel relations is facilitated by three key features of derived relational responding, known as mutual entailment, combinatorial entailment and transfer of function. Mutual entailment refers to the reversal of a specified relation between any two relata. In the case of symmetrical bidirectional relational frames (e.g. coordination, opposition, distinction), the mutually-entailed relation between the second and first relatum is equivalent to the original relation. For example, if relatum A is the same as relatum B, the mutually-entailed relationship is identical (i.e. relatum B is the same as relatum A). However, in the case of an asymmetrical unidirectional relationship (e.g. comparison, hierarchy), an inverse relationship is mutually-entailed between relata. For example, if relatum A is bigger than relatum B, the mutually-entailed relationship is that relatum B is smaller than relatum A.

Combinatorial entailment refers to the derivation of an unspecified relationship between two relata in accordance with each relatum's relationship with a mutual relatum. For example, if relatum A is opposite to relatum B and relatum C is the same as relatum B, the relationship between relata A and C is combinatorially entailed via their respective relations to relatum B (i.e. relatum A is opposite to relatum C).

Transfer of function is the process by which the behavioural function of a given stimulus is informed by its relation to another stimulus. For example, if a child shows fear towards a particular domestic cat, this fear may then extend to a neighbour's pet once the child learns that this animal is of the same species to the feared animal (i.e. also a cat). Such transfer has been demonstrated experimentally in studies such as Boyle, Roche, Dymond, & Hermans (2016) who report that after conditioning a fear response to the word 'broth', the conditioned response was then generalised to semantically synonymous terms (e.g. 'soup') but not to semantically distinct terms (e.g. 'help'). In this example, the behavioural response to the word 'soup' is governed not by any property of the term itself, but rather, by its conceptual relation to the word 'broth' and its entailed learning history. Furthermore, the child's level of fear to other stimuli may be attenuated by the nature of the relation between the feared stimulus and novel stimuli. If the child learns that a lion is a larger type of cat, the child may show a greater level of fear to lions compared to domestic cats. Conversely, if the child learns that a kitten is a smaller type of cat, the fear may be diminished somewhat.

RFT proposes that sophistication in derived relational responding in accordance to the small collection of relational frames may adequately account for much of what has been considered to constitute intellectual performance (e.g. vocabulary, numeracy, inductive and deductive reasoning, analogy; Barnes-Holmes et al., 2001). Indeed, generalised operant classes have served as the basis for a wide range of behaviour-analytic accounts of complex cognition, such as thinking (Hayes et al., 2001; Maltzman, 1955, 1962), attention (McIlvane, Dube, Kledaras, Iennaco, & Stoddard, 1990; McIlvane, Dube, & Callahan, 1996), problem-solving (Hayes et al., 2001; Skinner, 1984), insight (Epstein, 1987; Epstein, Kirshnit, Lanza, & Rubin, 1984; Epstein, 1990), creativity (Campbell & Willis, 1978; Reese & Parnes, 1970; Hayne Reese, Parnes, Treffinger, & Kaltsounis, 1976), perspective-taking (Barnes-Holmes et al., 2001;

McHugh et al., 2004), deception (McHugh, Barnes-Holmes, Barnes-Holmes, Stewart, & Dymond, 2007) and decision-making (Fantino, 1998, 2004; Fantino & Stolarz-Fantino, 2005). In the context of intelligence, RFT employs such operant classes in a similar fashion in accounting for a number of important intellectual skills.

1.4.1 Relational Responding and Verbal Intelligence

Perhaps the most readily available example of the relevance of relational responding to intelligence is found in the domain of verbal intelligence. Verbal intelligence is generally defined as the ability to comprehend and analyse verbal information, and to use verbal reasoning to solve problems (Wechsler, 1997). Most definitions of verbal intelligence, therefore, implicate vocabulary, verbal knowledge, verbal reasoning (including analogical reasoning) and numeracy as key components of this capacity; all of which can be understood from an RFT perspective. Vocabulary is traditionally viewed as one of the primary predictors of general intelligence (Marchman & Fernald, 2008; Smith, Smith, Taylor, & Hobby, 2005; Vetterli & Furedy, 1997; Wechsler, 1949, 1955, 1974, 1991, 2011), and can be viewed in RFT terms as simply a network of word-word and word-object relations (Cassidy et al., 2010). The extent and complexity of this network are explained by an individual's learning history in regard to explicitly-taught verbal relations, and the novel, untrained derived relations that are thus facilitated. For example, from an early age, word knowledge is acquired through simple interaction with adults, which, in its most basic form involves the pairing of a physical object to a word (e.g. pointing to a dog and telling the child that "this is a dog"). Thus, a simple word-object coordination relation is established, whereby the child learns to respond verbally with the word "dog" when presented with the associated physical stimulus. This coordination relation is then further reinforced by reversing the direction of this relation (i.e. rather saying "this is a dog", the teacher points to a dog and asks, "what is that?"). This relational network can be further extended if the child subsequently encounters a novel word (e.g. “pooch” or “mutt”) and is told that this word means the same as “dog”, thus establishing a word-word coordination relation to be integrated into the child’s existing relational network. This then allows the derivation of the untrained word-object relation between “pooch” and the physical stimulus. Similarly, further word-object coordination relations can be derived when the child encounters a different breed of dog for the first time and is told that this animal is the same species as the previously encountered breed (i.e., therefore, this new breed is also a type of dog, establishing a hierarchical relation).

While the establishment of word-word and word-object coordination relations provide the simplest explication of the relevance of relational responding to vocabulary, the wider collection of relational frames also play a role in facilitating vocabulary expansion. For example, an individual may understand that “fast” means the opposite of “slow” (i.e. a word-word opposition relation). When this individual is told that “swift” means the same as “fast”, he/she can infer both the meaning of the former, as well as its entailed relations to words previously associated with the latter (e.g. “swift” therefore is the opposite of “slow”). At a more complex and abstract level, an individual may have learned the meaning of the prefix “anti”, understanding that the addition of this prefix to a word confers a definition that is semantically opposite to the original word (e.g. forms an opposition relation). For example, once an individual learns the meaning of the prefix “anti”, the individual can thus infer the meaning of the word “antithesis” if he/she has already acquired the meaning of the words “thesis” (and vice-versa). Even without knowledge of the transformational effect of this prefix, if told that “antithesis” means the opposite of “thesis”, the individual can derive the meaning of one if knowledge of the other has already been acquired.

Finally, relational networks may also account for how word knowledge may be learned through analogical reasoning (itself a form of relational responding). When encountering a new word in casual conversation, one may learn that this novel word is "X's version of Y" or "the X of a particular context". For example, an individual may be told that "Fender are to electric guitars as Ferrari are to sports cars", "the Tony awards are the Oscar's of live theatre" or "Jamhuri Day is like Kenya's Fourth of July". In each example, such statements allow the inference of new word knowledge founded on proxy comparisons based on previous word knowledge which may have been explicitly trained, or at least derived from directly reinforced contingencies. More specifically to vocabulary, an individual can infer meanings of new words when presented within analogical premises alongside known words. For example, if told that "a milliner is to hats, as a jeweller is to jewellery", the definition of milliner can be inferred analogically as the relation between milliner to hats must be the same as the relation between jeweller and jewellery (i.e. a milliner must make and sell hats). It is in simple learning experiences and behavioural mechanisms such as these that therefore allow the formation of deep, expansive and sophisticated vocabularies.

In addition to its role in vocabulary expansion, analogical reasoning in itself is viewed as a key component of verbal intelligence. The importance of this form of reasoning is highlighted by the inclusion of analogical tasks in numerous gold-standard intellectual assessments across several iterations, such as the Stanford-Binet (Roid, 2003) and Woodcock-Johnson (Schrank, McGrew, Mather, & Woodcock, 2014). As outlined previously, analogical reasoning is inherently relational in nature, as it entails relating between relations. Most commonly, this implicates coordination relations between relational statements involving four relata (e.g. A is to B as C is the D). In their classic format, analogical reasoning trials typically require identification of one 'missing' relatum when presented with the three other relata within a two-premise relational network. For example, in order to correctly respond to the following trial: "dog is to mammal, as magpie is to what?", an individual must first identify the relation specified in the complete premise (i.e. a dog is a member of the biological class of mammals), and then apply this relation to the incomplete premise (i.e. what biological class does magpie belong to?) to identify the missing relatum (i.e. "bird").

1.4.2 Relational Responding and Numeracy

Numeracy can be defined as the ability to reason with numbers and complete mathematical operations (M. Brown, Askew, Baker, Denvir, & Millett, 1998). From an RFT perspective, any number is simply a relatum which represents a quantity within an ordinally-ranked series that clearly specifies the relations organising these relata (i.e. 2 is more than 1, but less than 3). At its most basic level, the collection of natural numbers essentially consists of a network of symbol-quantity coordination relations (and their associated verbal expressions) that 'maps' a collection of arbitrary abstract symbols (i.e. '1', '2', '3', etc.) to real-life, physical quantities. From an early age, knowledge of this network is acquired in a similar fashion to vocabulary, whereby an individual is taught that a given numeral or number word represents a given quantity of physical stimuli. This learning process may represent one of the earliest-developing examples of generalised relations, as the child is typically shown a given number of topographically distinct stimuli (e.g. "there are 2 footballs", "he has 2 crayons", "there are 2 dogs in the park"), that encourages the child to abstract the meaning of 'two' and then arbitrarily apply this newly-acquired term to novel contexts. This basic process extends further to higher levels of numerical operations, as children are taught the numerical equivalent of symbols that denote mathematical constants such as *pi* ($\pi = 3.14$) and Euler's number ($e = 2.72$).

Arithmetic is one of the primary branches of mathematics and involves the analysis of number and the computation of numerical operations such as addition, subtraction, multiplication and division. The symbols used for these four elementary arithmetic operations (+, -, x and ÷) can be considered as contextual cues that control a particular type of response (i.e. the symbol '+' between two numbers indicate that their quantities should be combined). To recycle an example used by Thirus, Starbrink, & Jansson (2016), in order for an individual to respond correctly when asked "What is π + 2?", he/she must have developed an understanding of a number of important relations. As described by Thirus et al., the individual must have established the three-term coordination relation between spoken word- mathematical symbol-numerical quantity for both '*pi*' and 'two' (i.e. *pi* is depicted symbolically as π, which equals 3.14). If, for example, the child has not been taught, or been able to derive that the symbol π is expressed verbally as '*pi*', he/she will not understand the question and will therefore be unable to respond accurately. If, on the other hand, he/she has not established that π equals 3.14, he/she will not be able to compute the answer. In addition to the prerequisite relations outlined by Thirus and colleagues, the child must also be able to respond in accordance to the numerical operation symbol '+', which functions as a contextual cue controlling behaviour (i.e. prompting the individual to add the quantities either side of the '+' symbol). Furthermore, if the child will not be able to compute the correct answer if he/she is unaware of the relationships of the number 2 (an integer) to each digit of the number 3.14 (i.e. 3 is an integer, 1 represents the tenths place value and 4 represents the hundredths place value). As the integers must be added together, the child must be able to identify which of the digits is an integer before computation, requiring knowledge of the hierarchical relational classification of numbers (e.g. what numbers are integers). If the child has not established this relationship, he/she may erroneously add 2 to a digit other than *pi*'s integer, resulting in an incorrect response.

While relatively few studies within the RFT literature have focussed on the relevance of relational responding to numeracy, parallel research streams have identified relational reasoning as a key contributor to mathematical fluency and sophistication (Carpenter, Franke, & Levi, 2003). Much research has focussed on “relational thinking”, defined as the analysis of relationships specified in mathematical problems carried out before mathematical computation is conducted (Molina, Castro, & Ambrose, 2005). Such analysis has been proposed to foster a deeper and more meaningful understanding of basic arithmetic, as well as complex algebraic operations (Carpenter et al., 2003, 2005; Kızıltoprak & Köse, 2017; Molina et al., 2005; Stephens, 2006; Stephens & Ribeiro, 2012). As Stephens & Ribeiro (2012) outline, the importance of relational thinking can be witnessed when faced with common missing-number sentences, such as “$23 + 15 = 26 + x$”. In order to complete this sentence, a number of strategies can be adopted. For one, the missing number can be identified through simple computation (i.e. $(23 + 15) – 26 = 12$). Alternatively, the solution can be reached by engaging in relational thinking by identifying the relationships between the numbers on either side of the equals sign. In this way, by analysing the relationship between the first numbers on each side of the equals sign (i.e. that 23 is three less than 26), this relationship can be applied to each side’s second numbers to identify the missing number. As the equals sign dictate that each side of the equation must be equivalent, if the right side’s first term is three greater than the left side’s first term, the right side’s second term must be three less than the left side’s second term (i.e. 12). In this way, the equals sign is viewed as a relational symbol (Stephens, 2007, 2008), which, from an RFT viewpoint represents the relational frame of coordination. Indeed, numerous researchers outside the field of RFT (e.g. Kieran, 1981; Molina et al., 2005; Stephens & Ribeiro, 2012) have encouraged this reinterpretation of the equal sign from an operational into a relational function of equivalence, as a means of facilitating improved performance. The processes involved in relational thinking, therefore, display clear similarities to the RFT viewpoint, as this form of reasoning relies heavily on various relational frames, such as coordination and comparison.

Irwin & Britt (2005) further suggest that the direction of relations specified in mathematical operations is also of key importance. In the example missing number sentence used above, it is important to identify the direction of the compensation required to 'balance' each side of the equation (i.e. knowing to subtract 3 from the left side's second term, rather than add 3). Essentially, this implicates either the reversal of a bidirectional relationship (i.e. symmetry; if A is the same as B, then B is the same as A) or the derivation of a unidirectional relationship (i.e. mutual entailment; if A is X more than B, B is X less than A). In the example, after identifying that the left side's first term is 3 more than the right side's first term, this relation must be reversed and applied the right side's missing second term (i.e. it must be 3 less than the left side's second term).

Furthermore, relational responding appears to be a key contributor to understanding fraction-decimal equivalence and graphical representation in mathematics (Leader & Barnes-Holmes, 2001; Lynch & Cuvo, 1995). Lynch & Cuvo (1995) administered a simple matching-to-sample protocol (MTS) to train symmetry and equivalence responding for fraction-to-graph and graph-to-decimal relations in fifth- and sixth-grade students and demonstrated that such a protocol facilitated generalised responding to untrained relations (e.g. fraction-to-decimal equivalence). Leader & Barnes-Holmes (2001) implemented a similar procedure, whereby 24 five-year-old children were exposed to a three-stage training intervention to establish fraction-graph relations. In this design, participants were first administered a series of demonstration trials, during which they were explicitly shown how to correctly respond to comparison fraction stimuli by pointing to the graphical representation of that fraction. For example, when presented with the fraction "$^3/_4$", the participants would be instructed to select the graphical representation that depicted a circle with three quadrants shaded in. Following this phase, this procedure was repeated with reinforcement for correct responding, but with guidance from the researcher withdrawn. In the final stage, this procedure was repeated once again, but with reinforcement also withdrawn. Results indicated that not only was this procedure effective in establishing fraction-graph equivalence, but that responding generalised to more complex pictorial representations of fractions.

A number of studies (McGinty et al., 2012; Ninness et al., 2006, 2009; Ninness, Rumph, McCuller, Harrison, et al., 2005; Ninness, Rumph, McCuller, Vasquez, III, et al., 2005) have displayed that such MTS procedures display considerable efficacy even at higher levels of mathematical complexity, proposing that such protocols can be used to train a range of proficiencies, including transformation-of-function graphs, algebraic functions, trigonometric functions and precalculus. Ninness, Rumph, McCuller, Harrison, et al., (2005) showed that MTS can be used to train formula-to-formula and formula-to-graph relations in relation to reflections and verbal and horizontal shifts about the coordinate axes, in a sample of participants' naïve to algebraic and trigonometric transformations. Ninness, Rumph, McCuller, Harrison, et al., (2005) replicated this design to train factored-formula-to-graph relations for vertical and horizontal shifts on the coordinate axes to similar success. Finally, Ninness et al. (2009) reported significant improvement in complex trigonometric formula-to-graph relations in accordance with coordination and opposition relational frames following training of mutually- and combinatorially-entailed relations. In all three experiments, participants displayed generalised, derived responding to novel, untrained tasks.

As evinced by the theoretical considerations and empirical investigations outlined above, relational responding proficiency can be considered as being of central importance to a gamut of numerical reasoning and mathematical operations, underlining the relevance of this repertoire to general intellectual and cognitive performance. The contribution of relational skill to various intellectual capacities can be further highlighted through analysis of widely-administered IQ tests, as many such testing batteries comprise a number of items which appear to “tap” relational responding repertoires.

1.4.3 Relational Responding and IQ Test Items

In the previous section, the relevance of relational responding to many facets of verbal intelligence and numeracy was outlined. However, the contribution of relational responding to general intellectual function is drawn into even sharper relief upon analysis of the types of tasks included in many traditional IQ assessments, as many such test items can be considered as assessments of relational skill.

1.4.3.1 Coordination and Opposition Relations.

Perhaps the most evident example of IQ test items invoking coordination relational responding comes in the form of assessments of vocabulary, in which participants are required to define a given word. As outlined previously, an established proficiency in coordination relational responding facilitates vocabulary acquisition and serves as the basis of linguistic reference (Stewart, Tarbox, Roche, & O’Hora, 2013), and is therefore a key contributor to performance on assessments of vocabulary. Such subtests are widely administered in a number of gold-standard IQ assessments, such as the Wechsler Adult Scale of Intelligence (WAIS; Wechsler, 1955, 1981, 1997b, 2008), Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 1999, 2013), Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949, 1974, 1991, 2003), Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 2002), Stanford-Binet Intelligence Scale (SB; Roid, 2003; Terman, 1916; Terman & Merrill, 1937, 1960; Thorndike et al., 1986), Woodcock-Johnson Tests of Cognitive Abilities (WJ; Woodcock & Johnson, 1977, 1989; Woodcock, McGrew, & Mather, 2001), Kaufman

Brief Intelligence Test (K-BIT; Kaufman & Kaufman, 1990, 2004) and the Differential Ability Scales (DAS; Elliot, 1990, 2007).

Vocabulary test items generally can be perceived as assessments of either word-word or word-object coordination relations. In the case of word-object coordination relations, participants may be asked to either to provide a verbal label for a visual stimulus or to select a picture when provided with such a verbal label. For example, a participant may be asked: “Show me the picture of the bird” in the presence of multiple other visual stimuli, or be shown a picture of a bird and asked to identify what is being depicted. Such object-word coordination relations are assessed in subtests such as the WAIS Vocabulary, WISC Vocabulary, WASI Vocabulary, SB 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 subtests.

Word-word coordination relations are also commonly assessed, generally as a more advanced measure of verbal ability, due to the level of abstraction involved. In these tasks, participants are provided with a sample word and asked to define it using a word (or collection of words) that is synonymous or equivalent in meaning. For example, the WASI, which includes a Vocabulary subtest as one of its four components (typifying its central status within the wider Wechsler IQ testing battery), probes for word-word relations in questions such as: “What does reveal mean?”. From an RFT perspective, any correct response to such a question (e.g. “show”, “expose”, “uncover”) represents a word (or set of words) which forms a relational frame of coordination with the sample word. Such word-word relations are assessed in a number of traditional IQ subtests, including WAIS Vocabulary, WISC Vocabulary, SB Vocabulary, WJ Verbal Comprehension and DAS Word Definitions.

Another example of a common IQ task which invokes coordination relational responding are subtests in which participants are required to identify the manner in which two words are related (e.g. “How are a pen and a pencil alike?”). For these test items, included in subtests such as WAIS/WASI/WISC Similarities, WJ Verbal Comprehension and DAS Similarities, participants must display knowledge of the shared characteristics of each word, and/or common categorisations they may fall into. Indeed, such test items are a relatively direct assessment of coordination relations, as participants must identify in what way, or along which continuum, are two words equivalent. For example, pens and pencils are both members of the categories “stationary” or “writing utensils”. Correct answers may be facilitated by both non-arbitrary coordination relations based on physical form, as well as arbitrary relational responding based on use or verbal categorisation. As such, this is both a test of established verbal coordination and hierarchical relational responding. That being said, at least in the case of the Wechsler Similarities subtest, answers based on the abstraction of relationships between two stimuli is favoured, as such responses are awarded more points than those which are based on physical similarities. This emphasis placed on abstraction as an indication of higher intellectual performance is commensurate with the RFT view of intelligence (Roche, Cassidy, & Stewart, 2013).

While assessments of verbal coordination relations may be among the most clearly evident examples of the relevance of relational responding to psychometric measures of intelligence, a number of IQ subtests measure the ability to identify and derive non-verbal coordination relations based on physical characteristics. For example, three separate subtests included in the Woodcock-Johnson Test of Cognitive Abilities’ Processing Speed section (Visual Matching, Decision Speed & Cross-out), as well as WISC Cancellation all involve the identification of physically identical or similar abstract shapes as quickly as possible. As such, in subtests such as these, fluency of relational responding is emphasised, as opposed to a singular interest in responding accuracy. The Non-Verbal General Ability subtest of the Alice Heim Group Abilities Test (AH4; Heim, Watts, & Simmonds, 1968) involves the selection of one geometric stimulus amongst an array that matches a sample stimulus exactly, thus probing for physical coordination relations. The Picture Concepts subtest, part of the WISC is a further example of an IQ subtest which relies on physical coordination relational responding. In this subtest, an array of visual stimuli in rows is presented to the individual. In order to respond correctly, individuals must select one option from each row that match due to similarity in some dimension. For example, in an early trial which presents one row including a glove and a banana and another row that includes a lightbulb and a strawberry, the correct answer would be to select the banana and strawberry, as both are fruit. Later trials involve responding based on arbitrary, rather than non-arbitrary features of the stimuli. An example of this would be the last trial of this subtest included in the WISC-IV, in which three rows are presented which include an array of items which differ greatly in physical form. The first row includes an opened tin can, a toothbrush and binoculars. The second includes a sledge, a microscope and a metal clamp, while the final row displays an umbrella, a shoe and a hair comb. The correct answer, in this case, would be to select the tin can, the metal clamp and the umbrella, as each of these can be firmly closed and locked in place. As many trials involve the identification of category membership for multiple stimuli (e.g. a banana and a strawberry are both types of fruit), tasks of this nature can also be considered as assessments of hierarchical relations.

Many IQ test items also probe for frames of opposition as well as coordination. The AH4 involves tasks which ask participants to identify the antonym of a given word. For example, participants may be asked; “Easy means the opposite of _____?” and provided with the following response options: “problem”, “simple”, “difficult”, "always" and "cannot". The Otis-Lennon School Ability Test (OLSAT; Ahmann, 1985) is a standardised assessment of abstract reasoning, and also employs an antonym subtest in this form, as a measure of verbal ability. In addition, the OLSAT also includes "odd-one-out' subtests such as Figural Classification and Picture Classification, in which the participants must identify which stimulus out of an array is least similar to the other stimuli. For example, the Figural Classification presents five images consisting of a butterfly, bird, dragonfly, ladybird and bee, and asks participants which is not the same as the others (i.e. bird as it is not an insect).

1.4.3.2 Comparison, Temporal, and Spatial Relations.

In terms of comparative relations, such responding is assessed most readily by assessments of arithmetic or general numeracy. In fact, many such test items are entirely relational in nature. For example, the WAIS-III includes questions such as: "Chris has two times as much as Robert. Chris has 99 pounds, how much money does Robert have?". To complete this test item, after the relation between Chris and Roberts respective quantities of money has been stated, the individual must then reverse this bidirectional relation and apply it quantity specified for Robert (i.e. Chris has twice as much as Robert, therefore 99 should be halved). The Otis-Lennon School Ability Test includes similar tasks in its Arithmetic Reasoning subtest, in which participants must compare relative quantities and perform various computations on such quantities. For example, participants will be shown a picture depicting "Francesca's frogs", and are told that Francesca's friend Chris has two frogs for every one frog she has. Participants must then select from four images each depicting various quantities of frogs, how many frogs Chris has.

Many IQ subtests designed to assess working memory can be considered to be assessments of temporal relational responding, especially in cases in which the information to be retained must be transformed or reordered in some way. The relevance of temporal relational responding in this regard is perhaps most marked in subtests such as WAIS/WISC Letter-Number Sequencing and WJ Auditory Working Memory, in which a mixed list of numbers, words and/or letters must be memorised and rearranged in accordance with a given rule (e.g. words in alphabetical order, followed by numbers in chronological order). Such subtests, therefore, rely heavily on temporal relational responding due to the fact that success on such trials hinges on the sequencing of stimuli in adherence to established temporal relational networks (i.e. alphabetical or chronological sequence). In addition, a number of other working memory subtests, such as WAIS/WISC Digit Span, WJ Numbers Reversed, WJ Memory for Words, WJ Memory for Sentences, SB Memory for Sentences and SB Block Span, all place an emphasis on sequence of response (either in terms of stimulus presentation or in accordance to a given rule), and therefore tap into temporal relational responding

A small number of IQ subtests, such as Wechsler Picture Arrangement may also be conceptually relevant to temporal relational responding by focussing on the logical sequencing of events. In such tasks, participants may be asked to rearrange storyboard-type images in order to compose a coherent narrative or may be asked to propose what may happen next in a given sequence. In one such trial of Wechsler Picture Arrangement subtest, the participant is presented with a number of cards, each of which depicts one step in the process of washing and drying laundry. In order to successfully complete this task, the participant must arrange these cards in a logically coherent order (e.g. the story must show clothes being washed before being dried and folded). Other examples of temporal relational responding trials present in IQ tests can be found in general knowledge subtests, which can include items based on sequence or temporal order. For example, the Wechsler Information subtest asks questions such as: “What day comes after Friday?” and “Which month comes next after April?”.

While spatial relations are not traditionally targeted in IQ assessments, a small number of tasks involve this form of relational responding. The Position & Direction subtest, part of the SB Visual-Spatial Processing section, is an explicit measure of spatial orientation across multiple task types. For example, one task type requires participants to locate items when given specific contextual cues based on spatial relationships (e.g. X is behind Y, Y is to the left of Z). Furthermore, participants may be asked to indicate direction and position in relation to a specified reference point, outline directions from an origin point to a destination point and to chart the position of an individual after walking a specified course.

1.4.3.3 Analogical Relations.

Analogical reasoning skill is regarded as a core facet of higher intellectual ability, and as such, many IQ tests include verbal analogy tasks as measures of verbal knowledge and abstract reasoning. Verbal analogies are included in a range of intelligence assessments, such as SB Verbal Analogies, WJ Number Series, WJ Number Matrices, WJ Verbal Comprehension, OLSAT Verbal Analogy, CAT Verbal Analogy and AH4 Verbal General Ability. Such subtests usually consist of an incomplete relational statement, in which the participants must identify a missing word based on the other relationships provided in the statement. For example, participants may be asked to complete the following statement: "Hand is to glove, as foot is to ____?". In order to identify the missing word and complete this task, the participant must identify the relationship between the first two stimuli (i.e. a glove is worn on a hand) and then identify a stimulus that shares this same relationship with the third stimulus (i.e. a sock or shoe is worn on the foot). For more basic trials, the relationships specified in each trial tend to be based on categorisation, function or topographical similarity. More advanced trials are more arbitrary in nature and may be based on more symbolic or semantic relationships (e.g. poetry is to rhyme, as philosophy is to theory).

While this type of analogical reasoning task tends to utilise words as stimuli, there are some subtests which employ numerical stimuli as relata, most notably WJ Number Series, WJ Number Matrices, DAS Sequential and Quantitative Reasoning, OLSAT Number Series, CAT Number Analogy, CAT Number Series and AH Verbal General Ability. These subtests generally involve identifying a number which belongs in a numerical sequence (e.g. "2, 4, 6, __, 10), or require participants to identify a number which adheres to a rule specified by sample stimuli. However, regardless of the stimuli used, or the nature of the relationships being specified, all analogical tasks require the inference and derivation of relationships and thus, relational responding is a key contributor to performance on such tasks.

Matrix reasoning subtests, measures of visuospatial analogical reasoning, are one of the most commonly employed protocols in the field of intelligence testing, so much so that many proponents view the various forms of Ravens Matrices as the purest measurement of general intelligence (Jensen, 1998; Spearman, 1946; Thorndike et al., 1986). Matrix reasoning subtests are a mainstay of Wechsler and Stanford-Binet IQ assessment batteries, and are also used in various short-form and proxy assessments such as the Bochumer Matrices Test (BOMAT; Hossiep, Turck, & Hasella, 1999), Comprehensive Test of Nonverbal Intelligence (C-TONI; Hammill & Pearson, 2017), Kaufman Brief Intelligence Test, Differential Ability Scales, Cognitive Abilities Test (CAT; Kerr & Lohman, 2012) and Otis Lennon School Ability Test. While such tests vary in duration and complexity, matrix reasoning tests generally involve the presentation of a sequence of abstract visual stimuli (usually arrangements of geometric shapes) and require individuals to select from an array the visual stimuli which 'belongs' or 'comes next' to the sample sequence. For example, being presented with a sample array of a triangle, a square and then a pentagon, the individual should select a hexagon in accordance to the 'rule' espoused by the sample (i.e. progressive increase in the number of angles in each successive shape). Matrix reasoning is, in essence, a measure of visuospatial analogy, which can be expressed verbally as: “Stimulus A/B is to Stimulus B/C, as Stimulus C is to what?”. Indeed, the AH4’s Non-Verbal General Ability Test presents matrix reasoning trials in exactly this format (i.e. Shape 1 is to Shape 2 as Shape 3 is to what?). Individuals must therefore identify which of the response options “fits” into the rule specified by the sample stimuli, which can only be derived by analysing the relationships between each of the sample stimuli as they are presented.

1.4.3.4 Hierarchical Relations

Hierarchical relations are also regularly implicated across various assessment types and comprise part of numerous IQ tests such as the Stanford-Binet, Woodcock-Johnson Tests of Cognitive Abilities, Comprehensive Test of Nonverbal Ability and the OTIS-Lennon School Ability Test. Hierarchical relations, and more specifically, member-to-class relations are assessed by “odd-one-out” tasks, notably in the subtests such as OLSAT Pictorial Classification, OLSAT & CAT Figural Classification and OLSAT & CAT Verbal Classification. In these three separate subtest types, the participant is presented with either five pictures (Pictorial Classification), five abstract geometric shapes (Figural Classification), or five words (Verbal Classification), and must select which of the five ‘belongs’ with the others, in accordance to specific categorical classification criteria. For example, in the Pictorial Classification subtest, participants may be presented with five images comprising a balloon, a basketball, an orange, a gift-wrapped box and a beach. In order to successfully respond to this trial, participants must identify which categorisation four of the images fall into (i.e. spherical objects), and which image is not a member of this group (i.e. gift-wrapped box). The WJ Concept Formation subtest is a measure of categorical reasoning and inductive logic that bears some similarity to the OLSAT and CAT Classification subtest items. In this task, participants are presented with a complete set of geometric shapes and must infer the rule that groups all but one of the stimuli together (e.g. common shape/colour/quantity), and then subsequently select the shape that does not belong.

A significant portion of the Comprehensive Test of Nonverbal Intelligence assesses hierarchical relational responding, as 2 of 6 subtests comprising its testing battery are direct assessments of these relations. During these subtests, Pictorial Categories and Geometric Categories, participants must select from an array of 5 stimuli (either pictures or geometric designs), which stimulus shares a relationship with two target stimuli. In the Pictorial Categories subtest, participants may be exposed to 5 response pictures consisting of a horse, rabbit, cat, fish and snake, alongside two target images of two different fish. The participant is then asked: "Which one of these (examiner points to 5 response options) is related to those (examiner points to target stimuli)?". In most cases, the correct answer can be achieved by selecting the response option which is the same type of thing (e.g. all are fish/fruit/circles). As is the case with previously discussed assessments of coordination relations, subtests such as Pictorial Categories and Geometric Categories may tap both coordination and hierarchical relations in this manner.

While several IQ tasks implicate the identification of physical or semantic categories among arrays of stimuli, a number of further subtests require participants to demonstrate knowledge of multiple members of a specified verbal category. The Woodcock-Johnson Tests of Cognitive Abilities includes two components that contain such tasks, Phonological Processing (an Auditory Processing subtest) and Retrieval Fluency (a Long-term Retrieval subtest). The Phonological Processing subtest contains three sections, the first two of which (Word Access & Word Fluency), both involve naming words which begin with a particular letter or phoneme (essentially probing for exemplars of a given verbal category, e.g. words beginning with 'a' or the phoneme ‘buh’). The Retrieval Fluency subtest, a measure of long-term memory, is similar as participants are asked to provide as many members of a given category (e.g. things to eat) within a one-minute time period. Therefore, for such a subtest, a more extensive relational network for the given category provides a significant advantage for performance.

The Incomplete Words subtest included in the WJ-IV may also be considered to require a degree of hierarchical relational responding proficiency, as participants are aurally provided with words in which phonemes are missing (e.g. muting the third syllable in the word ‘television’). In order to complete this task, the participant must analyse the phonemes provided, and along with the missing phoneme, identify which word they constitute (i.e. phoneme-to-word hierarchical relations). Similarly, in the CAT Figure Recognition, participants are shown a geometric component shape and must choose which from a selection of five complete geometric shapes it is a part of (i.e. component-to-whole hierarchical relations).

1.4.4 AARR-IQ Correlations

While the relevance of relational responding repertoires to intellectual performance may be highlighted by analysing the content of traditional IQ assessments, a small number of analyses (e.g. Colbert et al., 2017; O’Hora et al., 2005, 2008; O’Toole & Barnes-Holmes, 2009) have attempted to substantiate this theoretical assertion empirically by investigating the correlation between relational ability and scores on traditional IQ tests and subtests.

O’Hora et al. (2005) analysed the relationship between performance on three WAIS-III subtests (Vocabulary, Arithmetic & Digit-Symbol Coding) and scores on an assessment of derived temporal and distinction relations. Correlational analyses reported weak, but significant correlations between temporal relation task performance and both the Vocabulary (r =.34) and Arithmetic subtests (r = .23), but not Digit-

Symbol Coding. Indeed, as outlined in previous sections, the relationship found between relational skill and the former two subtests is predicted by theoretical accounts proposing the relevance of relational responding to vocabulary acquisition and mathematical operations. The relevance of relational responding to the third subtest, Digit-Symbol Coding, is less readily elucidated, which may explain the lack of correlation in this regard. In addition to the significant correlations reported, those who reached mastery criteria (score ≥ 20) on the 24-item derived relational responding assessment recorded significantly higher scores on both Vocabulary and Arithmetic, indicating a potential relationship between relational proficiency and higher levels of intellectual performance.

In an extension of the previous investigation, O'Hora et al. (2008) administered a full-scale WAIS-III IQ assessment and assessed the relationship between IQ scores and performance on a temporal relations task. The temporal relations task implemented required participants to demonstrate learning of the temporal relational functions of two abstract symbols (before: ()() and after: ::::) within 12 blocks of 16 trials. In this task, two relational statements, composed of two geometric shapes (square & circle) separated by one of the two abstract symbols were presented in either bottom corner of the screen. Two matching geometric shapes were then displayed in varying sequential orders on the top of the screen. In accordance with the order of the shapes being presented, participants would have to select one of the two relational statements at the bottom of the screen. Through multiple exemplars and corrective feedback, the abstract symbols should establish contextual control over responding, as participants learn the "meaning" of such symbols. The success criterion for this task was 15 correct responses out of 16. Those who successfully reached this criterion within 12 blocks (56% of participants) were found to have significantly higher Full-Scale, Verbal and Performance IQs when compared to those who failed the relational task. In addition, successful participants recorded significantly higher scores on two of four WAIS-III subindices: Verbal Comprehension and Perceptual Organisation. Overall performance on the temporal task (operationalised as correct response percentage), was found to show medium-strength correlations with all three IQ indices (Full-Scale, Verbal and Performance IQ), two of four IQ subindices (Verbal Comprehension & Perceptual Organisation) and five of thirteen IQ subtests (Vocabulary, Similarities, Information, Block Design & Symbol Search).

To complement previous correlational analyses which focussed on relational responding proficiency, O'Toole & Barnes-Holmes (2009) demonstrated that relational flexibility (i.e. the ability to adjust responding to changing contextual cues) may also be an important contributor to intellectual performance. Relational flexibility was measured using the Implicit Relational Assessment Procedure (IRAP; Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010), a computer-based task in which the rules governing participants' responses are switched on alternate blocks between relations that are consistent and inconsistent with previous learning, requiring participants to repeatedly adapt to new task demands. In O'Toole et al.'s iteration, participants would be presented with a relational premise in accordance with either the cues 'same/different' or 'before/after' (e.g. 'spring before summer', 'engagement before marriage', 'crawl before walk'), and would be required to respond by clicking either 'true' or 'false' onscreen, based on a specified rule for that given block. This rule would alternate between rewarding responding either consistent (e.g. choosing 'true' when presented with 'child before adult') or inconsistent (e.g. choosing 'true' when presented with 'adult before child') with common verbal practices and knowledge. As response time was predicted as being quicker for consistent trials due to their congruence with established knowledge, relational flexibility was operationalised by measuring the speed of response for inconsistent trials, as this represented a metric of the ease at which participants could adhere to novel rules governing behaviour. Results indicate that relational flexibility predicted intellectual performance (as measured by the Kaufman Brief Intelligence Test), a finding perhaps supported by the proposal that cognitive flexibility is regarded by some theorists as a key facet of higher-level cognition (Cattell, 1971; Jensen & Cattell, 2006; Kyllonen, Lohman, & Woltz, 1984; Premack, 2004).

Gore, Barnes-Holmes, & Murphy (2010) investigated the relevance of deictic relational responding to intelligence by administering the Relational Frame Theory Perspective Taking Protocol (RFT-PT, McHugh et al., 2004), a 62-item perspective-taking assessment and the WASI to sample of 24 adults diagnosed with mild-to-moderate intellectual disabilities. Results indicated moderate correlations between performance on the RFRT-PT and WASI Full-Scale, Verbal and Performance IQ.

The Promoting the Emergence of Advanced Knowledge Relational Training System (PEAK; Dixon, Whiting, Rowsey, & Belisly, 2014) is an autism evaluation and treatment program based on behaviour-analytic and RFT principles which targets and improves fluency in derived relational responding as a means of ameliorating basic verbal, social and cognitive skills. The PEAK consists of an extensive compendium of 184 training procedures across four modules (Direct Training, Generalization, Stimulus Equivalence and Transformation of Function), which target behaviours ranging from fundamental learning skills such as eye contact and object permanence to advanced verbal and social skills, such as understanding metaphor and sarcasm. As part of the evaluation process, the training program also includes a PEAK assessment, which provides a metric of derived relational responding proficiency based on the individual's ability to successfully complete sample trials from each of the 184 training modules. Correlational analyses have indicated that the PEAK assessment displayed strong correlations with both measures of verbal ability, Peabody Picture Vocabulary Test

(Dunn & Dunn, 2007) and Illinois Early Learning Standards Test (Illinois State Board of Education, 2013) and intellectual performance as assessed by various widely-administered IQ assessments (Dixon et al., 2014). While such results are promising, this promise is somewhat tempered by inconsistency regarding the range of IQ assessments used to measure intellectual ability in Dixon et al.'s analysis (2014). Approximately 30% of individuals' IQ scores were derived from measurements of adaptive behaviour such as the Vineland Adaptive Behaviour Scale (Sparrow, Balla, & Cicchetti, 1984; Yang, Paynter, & Gilmore, 2016) and academic achievement, such as Wechsler Individual Achievement Test (Wechsler, 2005), rather than more traditional intellectual assessments. However, despite this, the high level of correlation between PEAK assessment scores and measures of key cognitive skills, further suggests the importance of DRR to intelligence.

The TARPA (Training & Assessment of Relational Precursors and Abilities; Moran, Stewart, McElwee, & Ming, 2010) is a behaviour-analytic training intervention designed to facilitate the establishment and improvement of relational responding repertoires, as a means of improving generative language skills, predominantly in autistic populations. As part of this training protocol, an assessment is included which provides a metric of relational responding fluency across a number of forms, such as basic discrimination, non-arbitrary conditional discrimination, arbitrary relational responding, mutually-entailed relational responding, combinatorially-entailed relational responding and transfer of function. In early analyses of the program using a small samples of children diagnosed with autism and normally-developing children, strong significant correlations were reported between TARPA assessment scores and the Vineland Adaptive Behaviour Scales (Moran et al., 2010) and the Preschool Language Scale 4th edition (Moran, Stewart, McElwee, & Ming, 2014). In a larger follow-up investigation of this assessment using a comprehensive battery of intelligence, verbal ability and adaptive behaviour assessments, using a larger sample of children diagnosed with autism (Moran, Walsh, Stewart, McElwee, & Ming, 2015), TARPA scores were shown to display high levels of correlation with Stanford-Binet Abbreviated Battery raw IQ scores ($\rho = .74$), as well as scores for each of its two subscales: non-verbal ($\rho = .65$) and verbal ($\rho = .7$). TARPA scores were also found to correlate significantly with a standardised test of language ability, the Pre-school Language Scale ($\rho = .73$), alongside its two subscales, auditory comprehension ($\rho = .68$) and expressive communication ($\rho = .73$). Finally, TARPA scores displayed significant correlations with the Vineland Adaptive Behaviour Scales ($\rho = .64$), replicating the results of Moran et al., (2010).

In perhaps the most comprehensive psychometric analysis of the relationship between relational responding and intellectual performance, Colbert et al. (2017) conducted a correlational analysis of Relational Ability Index Scores with an extensive battery of cognitive ability metrics and a gold-standard IQ assessment, the Wechsler Adult Intelligence Scale (WAIS-III; Wechsler, 1997a). In the first of their two studies, Colbert et al. report significant correlations between RAI scores and performance on measures of literacy (National Adult Reading Test, Nelson, 1982; $r = .58$), general memory (Rey Auditory Verbal Learning Test, Rey, 1958; English version: Taylor, 1959; $r = .7$) and visuomotor tracking and divided attention (Trail Making Test; Lezak, 1995, $r = .36$). In their second study, RAI scores displayed medium-to-strong correlations with all three main WAIS IQ indices (Full-Scale IQ: $r = .74$, Verbal IQ: $r = .78$, Performance IQ: $r = .55$), at a rate commensurate, and in some cases exceeding, those reported for more traditional short-form and proxy IQ metrics. RAI scores were also shown to predict all four WAIS subindices, Working Memory ($r = .64$), Verbal Comprehension ($r = .61$), Perceptual Organisation ($r = .53$) and Processing Speed ($r = .43$). Finally, significant correlations were found between RAI scores and 10 of 13 WAIS IQ subtests. Such widespread correlations, at considerable levels of strength, provide strong evidence to propose that relational responding proficiency may be closely related to intellectual performance.

In sum, the discussed investigations reflect a growing literature base that proposes a significant degree of covariance between relational and intellectual ability, thereby supporting assertions of the close relationship between both repertoires. Furthermore, such close levels of correlation tentatively suggest that the amelioration of relational responding proficiency may potentially result in increments in intellectual ability. Research which addresses such a proposal will be discussed in the following section.

1.4.5 Training Relational Skills

Given RFT's conceptualisation of derived relational responding as a generalised operant, such responding is inherently flexible and can therefore be shaped by reinforcement and brought under contextual control (Hayes et al., 2001). As generalised operant classes are defined by their function rather than their topography, the form of such behaviours vary across different contexts (Barnes-Holmes & Barnes-Holmes, 2000). The establishment of such a class of behaviour therefore necessitates learning experiences involving a large number of exemplars which vary in their topography, as specific contextual cues must be eventually abstracted as discriminative for the operant behaviour (Barnes-Holmes et al., 2000). As outlined by Hayes & Wilson (1996), in the context of establishing relational responding, it is the specific formal characteristics of each relatum and not the contextual cue that dominates in early learning experiences (e.g. "the truck is bigger than the car"). However, by sufficiently altering the relata included in these explicitly-trained relations, the importance of relatum's form is gradually diminished, and the contextual control of the relational cue ('bigger than') is then abstracted and generalised.

Multiple Exemplar Training (MET) has therefore been identified as an efficacious means of establishing and improving relational responding repertoires by exposing participants to numerous trials which isolate a particular form of relational responding (Barnes-Holmes & Barnes-Holmes, 2000; Catania, 1992; Hayes et al., 2001). In such training protocols, a specific relational frame (e.g. coordination, comparison, opposition etc.) may be targeted to be established or improved. This protocol will expose the participant to multiple trials which provide the opportunity to demonstrate the desired response, providing immediate feedback to each response in order to establish correct responding. For example, in order to establish mutually-entailed coordination relations, the MET protocol would present the participant with a large number of trials in the form of "A is the same as B, is B the same as A?". Across trials, the relata used will vary, in order to establish the centrality of the contextual cue (i.e. "same as") rather than the topography of the relata. While the topography of the relata is modified for each trial, the conditions required to obtain reinforcement (i.e. responding that A is the same as B) remains constant, thereby shaping the desired responding. Through this process of directly reinforcing a particular type of relational responding across a large array of stimuli, this form of responding is established as a generalised operant and can be applied to novel contexts.

Dymond & Barnes (1995) conducted an early demonstration of the efficacy of multiple-exemplar training by administering this protocol to establish derived relational responding in accordance to the frames of coordination, opposition and comparison. Across two virtually identical experiments, four adult subjects were exposed to a training program which isolated nonarbitrary same, opposite, more than, and less than trials, while two others did not receive this training. In the pre-training phase, participants were presented with a simple match-to-sample procedure in which they were presented with a sample stimulus (comprising a simple geometric shape, e.g. a short line) and three comparison stimuli (e.g. a short line, a medium line, and a long line). Alongside these stimuli was a three-letter nonsense word, which functioned as a contextual cue to control behaviour. For example, a randomly-selected nonsense word, such as “CUG”, functioned as the contextual cue for “same”. Therefore, when presented with this contextual cue, participants were required to select the sample stimulus which was the same as the target stimulus. The contextual function of this nonsense word was established through corrective feedback following the emittance of a response. Another randomly-selected nonsense word served as the contextual cue for “opposite”, and when presented, would require the participant to select the comparison stimulus that was in opposition to the sample. Following pre-training, all 3 participants completed training for six arbitrary relations. This training replicated the format of the previous pre-training stage, with the only modification being that the trial stimuli were now 13 arbitrary alphanumerics (e.g. A1, B1, A2, and B2). The six relations that were trained in this phase were: A1 same as B1, B1 same as C1, A1 less than B2, A1 more than C2, N1 more than B2, and N3 less than C2. Participants were exposed to 10 trials assessing each of these relations, and were required to emit 9 correct responses for each. Following this, participants were administered an assessment of seven derived relations based on the relations specified in the last stage. Result indicated that the MET protocol implemented was successful in establishing derived coordination relational responding in the four pre-trained participants, but not the non-pre-trained participant. In addition, subsequent studies conducted as part of this investigation found that such an MET protocol was effective in establishing derived relational responding in accordance to opposition and comparison frames.

Barnes-Holmes, Barnes-Holmes, Roche, & Smeets (2001) demonstrated that MET facilitated the transformation of function in accordance with symmetry in a sample of four- and five-year-old children (n =16). In this design, participants were equally divided into four distinct experimental conditions. In the first condition, participants were first trained to demonstrate listening, echoic, and tacting behaviours in response to specific actions and objects (i.e. name training). For example, when the participant heard their name, the experimenter reinforced pointing to one of the objects (e.g. car). Following this training, participants were then trained to select one of two objects in response to the experimenter engaging in a given action (e.g. selecting a toy car when the experimenter waved their hand, but not when the experimenter clapped) in a conditional discrimination task. Name training was then administered once again, before the participants completed an assessment of derived symmetrical relational responding which required the reversal of the previously reinforced object-action relation (i.e. when presented with the toy car, the child would wave). If the participant did not demonstrate derived relational responding at this point, MET for symmetrical relations was administered. The second condition replicated the first, but removed the second name training session. The third condition also replicated the protocol for the first condition, with the only modification being that participants were trained to tact all of the objects and actions. Finally, the fourth condition replicated the first condition, but reversed the direction of the relations trained and tested (i.e. first trained object-action relations, then tested action-object relations). Results indicated that 3 participants displayed derived symmetrical relational responding upon the first assessment. The remaining participants did not immediately derive the appropriate relations, but did so following explicit MET for symmetry.

Gomez, Banos-Martin, Barnes-Holmes, & Barnes-Holmes (2007) extended upon these findings by administering the same training protocol to a sample of normally-developing 4-year-old children. In addition, in the second of their experiments, Gomez and colleagues modified the conditional discrimination tasks so that two actions, rather than one, were associated with each object, thereby establishing a three-term contingency. Participants were then required to demonstrate action-action equivalence responding when prompted, by deriving the untrained relationship between each action, based on their common relation to an object. For three of four participants, these relations were only derived successfully following MET for equivalence responding. The authors therefore conclude that MET is an efficacious means of establishing symmetrical and equivalence responding in cases in which it may be absent.

Barnes-Holmes et al. (2004) utilised MET to establish arbitrarily-applicable comparison relational responding in a sample of three children between the ages of four and six. In this design, participants were presented with two or three paper coins which were identical in size but not colour. Placed between these coins were arrows pointing either to the right or to the left, with either “BUY MORE” or “BUY LESS” printed above the arrow. Participants would then be asked to select which of the coins would “buy as many sweets as possible”. Without explicit training for comparison relational responding, all three participants failed to display AARR in accordance with comparison relations. However, following MET, all participants were able to derive untrained relations within three-coin contingencies. In addition, derived responding was shown to generalise to novel stimulus sets, demonstrating that relational responding can be trained as a generalised operant class of responding.

In an extension of the study conducted by Barnes-Holmes et al. (2004), Berens & Hayes (2007) further examined the utility of MET in establishing derived relational responding by administering an intervention to establish arbitrarily applicable comparison relational responding. This study recruited four 4- to 5-year-old children who failed an assessment of AARR in accordance with comparison relations, and administered a more experimentally-rigorous MET protocol to establish such responding within a multiple-baseline and multiple-probe design. Unlike Barnes-

Holmes et al.'s design, Berens and Hayes trained mutual- and combinatorial-entailment independently and did not rely on purely linear trial types (e.g. A > B > C) by modifying the arrangement of relata in these three-term contingencies in order to present them in a non-linear fashion (e.g. A > B & C < B). The results of this study further underline the positive impact of MET on relational responding repertoires.

MET has been found to be effective in establishing relational responding in children as young as 15 months (Luciano, Becerra, & Valverde, 2007). In this experiment, an infant named Gloria, who showed no evidence of receptive symmetry and naming at baseline, was exposed to MET in immediate and delayed receptive symmetrical responding (from object-sound to sound-object selection). The methodology employed in this study essentially mimicked common learning experiences in naturalistic settings, by presenting the child with ten physical objects (e.g. a wooden puzzle piece) and paired each object with a verbal utterance (e.g. "puzzle"). Symmetrical auditory-visual relations were tested after either a 1- or 30-minute delay by pointing to an object and asking Gloria to name that object. Gloria displayed receptive symmetry with a 3-hour delay following this MET. In the second and third of their experiments, conducted when Gloria was 17- and 22-months old respectively, visual-visual equivalence relational responding was established using two- and three-comparison matching-to-sample tasks.

Rosales, Rehfeldt, & Lovett (2011) investigated the impact of MET in establishing derived relations in typically developing preschool children currently learning English as a second language. Rosales and colleagues isolated and trained symmetrical relations between objects and their English names (object-word relations), and then tested for derived tacts (word-object relations). MET interventions using novel stimulus sets were administered to those who failed to derive such relations, before reassessing derive relations between objects and words from the originally tested stimulus set. Results indicated significant improvements in tacting behaviour for all participants following MET.

As outlined, there is a large number of studies that propose that relational responding is readily amenable to establishment and improvement via MET intervention, across a number of relational frames, most notably coordination (Dymond & Barnes, 1995; Gomez et al., 2007; Luciano et al., 2007; Rosales et al., 2011) and comparison (Barnes-Holmes, Barnes-Holmes et al., 2001; Berens & Hayes, 2007). Taking into account the theoretical and empirical work which converges to propose that these relational skills show a clear relationship with traditional metrics of intellectual performance, a number of investigations have analysed the utility of MET interventions designed to enhance relational responding proficiency in improving intellectual performance. Such investigations will be discussed in the next section.

1.4.5.1 SMART & Other MET Interventions.

Given the success of MET interventions in improving and establishing various forms of relational responding (Barnes-Holmes et al., 2001; Barnes-Holmes et al., 2004; Berens & Hayes, 2007; Dymond & Barnes, 1995; Gomez et al., 2007; Luciano et al., 2007) and the collection of correlational research proposing a close relationship between relational responding and intellectual performance (see Colbert et al., 2017 for a complete overview), it therefore stands to reason that, improving proficiency in relational responding may catalyse improvements in intellectual performance. Such an assertion of the potential malleability of intelligence represents a clear departure from more traditional, essentialist views of a stable intelligence (Ramsden et al., 2011; Spearman, 1904; Symonds & Spearman, 1928), but nonetheless reflects an allegiance to the foundational aims of intelligence testing (Binet, 1904a, 1904b; Nicolas et al., 2013; Siegler, 1992), and is supported by growing evidence proposing that intelligence can be improved (Aberg et al., 2009; Au et al., 2015; Buschkuehl & Jaeggi, 2010; Cohen, Amerine-Dickens, & Smith, 2006; Eikeseth et al., 2002a; Eikeseth, Smith, Jahr, & Eldevik, 2007; Jaeggi et al., 2008, 2010; Lovaas, 1987; Remington et al., 2007). In light of an extensive research base reporting positive correlations between intellectual ability and a number of socially-desirable variables, including educational attainment (Bourneville, 1895; Deary, Strand, Smith, & Fernandes, 2007; Jensen, 1998; Laidra, Pullmann, & Allik, 2007; Neisser et al., 1996; Roth et al., 2015), job performance (Gottfredson, 2003; Ones, Viswesvaran, & Dilchert, 2005; Schmidt & Hunter, 2004), income (Jencks, 1972; Lynn, 2010; Meisenberg, 2012) and self-reported happiness (Ali et al., 2013), the possibility of training relational responding proficiency as a means of improving intellectual performance therefore represents an extremely exciting avenue for research; one which may harbour significant implications for not only behaviour analysis and intelligence, but for society more generally.

One of the seminal studies in this research stream, Cassidy et al. (2011), investigated the utility of multiple exemplar relational training as a means of ameliorating intellectual performance across two studies. In the first of their two studies, four normally-developing children (mean age: 10 years 3 months) completed a five-phase systematic training regimen to improve relational responding proficiency. These five training phases comprised: (1) stimulus equivalence testing and training, (2) multiple exemplar training to establish coordination relational responding, (3) multiple exemplar training to establish opposition relational responding, (4) multiple exemplar training to establish the comparison relational responding. In addition to the four experimental participants, a further four control participants were required who completed only the first of these five training phases (i.e. stimulus equivalence training).

The first phase comprised two administrations of a computerised standard one-to-many matching-to-sample procedure in which conditional stimulus relations were trained between 6 nonsense syllables (A1 $\rightarrow$ B1 (not B2), A1 $\rightarrow$ C1 (not C2), A2 $\rightarrow$ B2 (not B1), A2 → C2 (not C1)), each followed by a testing stage. The 16-item matching-to-sample procedure therefore included four unique trials specifying relations between 6 stimuli, each of which was repeated four times. Corrective feedback was provided onscreen for all trials. Upon successful completion of this training procedure, participants were administered the first of two testing protocols. The first testing stage probed for the symmetrical relations (i.e. mutually entailed relations) specified in the training stage (e.g. B1 → A1, C1 → A1). Upon successful completion of this testing stage, the same matching-to-sample procedure was repeated, followed in this instance by a test for transitive relations (e.g. the combinatorically entailed relations, such as B1 → C1, derived by each relatum's common relation to A1). In order to progress onto each successive stage of this training phase, 100% correct responding was required in all cases.

The second phase involved multiple exemplar training and testing for stimulus equivalence (i.e. symmetry and transitivity). During this phase, corrective feedback was provided and then withdrawn on alternate training administrations until participants demonstrated symmetry and transitivity without the need for this feedback (i.e. performance became generalised). Five additional stimulus sets were integrated, with every participant being exposed to these novel sets irrespective of when stimulus equivalence performance generalised.

The third phase comprised of relational pre-training and multiple exemplar training for coordination relational responding. During this phase, participants were first exposed to a conditional discrimination pre-training task, contingent on physical properties of one sample stimulus and three comparison stimuli. In each pre-training trial, participants were exposed to a sample stimulus (e.g. a short horizontal line), preceded by a contextual cue (same or opposite), and succeeded by three comparison stimuli (e.g. a longest line, a longer line, and an identical line). When presented with the "same" contextual cue, participants were required to select the matching comparison stimulus. When presented with the "opposite" contextual cue, participants must pick the comparison stimulus that was most different to the sample stimulus (in the example above, the longest line). Corrective feedback was present for all trials. This stage consisted of four stimulus sets, repeated four times each.

In order to continue on to multiple exemplar training, 100% correct responding on this pre-training stage and a pre-training test stage (in which feedback was withdrawn) was required. Multiple exemplar training for this phase involved the training of three interrelated two-stimulus arbitrary coordination relations, which facilitated the emergence of a four-member relational network (i.e. A same as B, B same as C, C same as D). To establish this relational network, participants were exposed to a single relational statement (e.g. A same as B), and required to select either a "Yes" or "No" response onscreen. Corrective feedback was provided for this response, which, following multiple trials trained correct responding to each relational statement. In order to avoid direct control of the contextual cue over responding (i.e. establishing that the presence of "same" requires a specific invariant response), a fourth relational statement was included as part of training for which the correct answer was "no". In order to pass this stage, 100% correct responding across 20 trials was necessary. Once again, the testing stage for MET training mimicked the procedure, stimuli and passing criterion of the testing level, but withdrew corrective feedback.

The fourth training phase replicated the MET protocol of the previous phase by training a further relational network composed of three two-stimulus arbitrary opposition relations (i.e. A opposite B, B opposite C, C opposite D). The five training phase also replicated the relational pre-training and MET protocol employed in the previous two stages, but replaced the previously-used contextual cues of 'same' and 'opposite' with 'more than' and 'less than'. For this stage however, the passing criterion was 100% correct responding across 30 trials (five exposures to each of the three relational statements for both ‘more than’ and ‘less than’ contextual cues).

Results indicated a significant effect of training on all three WISC IQ indices (Full Scale, Verbal and Performance IQ). Mean score increases for the training group for Full Scale (27.3 points), Verbal (17.8 points) and Performance IQ (32.5 points) showed marked contrast to score changes displayed by the control group (-2.25, .25 & -4 respectively). Furthermore, the extent of IQ changes is further underlined by the finding that for each of the 3 main IQ indices for the 4 experimental participants, 11 of these 12 indices increased by at least 2 standard deviations.

In the second of Cassidy et al. (2011) studies, the pre-training and MET protocol for coordination, opposition and comparison relations virtually identical to that implemented in the first of their studies was administered to a sample of 8 children experiencing educational difficulties. The only modification to this protocol was the addition of two further control trials and an additional 10 trials total in the ‘same’ training phase. In addition, a remedial training program was devised for those who failed to reach criterion on coordination training following seven cycles of training. This program was identical to the main training protocol, but replaced the nonsense word stimuli used in relational statements for non-arbitrary stimuli (e.g. lines, circles and boxes). In order to provide an accurate metric of relational responding proficiency for this sample, the Relational Abilities Index (RAI) was developed which consisted of 20 coordination, 20 opposition and 20 comparison trials. For each trial, a relational statement consisting of a two nonsense stimuli separated by a contextual cue in accordance to the relational frames of ‘same’, ‘opposite’ , ‘more than’ and ‘less than’ (e.g. ‘TUF same as FEG’) was presented to the participant, followed by a relational question probing for the relations specified (e.g. ‘is FEG same as TUF?’). No nonsense stimuli appeared more than once during assessment, and feedback was not provided for any trials. Following relational skills training, it was found that RAI scores increased significantly from a mean of 58.5% correct responding at baseline to 92.4% at follow up. In addition, there was a significant improvement in mean scores for WISC-IV Full-Scale IQ (13.1 points), as well as for three of four IQ subscales: Verbal Comprehension (10.2 points), Perceptual Reasoning (12.4 points) and Processing Speed (16.2 points).

Cassidy et al. (2016) conducted two further investigations into the efficacy of relational skills training in improving intellectual performance by conducting the first formal analysis of the SMART program (Strengthening Mental Abilities with Relational Training). The SMART program was based on the Cassidy et al. (2011) MET training protocol, and consists of 55 training levels designed to improve relational responding in accordance to the frames of coordination, opposition and comparison. As such, the program is divided into two blocks, with the first block training Same/Opposite relational responding (29 trials) and the second training More/Less relational responding (26 levels). Each training level comprised of 16 trials devised to isolate and increase fluency in a specific form of relational responding (e.g. combinatorially-entailed relations first term-third term relations). All trials involved the presentation of a number of relational premises using nonsense stimuli (e.g. 'ZIG is same as DEG') followed by a relational question based on the premise(s) specified. All stimuli consisted of nonsense words in the form of consonant-vowel-consonant. No nonsense item was repeated during training or testing, and a time limit of 30 seconds was imposed for all trials. Each SMART level involved a training and a testing stage, in which 100% correct responding was required to move on to successive levels. All responses were followed by corrective feedback during the training phase, but not the testing phase. If participants failed to meet criterion for a given training or testing stage, they would be re-administered this stage until completion.

As participants progress through each training block, task difficulty increases, as the complexity of the relational skills being trained increases. Task complexity was therefore controlled by modifying; 1) the number of relational premises (1-3); 2) the order of relational premises (sequential or random); 3) the directionality of the relational question (i.e., whether or not the relational question probes for first term-last term relations, or last term-first term relations as specified in the premises); 4) the number of relation types presented in each trial (e.g., only "same" relations, or a combination of "same" and "opposite"); and 5) the presence/absence of the relational cue used in the question in the relational premise(s), (e.g., CUG is same as LER, is CUG same as LER?"). In line with the development of the SMART program, the RAI assessment was modified in order to match the training program's 55 level structure. As such, the RAI included 55 trials, each of which reflected one exemplar of the type of relational responding targeted by each of the SMART program's training levels.

The first of the Cassidy et al. (2016) studies administered the SMART program to an entire class cohort of students attending primary school (n = 15, mean age: 11 years, 1 month) over the course of approximately 3 months. Post-intervention analyses of RAI scores indicated a significant rise from baseline (*M* = 33.8 out of 55, *SD* = 8.11) to follow-up (*M* = 48.5, *SD* = 5.54), supporting the effectiveness of the SMART program in improving relational responding proficiency. Furthermore, a statistically significant rise of 23 points was found for mean WISC-IV Full Scale IQ scores.

In their second study, Cassidy et al. investigated the benefit that SMART training may provide for performance on a widely-used assessment of scholastic ability, the Differential Aptitude Test (DAT) in a sample of secondary school students (n = 33, mean age: 16 years, 4 months). Results found significant increases in all three DAT indices: Verbal Reasoning, Numerical Reasoning and Educational Aptitude composite scores following relational training. In addition, RAI scores rose significantly from 44.7

to 51.3 out of 55. An additional correlational analysis indicated that baseline RAI scores predicted scores for all three DAT indices.

Hayes & Stewart (2016) carried out a similar analysis of the SMART program, in conjunction with the Non-arbitrary Same & Different Relational Evaluation Procedure (NSD-REP) using an extensive battery of widely-administered intellectual and scholastic ability assessments in a sample of primary school children aged between 10 and 11 years old (n = 28). SMART was administered in 1-hour biweekly sessions over the course of 15 weeks, while an ability-matched control group completed a computer-coding training program called Scratch. The NSD-REP involved three training levels in which non-arbitrary same/different relations are trained using identical and non-identical visual stimuli. In all levels, participants are presented with two images, and required to respond to a relational question (Level 1 & 2: 'Are these the same/different?; Level 3: 'Are these not the same/different?'). In the first level, participants respond by selecting either the "Same" or "Different" buttons for Level 1 trials, and "Yes" or "No" for Levels 2 and 3. The testing battery implemented included the Relational Abilities Index, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2013), two Wechsler Intelligence Scale for Children subtests (WISC; Wechsler, 2003), three Wechsler Individual Achievement Test scales (WIAT; Wechsler, 2005), the Drumcondra Primary Reading Test Revised (DPRT-R; Educational Research Centre, 2007) and the Drumcondra Primary Mathematics Test Revised (DPMT -R; Educational Research Centre, 2006). Significant improvements were found for only the SMART group in scores for WASI Block Design, WISC Digit Span and Letter-Number Sequencing, all three WIAT indices and Drumcondra Primary Mathematics Test Revised. In addition, while the scores increase witnessed in the SMART group for the other measures did not reach statistical significance, the magnitude of these rises for all measures were greater than those displayed by the Scratch group. Finally, baseline RAI scores were found to display moderate significant relationships with DPMT-R (*r* = .69), DPRT-R (*r* = .59), WIAT Spelling (*r* = .52), WIAT Reading (*r* = .49), WIAT Numerical Operations (*r* = .4), WASI Full-Scale IQ (*r* = .44), WASI Block Design (*r* = .45) and WISC Letter-Number Sequencing (*r* = .41).

In an investigation of the impact of SMART training on mathematical reasoning and intellectual performance, Thirus et al. (2016) administered the training to a sample of 21 Swedish high school students aged between 16 and 18 years old. Training took place over the course of 8 to 10 weeks, with approximately half of the sample completing SMART training during this period. Results found that SMART training did not significantly improve performance on an assessment of mathematical ability based on the Swedish high-school curriculum. However, scores on the assessment of intellectual ability, Raven's Standard Matrices (Raven & Court, 2000) were found to increase significantly following relational training. In addition, analyses of variance indicated that RAI scores rose significantly for the experimental group, but not for the control group.

Amd & Roche (2018) found that completion of SMART resulted in significant increases in Raven's Progressive Matrices Score (RPM; Raven & Court, 2000) in a sample of 35 underprivileged children in Bangladesh. SMART Training was delivered over a period of 12 weeks in bi-weekly group sessions of 30 minutes each. Of note is Amd & Roche's finding of a clear 'dosage effect', as post-intervention increases in RPM scores were strongly predicted by number of training levels completed. Results show a significant difference in improvements in RPM score between high engagement (training levels competed > 24) and moderate engagement groups (training levels completed; 13-22), and between moderate engagement and low engagement groups (training levels < 7). In addition, it was found that baseline RAI score did not predict post-intervention score changes.

A recent study conducted by McLoughlin, Tyndall, & Pereira, (2018) investigated the impact of SMART training, alongside a supplementary module training analogical relational proficiency, on intellectual performance in a small sample of adults (n=8). The additional analogical relations module, termed SMARTA (Strengthening Mental Abilities with Relational Training: Analogy), retained the general protocol employed by the main SMART program, but required participants to derive the relationship between relations specified in two sets of relational premises. For example, participants are presented with four two-relata relational premises specifying a 5-member relational network (e.g. "NEP more than EFA. EFA more than FOP. FOP less than ENE. ENE less than ANJ") alongside a relational question probing for the relation between relations specified in each network (e.g. "Is ENE to FOP opposite to ANJ to ENE?"). As in the main program, participants recorded their response by selecting either 'Yes' or 'No' buttons onscreen. Participants completed a 48-item SMARTA assessment with no feedback provided, which identified the trial types which were responded to incorrectly. Participants were then instructed to train proficiency in such trials in 16-trial training and testing blocks (as in the main SMART program). Passing criterion was 16 correct responses in training and testing blocks. Participants were then re-administered the SMARTA, and unless participants recorded 100% correct responding, the training process was repeated.

Participants were tested with the Kaufmann Brief Intelligence Test four times, each spaced one week apart. No intervention was administered during the first testing interval (between test administrations 1 & 2). SMART and SMARTA training was administered in the second and third testing intervals respectively. Analyses of variance indicated no significant differential effect of training intervention on overall K-BIT scores, or any of its three subtests (Verbal Knowledge, Riddles & Matrices). T-tests indicated that there were significant increases in fluency scores for Verbal Knowledge and Riddles following SMART training, and KBIT-2 Fluency following SMART and SMARTA training. These results come with major caveats, as these score increases were due to significant reductions in response time, rather than any increase in response accuracy. As such, participants simply responded more quickly to subtest items, but accuracy did not significantly improve. Due to the intensive testing schedule (four K-BIT administrations within three weeks), the probability of witnessing practice effects, particularly in content-based subtests (i.e. Riddles and Verbal Knowledge) is increased significantly. This point is substantiated by significant K-BIT fluency score increases found for the control group following a period of no intervention, and the finding that only content-based subtests were found to rise significantly following intervention. As such, it may be the case that such short training periods are not sufficient to increase intellectual performance in any meaningful way.

In a single-case study, Vizcaíno-Torres et al., (2015) administered MET training for coordination, opposition, and comparison to a 4-year-old child for a total of 12 hours over the course of 5 and a half months. Adhering to guidelines proposed by Barnes-Holmes et al. (2004), Berens & Hayes (2007) and Luciano et al. (2007), the training was comprised of six phases, in which Phases 1,3 & 5 evaluated, trained and tested coordination, opposition and comparison respectively. For Phases 2, 4 & 6, fluency and flexibility was subsequently trained for the relational frames of coordination, coordination/opposition, and comparison/opposition, respectively. Results indicated an increase of 25 IQ points, as measured by the General Cognitive Index (GCI), a scale derived from scores on three of the four subscales (Verbal, Perceptual-Manipulative, & Numerical) included McCarthy's Aptitudes and Psychomotricity Scale's (MCSA, McCarthy, 1988).

In another single-case design, Ruiz, Suarez, & Lopes (2012) also reported a considerable increase in MCSA General Cognitive Index score of 35 points for a 4-

year-old participant diagnosed with autism, following bi-weekly MET intervention sessions to establish fluency and flexibility in fundamental relational skills over the course of 6 months.

Parra & Ruiz (2016) carried out a further investigation in which fluency in relational responding in accordance to the frame of coordination was trained in a single 4-year-old participant, with an age- and ability-matched control participant receiving no intervention. Phase 1 of the intervention trained visual-auditory coordination relations, and involved 16 trials of receptive naming (i.e., object-word coordination relations), in which the experimenter presented the participant with a target object (e.g. hairpin) and stated the name of the object, before asking the child to subsequently name the presented object. Finally, the target object was placed among an array of other objects, and the participant would be required to retrieve the target when prompted (i.e. when told "give me the hairpin"). Corrective feedback was provided for each trial during training, and the participant could only progress onto Phase 2 following responding correctly to at least 15 of 16 trials across two consecutive sets. Phase 2 trained visual-visual coordination relations, comprising of a many-to-one matching-to-sample procedure following training of four conditional discriminations between abstract shapes (B1→A1, C1→A1, B2→A2 and C2→A2). The aim of this phase was to establish the mutually entailed (e.g. A1→B1) and combinatorially-entailed (e.g. B1→C1) relations facilitated by the trained relational network. Mastery criterion for this phase was 7 correct responses out of 8 assessment trials for both mutual and combinatorial relations. Finally, phase 3 trained auditory-auditory coordination relations, in which two stimulus sets were used: eight short stories involving children and toys, and seventeen sets of synonyms. For the synonym set, the participant would be asked to reverse the relation being two synonyms after being told that they are the same (i.e. mutual entailment) and to derive to a relation of coordination between two synonyms due to their common relation to a third synonym. The short story set would establish coordination relations between sets of three actions (e.g. when teacher draws a square on blackboard (A) → children raise hands (B) , when children raise hands (B) → violin plays (C)), and to be able to demonstrate establishment of the mutually-entailed relations (e.g. "why did the children raise their hands?" B → A) and combinatorially-entailed relations (e.g. "what did the teacher do to make the violin sound?" A→C). Following intervention, the experimental participant displayed significant rises on all four subscales of McCarthy's Aptitudes and Psychomotricity Scale (Verbal, Perceptive-manipulative, Memory and Motor). In terms of overall cognitive ability, significant rises of 26 standardised points were found for the experimental participant, General Cognition Index, with a 10-point increase for the control participant.

Training interventions which target and train relational responding proficiency, therefore, show considerable promise as an efficacious means of improving intellectual performance. However, such promise is tempered by the relatively modest empirical evidence produced thus far, as much remains to be explored in order to substantiate the tentative support provided by the above investigations.

1.5 Other 'Brain Training' Interventions

While a number of training interventions designed to improve cognitive function have emerged from a behaviour-analytic paradigm, there has been a increase in interest in such training interventions more generally, with a wide variety of protocols being developed outside the field of behaviour analysis. Such protocols vary widely in the skills and proficiencies targeted, and have aimed to improve intelligence and cognitive ability through the use of video games (e.g., Ballesteros et al., 2015, 2014; Dye, Green, & Bavelier, 2009; Green & Bavelier, 2012; Toril, Reales, & Ballesteros, 2014; Toril, Reales, Mayas, & Ballesteros, 2016; Wang et al., 2016), chess (e.g., Aciego, Garcia, &

Betancort, 2012; Kazemi, Yektayar, & Abad, 2012; Sala & Gobet, 2016) and music instruction (e.g., Benz, Sellaro, Hommel, & Colzato, 2016; Bergman Nutley, Darki, & Klingberg, 2014; Franklin et al., 2008; Miendlarzewska & Trost, 2014). While a number of such interventions have displayed preliminary promise, it appears that claims of efficacy for many of these interventions seem premature when placed under more rigorous methodological and statistical scrutiny (Sala & Gobet, 2017; Haier, 2014; Melby-Lervag & Hulme, 2013; Owen et al., 2010; Simons et al., 2016).

Some of the most noteworthy research investigating cognitive training programs and their impact on intellectual performance has focussed on the implementation of various interventions designed to improve working memory (e.g. Jaeggi et al., 2008, 2010; Rudebeck, Bor, Ormond, O'Reilly, & Lee, 2012; Westerberg & Klingberg, 2007). Working memory (WM) is defined as the ability to maintain and engage with information over a short period of time (Heitz, Unsworth, & Engle, 2005), and has been proposed as a contributor to various cognitive functions, such as learning (Cain, Oakhill, & Bryant, 2004; Cowan & Alloway, 2008); reasoning (Kyllonen & Christal, 1990), arithmetic (Geary, Hoard, & Hamson, 1999; Swanson & Sachse-Lee, 2001), reading comprehension (Cain et al., 2004; Carretti, Borella, Cornoldi, & De Beni, 2009; Daneman & Carpenter, 2004), writing (McCutchen, 1996), scholastic aptitude (Alloway & Alloway, 2010) and fluid intelligence (Ackerman et al., 2005). Such is the impact of these interventions, Sala & Gobet (2017) have identified working memory training protocols as the most discussed and studied form of cognitive training.

In the seminal emanating from this research stream, Jaeggi et al. (2008) investigated the efficacy of dual *n*-back training in increasing fluid intelligence, defined as the ability to adapt and apply our cognitive resources to new problems and situations (Carpenter, Just, & Shell, 1990), following dual *n*-back working memory training. This training task presented participants with a series of eight squares being sequentially displayed onscreen for 500 milliseconds at eight different locations with each stimulus presentation spaced 2.5 seconds apart. In addition, the presentation of each square was paired with playback of an audio recording of one of eight consonants delivered through headphones. Participants were then required to respond when a stimulus (auditory or visual) was presented a specified number (*n*) of stimulus presentations previously. As such, in a 3-back trial, participants would be required to respond when an auditory or visual stimulus reappeared after 3 subsequent stimulus presentations. The value of *n* was then incrementally increased following successful completion of a training block. Training blocks consisted of six visual and six auditory stimulus presentations, with four single modality presentations and two dual-modality presentations. In this experiment, four groups were exposed to *n*-back working memory training across 8, 12, 17 and 19 days respectively. Each training group was matched with a control group that completed baseline and follow-up assessments in accordance to the same schedules. Jaeggi and colleagues report significant gains in fluid intelligence, as measured by matrix reasoning tests that increased linearly as a function of the amount of dual *n*-back training (i.e. a dosage effect).

Jaeggi et al. (2010) extended upon these findings by conducting a further investigation whereby the differential effects of single modality versus dual modality *n*-back on fluid intelligence was analysed. In this design, two matched groups conducted either the dual *n*-back protocol administered in Jaeggi et al. (2008) or a single *n*-back protocol in which only visuospatial stimuli were utilised over the course of 20 sessions spread over 4 weeks. Baseline and follow-up assessments consisted of a single *n*-back assessment, the automated operation span task (OSPAN; Unsworth, Heitz, Schrock, & Engle, 2005), the BOMAT and one half of the RAPM. A no-contact control group completed identical baseline and follow-up assessments but received no intervention. Analyses of near-transfer (operationalised by scores in a single *n*-back assessment) indicated that both training groups displayed significant score increases, with no significant differences in between-group post-training gains. In terms of far-transfer measures, all three groups displayed significant rises on BOMAT scores with only the two training groups recording significant post-intervention increases in RAPM performance. Across both training groups, participants scores improved by 1.76 trials (out of 19 trials) following training. Operation span scores for both training groups decreased slightly following intervention, but this decrease was not significant.

Rudebeck et al. (2012) also reported improvements in fluid intelligence following a similar, 20-day program of dual *n*-back training. Mean BOMAT scores for the experimental group increased from 7.6 to 9.5 out of a possible 29, representing significant, but relatively modest gains in fluid intelligence. Correlational analyses indicated that post-intervention BOMAT score changes were predicted by baseline BOMAT performance, but not training task improvement.

While such increases in fluid intelligence resulting from working memory training have received considerable interest (labelled "a landmark result", Sternberg, 2008), the validity and applicability of these findings have come under criticism due to issues regarding the methodological rigour and consistency of the experimental designs utilised (Melby-Lervag & Hulme, 2013; Moody, 2009; Redick et al., 2013; Sala & Gobet, 2017; Shipstead, Redick, & Engle, 2012). In a response to the Jaeggi et al. (2008), Moody (2009) highlights a number of procedural inconsistencies that may burden such results with major caveats, perhaps the most delimiting of these arising from variation in the composition of testing batteries administered across groups. While the 8-day training group were tested using a widely-administered and well-validated assessment of fluid intelligence, Raven's Advanced Progressive Matrices (RAPM; Raven & Court, 2000), the three other experimental groups completed the Bochumer

Matrices Test (BOMAT; Hossiep et al., 1999). Both the BOMAT and RAPM are assessments of visuospatial analogy, in which participants are required to select from an array a figure that matches or 'belongs' in a sample matrix of figures that allow the inference of a given 'rule'. However, a key difference between the testing batteries is that RAPM presents tasks in a 3x3 figure format, while the BOMAT involves trials in a 5x3 format. In light of such inconsistency, it is illuminating to note that significant improvements in performance were only found for the BOMAT-tested groups. As such, Moody (2009) suggests that the 15-figure trials presented during the BOMAT may place a greater load on working memory than would be expected for the 9-figure RAPM trials capacity. Therefore, the increases in fluid intelligence that BOMAT score increases are proposed to espouse may be due to specific improvements in working memory (Shipstead, Redick, et al., 2012), rather than far-transfer effects on fluid intelligence *per se*.

Moody (2008) also highlights the reduced time limit (from a recommended 45 minutes to 10 minutes) afforded to Jaeggi and colleagues' administration of the BOMAT in their 2008 study. Both replications discussed above (Jaeggi et al., 2010; Rudebeck et al., 2012) adopted this shortened time, due to "time restrictions and the possibility of ceiling effects" (Rudebeck et al., 2012, p. 3). However, the BOMAT is constructed as a Guttmann scale (as trials progressively increase in difficulty), and as such, in order to provide a sensitive measurement of performance, it is imperative that participants are given the opportunity to complete as many trials as they are capable of. As expressed by Moody, this truncated administration time precluded the possibility of participants reaching the later, more difficult trials, and thus served as a metric of speed of response rather than proficiency or ability. In fact, across the three interventions, no mean group score for correct BOMAT trials at pre- or post-intervention was above 50%. Additionally, one's ability to respond quickly to such early trials may not represent an appropriate means of "tapping" fluid intelligence due to the relative lack of complexity these early trials constitute.

However, a number of analyses on the relationship between WM tasks and fluid intelligence may at least partially assuage such concerns regarding these shortened test administrations. Such accounts (e.g. Salthouse, 2014; Salthouse & Pink, 2008; Unsworth & Engle, 2006) indicate that there may be a relatively constant correlation between WM and fluid intelligence from low to high levels of cognitive load and complexity. That being said, there still remains a lack of clarity regarding the rationale behind Jaeggi and colleague's use of such unconventional assessment procedures.

Of relevance to this debate is the recurrent finding that working memory interventions often report increases in fluid intelligence that are not accompanied by gains in near-transfer working memory measures (Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2008; Harrison et al., 2013). Such a lack of coherence between post-intervention near- and proposed far-transfer would appear to preclude the possibility of any meaningful interpretation of results, may undermine the reliability of the assessments being implemented and potentially be caused by Type I error (Melby-Lervag & Hulme, 2013). Perhaps more fundamentally, such results, alongside concerns regarding the ability of WM interventions to produce reliable improvements in general intellectual function, may put in question the pertinence of improving WM as means of achieving this end. While various correlational analyses have indicated that the strength of the relationship between working memory and fluid intelligence lies somewhere between .35 and .65 (Ackerman et al., 2005; Kane, Hambrick, & Conway, 2005; Oberauer, Wilhelm, Schulze, & Süß, 2005), the exact nature of this relationship is yet to be fully explicated (Ackerman et al., 2005; Salthouse & Pink, 2008), therefore leaving the precise mechanism of any possible post-intervention score changes unelucidated.

Indeed, the relevance of this point extends to the wider gamut of 'cognitive enhancement' interventions, as a common criticism levelled against such protocols is that they fail to adequately explain the means by which a given training program is proposed to facilitate improvement in intellectual function (Schubert, Strobach, & Karbach, 2014; Simons et al., 2016). Many accounts appear to rely heavily on correlations between the skills being trained and the outcome variable, with little attention paid to discussing how improving the former may catalyse change in the latter. Indeed, as Deary (2012) outlines, one can expect a reasonable degree of correlation between any two cognitive tasks. Thus, while the finding of a significant correlation may potentially indicate a functional relationship, it does not necessarily entail that training one repertoire produces improvements in the other (see Ball, Edwards, Ross, & McGwin, 2010; Willis et al., 2006). Melby-Lervag and Hulme (2013) propose that despite the expressed promise of cognitive training systems, the exact mechanisms by which many such programs may actually exert their beneficial effects on intelligence/cognitive function is yet to be fully explained. The authors infer that such programs, in the absence of any detailed task analysis or theoretical accounts, implement training in order to repeatedly "load" a limited cognitive resource with a view of eventually increasing its capacity. The authors argue that this "physical-energetic" model is analogous to "strengthening a muscle by repeated use" (p. 273). It is essential, therefore, that such 'brain training' programs fully delineate the means by which these post-intervention outcomes are produced by training a given repertoire.

The importance of this failure to explain the mechanisms of training effect on intellectual performance is thrown into sharp relief by perhaps the most prevalent criticism of cognitive training programs: a scarcity of empirical evidence reporting clear far-transfer effects (Dahlin, Nyberg, Bäckman, & Neely, 2008; Harrison et al., 2013; Holmes, Gathercole, & Dunning, 2009; Holmes et al., 2010; Moreau, McNamara, &

Hambrick, 2018; Redick et al., 2013; Simons et al., 2016; Thompson et al., 2013; Van der Molen, Van Luit, Van der Molen, Klugkist, & Jongmans, 2010; Waris, Soveri, & Laine, 2015). As expressed by Simons et al. (2006), the general consensus appears to be that:

> We find extensive evidence that brain-training interventions improve performance on the trained tasks, less evidence that such interventions improve performance on closely related tasks, and little evidence that training enhances performance on distantly related tasks or that training improves everyday cognitive performance. (p.1)

For instance, in one of the few such analyses to administer a comprehensive IQ assessment following working memory training, Holmes, Gathercole, & Dunning (2009) reported clear evidence of improvements in working memory performance, but these gains did not result in significant increases in WASI Verbal or Performance IQ scores (results for Full Scale IQ were not reported). In the case of far-transfer effects being reported, Melby-Lervag and Hulme’s analysis (2013) shows that transfer effects tend to either be restricted to uncontrolled designs or are rendered non-significant a number of months after training has been completed. Without a clear demonstration of a given program’s effectiveness in producing genuine, consistent improvements in daily functioning and general intellectual/cognitive performance, many of the cognitive training interventions fail to substantiate preliminary evidence of their potential efficacy. While this may represent a somewhat pessimistic view of cognitive training programs, a considerable collection of meta-analyses and replication studies (e.g, Chooi & Thompson, 2012; Melby-Lervag, Hulme, Melby-Lervåg, & Hulme, 2013; Miendlarzewska & Trost, 2014; Moody, 2009; Moreau, Macnamara, & Hambrick, 2019; Owen et al., 2010; Redick et al., 2013; Sala & Gobet, 2017; Shipstead, Hicks, & Engle, 2012; Shipstead, Redick, et al., 2012; Simons et al., 2016; Soveri, Antfolk, Karlsson, Salo, & Laine, 2017a; Waris et al., 2015) appear to burden claims of such programs' efficacy with serious caveats at best, and call for further experimental evidence in support of arguments towards their benefits for general intelligence and cognitive performance.

1.6 The Current Thesis

The current thesis aims to extend upon emerging evidence proposing the close relationship between relational responding and intelligence (e.g. Colbert et al., 2017; O'Hora et al., 2005, 2008) as well as further investigating the preliminary utility reported for relational skills intervention in improving intellectual function (e.g. Cassidy et al., 2016; Cassidy, Roche, & Hayes, 2011; Hayes & Stewart, 2016). In regards to the latter pursuit, a series of investigations are planned which will further explore the utility of the SMART program, as implemented by Cassidy et al. (2016), as it has thus far demonstrated considerable promise as an efficacious means of producing demonstrable improvements in intellectual performance. The first experiment will essentially serve as a large-scale manipulation check of the effectiveness of this program in improving proficiency in the skillset it targets, namely AARR in accordance with the frames of coordination, opposition and comparison, in a large sample of 12- to 14-year-old Irish secondary school students. By either confirming or disconfirming this efficacy, Experiment 1 will serve as the basis for subsequent studies which will endeavour to ascertain the potential effect of relational skills training on intellectual performance and academic aptitude.

Following this initial analysis, subsequent investigations will aim to identify whether relational skills training can produce improvements in intellectual performance, as assessed by gold-standard IQ measures, and academic performance measures across a variety of age and ability levels. In addition, such experiments will aim to improve upon previous studies by increasing the level of methodological rigour applied in analysing the effect of relational skills training. Experiment 2 will consist of a single-blind randomised control trial of SMART training in a sample of Irish secondary school students aged between 15 and 17. Experiment 3 will comprise the first analysis of the impact of SMART on a sample of normally-developing pre-adolescent children using a crossover design, as a means of investigating whether SMART is appropriate and accessible for younger children and those with lower levels of pre-intervention intellectual ability and relational skill. Experiment 4 will build upon the previous experiment by administering a purposely-developed relational skills training intervention designed to establish the prerequisite relational skills required to complete the main SMART program (i.e. AARR) in a sample of young children presenting with additional educational needs and below-average levels of intellectual ability. Experiment 5 will then seek to study the potential effect of relational skills training on academic aptitude, as measured by the Department of Education's assessment of choice (the Drumcondra Reasoning Test) in a sample of Irish secondary school students.

The final two experiments of the current thesis will be correlational in nature, and will endeavour to further elucidate the relationship between intellectual performance and proficiency in responding in accordance with the various relational frames. It is hoped that such an analysis will not only lead to greater insights regarding the nature of this relationship, but will also identify which frames may be most pertinent to intellectual function, thereby potentially allowing the development of more accurate relational responding metrics. By conducting such an analysis, it may be possible to modify the Relational Abilities Index to include additional frames as a means of extending its validity and utility as a functional alternative to traditional IQ assessments. Furthermore, this analysis may also highlight which relational frames may be integrated into current relational skills training interventions in order to potentially catalyse even greater benefits for intellectual performance. To this end, Experiment 6 will assess the level of correlation between performance on a gold-standard measure of IQ (WASI) and the Multiple Relations Assessment Procedure (MRAT), a measure of coordination, distinction, opposition, temporal, analogical and deictic relational responding. Experiment 7 will build upon these findings by analysing the relationship between intellectual performance and academic aptitude with another extended assessment of relational responding, the RAI+, which assesses proficiency in coordination, opposition, distinction, temporal and analogical relational responding.

Chapter 2:

Assessing the efficacy of the SMART program in improving relational responding proficiency

2.1 Introduction

Due to the extensive research literature proposing the relevance of relational responding to intellectual performance (Cassidy et al., 2010; Colbert et al., 2017; O'Hora et al., 2005; O'Hora et al., 2008; Roche et al., 2013), as well as empirical support for the malleability of this form of responding (Barnes-Holmes et al., 2001; Barnes-Holmes, Barnes-Holmes, Smeets, Strand, & Friman, 2004; Berens & Hayes, 2007; Leader & Barnes-Holmes, 2001; Luciano et al., 2007) a number of relational skills training interventions have been developed to improve intellectual performance, such as the SMART program (Cassidy et al., 2016; Cassidy, Roche, & Hayes, 2011). The SMART program has thus far received considerable support in a number of small-scale and pilot studies (Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; Luciano et al., 2007; Vizcaíno-Torres et al., 2015). The reliability of this SMART effect, however, is yet to be established in a large-scale investigation which captures a broad spectrum of intellectual ability levels.

As one of the primary objectives of the current thesis is to conduct an analysis of the efficacy of relational skills training in improving intellectual and academic performance, it is important to first establish if, and by how much, relational responding proficiency itself (rather than broader intellectual abilities) can be improved via intervention. Furthermore, by assessing the distribution of pre- and post-intervention scores on a measure of relational responding, insights can be gleaned regarding the nature and magnitude of skill changes, as well as the potential impact of variations in baseline ability levels and amount of training completed on these skill improvements. As such, the current study aims to investigate the efficacy of the SMART program in improving relational responding proficiency in accordance with the frames of coordination, opposition and comparison. While there are a number of alternative relational skills training interventions that could have been selected for the purpose of the current analysis, the SMART program is the most widely reported and will be used in subsequent studies. As such, this will essentially serve as an in-depth manipulation check of the SMART program in improving accuracy and fluency in the relational skills it targets and trains. In addition, the current analysis will be the first to compute a Full-Scale IQ estimate derived from RAI scores via a linear regression equation obtained from previous correlational analysis of performance on each of these scales. By estimating Full-Scale IQ scores, and more importantly, changes in scores post-intervention, it may be possible to give an approximation of the impact of the SMART program in improving intellectual performance as well as the influence of baseline intellectual ability on training progress and outcomes.

To this end, the current study comprises of an analysis of pre- and post-intervention performance on a measure of relational responding, the Relational Ability Index (RAI), in a sample of Irish secondary school students following the administration of the SMART program over a period of 3-4 months. Post-intervention RAI score changes will function as the primary focus of this analysis, but the influence of baseline levels of ability and number of training levels completed will also be investigated.

2.2 Method

2.2.1 Participants

The current sample consisted of the entire cohort of 1st year students (n =168) attending a secondary school in Sligo, Ireland. Mean age for this sample was 13 years and 6 months (*SD* = 7.8 months). All students completed baseline and follow-up RAI assessments at the beginning (September) and end (May) of one school year, following 3-4 months of relational training in the interim.

2.2.2 Settings and Materials

All RAI assessments and SMART sessions took place in a school-based computer lab in Summerhill College, Sligo which had the capacity for approximately 30 children, who completed relational skills training under supervision by a member of school staff. As such, the current sample was divided into a series of groups who trained individually, each of which received the same number of training sessions.

2.2.2.1 Relational Abilities Index.

The Relational Abilities Index is an online assessment of relational responding proficiency. The RAI assessment used here was precisely as employed by Cassidy et al. (2016) and Colbert et al. (2017) via the website RaiseYourIQ.com, The RAI consists of 55 syllogistic relational network problems which involve the presentation of 1-3 relational premises involving three-letter nonsense syllables (e.g. CUG is more than BEF), followed by a relational question (e.g. is BEF more than CUG?) which requires the participant to click either a ‘Yes’ or ‘No’ onscreen button in response. The location of these binary response options was alternated randomly throughout the assessment, in order to control for positional responding. As trial stimuli consisted of 248 randomly-generated nonsense syllables, no stimulus appeared more than once during the RAI assessment. All nonsense syllables took the form consonant-vowel-consonant to ensure they were readily pronounceable. A 30-second time limit was imposed for all trials, and failure to respond within this window was treated as an incorrect response.

The RAI assesses ability in responding to coordination/opposition (same/opposite) and comparison (more than/less than) relational frames. As such, Block 1 consists of the first 29 RAI trials, all of which assessed responding in accordance to coordination and/or opposition relational frames, while Block 2 comprises 26 more than/less than tasks. Task difficulty increased progressively throughout each block, with trial difficulty controlled by modifying; 1) the number of sample relational statements presented (1-3); 2) the order in which these statements are presented (i.e., in a sequential or random order); 3) the directionality of the relational question (i.e., probing for first term-last term relations, or last term-first term relations as specified in the premises); 4) whether or not the relational statement or question utilised more than one relation type (e.g., inclusion of only "same" relations or a combination of "same" and "opposite" relations); 5) whether or not the relational term presented in the question was present in any of the premises (e.g., "BEF is more than TIF. Is BEF more than TIF?").

The RAI computes a number of metrics, such as total correct responses (RAI score) and time taken to complete the assessment. These variables allow the computation of an RAI Fluency score, which takes both speed and accuracy of responding into account, computed as follows:

$$60000\left(\frac{RAI\ score - no.of\ incorrect\ responses}{Time\ to\ complete\ RAI}\right)$$

This formula weighs the number of correct responses (represented by RAI score) against incorrect responses, essentially giving a correct-incorrect differential score per minute. This is designed to penalise random responding which can be expected to produce a 50% correct response in a very short period of time. In addition to these various metrics, RAI scores allow the computation of a Full-Scale IQ estimate, based on the linear regression equation computed by analysing IQ and RAI score distributions collected as part of previously published investigations (Cassidy et al., 2011; Colbert et al., 2017; Hayes & Stewart, 2016). This equation is displayed below:

$$Estimated\ IQ = RAI\ score(1.14) + 60.192$$

As the estimation of IQ scores in this way assumes a normal distribution of RAI scores, the nature of the sample used (i.e. an entire year cohort of secondary school students) should provide a dataset suitable to this end.

2.2.2.2 Relational Skills Training Protocol.

The relational skills training protocol (SMART) mirrored the structure of the RAI assessment insofar that for each of the 55 RAI assessment trial types, there was a corresponding training level that isolated, expanded upon and trained responding of that type. As such, the types of relational responding proficiency required in order to respond correctly to each of the 55 RAI test items were individually targeted and trained across multiple exemplars in a training stage of potentially infinite length. In line with the RAI format, the relational training protocol consisted of 29 Same/Opposite and 26 More than/Less than training levels. Each training level comprised training and test phases of potentially infinite length, that would repeat until success criteria were achieved. During the training phase, participants were required to respond to syllogistic relational puzzles in exactly the same format as the RAI. In order to complete this phase and move on to the testing phase, participants must produce 16 consecutive correct responses. If an incorrect response is made, the participant must begin again until this passing criterion is met. During the testing phase, participants are exposed to 16 relational tasks of the type trained in the previous phase. 100% correct responding is required to move onto the next training level. If this is not achieved, the participant begins the training phase once again, with this cycle repeating until passing criteria for both phases are met. Sample trials from the SMART program/Relational Abilities Index are displayed in Figure 2.1. For a complete description of the relational training protocol employed in this study, see Cassidy et al. (2016).

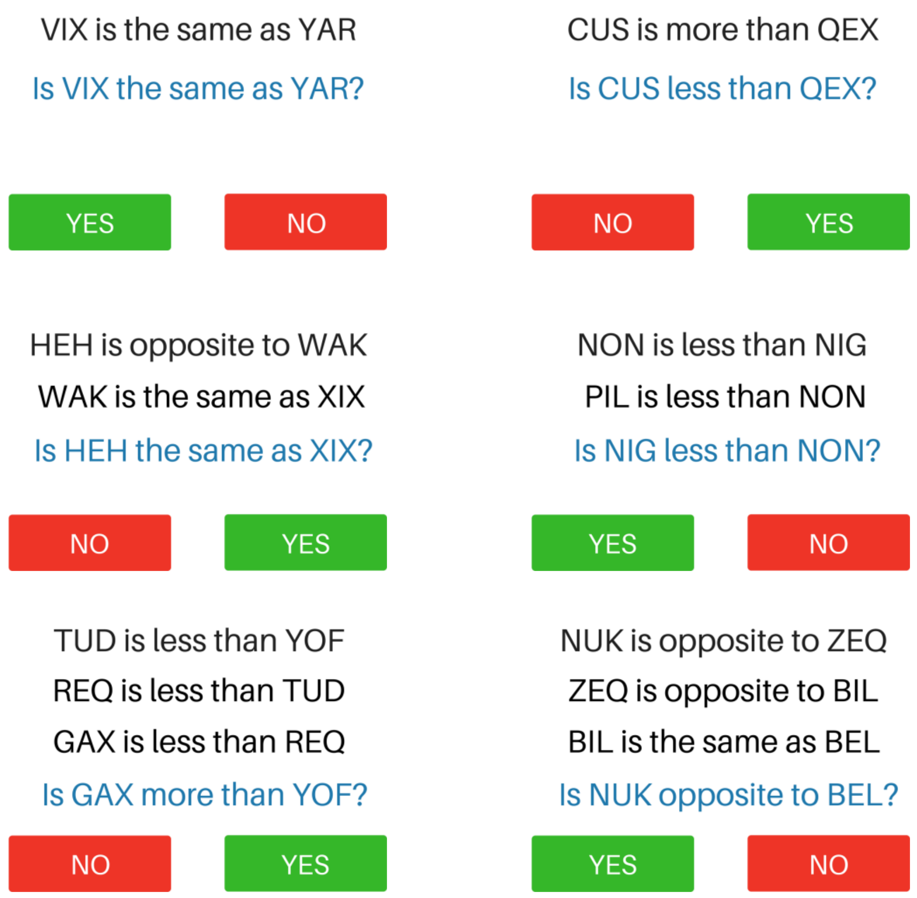


Figure 2.1. Sample trials from the SMART program/Relational Abilities Index.

2.2.3 General Procedure

Participation began with the administration of the RAI for all participants, delivered online via personal computer. Following the completion of this assessment, relational skills training was then administered within school hours in biweekly sessions of approximately 45 minutes across a 12-week period. As the training program permitted students to complete a maximum of 5 new levels each day, the minimum number of sessions required to complete all 55 levels was 11. Once these 55 levels were complete, a follow-up RAI assessment was administered immediately. Due to capacity constraints, only 30 students could complete their assessments/training at any one time, and therefore, the sample was divided into 6 groups, all of whom received the same amount of training.

2.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland**.** Parental consent and student assent were obtained for all participants.

2.3 Results and Discussion

2.3.1 Descriptive Statistics

Table 2.1 displays mean baseline descriptive statistics for RAI scores, RAI Fluency scores, RAI completion time and estimated Full-Scale IQ. Mean RAI scores were 38.8 out of 55 (*SD* = 7.58) at baseline. Mean scores for RAI Fluency was 2.21 (*SD* = 1.88). On average, it took participants 10.4 minutes to complete the RAI assessment (*SD* = 4.5 minutes). Mean Full-Scale IQ (FSIQ), estimated from participants' RAI scores, was in the average range (*M* = 102.7, *SD* = 8.6). Kolmogorov-Smirnov tests of normality indicated that baseline scores for RAI were normally distributed, while scores for RAI Fluency, time to complete training and Full-Scale IQ estimate were not normally distributed.

Table 2.1

Mean Descriptive Statistics at baseline and follow-up

	Baseline		Follow-up	
Variable	M	SD	M	SD
RAI	38.8	7.58	46.8	7.37
RAI Completion time (mins)	10.4	4.52	9.3	2.92
RAI Fluency	2.21	1.88	4.35	2.37
Full Scale IQ estimate	102.7	8.6	113.5	8.4

In total, 134 of 169 participants (79.3%) completed all 55 training levels. On average, participants completed 50.4 training levels (*SD* = 10.6). An independent samples t-test indicated that those who completed training had significantly higher baseline RAI scores (*M* = 38.9, *SD* = 7.18) compared to those that did not (*M* = 31, *SD* = 5.67), $t(166) = -6$, $p < .001$. Figure 2.2 displays the distribution of RAI scores before and after relational skills training.

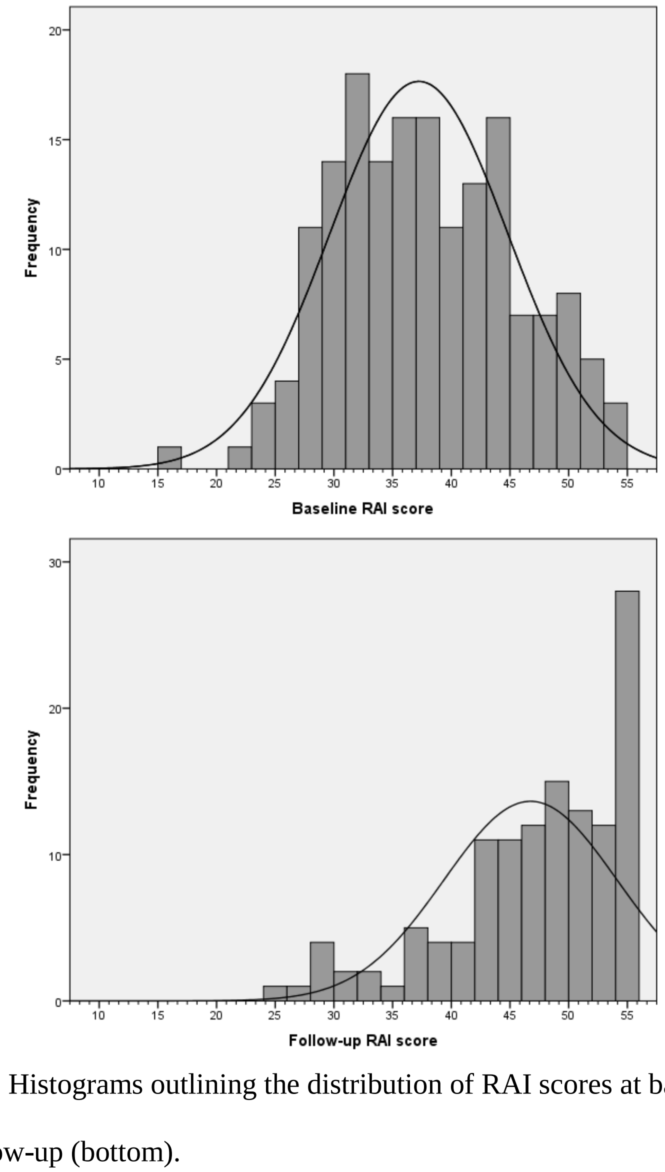


Figure 2.2. Histograms outlining the distribution of RAI scores at baseline (top) and at follow-up (bottom).

2.3.2 Analysis of post-intervention changes in RAI and IQ estimate scores

In order to investigate changes in scores for RAI following relational training, a paired samples t-test was conducted. To assess changes in RAI fluency scores, time to complete RAI assessment and FSIQ estimate scores, a series of Wilcoxon Signed Rank Tests were conducted. In line with Bonferroni procedures, an alpha level of .01 was set for these analyses. As the post-intervention RAI is administered only after completing training, participants who did not complete all 55 levels (n = 33) are excluded from this analysis.

RAI scores increased significantly following relational training, rising almost 8 points from Time 1 (M = 38.8, SD = 7.48) to Time 2 (M = 46.8, SD = 7.37), $t(125) = -13.66$, $p < .001$. The Cohen's d effect size was very large at 1.22. In terms of RAI Fluency scores, there was a significant increase in scores from Time 1 (M = 2.21, SD = 1.88) to Time 2 (M = 4.35, SD = 2.37), $z = -8.262$, $p < .001$, with a medium effect size (r = .45). RAI completion time decreased significantly by just over one minute from Time 1 (M = 10.4, SD = 4.52) to Time 2 (M = 9.3, SD = 2.92), $z = -3.03$, $p = .002$, with a small effect size (d = .17). Finally, there was a significant rise in estimate FSIQ scores from Time 1 (M = 102.6, SD = 8.6) to Time 2 (M = 113.5, SD = 8.4), $z = -8.88$, $p < .001$. The Cohen's d effect size statistic was very large at 1.3.

2.3.2.1 Effect of baseline ability on post-intervention score changes.

To further investigate the utility of the current intervention in improving relational responding proficiency, the current sample was separated into four subdivisions based on baseline estimated FSIQ score: Below Average (FSIQ < 90, n = 9), Low Average (FSIQ = 90-99, n = 57), High Average (FSIQ = 100-109, n = 66) and Above Average (FSIQ > 110, n = 36). Upon dividing the sample in this manner, it was found that only 3 participants in the Below Average group completed training. As such, this group was omitted from the following analysis. As the combined number of paired samples t-tests and Wilcoxon Signed Ranks tests was 9 in total, an alpha level of .004

was set, in line with Bonferroni procedures. Figure 2.3 displays mean score increases for RAI, RAI Fluency and IQ estimates scores for each ability grouping.

For the low average group, analyses of normality indicated that scores for RAI score, RAI Fluency score and IQ estimates were not normally distributed. RAI scores rose significantly by 5.7 points from baseline (M = 36, SD = 2.15) to follow-up (M = 41.7, SD = 7.45), z = -4.8, p < .001, with a large effect size (d = 1.04). RAI Fluency scores did not increase significantly from baseline (M = 1.6, SD = 1.8) to follow-up (M = 2.83, SD = 6.67), z = -2.97, p = .003, with a small effect size (d = .25). Full Scale IQ estimate scores rose 12.6 points from baseline (M = 95.1, SD = 2.38) to follow-up (M = 107.7, SD = 8.5), z = -4.8, p < .001, with a very large effect size (d = 2.02).

For the high average group, analyses of normality indicated that RAI, RAI Fluency and IQ estimate scores were not normally-distributed. RAI scores rose by 7.8 points from baseline (M = 38.9, SD = 2.69) to follow-up (M = 46.7, SD = 6.2), z = -5.76, p < .001, with a large effect size (d = 1.6). RAI Fluency scores increased significantly from baseline (M = 2.27, SD = 1.04) to follow-up (M = 4.14, SD = 1.57), z = -5.75, p < .001, with a very large effect size (d = 1.4). Full Scale IQ estimate scores increased by 8.8 points from baseline (M = 104.6, SD = 3.06) to follow-up (M = 113.4, SD = 7.09), z = 5.76, p < .001. The Cohen's d effect size for this increase was very large at 1.61.

For the above average group, analyses of normality indicated that scores for RAI, RAI Fluency and IQ estimate scores were normally distributed. RAI score rose significantly by 4.6 points from baseline (M = 48.1, SD = 3) to follow-up (M = 52.7, SD = 2.49), t(32) = -9.43, p <.001. The Cohen's d effect size for this increase was very large at 1.67. RAI Fluency scores increased significantly from baseline (M =3.32, SD = .9) to follow-up (M = 6.28, SD = 2.11), t(32) = -11.14, p < .001. The Cohen's d effect size for this increase was very large at 1.82. Full Scale IQ estimate scores increased by

5.2 points from baseline (*M* = 115.1, *SD* = 3.43) to follow-up (*M* = 120.3, *SD* = 2.84), *t*(32) = -9.43, *p* < .001, with a large effect size (*d* = 1.65).

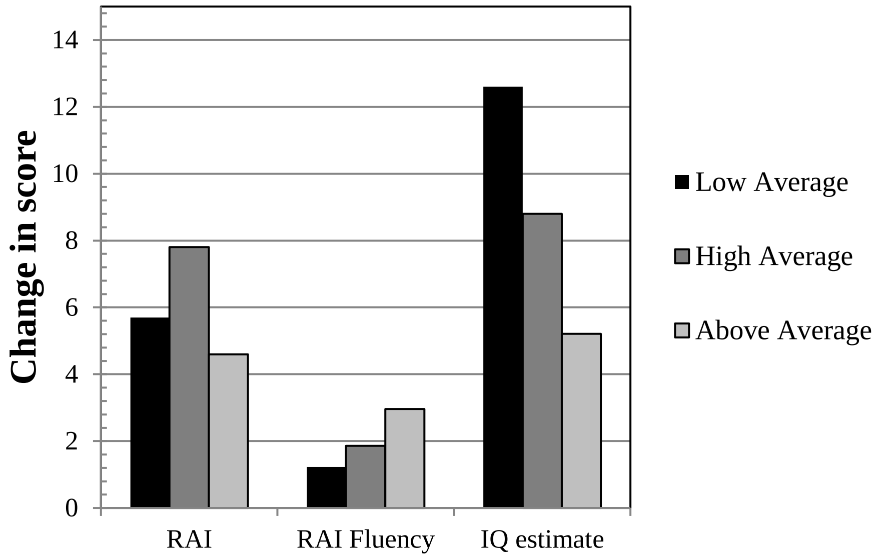


Figure 2.3. Bar charts displaying mean score changes for RAI scores, RAI fluency scores and IQ estimates scores across the three ability levels.

2.3.3 Correlational Analyses

Baseline RAI scores were found to display a medium-strength positive correlation with RAI Fluency (*r* = .63, *p* < .001) and a strong positive correlation with time taken to complete the RAI assessment (*r* =.62, *p* < .001). At follow-up, this pattern of correlation was similar, as post-intervention RAI scores displayed strong correlations with post-intervention RAI Fluency (*r* = .76, *p* < .001) and a weak correlation with time taken to complete the post-intervention RAI (*r* = .27, *p* = .01).

RAI scores at baseline also showed a moderate inverse relationship with post-intervention RAI score change (*r* = -.45, *p* < .001), indicating that participants with higher baseline RAI scores showed smaller RAI score increases following intervention and vice-versa. For instance, while the mean increase in RAI score for the entire sample was 8 points (*SD* = 7.5), the average rise in scores for those who scored in the top 20% of RAI scores was just over half of this figure (*M* = 4.6, *SD* = 2.8). This may be explained by the fact that participants with higher RAI's at baseline have already demonstrated a degree of proficiency in most of the relational skills that will subsequently be trained in the SMART program. For example, a participant with a baseline RAI score of 50 has already responded correctly to examples of all but 5 of the types of tasks which will be trained in the 55 training levels. As such, this participant stands far less to 'gain' from training these skills (as most have already been established) than a participant with a baseline score that is in the average or below average range for this sample.

The primary aim of the current investigation was to investigate the effectiveness of a relational skills training program (SMART) in improving proficiency in relational responding in accordance to the frames of coordination, opposition and comparison, as assessed by the Relational Abilities Index (RAI). Analyses of pre- and post-intervention RAI scores indicated that relational training produced very large, statistically significant improvements in RAI scores for the current sample. As such, it can be concluded that the SMART program is effective in improving fluency in the relational skills it targets. Pertinent to this discussion is the fact that the RAI score improvements witnessed cannot be explained by practice effects, as due to the construction of the RAI, every administration utilises a randomly-generated stimulus set from a virtually infinite number of stimuli and questions vary slightly for every run of the test insofar as they are taken randomly from each one of the possible question formats presented during each of the 55 stages. Therefore, while there is complete uniformity across administrations in terms of the function of each of the 55 RAI tasks, the physical form of these tasks is unique to each administration. This represents a key advantage of using the RAI to measure relational responding (and potentially intellectual performance), as improvements on this scale reflect genuine improvements in skill rather than an artefact of retesting. As such, the RAI holds the distinction that it assesses a skill which is tightly controlled in terms of function but is not vulnerable to the practice effects that may be witnessed when re-administering other psychometric instruments. Furthermore, by estimating Full-Scale IQ scores for each participant, the current results also indicate that these improvements in relational skill may facilitate large scale improvement in intellectual performance, insofar as estimated Full-Scale IQ scores rose by over 11 points for this sample. The transfer of this effect to general intellectual performance, as assessed by traditional IQ assessment will be investigated in Experiment 2.

In addition to the score increases found for the sample taken as a whole, further analysis revealed significant post-intervention RAI score increases at all ability levels (low average, high average and above average) for the current age group. While such rises were significant for all ability groups, the magnitude of these score increases was shown to differ across ability levels, with those in the high average category showing the greatest score increases. However, while SMART led to widespread improvements for the current age group and ability level, the efficacy and application of the SMART program in raising scores on standardised tests and doing so within populations of younger individuals and individuals with lower levels of baseline relational and/or intellectual performance warrants further study and will represent a novel investigation programme within the relevant literature. These issues will be pursued in the chapters to follow.

Chapter 3:

Assessing the efficacy of the SMART program in improving intellectual and academic performance in a sample of Irish secondary school students

3.1 Introduction

Experiment 1 essentially functioned as a large-scale manipulation-check of the SMART system, and further underlined the efficacy of this intervention in improving relational responding proficiency. However, while significant improvements were found for Full-Scale IQ scores estimated from RAI scores by means of linear regression, the impact of relational skills training on actual IQ scores requires further investigation. Several studies have now shown that when relational skills repertoires are enhanced, large gains in intelligence quotients, and scores on other tests of general cognitive functioning, are observed (Amd & Roche, 2018; Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; Thirus et al., 2016; Vizcaíno-Torres et al., 2015).

While such studies have produced promising results, reports of such transfer effects deserve a special kind of critical attention. That is, spurious claims that various training methods or practices can increase intelligence (e.g., The "Mozart" effect) have plagued psychology for decades and the popularity of such methods usually outlives emerging evidence that no such effects can be substantiated. With regard to SMART, studies from a small number of separate laboratories have been published, but each of these studies suffers from various methodological limitations. Specifically, all, barring Cassidy et al. (2016; Experiment 2), involved non-blinded and non-independent testers pre- and post-intervention. Indeed, in a comprehensive meta-analysis of the effect of blindedness on post-intervention treatment effects in randomised controlled trials, Schulz, Chalmers, Hayes, & Altman (1995), report that such effects are exaggerated by approximately 17% due to non-blind tester bias. There is also an absence of control groups in both of the Cassidy et al. (2011, 2016) studies and the Amd and Roche (2018) study, which has been identified as a key criticism of many intervention studies attempting to increase intelligence (Melby-Lervag & Hulme, 2013; Redick et al., 2013; Shipstead et al., 2012; Simons et al., 2016). Other issues include the use of single-case designs (Luciano et al., 2007; Vizcaíno-Torres et al., 2015) and insufficient training periods and/or test-retest intervals (Amd & Roche, 2018; McLoughlin et al.,2018). As such, a further and improved replication of the reported 'SMART' effect is required, with random participant assignment, blinded testers and extended training periods.

The current study is the first to implement blind testing in the study of relational skills training programs in a randomised controlled trial of the SMART method. Thus, it does not aim to replicate previous studies precisely, but aims to interrogate the reported effects using more stringent methodologies. Indeed, this is the optimal way in which to test the theoretical hypotheses underlying an intervention rather than the methodologies *per se* (see Crandall & Sherman, 2016 for a more complete discussion of the relative merits of direct and conceptual replication). In this study, the relational skills training intervention used by Amd and Roche (2018), Cassidy et al. (2016), Hayes and Stewart (2016), and Thirus et al. (2016) was administered to a group of 15 to 17-year-old children over a period of three months using a single-blind randomised controlled design. Scores on a standardized assessment of intelligence (Wechsler Abbreviated Scale of Intelligence, WASI; Wechsler, 1999) were administered to all participants before and after completing the training program, in order to assess its impact on intellectual performance. The training program was administered entirely by independent parties (school teachers) and the researchers had no role in the administration of the training program or in participant assignment.

3.2 Method

3.2.1 Participants

A sample of 26 secondary school students (Mean age = 16.5 years, *SD* = 0.67; 11 male and 15 female) attending 4th year in an Irish secondary public school were included in the current study. As the school provides SMART training as part of its curriculum, all students in the current sample were scheduled to complete the training during the course of the school year. Following baseline IQ testing by the authors, participants were divided randomly into an Experimental (*n* = 12, Mean FSIQ = 99.2) and a Control group (*n* = 14, Mean FSIQ = 98.9). The allocation of students to their respective groups was carried out by the school, and no member of the research team was involved in this process, ensuring that the experimenters remained blind to group membership up to and including re-administration of follow-up measures. All participants in the control condition were given access to the training program following completion of the study.

3.2.2 Settings and Materials

All WASI assessments took place in a small room (3m x 3m approx.) within one of the school's two main buildings. All RAI assessments and SMART sessions took place in the school's computer lab.

3.2.2.1 Wechsler Abbreviated Scale of Intelligence.

The Wechsler Abbreviated Scale of Intelligence (WASI, Wechsler, 1999) is a widely administered, short-form assessment which gives an approximation of an individual's intellectual performance relative to his/her peers. For the purpose of the current analysis, the full WASI test battery (the Vocabulary, Similarities, Block Design and Matrix Reasoning subtests) was administered, allowing the derivation of scores for Full-Scale IQ, Verbal IQ and Performance IQ. Administration time for the WASI is approximately 30 minutes.

3.2.2.1 Relational Abilities Index.

The Relational Abilities Index administration replicated the assessment implemented previously by Cassidy et al. (2016) and in Experiment 1.

3.2.2.2 Relational Skills Training Protocol.

The relational skills training intervention replicated the online SMART program implemented previously by Cassidy et al. (2016) and in Experiment 1.

3.2.3 General Procedure

All participants were administered WASI IQ assessments at baseline. To ensure that the experimenters were blind to group membership, participants were then divided into two IQ-matched groups by school staff. The experimental group was then administered the SMART program in bi-weekly, 45-minute sessions within school hours over a 12-week period. During these sessions, the control group continued with their regular classroom activities. Following this training period, all participants were then retested using the WASI. Once the study was completed, access to the SMART program was offered to the control group for ethical reasons (i.e., not to deny treatment).

3.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland**.** Parental consent and student assent were obtained for all participants.

3.3 Results and Discussion

3.3.1 Descriptive Statistics

In total, just over half the experimental group (n =7) completed all training levels, with the mean number of completed levels being 41.6 out of 55. In terms of RAI test performance, mean baseline score (for the experimental group) was 39.3 out of a possible 55. Due to technical issues affecting the host site for RAI testing and SMART training, follow-up RAI test statistics were not recorded successfully, and as such, analyses of post-intervention RAI test scores were not possible. Mean Full Scale IQ scores were in the average range at baseline for both the Experimental (*M* = 99.2, *SD* = 16.25) and the Control groups (*M* = 98.9, *SD* = 8.4). Full descriptive statistics for IQ scores at baseline and follow-up are displayed in Table 3.1.

Table 3.1

Mean IQ index and subtest scores at baseline and follow-up for the experimental and control groups. Standard deviations are displayed in brackets

	Experimental		Control	
Measure	Baseline	Follow-up	Baseline	Follow-up
Full Scale IQ	99.2 (16.3)	117.9 (15.7)	98.9 (8.4)	99.1 (8.2)
Verbal IQ	100.6 (17.8)	120.6 (16.5)	98.7 (8.9)	98.4(8)
Vocabulary	49.7 (11.6)	60.3 (11.6)	49.5 (6.5)	48.6 (5.6)
Similarities	49.5 (9.8)	62.3 (8.1)	49.2 (6.5)	49.4 (7.6)
Performance IQ	97.8 (13.3)	111.3 (13.5)	98.1 (9.2)	99.9 (8.6)
Block Design	49.3 (9.9)	56.7 (9.4)	52.5 (9.7)	53 (8.9)
Matrix Reasoning	48.5 (8.4)	56.8 (6.4)	45.6 (6.2)	47.6 (5)

3.3.2 Correlational Analysis

To investigate the relationship between relational responding and intellectual performance, a correlational analysis of RAI scores and IQ indices and subtest scores was conducted. At baseline, moderate-to-strong correlations were found for RAI scores and each of the three main IQ indices: Full-Scale (r = .64, p = .03), Verbal (r = .59, p = .04) and Performance IQ (r = .58, p = .048). In terms of IQ subtests, baseline RAI scores displayed moderate-to-strong correlations with both Verbal IQ subtests, Similarities (r = .62, p = .03) and Vocabulary (r = .59, p = .046). RAI scores also correlated significantly with one of the two Performance IQ subtests, Matrix Reasoning (r = .68, p = .02), but not with Block Design. In analyzing training progress, it was found that baseline Full Scale IQ was a strong predictor of the number of training levels completed by participants (r = .67, p = .03). Baseline RAI scores did not correlate with number of training levels completed.

3.3.3 Analysis of IQ score changes following intervention

3.3.3.1 Full-Scale IQ.

A two-way (condition x time) mixed ANOVA found a within-subjects effect of time on Full-Scale IQ, $F(1, 24) = 149.81$, $p < .001$, $\eta_p^2 = 0.862$, and an interaction effect of time*condition, $F(1, 24) = 140.95$, $p < .001$, $\eta_p^2 = 0.854$. The between groups effect of condition did not reach statistical significance, $F(1,24) = 2.98$, $p = .06$, $\eta_p^2 = 0.14$. For the Control Condition, a paired samples *t*-test found that there was no difference between Full-Scale IQ score at baseline (M = 98.86, SD = 8.44) and at follow-up (M = 99.14, SD = 8.18), $p > .05$. A further paired samples t-test found a significant increase for the Experimental Condition (i.e., SMART intervention group) in Full-Scale IQ scores from baseline (M = 99.17, SD = 16.25) to follow-up test (M = 117.92, SD = 15.7), $t(11) = -16.23$, $p < .001$, 95% CI [16.21, 21.29]. An independent samples t-test comparing the groups for change in IQ score from baseline to follow-up test found a significant difference, $t(24) = 11.87$, $p < .001$, 95% CI [15.25, 21.67]. Rises in Full Scale IQ for both the experimental and control participants can be seen in Figure 3.1.

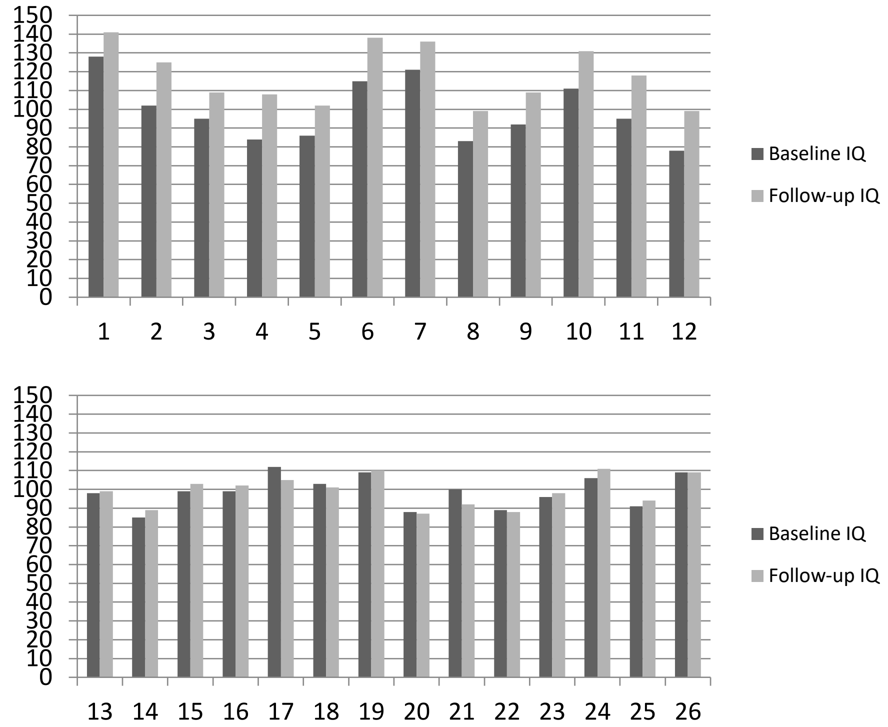


Figure 3.1. Histograms displaying WASI Full-Scale IQ scores for SMART participants (top) and control participants (bottom) at baseline and follow-up.

On average, Full Scale IQ scores recorded for the SMART intervention group increased by more than one full standard deviation (i.e., 18.8 points > 15 points) post-intervention, which demonstrates a mean percentile rank increase of over 31% from approximately the 47th percentile (M = 99.14) at baseline to the 88th percentile (M = 117.9) at follow-up. This increase moved the average group IQ classification band from 'average' to 'high average'. There was no significant correlation between Full-Scale, Verbal or Performance IQ score at baseline and the change IQ score which suggests that pre-training IQ score did not predict or account for the change in IQ score. For the experimental group, number of training levels completed did not correlate with subsequent IQ change. Mean scores for Full-Scale, Verbal and Performance IQ at both baseline and follow-up are displayed in Figure 3.2.

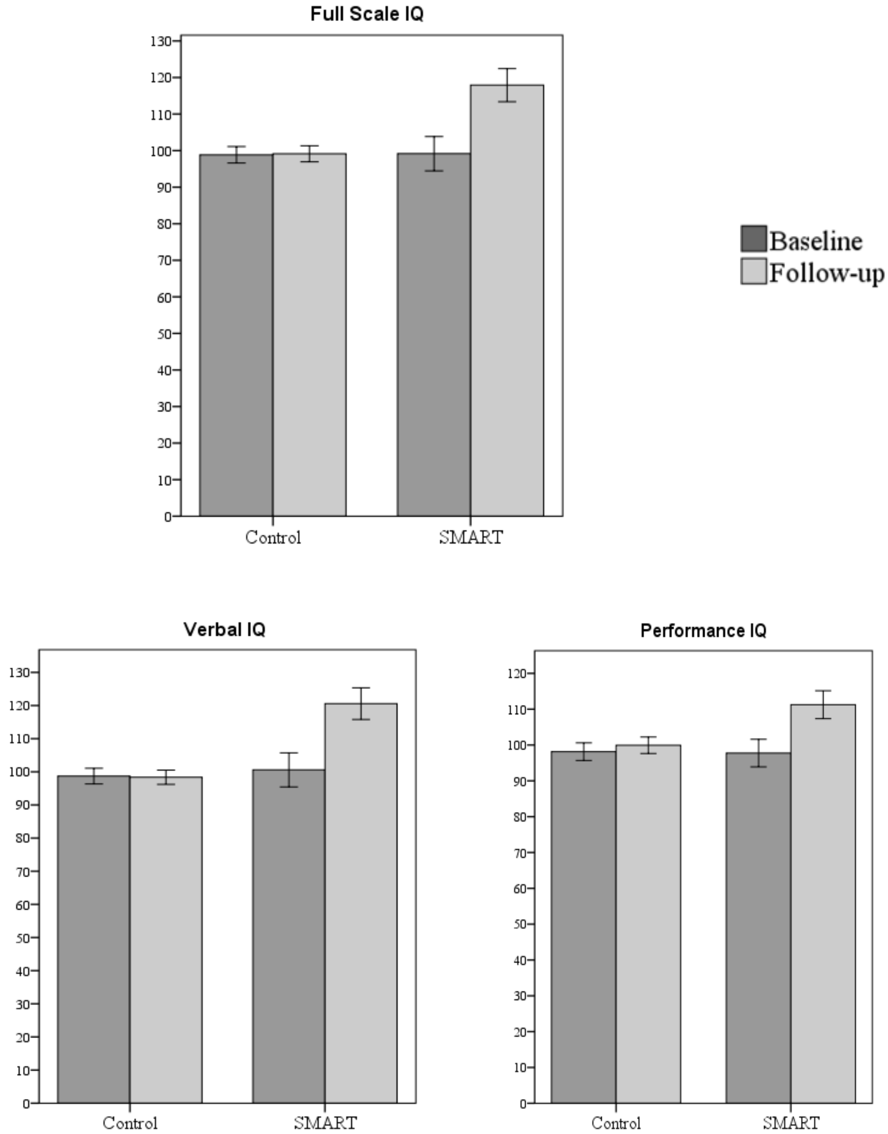


Figure 3.2. Mean IQ scores at baseline at follow-up for both SMART and control participants. Error bars represent standard error from the mean.

3.3.3.2 Verbal IQ and Performance IQ.

A two-way (condition, time) mixed ANOVA was conducted for the Verbal IQ (VIQ) composite scores at Time 1 and Time 2. There was a within-subjects effect of time, $F(1, 24) = 41.17$, $p < .001$, $\eta_p^2 = .632$, and interaction of time*condition, $F(1, 24) = 44.22$, $p < .001$, $\eta_p^2 = .648$. The between-subjects effect of condition reached statistical significance, $F(1, 24) = 5.89$, $p = .023$, $\eta_p^2 = .197$. Follow-up paired samples t-tests found a significant increase in VIQ for the Experimental Condition from pre- (M =

100.58, SD = 17.83) to post-intervention (M = 120.58 , SD = 16.54), $t(11) = -7.78$, $p <$ .001, 95% CI [-25.66, -14.34], but no significant difference between Time 1 (M = 98.71, SD = 8.89) and Time 2 (M = 98.4, SD = 8) VIQ for the Control Condition, $t(13) = 4.22$, $p = .845$, 95% CI [-3.5; 4.22].

A further two-way (condition, time) mixed ANOVA was conducted to examine differences in Performance IQ composite scores at Time 1 and Time 2. There was a significant within-subjects effect of time, $F(1, 24) = 88.95$, $p < .001$, $\eta_p^2 = .788$, and an interaction of time*condition, $F(1, 24) = 52.24$, $p < .001$, $\eta_p^2 = .685$, but no between-subjects effect of condition $F(1, 24) =1.59$, $p = .219$, $\eta_p^2 = .062$. Follow-up paired samples t-tests found a significant increase in PIQ from Time 1 (M = 97.75 , SD = 13.34) to Time 2 (M = 111.25, SD = 13.49) for the Experimental Condition, $t(11) = -11.34$, $p < .001$, 95% CI [--16.12; -10.88], but no significant difference between Time 1 (M = 98.14, SD = 9.17) and Time 2 (M = 99.93 , SD = 8.61) for the Control Condition, $t(13) = -1.62$, $p = .129$, 95% CI [-4.16; .591].

3.3.3.3 IQ subtests.

To gain a deeper understanding of the specific increases that mediate the observed effects of relational training on Full-Scale IQ indices, a series of mixed between-within ANOVAs were conducted to assess changes in performance for each of the IQ subtests. For the Vocabulary subtest, there was a significant within-subjects effect of time, $F(1, 24) =22.1$, $p < .001$, $\eta_p^2 = .479$, and an interaction of time*condition $F(1, 24) = 29.8$, $p < .001$, $\eta_p^2 = .554$. The between-subjects effect of condition was also significant, $F(1, 24) = 5.47$, $p = .028$, $\eta_p^2 = .185$.Paired samples t-tests found a significant increase in Vocabulary subtest scores from Time 1 (M = 50.5 , SD = 11) to Time 2 (M = 62, SD = 9.64) for the Experimental group, $t(11) = -6.19$, $p < .001$, 95% CI [-15.59; -7.41], but the Control group failed to display significant rises from Time 1 (M

= 49.5, *SD* = 6.55) and Time 2 (*M* = 48.64 , *SD* = 5.62), *t*(13) = .627, *p* = .54, 95% CI [-2.09; 3.81].

For the Block Design subtest, there was a significant within-subjects effect of time *F*(1, 24) = 29.29, *p* < .001, η_p^2 = .55, and an interaction of time*condition, *F*(1, 24) = 22.15, *p* < .001, η_p^2 = .48. There was no significant between-subjects effect of condition, *F*(1, 24) = .002, *p* = .965, η_p^2 = .000. Follow-up paired samples t-tests found a significant increase in Block Design scores from Time 1 (*M* = 49.33 , *SD* = 10.53) to Time 2 (*M* = 56.5, *SD* = 10.2) for the Experimental Condition, *t*(11) = -7.29, *p* < .001, 95% CI [-9.33; -5], but none between Time 1 (*M* = 52.5, *SD* = 9.67) and Time 2 (*M* = 53 , *SD* = 8.87) for the Control Condition, *t*(13) = -.498, *p* = .627, 95% CI [-2.67; 1.67].

For the Similarities subtest, there was a significant within-subjects effect of time *F*(1, 24) = 35.52, *p* < .001, η_p^2 = .61, and an interaction of time*condition, *F*(1, 24) = 35.84, *p* < .001, η_p^2 = .599. The between-subjects effect of condition was not statistically significant, *F*(1, 24) = 3.99, *p* = .057, η_p^2 = .142. Follow-up paired samples t-tests found a significant increase in Similarities scores from Time 1 (M = 49.17, SD = 10.52) to Time 2 (M = 61.67, SD = 8.16) for the Experimental Condition, *t*(11) = -8.14, *p* < .001, CI [-15.89; -9.12], but no significant difference between Time 1 (*M* = 49.21, *SD* = 6.47) and Time 2 (*M* = 49.36, *SD* = 7.57) for the Control Condition, *t*(13) = -.103, *p* = .919, 95% CI [-3.14; 2.85].

For Matrix Reasoning scores, there was a significant within-subjects effect of time *F*(1, 24) = 32.98, *p* < .001, η_p^2 = .579, and an interaction of time*condition, *F*(1, 24) =13.58, *p* < .005, η_p^2 = .361. The between-subjects effect of condition was significant, *F*(1, 24) = 5.08, *p* = .034, η_p^2 = .175. Follow-up paired samples t-tests found a significant increase in Matrix Reasoning scores from Time 1 (*M* = 47.92 , *SD* = 9.08) to Time 2 (*M* = 56.75, *SD* = 7.03) for the Experimental Condition, *t*(11) = -5.1, *p* < .001, CI [-12.64; -502], but no significant difference between Time 1 (*M* = 45.64, *SD* = 6.2) and Time 2 (*M* = 47.57 , *SD* = 5) for the Control Condition, $t(13) = -2.13$, $p = .053$, 95% CI [-3.88; .026].

The purpose of the current experiment was to investigate the effectiveness of a relational skills training intervention in improving intellectual performance as assessed by a traditional metric of IQ. In this regard, the results of the current investigation appear to further underline the efficacy of the SMART program, supporting previous findings. Results indicate that there was a statistically significant increase in Full-Scale IQ for Experimental participants (*M* = 18.4 points), while the mean score for Control group remained virtually unchanged. In addition, baseline Full-Scale IQ scores were not found to predict or account for subsequent post-training Full-Scale IQ scores, indicating that the SMART training program may be an effective means of increasing intellectual performance across a range of intellectual levels. Similar score increases were found for Verbal IQ scores, with Experimental participants displaying a mean rise of 19.7 points, while the Control group's score dropped by just under half a point. Performance IQ scores increased significantly only for the Experimental group, although the between-group difference was not found to be statistically significant following a mixed between-within ANOVA. Finally, results indicated significant improvements on all four IQ subtests following training for the Experimental group, while the Control group did not show significant improvements for any subtest. As such, the results of the current analysis appear to further underline the proposition that relational skills training interventions may be a reliable means of increasing general intelligence, at least for normally-developing, adolescent participants.

In addition, complementing previous findings (e.g. Colbert et al., 2017; O'Hora et al., 2005; O'Hora et al., 2008) relational skill was found to correlate significantly with intellectual performance. Baseline RAI scores displayed moderate-to-strong correlations with Full Scale, Verbal and Performance IQ. In addition, RAI scores correlated significantly with three of four IQ subtests (Similarities, Vocabulary and Matrix Reasoning). Such findings highlight the relevance of relational responding to intelligence, and further emphasise the potential validity of the RAI as a functional alternative to traditional assessments of intellectual performance.

In support of the findings of Experiment 1, the number of SMART training levels completed by participants was predicted by baseline Full-Scale IQ. For the current sample, all participants with below average or borderline baseline Full-Scale IQs failed to complete the training program. Indeed, this discussion highlights an issue that requires further elucidation in relation to the SMART program: What is the range of ability levels of individuals that can be expected to complete the training program, and access its benefits? In addition, a further issue is logically entailed due to the relationship between pre-intervention intellectual performance and training progress, that is, how accessible is the current SMART program for younger children? There is a relative dearth of research into the developmental 'milestones' of relational responding establishment (e.g. at what age do children usually learn to derive equivalence relations? When is AARR usually established?). Previous research has reported positive effects of SMART training in normally developing samples of 11- and 12-year-olds (Cassidy et al., 2016; Cassidy et al., 2011), and 15-17-year-olds (Cassidy et al., 2016). Studies examining the effect of relational training on younger kids have mainly studied children with learning or behavioural difficulties (Cassidy et al., 2011; Hayes & Stewart, 2016). Therefore, the impact of SMART training in samples of normally-developing pre-adolescent children requires further attention. In Experiment 3, this question will be at least partially addressed, by administering the SMART program to a group of 10- and 11-year-old primary school students. This investigation will also consist of further evaluation of the impact of pre-intervention ability level on post-intervention outcomes and training level progression. This experiment will utilise a crossover design, which will further increase the degree of experimental rigour applied to the study of the SMART program. This represents an important progression for this stream of research, as the use of crossover designs minimise the potential influence of non-specific, confounding and extraneous variables (Kazdin, 1980; Stoney & Lee Johnson, 2012). As such, the subsequent analysis should provide a greater degree of precision in analysing the specific influence of relational skills training, while reducing the impact of other, potentially confounding factors.

Chapter 4

Implementing a crossover design to evaluate the impact of the SMART program in improving intellectual performance in a sample of Irish primary school students

4.1 Introduction

In Experiments 1 & 2, SMART was administered to two samples of secondary school students, resulting in significant improvements in relational responding proficiency. In Experiment 1, it was found that pre-intervention levels of relational responding proficiency predicted training progress in a sample of 12- to 14-year-olds as those with higher baseline RAI scores were more likely to complete training. Such a finding suggests that in order to complete SMART training, there may be a prerequisite level of relational skill necessary to complete the program. In Experiment 3, which used an older sample (16-17 years old) that displayed higher mean baseline RAI score, this relationship between baseline relational skill and training progress was not replicated. It is proposed that this is possibly explained by the fact that as a group, this older sample displayed more proficient relational responding repertoires before the intervention, and therefore, for the most part, these participants displayed the prerequisite ability level that may be required to complete the training program in a 3- to 4-month period. The finding that baseline ability correlated with training progress in one sample and not the other, however, raises the possibility that pre-intervention ability may preclude program completion at some ages and ability levels, but not at others. As such, the accessibility of the SMART program for younger children and those with lower levels of relational skills warrants further investigation.

The current experiment aims to elucidate the relationship between SMART training progress and baseline ability, by administering the SMART program to an entire class of 10- and 11-year-old primary school students. This is of interest, because, as of yet, the only analyses of the SMART program in pre-adolescent children have focussed on those diagnosed with various behavioural and learning difficulties (Cassidy et al., 2011; Hayes & Stewart, 2016). Therefore, given the range of ability levels, we can expect from a class cohort of students, it is hoped this analysis will provide further insight into the role of baseline ability in predicting both training progress and post-intervention changes in relational and intellectual ability. The current study will also serve as an assessment of the applicability and efficacy of the SMART program using a sample of normally-developing children younger than that used in any published SMART study to date. Indeed, while post-SMART IQ gains have been found in older samples in Experiment 2 and Cassidy et al. (2016), the effect of the program in improving the intellectual performance of a representative group of 10- and 11-years has not been assessed.

The current investigation will also be the first to employ a crossover design in assessing the impact of relational skills training in improving intellectual performance. This represents an important improvement to the experimental rigour and control employed in this research stream, as crossover designs entail increased power over other research designs as they significantly reduce the impact of non-specific, confounding and extraneous factors (Kazdin, 1980; Stoney & Lee Johnson, 2012). Furthermore, crossover designs provide increased experimental control, because each participant essentially acts as his/her own control (Louis, Lavori, Bailar, & Polansky, 1984). Indeed, for many of the studies proposing the benefits of relational skills training (Amd & Roche, 2018; Cassidy et al., 2011, 2016; McLoughlin et al., 2018), control groups were not utilised. By using a crossover design, not only will a control group be present, but the durability of any post-intervention score changes can also be assessed, as the group who receives relational skills training in the first phase will be administered IQ assessments immediately after their training period and also after a 3-4 month period of no-intervention (while the other group receives training). Such data will assess whether post-training score increases are transient or durable following the cessation of relational skills training.

The current experiment will, therefore, comprise a crossover design in which the efficacy of SMART training in improving relational skill and intellectual performance (as assessed by the WISC-IV, Wechsler, 2003) in a sample of normally-developing 10- and 11-year-old children. In addition, relational skill will be assessed by administering the Relational Abilities Index before and after relational skills training.

4.2 Method

4.2.1 Participants

The current sample consisted of an entire class (n = 28) of boys attending 5th class in Drimnagh Castle CBS in Dublin, Ireland. As such, a convenience non-random sampling method was employed. Mean age for this sample was 10 years and 0 months (range: 9 years, 10 months – 11 years, 1 month). Prior to participation, guardians of all potential participants were informed that they should not volunteer their child for the study if he has at any point attended a school of special education outside of the mainstream school system or suffers from any intellectual problems that they know to or feel constitute an intellectual or learning disability.

4.2.2 Settings and Materials

All assessments took place in a private room intended for psychometric/educational assessments in the host primary school. Training sessions took place in a school classroom, supervised by either the primary researcher or a designated member of school staff. Students completed training on internet-connected personal tablets.

4.2.2.1 Wechsler Intelligence Scale for Children IV.

Each participant was administered the full battery of the Wechsler Intelligence Scale for Children IV (David Wechsler, 2003), an individually administered assessment of general intellectual performance. The full WISC administration consists of 10 core subtests, nine of which were employed for the purpose of this study (Block Design, Similarities, Digit Span, Picture Concepts, Coding, Vocabulary, Matrix Reasoning, Comprehension & Symbol Search). One alternate subtest, Arithmetic was substituted for the tenth core subtest, Letter-Number Sequencing, as a number of students had difficulty understanding what was required in the latter subtest. The administration of these subtests allowed the computation of Full-Scale IQ, as well as its four subscales, Verbal Comprehension, Perceptual Reasoning, Working Memory and Processing Speed. The Cancellation, Information and Word Reasoning subtests were not administered, as these represented supplementary procedures and were not required for the computation of the aforementioned indices. Administration time for the WISC-IV was approximately 90 minutes.

4.2.2.2 Relational Abilities Index.

The Relational Abilities Index administered replicated the assessment utilised by Cassidy et al. (2016) and in Experiments 1 & 2.

4.2.2.3 Relational Skills Training Protocol.

The relational skills training intervention replicated the online SMART program implemented previously by Cassidy et al. (2016)and in Experiments 1 & 2.

4.2.3 General Procedure

After written consent was obtained from participants' parents/legal guardians, all participants were administered a baseline WISC assessment individually. Following the baseline assessments, the sample was divided into two groups matched for IQ. Group 1 completed the RAI assessment and began training immediately, continuing to do so for 3 months, while Group 2 functioned as a control group and did not receive any training during this period. While Group 1 was attending training sessions, Group 2 continued with their regular classroom activities. Following this three-month training period, follow-up WISC assessments were administered to both groups. The roles of each group were then switched, with Group 2 training for 3 months, while Group 1 returned to their regular classroom activities during training periods. Following this three-month training period, a third WISC assessment was administered to all participants. In addition, all participants completed the RAI assessment immediately before and after completing their training period.

4.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland. Parental consent and student assent were obtained for all participants.

4.3 Results and Discussion

4.3.1 Descriptive Statistics

Baseline scores for Full-Scale IQ (*M* = 94.1, *SD* = 10.9), Verbal Comprehension (*M* = 94.3, *SD* = 11), Perceptual Reasoning (*M* = 95.2, *SD* = 12.4), Working Memory (*M* = 98.8, *SD* = 10.6) and Processing Speed (*M* = 93.3, *SD* = 11) were all in the average range. The mean Relational Ability Index score was 32.9 out of a maximum score of 55. Independent samples t-tests indicated that there were no significant differences between groups for scores for age, RAI, Full-Scale IQ, Verbal Comprehension, Perceptual Organisation, Working Memory or number of training levels completed. Between-group differences in scores for Processing Speed, on the other hand, were statistically significant, although in terms of standardised scores, this represented a difference of just over one point. Table 4.1 below displays baseline descriptive statistics for RAI, IQ and training levels completed for both groups.

Table 4.1

Baseline Descriptive Statistics for both groups

Measure	Group 1		Group 2	
	M	SD	M	SD
RAI	32.5	6.8	33.4	5.5
Full Scale IQ	94.1	12	94.1	10.1
Verbal Comprehension	95.8	11.4	92.9	10.9
Perceptual Reasoning	93.7	11.8	96.7	13.3
Working Memory	98.4	8.7	99.2	12.5
Processing Speed	93.9	14.3	92.6	6.7
Training Levels Completed	39.4	12.3	35.1	14.2

Due to the time constraints involved in scheduling a crossover design within one academic year, most of the sample (n = 23) did not complete the full 55 level relational skills training program, with the mean number of levels completed being 37.2 (*SD* = 13.2). This reflects that on average, students completed all of Block 1 (coordination/opposition trials), but just 8 of the 26 Block 2 trials (comparison). In all, 17 of 28 participants completed Block 1, with a further 3 completing all but the final training level in this block.

4.3.2 Correlational Analysis

Correlational analyses revealed a strong, significant correlation between pre-intervention RAI and Full-Scale IQ scores ($r = .69$, $p < .001$). This was replicated, albeit with diminished strength, with post-intervention scores for RAI and Full-Scale IQ scores ($r = .49$, $p = .01$). In terms of the relationship between relational ability and IQ subindices, moderate significant correlations were found for RAI scores and Verbal

Comprehension (*r* = .56, *p* = .002), Perceptual Reasoning (*r* = .58, *p* = .001) and Working Memory (*r* = .44, *p* =.02), but not for Processing Speed. It is also illuminating to note that RAI score (*r* = .69) was actually a better predictor of FSIQ than Processing Speed (*r* = .55) before relational training. In addition, RAI scores also correlated with 7 of 10 WISC subtests. Full results for correlational analyses of RAI scores with Full-Scale IQ, IQ subindices and IQ subtest scores are displayed in Table 4.2.

Table 4.2

Correlations between RAI scores and WISC-IV index and subtest scores

Measure	Correlation coefficient	Significance level
Full Scale IQ	.69**	<.001
Verbal Comprehension	.56**	.002
Similarities	.5**	.007
Vocabulary	.57**	.002
Comprehension	.35	.066
Perceptual Reasoning	.58**	.001
Block Design	.4*	.033
Picture Concepts	.46*	.014
Matrix Reasoning	.39*	.038
Working Memory	.44*	.018
Digit Span	.39*	.04
Arithmetic	.77**	<001
Processing Speed	.25	.195
Coding	.18	.366
Symbol Search	.19	.34

* indicates correlations significant at p < .05 level
** indicates correlations significant at p < .01 level

4.3.3 Analysis of IQ Score Change

A mixed between-within ANOVA was conducted to assess the impact of relational skills training versus no intervention on participants' Full-Scale IQ scores across the three WISC-IV administrations. There was no significant interaction effect between intervention type and time, Wilks' Lambda = .93, $F(1,26) = .98$, $p = .39$, partial eta squared = .07. There was a large main effect for time, Wilks' Lambda = .27, $F(1,26) = 33.12$, $p < .001$, partial eta squared = .73, with both groups showing an increase in Full IQ scores across the three time periods. The main effect comparing the two types of intervention was not significant, $F(1,26) = .007$, $p = .93$, partial eta squared = .000. Figure 4.1 displays the change in participants' Full-Scale IQ, as well as the four IQ subindices, across the three testing periods for both groups.

There was a significant interaction effect between intervention type and time for Verbal Comprehension, Wilks' Lambda = .73, $F(1,26) = 4.55$, $p = .02$, partial eta squared = .27. There was a large main effect for time, Wilks' Lambda = .48, $F(1,26) = 13.73$, $p < .001$, partial eta squared = .52, with both groups showing an increase in Verbal Comprehension scores across the three time periods. The main effect comparing the two types of intervention did not reach significance $F(1,26) = 1.6$, $p = .22$, partial eta squared = .06.

For Perceptual Reasoning scores, there was a significant interaction effect between intervention type and time, Wilks' Lambda = .99, $F(1,26) = .07$, $p = .005$, partial eta squared = .07. There was a large main effect for time, Wilks' Lambda = .29, $F(1,26) =31.35$, $p < .001$, partial eta squared = .72, with both groups showing an increase in Verbal Comprehension scores across the three time periods. The main effect comparing the two types of intervention was not significant, $F(1,26) = .253$, $p = .62$, partial eta squared = .01.

For Working Memory scores, there was no significant interaction effect between intervention type and time, Wilks' Lambda = .9, $F(1,26) = .1.39$, $p = .27$, partial eta squared = .1. There was no significant main effect for time, Wilks' Lambda = .8, $F(1,26) = 3.1$, $p < .06$, partial eta squared = .2, as neither group displayed significant increases in Working Memory scores over the three testing periods. The main effect comparing the two types of intervention was also not significant, $F(1,26) = .09$, $p = .77$, partial eta squared = .003.

For Processing Speed scores, there was not a significant interaction effect between intervention type and time, Wilks' Lambda = .92, $F(1,26) = 1.1$, $p = .35$, partial eta squared = .08. There was a large main effect for time, Wilks' Lambda = .7, $F(1,26) = 5.41$, $p < .01$, partial eta squared = .3, with both groups showing an increase in Processing Speed across the three time periods. The main effect comparing the two types of intervention was not significant, $F(1,26) = .218$, $p = .64$, partial eta squared = .01.

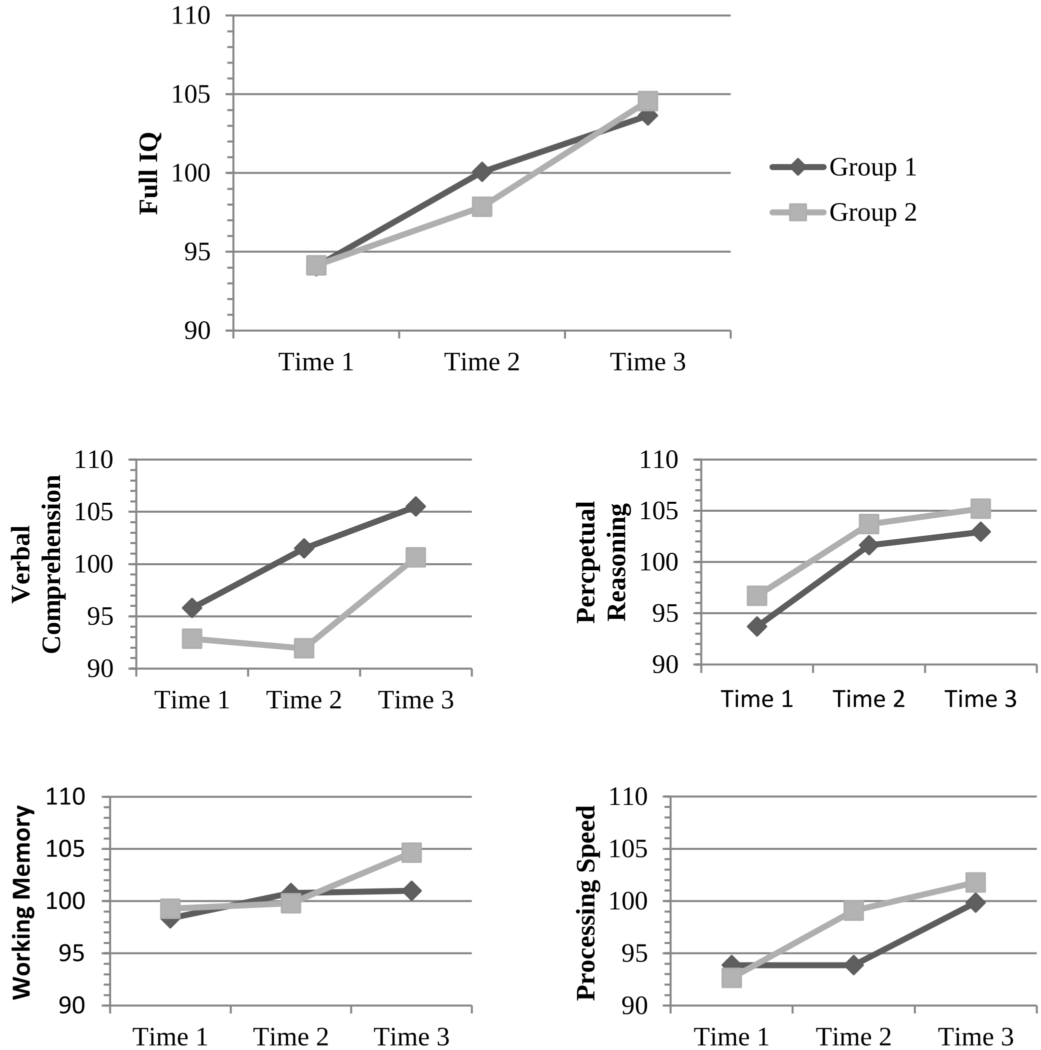


Figure 4.1. Line graphs displaying WISC-IV Full-Scale IQ and subindex score changes across the three test administrations.

4.3.4 Post-hoc analyses of IQ change

A series of post-hoc paired samples t-tests were then computed to further investigate the change in WISC-IV index and subtest scores separately for both groups. Without discounting the validity of the current ANOVA results, this was done to gain a greater insight into the IQ rises demonstrated by each group in isolation of each other. As such, paired samples t-tests for each group were computed for Full-Scale IQ and three of the four IQ subindices. Working Memory was excluded from this analysis due to a lack of significant increase in scores across the three testing periods, as indicated by the ANOVA. For these 8 pre- to post-intervention analyses (for each group respectively), an alpha level of $p < .003$, (two-tailed) was set in line with Bonferroni procedures.

Following their training period, Group 1 displayed a statistically significant increase of 6 points in Full IQ scores from Time 1 ($M = 94.1$, $SD = 11.96$) to Time 2 ($M = 100.1$ $SD = 15.6$), $t(13) = -3.34$, $p = .003$ (one-tailed). Similarly, Group 2 also displayed significant post-intervention score increases of 6.7 Full Scale IQ points from Time 2 ($M = 97.9$, $SD = 10.11$) to Time 3 ($M = 104.6$, $SD = 15.1$), $t(13) = -3.85$, $p = .002$ (two-tailed). In contrast, neither group displayed significant rises during their control periods.

In terms of Verbal Comprehension, Group 1 displayed a significant increase of 5.7 points from Time 1 ($M = 95.8$, $SD = 15.59$) to Time 2 ($M = 101.5$, $SD = 12.07$), $t(13) = -2.86$, $p = .01$ (two-tailed). Group 2 also showed a significant post-intervention increase of almost 9 points in Verbal Comprehension scores from Time 2 ($M = 91.9$, $SD = 11.95$) to Time 3 ($M = 100.6$, $SD = 15.65$), $t(13) = -5.14$, $p < .001$ (two-tailed). Neither group demonstrated significant rises following their respective control periods.

For Perceptual Organisation, there was a significant rise of almost 8 points for Group 1 from Time 1 ($M = 93.7$, $SD = 11.8$) to Time 2 ($M = 101.6$, $SD = 14.17$), $t(13) =$

-3.71, *p* = .003 (two-tailed). Group 2 did not show significant rises from Time 2 (*M* = 103.7, *SD* = 13) to Time 3 (*M* =105.2, *SD* = 15.8), t(13) = -.76, *p* = .46 (two-tailed). While Group 1 did not show a significant rise during their control period, Group 2 scores did in fact rise significantly from Time 1 (*M* = 96.7, *SD* = 13.2) to Time 2 (*M* = 103.7, *SD* = 13), *t*(13) = -3.83, *p* = .002 (two-tailed).

For Processing Speed scores following training, mean scores for Group 1 did not change from Time 1 (*M* = 93.9, *SD* = 14.3) to Time 2 (*M* = 93.9, *SD* = 16.2), *t*(13) = 0, *p* = 1 (two-tailed). Group 2 also failed to record significant rises in Processing Speed scores from Time 2 (*M* = 99.1, *SD* = 11.01) to Time 3 (*M* = 101.8, *SD* = 12.14), *t*(13) = -1.14, *p* = .28 (two-tailed). While Group 1 also did not show significant rises on this scale following their control period, Group 2 scores rose significantly from Time 1 (*M* = 92.6, *SD* = 6.7) to Time 2 (*M* = 99.1, *SD* = 11.01), *t*(13) = -2.38, *p* = .03.

4.3.5 Analysis of RAI Score Change & Training Levels Completed

A paired samples t-test was conducted to assess the efficacy of relational skills training in improving relational ability, as measured by the RAI. There was statistically significant increase in RAI scores from Time 1 (*M* = 33, *SD* = 6.2) to Time 2 (*M* = 36.3, *SD* = 8.7), *t(*26) = -2.46, *p* = .02. On average, RAI scores increased by 3.33 points, with a 95% confidence interval ranging from .55 to 6.12.

Further analyses revealed that post-intervention Full-Scale IQ change was predicted by relational training levels completed (*r* = .62, *p* < .001), but also correlated with pre-intervention Full-Scale IQ (*r* = .68, *p* < .001). These results would appear to suggest that while training progress was an important contributor to post-intervention IQ change, these changes were also dependent upon baseline levels of ability. Furthermore, there was also a strong, significant correlation between pre-intervention Full-Scale IQ and number of training levels completed (*r* = .68, *p* < .001), suggesting that higher-ability participants were more likely to reach the latter stages of training. Indeed, the pre-intervention FSIQ scores of those who completed training (*M* = 106.2, *SD* = 5.4) was significantly higher than those who did not (*M* = 93.7, *SD* = 11.5, *t*(26) = 2.33, *p* = .03. In fact, no participant with a FSIQ below 100 completed the relational skills training.

In order to explicate the complex relationship between these three variables (pre-intervention FSIQ, post-intervention FSIQ change & training levels completed), a partial correlation reported a medium-strength significant relationship between post-intervention IQ change and training levels completed while controlling for pre-intervention FSIQ (*r* = .55, *p* = .003). Furthermore, once the influence of training level completion was controlled for, there was no longer a significant correlation between pre-intervention FSIQ and post-intervention IQ change (*r* = -.12, *p* = .56). This result indicates that while pre-intervention FSIQ score appeared to predict post-intervention FSIQ score change, this relationship was in fact accounted for by training level completion. As such, withstanding the finding that the number of training levels completed was predicted by pre-intervention FSIQ, it was the former that was found to be the key determinant of post-intervention improvements in intellectual performance.

Upon closer inspection of the distribution of training levels completed, it was found that students who completed all 55 training levels (n = 5) displayed post-intervention FSIQ rises (*M* = 13, *SD* = 6.2) that were over 2.5 times greater than those who did not (*M* = 4.9, *SD* = 5.7). An independent samples t-test found that this difference was significant, t(26) = 2.82, *p* = .009. Indeed, students who completed the training program displayed IQ subindex rises multiple times greater than those found for participants who did not complete training, as displayed in Figure 4.2. Furthermore, when analysing completion rates for Block 1 of relational training, an independent samples t-test found that those who completed the 29 Block 1 levels showed a significantly greater post-intervention FSIQ rise (*M* = 8.88, *SD* = 6.52) than those who did not (*M* = 2.45, *SD* = 4.37), *t*(26) = 2.87, *p* = .008.

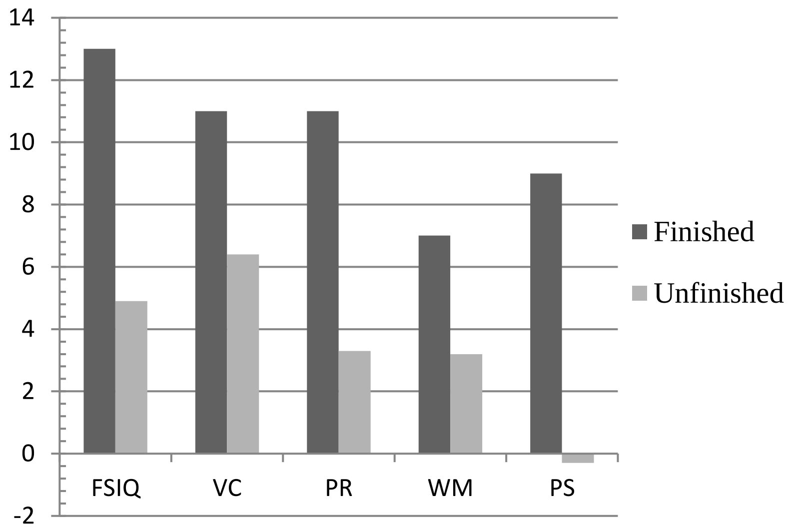


Figure 4.2. Bar chart displaying mean score changes for Full-Scale IQ and IQ subindex scores for participants who completed all 55 levels and those who did not.

4.3.6 Test-Training Discrepancy Analysis

As the RAI includes 55 trials, each of which will later be individually isolated and trained during the 55 training levels, it is interesting to compare baseline RAI scores with number of training levels completed, as a very rough index of how many additional relational trial types were established for participants as a result of training. As such, the discrepancy between baseline RAI test score and training levels completed was computed (i.e. training levels completed minus pre-intervention RAI score). Correlational analyses indicated a moderate relationship between this Test-Training discrepancy score (*M* = 4.9, *SD* = 11) and post-intervention IQ rise (*r* = .47, *p* = .01). Furthermore, those with a positive Test-Training discrepancy score (i.e., training levels completed > pre-intervention RAI score) displayed FSIQ rises that were over double (*M* = 7.7) those found for participants who had a negative discrepancy score (*M* = 3.6), indicating that higher IQ rises were found for those participants who completed training of a number of relational trial types which exceeded the number of trial types responded to correctly pre-intervention (i.e. how many additional or novel relational trial types participants could respond accurately to following training).

While there may be some value to this discrepancy analysis, it must be noted that RAI scores are not in the traditional Guttman-style scale, and as such, while the sample had a mean baseline RAI score of just under 33, it cannot be assumed that on average, participants responded to the first 33 trials correctly, but rather produced 33 correct responses across the 55 trials, according to no particular distribution.

The current analysis aimed to assess the utility of a relational skills training intervention in increasing intellectual performance by implementing a randomly controlled crossover design. In summary, on a group level, the current investigation failed to find a significant effect of relational skills training on intellectual performance as assessed by the WISC-IV. However, upon further inspection of the data, post-hoc t-tests indicated that following each training period, both groups displayed significant rises for FSIQ (M = 6.4 points) and Verbal Comprehension (M = 7.2), but did not display significant rises following their respective control periods. While this finding does not negate the result of the ANOVAs, this, taken together with a number of other results (including those of Experiments 1 & 2) indicate that this lack of effect may be at least partially mediated by the failure of the majority of participants to complete the entire program, insofar as there was a clear relationship between the number of training levels completed and post-intervention increases in IQ scores. Finally, correlational analyses showed that relational ability scores displayed moderate-to-strong relationships with WISC Full-Scale IQ, three of four WISC IQ indices (Verbal Comprehension, Perceptual Reasoning & Working Memory) and 7 of 10 WISC IQ subtests.

Of note was the finding that, in line with a multitude of previous studies (Colbert et al., 2017; Dixon et al., 2014; Gore et al., 2010; Moran et al., 2010, 2015; O'Hora et al., 2005, 2008; O'Toole & Barnes-Holmes, 2009), relational ability was found to exhibit a strong significant relationship with general intellectual performance. The profile of covariance found in the current study greatly resembles that of Colbert et al. (2017), who found significant correlations between RAI scores and 10 of 13 WAIS IQ subtests in an adult sample. In fact, of the 9 IQ subtests present in both the WISC and WAIS, significant correlations were found for 6 of these subtests (Vocabulary, Similarities, Arithmetic, Digit Span, Block Design and Matrix Reasoning) in both Colbert et al. (2017) and the current study. Of the remaining three shared subtests, neither study reported a significant relationship between RAI score and Coding. The only subtests which showed a discrepant significance between studies were Symbol Search and Comprehension, which were both significant in Colbert et al. (2017)'s analysis, but not the current study.

While RAI scores did not show a relationship with the fourth WISC subindex, Processing Speed, it is illuminating to report that this measure actually showed a weaker relationship to FSIQ (r = .55) than that found for RAI score (r = .69). In addition, the two subtests which comprise the Processing Speed index, Coding and Symbol Search, accounted for 2 of the 3 subtests which did not correlate with RAI score and did not correlate significantly with FSIQ. As such, the lack of relationship between RAI scores and Processing Speed test items may not be viewed as a construct failure of the RAI, but may instead represent theoretical divergence between a relational skills account of intelligence and the more traditional view espoused by Wechsler IQ tests. Furthermore, this lack of relationship may be explained by the fact that the RAI does not reward speed of response in the same way as Processing Speed items. While the imposition of a 30-second time limit per trial places some emphasis on speed of response, no additional reward is provided for response latency as long as a response is registered within this time limit. The focus on speed of response is far more pronounced on Processing Speed test items, as both Processing Speed subtests require participants to answer as many trials as possible within a 2-minute period. To put this contrast more clearly, the RAI does not differentiate between 5 or 25 correct responses across a 2-minute time period. Both Processing Speed subtests, on the other hand, would offer massively contrasting standardised scores in each case.

Upon more detailed investigation, it was found that a key issue with the current design (and a possible explanation of diminished training effects), was the fact that only a small portion of the sample (n = 5) completed all 55 training levels. In fact, on average, the sample completed only 37 of 55 levels, which represents all 29 Block 1 levels, but only 8 of 26 Block 2 levels. Therefore, in general, students received comprehensive training in coordination & opposition relational responding, but only rudimentary training in comparison relations. The importance of these reduced completion rates is brought into sharp focus due to the finding that post-intervention FSIQ rises were best predicted by the number of training levels completed. In addition, those who completed all 55 training levels, displayed FSIQ rises that were 2.5 times larger than those who did not. This general finding was replicated for each of the four IQ subindices, as participants who completed training displayed rises multiple times greater than those who did not. Such results are therefore in line with the "dosage effects" highlighted previously by Amd & Roche (2018). As such, while our main analyses did not identify a clearly significant effect of training on FSIQ, these results strongly suggest that this lack of effect may be mediated by a general failure to complete the entire training program, and therefore access the complete range of benefits it can provide.

Of great relevance to this issue is the mean baseline RAI (32.9) reported for this sample, which indicated that on average, participants were able to respond to about 33 of the 55 relational trial types that would be subsequently isolated and trained during the intervention period. As participants on average completed only 37 training levels, it stands to reason that many students did not access training on relational trial types far beyond their baseline level of proficiency (i.e. training improved sophistication in skills already established, rather than established new skills). This may have hindered the efficacy of the program, as theoretically, improvements in intellectual growth should be facilitated by training relational skills not already present in the participant's repertoire. In this way, students with lower baseline RAI scores stand more to gain from the SMART training program, as they display very limited proficiency in the skills being subsequently trained (i.e. there is more to learn). Indeed, a discrepancy analysis indicated that those with a positive Test/Training discrepancy score (i.e., training levels completed > pre-intervention RAI score) displayed FSIQ score increases that were more than double those with a negative Test/Training discrepancy score. Such a result indicates that if completion of all 55 levels cannot be achieved, post-intervention improvements may still be achieved if the number of training levels completed exceeds baseline RAI scores.

There are a number of possible explanations for why only a small number of students completed the entire program. The most likely of which would appear to be that training was too complex and/or too extensive to be completed by the average 10-year-old in the time frame allotted. While an extended training period may have led to greater completion rates, the finding that baseline FSIQ predicted participant's training progress is noteworthy. In fact, while training progress was found to be the key determinant of post-training IQ improvement, it was also found that those with higher baseline IQs completed more training levels. The implication of this result is that while higher IQ participants did not necessarily display the greatest post-intervention score IQ increases, they were more likely to complete more of the training. This would then further suggest that the program, in its current form may require a certain “basement” level of ability in order to engage effectively and successfully with it. Indeed, all students who completed the entire program presented with FSIQs of at least 100 at baseline.

Further to this point, the experimenter noted that during the supervision of training sessions, the progress of a considerable number of students was severely hindered by their difficulty in comprehending the purely arbitrary nature of relational tasks, and more specifically deriving arbitrary relations. This issue appeared to be more pronounced for students who presented with below-average IQ scores at baseline. As a remedial protocol, these arbitrary relata were often substituted for non-arbitrary physical aids (such as pieces of paper) and/or words in order to help students derive relations. Such difficulties grasping the “arbitrariness” of stimuli resulted in significant delays in training progress, and most students who displayed difficulty in this regard failed to complete even the first block of relational skills training. It appears, therefore, that for the age and ability level of the current sample (a group of 10-11-year-olds in the average IQ range), the level of proficiency in arbitrarily-applicable relational responding required to complete even the earliest levels of the SMART training may not already be established in all participants. As such, a remedial program which develops proficiency in non-arbitrary derived relational responding, and then further “phases-in” arbitrary stimuli may be conducive to progress on the SMART program, and overall intellectual functioning as a result. A training protocol specifically devised for this purpose will be analysed in Experiment 4.

Finally, the IQ score increases displayed following non-intervention periods requires further attention. While the score increases for Group 1 following their non-intervention period may potentially be explained by the positive impact an enhanced relational skills repertoire may continue to provide after cessation of training (as these skills may facilitate learning), such increases for Group 2 participants after their non-intervention period cannot be accounted for in this way. This pattern was most notable for Performance IQ subindices, as Group 2 displayed significant increases in Perceptual Reasoning and Processing Speed scores following their initial control period. The increases on these subindices, may at least be partially explained by the finding that across multiple test administrations. It appears that measures of intellectual and cognitive performance that include a timed element are more likely to display spontaneous score increases when compared to those that do not (Basso, Bornstein, & Lang, 1999; Dodrill & Troupin, 1975; L. J. Rapport, Brooke Brines, Axelrod, & Theisen, 1997; Sattler, 2001). As both Performance IQ subindices are computed (at least partially) based on performance on timed subtests (while Verbal IQ subindices are not), the presence of such score increases may possibly be explained as an artefact of retesting, rather than genuine improvements in performance. As intervention studies such as the current experiment necessitate multiple IQ administrations over relatively truncated periods, it may prove difficult to avoid such practice effects using traditional IQ testing batteries (as most involved timed performance in some regard). However, subsequent investigations may employ larger samples in an effort to ‘wash out’ such spontaneous score increases and help identify genuine improvements in intellectual performance.

The current study aimed to investigate the efficacy of the SMART program in improving intellectual performance in a group of 10- and 11-year-old primary school children, by implementing bi-weekly relational skills training sessions over the course of 3-4 months. The results of the current analysis indicate that this intervention was unsuccessful in significantly improving WISC FSIQ scores (or WISC IQ subindex scores) on the group level. However, upon further investigation, it appears that this lack of effect may have been at least partially accounted for by reduced completion rates and lower levels of training progress, as the majority of participants were unable to complete the training in the time period allotted. Analyses of the impact of training progress indicated a clear "dosage effect", as post-intervention IQ increases were found to be predicted by the number of SMART levels a participant completed. While insufficient time afforded to training may have been a contributing factor to this lack of training progress, it was also found that, for the current sample, the basement level of intellectual and/or relational ability required by the SMART may not have been present in all participants. As such, the current results hold important implications for the applicability of SMART training for younger age groups and those at lower levels of intellectual and relational ability. These implications will be further explored in Chapter 5.

Chapter 5

The SMART:Remedial system: A pilot analysis of a remedial relational skills intervention designed to increase basic relational skills and intellectual performance

5.1 Introduction

Following on from the results of Experiment 4, which found that the SMART protocol was ineffective in significantly improving intellectual performance in a sample of primary school children, the current analysis aimed to address some of the issues that may possibly explain this lack of effect. Correlational analyses revealed that while the number of training levels completed is the best indicator of post-intervention IQ rises, students with lower baseline IQs were considerably less likely to progress to the latter stages of the training program. Indeed, no student with an IQ below 100 completed the program during the 3-month training period. As such, those at the lower end of ability may not be able to access the benefits that SMART may provide. While an extended training period may have resulted in higher rates of training completion for some students, it was found that those at the lower end of the ability spectrum displayed considerable and delimiting difficulties with the arbitrary nature of relational tasks, and failed to demonstrate the baseline level of relational responding proficiency required to complete even the earliest levels of training. In Experiment 1, a similar result was obtained in regard to the effect of baseline relational ability on training progress, as lower baseline RAI scores were associated with fewer training levels completed.

Therefore, while the SMART program displays considerable promise as a means of reliably fostering genuine intellectual and academic improvement, the current protocol may not provide benefits for all age and ability levels due to the basic relational responding repertoires that are prerequisite for its completion. Chiefly, in order for a participant to complete even the earliest SMART levels, a basic foundation in arbitrarily applicable relational responding (AARR) is necessary. Numerous studies have suggested AARR repertoires may be weak or entirely absent in young children (e.g., Barnes-Holmes et al., 2004) and those diagnosed with various learning and/or developmental issues (Murphy & Barnes-Holmes, 2010; Murphy, Barnes-Holmes, &

Barnes-Holmes, 2005). Therefore, such individuals may not be capable of accessing the benefits of the SMART program. In order to address this issue, the SMART: Remedial (SMART:R) program has been specifically developed to incrementally build upon the most basic non-arbitrary relational skills in order to establish AARR at the level required to begin the main SMART program. Unlike the SMART program which exclusively uses nonsense words as relata, the SMART:R first targets more fundamental relational skills based on the formal properties of real-world stimuli (i.e. non-arbitrary relational responding), before progressively increasing the degree of "arbitrariness" in relational skills tasks through the use of monetary stimuli, familiar words, unfamiliar words and algebraic symbols.

The current study aims to investigate the efficacy of the SMART:R system in improving relational skill and intellectual performance in a sample of students attending additional educational support in an Irish primary school. Baseline and follow-up assessments of relational skills (SMART:R assessment) and intellectual ability (Wechsler Abbreviated Scale of Intelligence; Wechsler, 2013) were administered to an experimental group and an ability-matched control group. Following baseline assessments, experimental participants completed bi-weekly SMART:R sessions over the course of 16 weeks, while the control group received no specific intervention apart from their regularly-scheduled educational support classes.

5.2 Method

5.2.1 Participants

The current sample consisted of 22 male students recruited from a Drimnagh Castle CBS' 3rd and 4th year class cohorts, all of whom were attending additional educational support classes at the time of the study. Participants were chosen by members of school staff with no input from the experimenter and divided into two matched groups based on academic performance. Mean age for this sample was 9 years and 6 months (range: - 8 years 1 month - 11 years 4 months).

5.2.2 Settings and Materials

All assessments took place in a private room intended for psychometric/educational assessments in the host primary school. Training sessions took place in a school classroom, supervised by the primary researcher in all cases. Students completed training on internet-connected personal tablets.

5.2.2.1 Wechsler Abbreviated Scale of Intelligence.

The Wechsler Abbreviated Scale of Intelligence administration replicated the protocol employed in in Experiment 2.

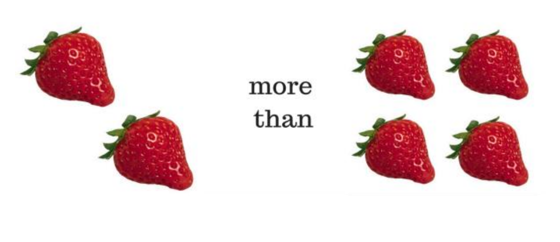
5.2.2.2 SMART:Remedial Training System.

The SMART: Remedial (SMART:R) program is a 23-level system designed to train basic fluency in non-arbitrarily- and arbitrarily-applicable relational responding. As such, SMART:R can be viewed as a prerequisite to the main SMART program by establishing the basic relational responding skills required to begin the main program. As such, the aim of the SMART:R is to assess and train non-arbitrary relational responding tasks, which progressively introduce a greater degree of arbitrary content as the individual moves through the program. As is the case with the SMART program, the SMART:R system consists of a Block 1 (levels 1-7), which trains More/Less relational tasks, and a Block 2 (levels 8 – 18), which trains Same/Opposite relational tasks. In addition, SMART:R also includes a third block, which trained AARR using purely abstract, algebraic-style symbols across both More/Less and Same/Opposite frames. All levels began with instructions as to what was included in the training level to follow, as well as explanations of what was required to complete the assessment.

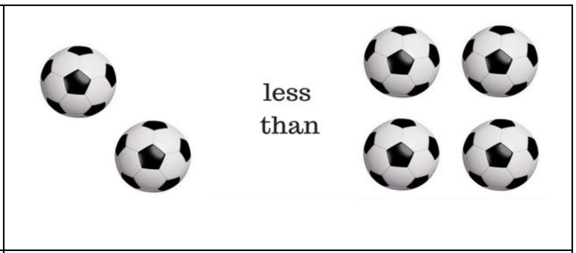
As individuals progress through each training block, tasks increase in difficulty. Task difficulty was controlled by modifying: 1. Number of relational premises (1-2); 2. Number of relational frames per task (single or mixed); 3; Directionality of relational question; 4. Order of relational premises; 5. Stimulus type (visual/verbal and arbitrary/non-arbitrary); 6. Presence of novel/unfamiliar stimuli. Participants were required to make 16 out of 16 correct responses in order to progress to the next training level. Following completion of a training level, participants were provided with their score, as well as corrective feedback for the trials that they did not respond correctly to.

Block 1 consisted of 7 levels designed to train both non-arbitrary relational responding within the frame of comparison (More/Less relations), while progressively increasing the degree of abstraction present in order to establish arbitrarily-applicable relational responding. Examples of each unique trial type included in Block 1 are displayed in Figure 5.1. Levels 1, 2 and 3 trained non-arbitrary relational responding by presenting participants with a single relational premise, outlining the relation between two sets of visual stimuli, followed by binary response options “Yes” and “No”. For example, participants may be presented with a picture of two pears to the left of the screen, a picture of one pear to the right of the screen, with the contextual cue “more than” in between the stimuli. Participants would then be required to click “Yes” if this relational premise was true, or “No” if the premise was false. All stimuli were composed of images of fruit, cutlery or sports balls, numbering between one and four examples of each for any given visual stimulus. For Levels 1 and 2, each relational premise would present two images depicting varying quantities of one class of visual stimuli (e.g., premises would include two images of either pears or footballs, but not one image of each). This was done in order to allow participants to focus solely on the physical properties of the images and therefore did not necessitate any form of abstract reasoning. Level 1 included only “more than” contextual cues, while Level 2 included “less than” trials, exclusively. Level 3 presented participants with a mixture of trials from the previous two levels, and essentially functioned as a revision level.

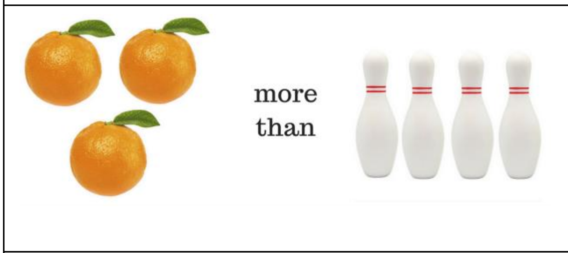
Level 4 utilised the same format as the previous two levels but included trials of mixed stimulus classes (e.g. “four footballs more than two forks”), which therefore required participants to focus on the quantity of the items presented in each set. Levels 5 and 6 then further heightened the level of abstraction required for completion by introducing monetary coins (all Euro coins ranging from 10 cent to 2 Euro) as stimuli. As such, participants were forced from this stage forward to forego responding in accordance to the physical quantity of stimuli present, and respond in accordance to the *arbitrary value* of the stimuli presented (i.e., because a lower quantity of high-value coins may be worth more than a larger quantity of low-value coins). Again, this was done in an effort to reduce the participants’ reliance on formal, physical characteristics (e.g., quantity or even size), and instead, focus on the arbitrary value assigned to the coins presented. In level 5, only single coins were presented as stimuli, in order to confirm that participants understood the assigned value of each coin. In Level 6, various assortments of coins were then presented, requiring students to compare the arbitrary value of each set of coins in order to respond accurately. Level 7 consisted of a revision level, whereby 16 trials are randomly selected from the question banks of the preceding six levels.



Level 1/3



Level 2/3



Level 4

Level 5

Level 6

Figure 5.1. Sample trial items from Block 1 of the SMART:R program.

Block 2 consisted of 10 levels which trained participants to respond to verbal stimuli within the frame of same/opposite. Sample trials for each trial type in Block 2 are displayed in Figure 5.2. In Levels 8, 9 and 10, participants were presented with pairs of words (e.g., “slow” and “fast”), alongside visual aids (e.g., the word “slow” being presented alongside an image of a tortoise, the word “fast” being presented alongside an image of a cheetah), followed by a relational question (e.g., “Are these opposite?”). In this way, participants were either presented with two identical images/words or two oppositional images/words (e.g., big/small, heavy/light, high/low). Level 8 included only questions probing for “sameness”, while Level 9 included only questions probing for “oppositeness”. Level 10 then mixed up the relational question across trials. Finally, Level 11 removed the visual aids, presented only word pairs and required participants to respond to a relational question (either: “are these same?” or “are these opposite?”).

Levels 12 and 13 presented participants with two premise relational tasks using common words as relata, followed by a relational question (e.g. "GIRL is the same as LASS, LASS is the same as GAL, is GAL the same as GIRL?"). As such, this was the first level that involved derived relational responding for completion, as participants must derive the combinatorially-entailed relation between stimuli "A" and "C" in a three-term (A-B-C) relational network, for example. However, for these two levels, every effort was made to ensure that the verbal stimuli included for all trials were sufficiently basic so that each participant would have already acquired the meaning of these words. As such, this was not a pure assessment of derived relational responding, as participants may have been able to respond correctly to trials by reading the relational question alone, as long as the words specified in that question were already present in their vocabulary. This was done deliberately, in order to introduce participants to DRR, by displaying the relations that can be derived from a three-term contingency within the frames of coordination and opposition.

Levels 14-17 retained the format of the previous two levels, but gradually introduced novel stimuli intended to force participants into engaging in derived relational responding. This was done by replacing familiar, commonplace words with unfamiliar, complex and/or archaic synonyms (e.g., replacing the word "bright" with "lucent" or "radiant"). As it can be assumed that most, if not all, of these words were not present in the participant's vocabulary, the participants would, therefore, be required to derive the meaning of these novel words by analysing their relationship to the other familiar words that were still in the three-term network. For example, when presented with the premises: "Bright is opposite to Dark, Dark is opposite to Lucent, followed by the question "Is Bright the same as Lucent?", the participant was likely to have to derive the meaning of the word "lucent" through analysing its relation to the familiar word "dark". As such, these novel words essentially function as nonsense words, which allow further abstraction to develop within the participant's repertoire of relational responding. This was the format used in Levels 14 and 15 (i.e., two familiar words, one novel) within the frames of coordination (Level 14) and opposition (Level 15).

In Levels 16 and 17 an additional novel word was introduced so that only the middle word within the three-term network was likely to be present in the participants' vocabulary (i.e., two novel words, one familiar). For example, "Lucent is opposite to Dark, Dark is opposite to Radiant, is Lucent the same as Radiant?". In this case, the participant is likely to rely on each word's relation to the familiar word "dark" in order to ascertain each word's meaning and the relationship between these words. Level 16 trained coordination relations, while Level 17 trained opposition relations in this way. Once again, this block concluded with a revision level, which chose 16 questions at random from the entire question bank for each level of Block 2.

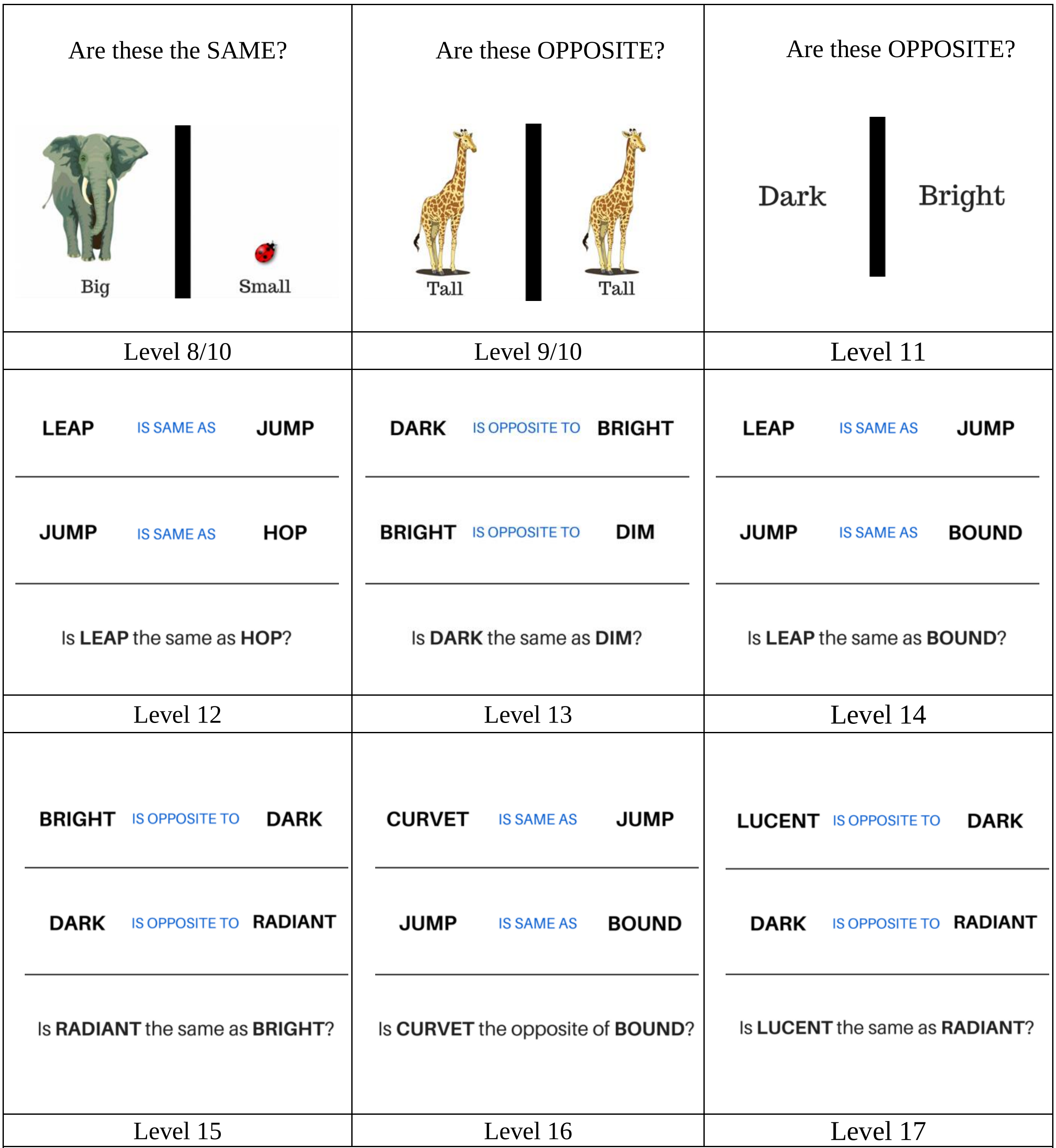


Figure 5.2. Sample trials from Block 2 of the SMART:R program.

Finally, Block 3 presented participants with two-premise, three-term relational trials which used only the algebraic symbols: "X", "Y" and "Z". As such, this block was a pure test of AARR, as participants could not rely on physical properties or prior knowledge of the relata in order to respond correctly. For example, "X is more than Y, Y is more than Z, is Z more than X?". Figure 5.3 displays examples of the trials included in this block. In Level 19, trials presented premises which included either "more than" or "less than" contextual cues, but not a mixture (e.g., "X is more than Y, Y is more than Z") followed by a relational question (e.g. "is Z more/less than X?"). Level 20 then presented premises with a mixture of "more than" and "less than" contextual cues (e.g., "X is less than Y, Y is more than Z"). Level 21 repeated the format of Level 19, presenting only "same" or "opposite" cues in the premises, but not a mixture of both (e.g., "X is same as Y, Y is same as Z"). Level 22 then introduced relational networks which included a mixture of "same" and "opposite" relations. Finally, Level 23 comprised a revision level, whereby a selection of 16 questions was chosen from the question banks of the previous Block 3 levels.

Figure 5.3. Sample trials from Block 3 of the SMART:R program.

5.2.2.3 SMART:R Assessment.

The SMART:R Assessment is a 56-item assessment of basic relational responding that requires 10-15 minutes to complete. The assessment comprises a sample of the types of relational tasks that constitute each of the SMART:R program's first 2 training blocks, with the exception of training levels which revised previous content (i.e., levels 3, 7, 10 and 18). As such, the assessment followed the same trial format as the training levels, whereby a participant was presented with one or two relational premises (e.g., 3 footballs are more than 1 football), followed by a relational question. In all cases, participants were asked to respond with a “Yes” or “No” response, by clicking buttons placed below the relational task. The SMART:R assessment included four trials each from Levels 1, 2, 4, 5, 6, 8, 9, 11,12, 13, 14, 15, 16 and 17. Overall SMART:R scores can be subdivided into two subscores, comprising Block 1 scores (trials 1-20) and Block 2 scores (trials 20-56) with the latter being further subdivided into non-AARR scores (trials 20-40) and AARR scores (trials 40-56).

5.2.3 General Procedure

All participants were first administered the full WASI battery at baseline. The SMART:R assessment was also administered via internet-connected tablets. Training was delivered in bi-weekly sessions of 45 minutes each, during school hours over a period of 16 weeks. During the training sessions, the control group continued with their regular classroom activities. Follow-up WASI and SMART:R assessments were administered to all participants upon completion of the study. In addition, all control participants were given access to the online training program after follow-up assessments were completed.

5.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University’s Social Research Ethics Committee and the Psychological Society of Ireland**.** Parental consent and student assent were obtained for all participants.

5.3 Results and Discussion

5.3.1 Descriptive Statistics

Table 5.1 displays baseline descriptive statistics for SMART:R and IQ test scores for both groups. After completing baseline assessments, one control group participant was excluded from the analysis, as his Full-Scale IQ (125) was an extreme outlier for the overall sample. In addition, one participant from the training group was excluded as he did not complete the training program or follow-up assessment, due to an extended period of school absence. Furthermore, follow-up data for individual SMART:R assessment trials were lost for two participants due to issues with the online host website. Overall SMART:R scores, however, were recorded for these participants and are included in analyses where relevant. In terms of the overall sample, mean scores for Full-Scale (M = 87.7, SD = 9.5), Verbal (M = 87.9, SD = 7.5) and Performance (M = 89.5, SD = 13.1) IQ were all in the low average category as expected. Independent samples t-tests indicated no significant difference in baseline scores between groups for SMART:R score, FSIQ, VIQ and PIQ.

Table 5.1

Baseline Descriptive Statistics for SMART:R and WASI IQ scores

Measure	SMART:R	Control	Total	Range
SMART:R	62.2% (10.3)	58.1% (8.4)	60.2% (9.5)	45 – 77
Full Scale IQ	91.5 (7.8)	83.5 (9.7)	87.7 (9.6)	68 - 105
Verbal IQ	89.1 (6.2)	86.5 (8.9)	87.9 (7.5)	73 - 103
Vocabulary	41.5 (6.6)	38.4 (8.5)	40 (7.5)	27 - 53
Similarities	44.1 (3.9)	43.3 (6.8)	43.7 (5.3)	29 - 51
Performance IQ	94.7 (10.7)	83.7 (13.5)	89.5 (13.1)	68 - 110
Block Design	47.5 (7)	42.5 (9.1)	45.1 (8.3)	34 - 60
Matrix Reasoning	45.8 (9.6)	35 (10.3)	40.7 (10)	21 - 58

5.3.2 Correlational Analysis

SMART:R scores displayed moderate significant correlations with Full-Scale (r = .45, p = .04) and Performance IQ (r = .53, p = .01). SMART:R scores also correlated significantly with one of the two Performance IQ subtests: Block Design (r = .53, p =.01), but not with Matrix Reasoning. SMART:R scores did not correlate with Verbal IQ or either of the Verbal IQ subtests. Additional analysis found that changes in FSIQ, VIQ, and PIQ scores were not predicted by baseline scores for FSIQ or SMART:R assessment. SMART:R scores at baseline however, did correlate significantly with post-intervention changes in SMART:R score (r = -.63, p = .04). In addition, the number of days required to complete training did not correlate with baseline SMART:R or FSIQ scores.

5.3.3 Analysis of post-intervention changes

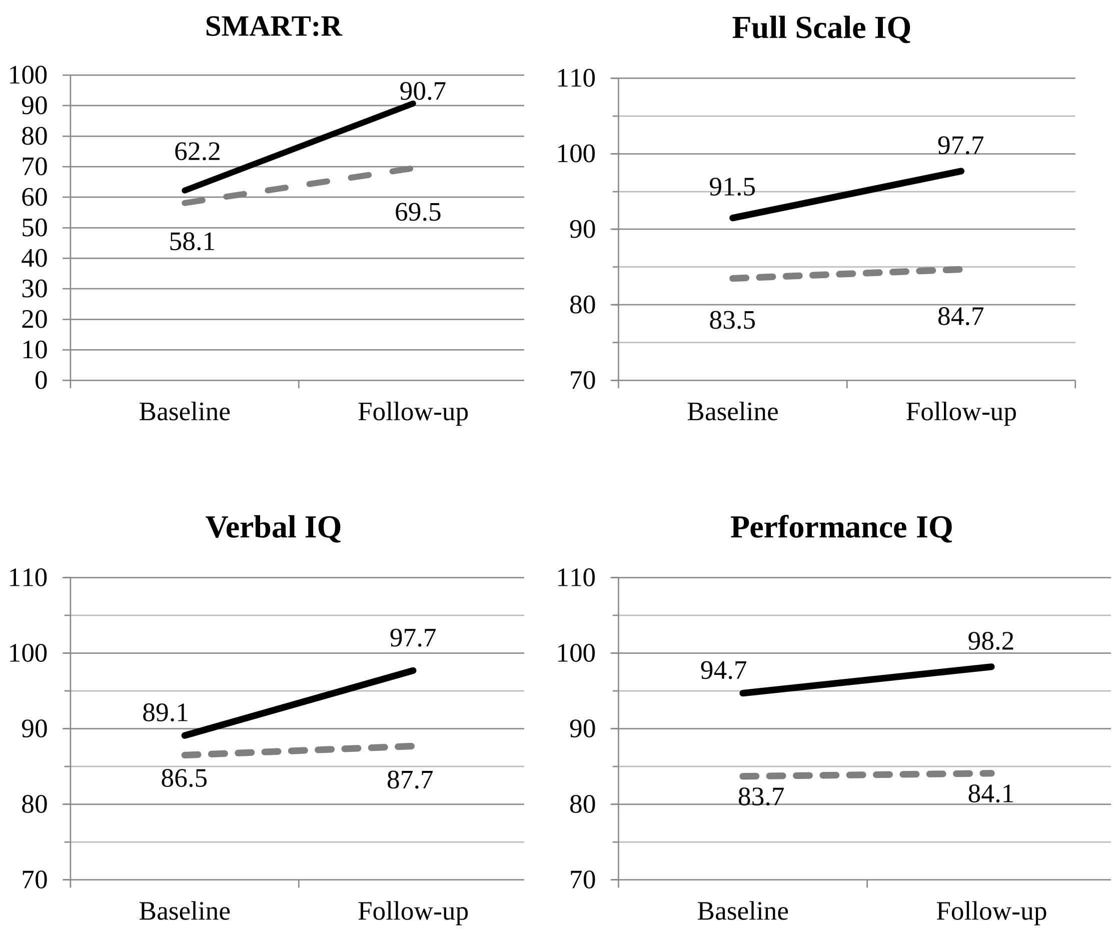


Figure 5.4. Line graphs depicting changes in SMART:R and WASI IQ index scores from baseline to follow-up for SMART:R (solid line) and control participants (dashed line).

Figure 5.4 displays changes in SMART:R assessment and IQ index scores from baseline to follow-up for both groups. A mixed between-within ANOVA was conducted to investigate the effectiveness of SMART:R (versus control) in increasing relational proficiency as assessed by the SMART:R assessment. There was a significant interaction effect between intervention type and time, Wilks' Lambda = .54, $F(1,19) = 16.44$, $p = .001$, partial eta squared = .46. There was a main effect for time, Wilks' Lambda = .18, $F(1,19) = 87.73$, $p < .001$, partial eta squared = .82, as both groups showed increased SMART:R scores following the intervention period. The main effect comparing the two types of intervention was also significant, $F(1,19) = 12.41$, $p = .002$, partial eta squared = .4, indicating that SMART:R displayed considerable efficacy in increasing relational responding proficiency.

Table 5.2 displays pre- and post-intervention SMART:R scores for both groups. In terms of SMART:R Block 1 scores, a mixed between-within ANOVA indicated that there was no significant interaction effect, Wilks' Lambda = .98, F(1,17) = .34, $p = .57$., partial eta squared = .02. There was a main effect for time, Wilks' Lambda = .47, $F(1,17) = 19.46$, $p < .001$, partial eta squared = .534. The main effect comparing the two types of intervention was also significant with a large effect of SMART:R, $F(1,17) = 4.85$, $p = .04$, partial eta squared = .22.

For SMART:R Block 2 scores, there was a significant interaction effect, Wilks' Lambda = .57, $F(1,17) = 12.96$, $p = .002$, partial eta squared = .43. There was also a main effect for time, Wilks' Lambda = .48, $F(1,17) = 18.1$, $p = .001$, partial eta squared = .52. The main effect comparing the two types of intervention was also significant, $F(1,17) = 12.18$, $p = .003$, partial eta squared = .42. The partial eta squared statistic indicated a very large effect of SMART:R on Block 2 scores.

As one of the specific aims of the SMART:R programs was to establish and/or improve AARR, a further mixed between-within ANOVA was conducted to assess changes in performance on SMART:R AARR trials specifically. Results indicated a significant intervention effect between interaction effect and time, Wilks' Lambda = .67, $F(1,17) = 8.39$, $p = .01$, partial eta squared = .33. There was also a main effect for time, Wilks Lambda = .6, $F(1,17) = 11.27$, $p = .004$, partial eta squared = .4, as both groups showed an increase in scores over time. The main effect comparing the two types of intervention was significant, $F(1, 17) = 7.37$, $p = .02$, partial eta squared = .3, suggesting that SMART:R was effective in improving AARR efficiency.

In terms of Block 2 Non-AARR trials, there was a significant interaction effect between the two main variables, Wilks' Lambda = .77, $F(1,17) = 5.08$, $p = .04$, partial eta squared = .23. There was also a main effect for time, Wilks' Lambda = .7, $F(1,17) = 7.39$, $p = .02$, partial eta squared = .3. The main effect comparing the two types of intervention was significant, $F(1,17) = 10.61$, $p = .005$, partial eta squared = .38.

Table 5.2

Mean baseline and follow-up statistics for SMART:R assessment scores

	SMART:R		Control	
	Baseline	Follow-up	Baseline	Follow-up
Block 1	71.4% (20.7)	99.1% (2)	61% (27.4)	80.7% (17)
Block 2	58.3% (14.7)	83.1% (12.5)	53.6% (6.9)	56.6% (8.9)
Non-AARR	69.4% (17.6)	90.4% (10.5)	63.7% (11.8)	65.8% (13.5)
AARR	51.9% (18.2)	81.8% (19.1)	48.2% (11)	52.2% (12.4)

A mixed between-within ANOVA was also conducted to assess the impact of relational skills training on participants' Full Scale, Verbal and Performance IQ increase across the two WASI administrations. For Full Scale IQ scores, there was a significant interaction effect between intervention type (experimental and control) and time, Wilks' Lambda = .81, $F(1,19) = 4.34$, $p = .05$, partial eta squared = .19. There was a main effect for time, Wilks' Lambda = .67, $F(1,19) = 9.42$, $p = .006$, partial eta squared = .33, with the combined participant cohort showing an increase in Full-Scale IQ scores across the two time periods. The main effect comparing the two types of intervention was significant, $F(1,19) = 7$, $p = .02$, partial eta squared = .27, indicating that SMART:R was significantly more efficacious in increasing Full-Scale IQ scores when compared to no intervention. The partial eta squared statistic indicated that the SMART:R training exerted a large effect on FSIQ.

For Verbal IQ, there was a significant interaction effect between intervention type and time, Wilks' Lambda = .67, $F(1,19) = 9.36$, $p = .006$, partial eta squared = .33. There was a main effect for time, Wilks' Lambda = .54, $F(1,19) = 16.37$, $p = .001$, partial eta squared = .46, with the combined participant cohort displaying a rise in Verbal IQ scores across the two time periods. The main effect of intervention type was not significant, $F(1,19) = 2.9$, $p = .1$, partial eta squared = .13. Post-hoc analyses of Verbal IQ scores indicated that while the increase in scores found for the control group (M =1.2, SD = 3.94) was not significant, there was a significant increase in the experimental group's scores from baseline (M = 89.1, SD = 6.2) to follow-up (M = 97.7, SD = 10.5), t(10) = -4.28, $p = .002$. The Cohen's d statistic (0.99) for this increase indicated a large effect of SMART:R training.

In the case of Performance IQ, there was no significant interaction effect, Wilks' Lambda = .97, $F(1,19) = .59$, $p = .45$, partial eta squared = .03. There was no main effect for time, Wilks' Lambda = .95, $F(1,19) = .49$, $p = .35$, partial eta squared = .05, with no significant rise in Performance IQ for the combined participant cohort across the two time periods. The main effect comparing the two types of intervention was significant, $F(1,19) = 5.15$, $p = .04$, partial eta squared = .21. The partial eta squared statistic indicated a large effect of SMART:R on PIQ scores.

5.3.4 Analysis of individual participants' scores

Analysis of individual WASI IQ score changes found that of the 11 experimental participants, every one displayed a higher FSIQ following intervention. Mean FSIQ increase for experimental participants was almost 7 points (*SD* = 6.4), with individual increases ranging from 1 to 20 points. This change is reflected in a significant increase in FSIQ percentile of 14 ranks from baseline (*M* = 30.6, *SD* = 17.3) to follow-up (*M* = 44.6, *SD* = 23.1), $t(10) = -3.2$, $p = .01$. In comparison, the mean change in FSIQ scores for the control group was 1.2 points (*SD* = 4.9), with score changes ranging from -6 to +8. A paired-samples t-test revealed that the change in FSIQ percentile scores for the control group (*M* = 2.4, *SD* = 9.4) was not statistically significant. Table 5.3 displays FSIQ standardised scores, percentile scores, and classification for all participants at baseline and follow-up.

Table 5.3

Baseline and follow-up Full-Scale IQ standardised and percentile scores for each participant

		Baseline			Follow-up		
No.	Group	FSIQ	Ptile	Category	FSIQ	Ptile	Category
1	SMART:R	87	19	Low Average	93	32	Average
3	SMART:R	98	45	Average	100	50	Average
4	SMART:R	87	42	Low Average	117	87	High Average
5	SMART:R	89	23	Low Average	104	61	Average
6	SMART:R	100	50	Average	103	58	Average
7	SMART:R	85	18	Low Average	86	21	Low Average
8	SMART:R	105	63	Average	108	70	Average
9	SMART:R	85	16	Low Average	87	19	Low Average
10	SMART:R	88	21	Low Average	97	42	Average
11	SMART:R	93	32	Average	95	37	Average
12	SMART:R	79	8	Borderline	85	16	Low Average
13	Control	100	50	Average	95	37	Average
15	Control	86	18	Low Average	93	32	Average
16	Control	80	9	Low Average	78	7	Borderline
17	Control	93	32	Average	96	39	Average
18	Control	71	3	Borderline	71	3	Borderline
19	Control	80	9	Low Average	74	4	Borderline
20	Control	82	12	Low Average	84	14	Low Average
21	Control	91	27	Average	99	47	Average
22	Control	68	2	Very Low	74	4	Borderline
23	Control	84	14	Low Average	83	13	Low Average

Verbal IQ score changes ranged from -1 to 20 points for the experimental group with a mean increase of just over 9 points. There was a statistically significant increase of 20 ranks in VIQ percentile scores from baseline (*M* = 24.9, *SD* = 13) to follow-up (*M* = 44.9, SD = 23.5), $t(10) = -3.2$, $p = .01$. Mean change in the control group's VIQ scores was an increase of 1.2 points (*SD* = 3.9), with changes ranging from -5 to +8. A paired-samples t-test revealed that the change in VIQ percentile scores for the control group (*M* = 2.6, *SD* = 7.5) was not statistically significant.

Changes in Performance IQs for the experimental group were more variable, with changes ranging from -8 points to +22 points (*M* = 3.4, *SD* = 11). Paired-sample t-tests indicated that the experimental group demonstrated a non-significant increase of just under 8 PIQ percentile ranks. Changes in PIQ were also highly variable for the control group (*M* = 0.4, *SD* = 7.3), with changes ranging from -9 to +15. Once again, a paired-samples t-test revealed that the control group change in PIQ percentile scores (*M* = 1.8, *SD* = 9.7) was not statistically significant. Bar-charts depicting baseline and follow-up SMART:R and WASI IQ scores for each individual participant can be found in Figure 5.5.

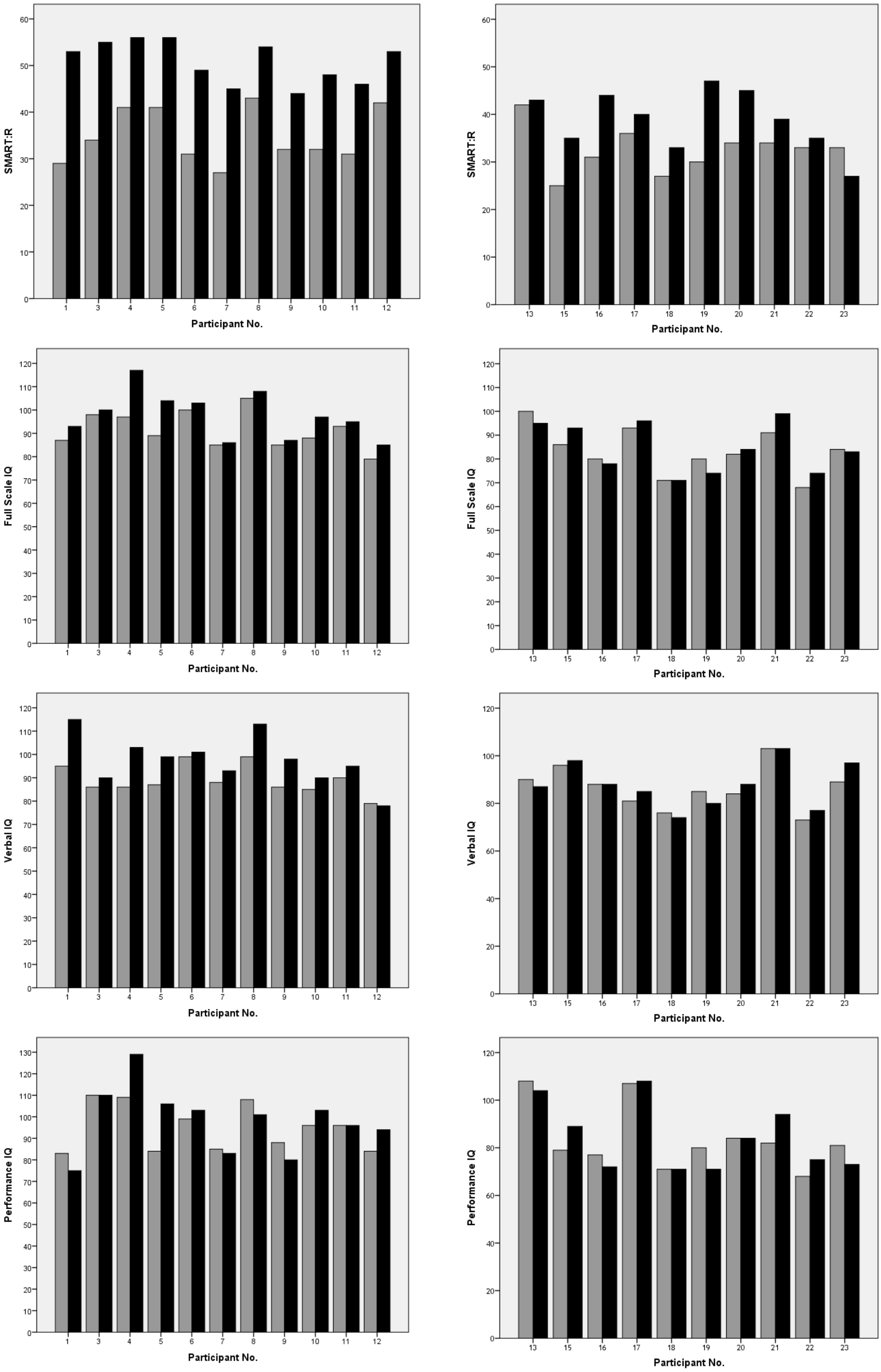


Figure 5.5. Bar charts displaying changes in FSIQ, VIQ, PIQ and SMART:R scores at for experimental participants (left) and control participants (right).

5.3.5 Analysis of IQ subtest score changes following intervention

In order to gain a more precise understanding of the nature of post-intervention IQ rises, a series of t-tests were computed to investigate how performance on each of the four WASI IQ tests (in terms of scaled scores) may have been affected by training. In line with Bonferroni procedures, an alpha level of 0.0125 was set for these analyses.

In terms of Verbal IQ subtests, standardised scores for the Vocabulary subtest rose significantly between baseline (M = 41.5, SD = 6.6) and follow-up (M = 50.2, SD = 8.6), $t(10) = -8.42$, $p < .001$. The Cohen's d statistic for this rise (1.14) indicates a large effect size. Scaled scores for the Similarities subtest did not increase significantly from baseline (M = 44.1, SD = 3.9) to follow-up (M = 47.2, SD = 6.6).

In terms of Performance IQ subtests, scaled scores for Block Design also did not increase significantly from baseline (M = 47.5, SD = 7) to follow-up (M = 49, SD = 8.8). Finally, Matrix Reasoning scaled scores did not change significantly from baseline (M = 45.8, SD = 9.6) to follow-up (M = 49.4, SD = 12.1).

The aim of the current analysis was to investigate the efficacy of a newly developed remedial program (SMART:R) to train basic relational responding proficiency and establish more fluent arbitrarily-applicable relational responding as a means of ameliorating intellectual deficits. In this regard, the current experiment appears to support the effectiveness of the SMART:R intervention. Analyses of variance indicated that the SMART:R exerted a significant effect on Full-Scale IQ, with a mean increase of 6.3 points following training, compared to an increase of 1.2 points for control participants. The SMART:R was also effective in increasing Performance IQ scores by 3.5 points, whereas scores for the control group remained virtually unchanged. Post-hoc analyses also indicated that Verbal IQ score increases (M = 8.6) were observed only for the experimental group. However, increases in Performance IQ, while larger for the experimental group than the controls, were non-significant.

While the SMART:R program significantly increased WASI IQ scores, the magnitude of these rises was smaller than those reported for the standard SMART program (e.g., Cassidy et al., 2011, 2016). Specifically, Cassidy et al., (2011) reported a mean 13-point Full-Scale IQ gain in a sample of children diagnosed with a variety of educational and behavioural difficulties. A further study by Cassidy et al. (2016) reported mean gains of 23 points, as assessed by the WAIS-III. Other studies, however, have reported IQ gains more in line with those of the current study (e.g., Amd & Roche, 2018, Thirus et al., 2016) and so variability in effects appears to be a feature of the research at this early stage. Of course, there is one obvious reason why this particular study may not have produced IQ gains in the double-figures, and this relates to the scope of the skills trained by the SMART:R program. Specifically, while the current study focused on remedial training to establish AARR, the training procedures utilised in the Cassidy et al. studies (i.e. the main SMART program), treated relatively fluent AARR as a prerequisite due to the exclusive utilisation of abstract relata (i.e., nonsense words). These studies trained sophistication only in this form of relational responding. As such, the main SMART program may train skills that are more germane to performance on standardised tests of intelligence.

In contrast, the current SMART:R assessment was focused at least partially on non-arbitrary relational responding skills, and these are not recruited widely across standardised IQ test batteries (see Cassidy et al., 2010). This suggestion is supported by the comparatively lower levels of correlation observed between IQ indices for the SMART:R assessment than those that have been reported between IQ indices and SMART relational skills assessment (Relational Abilities Index). In comparison, the aim of the SMART:R system is to usher in AARR and merely facilitate multiple exemplar training in AARR. In effect, it might be argued that it is, in fact, impressive that such IQ gains could be established by honing a skill that is a prerequisite to sophisticated AARR. It is also worth noting that both of the Cassidy et al. studies employed the WISC, a more comprehensive IQ test, whereas the current study employed the WASI, an abbreviated IQ measure. In effect, the more comprehensive WISC may be more sensitive to improvements in relational skills and therefore may indicate greater score increases following relational skills intervention.

While the main aim of the current analysis was to investigate the impact of the SMART:R program on IQ scores, a secondary goal was to ascertain whether it is an effective tool in increasing relational responding proficiency. A series of ANOVAs indicated a significant effect of the intervention on Block 1(More than/Less than) and Block 2 (Same/Opposite) trial scores. In terms of scores for Block 1 trials in the SMART:R assessment, mean scores rose significantly following training for the experimental group, with all experimental participants responding correctly to at least 95% of trials following intervention, compared to a mean score of 71.4% correct trials at baseline. In addition, correct responding rose from 58.3% to 83.3% for Block 2 trials for the experimental group but remained around chance levels for the control group. A further aim of the SMART:R program is to establish AARR in a sample in which it may be absent or weak, in order to build the skills prerequisite for completing the main SMART program. In this regard, analyses of variance indicated that there was a clear effect of the SMART:R intervention in increasing proficiency in AARR. At baseline, scores on AARR trials for both groups were approximately at chance levels (SMART:R = 51.9%, Control = 48.2%). However, at follow-up, scores for SMART:R participants rose significantly to 82%, while scores for control group participants remained virtually unchanged. Such results would, therefore, support the efficacy of the SMART:R program in improving proficiency in arbitrary and non-arbitrary relational responding according to the frames of coordination, opposition and comparison.

The considerable impact of the SMART:R program on Verbal IQ and the Vocabulary subtests scores is predicted by an extensive literature proposing the relevance of relational responding to language (Colbert et al., 2017; de Rose et al., 1992; Edwards et al., 2011; McHugh et al., 2004; Nippold & Sullivan, 1987). In addition, many of trials included in Block 2 of the program mimic naturalistic language acquisition as the individual is required to respond correctly to relational questions which probe for relations between words that are likely to be unfamiliar to them (e.g., "Is PURLOIN the same as PILFER?") In particular, Levels 14 to 17 provide an analogue of how the establishment of word-word relational networks facilitate language acquisition, as novel words may be integrated into existing relational networks, allowing the derivation of the novel word's definition based on their relation to other words already present in an individual's vocabulary. It is therefore unsurprising, given the generalised applicability of these relational skills, that of the four WASI IQ subtests, scores for the Vocabulary subtest were most improved following intervention.

While the SMART:R program facilitated significant increases in Performance IQ, the improvement in scores on this metric were less pronounced (M = 3.4 points) and more variable (SD = 11) than those found for Full Scale or Verbal IQ. The increase was also not statistically significant. This may be at least partially explained by a lack of specific relevance of SMART:R relational responding tasks to Performance IQ test items. While a clear similarity can be seen between the predominantly verbally-based trials of the SMART:R and both Verbal IQ subtests, the relevance of such trials to Performance IQ subtest items is less obvious. However, while both the Matrix Reasoning and Block Design subtests for Performance IQ can be viewed as a type of high-level relational responding termed *pragmatic verbal analysis* (Hayes, Gifford, Townsend, & Barnes-Holmes, 2001), the basic relational skills trained in the current analysis may not have exerted any influence on such skills.

In conclusion, based on the current results, the SMART:R program may offer a promising adjunct and/or alternative to the SMART program by providing remedial training to those who present with lower levels of relational responding fluency. Furthermore, the current analysis lends further weight to a growing body of research that proposes that intellectual performance can be improved via behaviour-analytic interventions based on Relational Frame Theory, specifically those which target and train relational skills. Given that the current thesis has added considerably to the current literature base proposing that relational responding training interventions may harbour implications for intellectual performance, the next experiment will investigate whether these implications extend beyond intellectual function and into an academic context by analysing the effect of the main SMART program on a gold-standard measure of academic aptitude.

Chapter 6

A large-scale analysis of the effectiveness of SMART in improving intellectual and academic performance in a sample of Irish secondary school students

-

6.1 Introduction

A number of research investigations (including Experiments 2 & 3 of the current thesis) have proposed that SMART may be an effective means of improving intellectual performance (Cassidy et al., 2011, 2016; Hayes & Stewart, 2016). However, despite the close relationship between intelligence and academic performance (Bourneville, 1895; Deary et al., 2007; Jensen, 1998; Laidra et al., 2007; Neisser et al., 1996; Roth et al., 2015), only two such studies (Cassidy et al., 2016; Hayes & Stewart, 2016) has addressed the potential impact of relational skills training on scholastic aptitude. In the second of Cassidy and colleagues' (2016) studies, significant improvements in Verbal Reasoning, Numerical Reasoning and Educational aptitude assessed by a widely-administered assessment of scholastic attainment were reported in a small sample of 15- to 17-year-old secondary-school students. However, as the effect size of these score increases were small, further investigation is required to gain a better understanding of the extent to which an improved relational responding repertoire may benefit school performance. As part of the Hayes & Stewart (2016) design, three subtests of the WIAT-II (reading, spelling & numerical operations) were included in an extensive test battery administered to assess the impact of the SMART program in a sample of 10- and 11-year-old children. Alongside increases in scores on measures of IQ, this study reported significant rises in scores for all three WIAT subtests administered.

Despite relatively little experimental analysis of the potential improvements in academic ability that may result from relational skills training, numerous correlational studies have proposed a close relationship between these two repertoires. For example, in terms of verbal attainment, various analyses conducted both inside and outside the rubric of RFT have reported that relational responding may be of central importance to reading (de Rose, de Souza, Rossito, & de Rose, 1992; Farrington-Flint & Wood, 2007; Goswami, 1986; Mackay, 1985; Sidman, 1971), vocabulary (Edwards, Figueras, Mellanby, & Langdon, 2011; McHugh et al., 2004; Nippold & Sullivan, 1987), grammar (Hock, 1991, 2008) and even spelling (Brown, Sinatra, & Wagstaff, 1996; Goswami, 1988; Mackay, 1985). Indeed, several intervention studies have produced considerable improvements in literacy and general verbal ability as a result of relational responding training interventions (Almeida-Verdu et al., 2008; de Rose et al., 1992; de Rose & de Souza, 1996; de Rose, Rossito, Rose, Peder, & Sao, 1985; Melchiori, 2000; Murphy & Barnes-Holmes, 2009, 2010; Murphy & Barnes-Holmes, 2009). Given the success of such interventions in improving verbal attainment, the specific impact of SMART on such repertories requires further investigation.

In terms of numeracy, relational responding and relational thinking have been found to contribute broadly to numerical operations (Carpenter et al., 2003; Carpenter, Levi, Franke, & Zeringue, 2005; Cassidy et al., 2016; Molina et al., 2005; Molina, Castro, & Castro, 2008; Stephens, 2007). Furthermore, the work of Ninness and colleagues (McGinty et al., 2012; Ninness et al., 2005, 2006, 2009) have demonstrated that training relational repertoires may harbour benefits for a range of mathematical competencies.

As such, the current investigation aims to analyse the impact of SMART in improving scholastic aptitude in a large sample of 12- to 14-year-old students (n = 174), as measured by the Drumcondra Reasoning Tests (DRT, Educational Research Centre, 2016), the Irish Department of Education's assessment of choice for use with second-level students. All students completed Drumcondra Reasoning Test and Relational Ability Index assessments at the beginning (September) and end (May) of one academic year, with the experimenter being completely blind to group membership. In addition, a subset of this sample (n = 38) will be administered three WASI IQ assessments in

September, January and May, as part of a crossover design in order to study the relationship between changes in intellectual performance and academic aptitude.

6.2 Method

6.2.1 Participants

The current sample consisted of 1st year students (n =174) attending Summerhill College, a secondary school in Sligo, Ireland for whom parental consent and student assent had been obtained. As the current analysis aimed to study a fully representative sample of 1st year students, no exclusion criteria were enforced for the current sample.

6.2.2 Settings and Materials

All RAI assessments, DRT assessments and SMART sessions took place in a school-based computer lab which had the capacity for approximately 30 children, who completed relational skills training under supervision by a member of school staff. WASI assessments took place in a private room intended for psychometric/educational assessments in the host primary school.

6.2.2.1 Drumcondra Reasoning Test.

The Drumcondra Reasoning Test (DRT; Educational Research Centre, 2016) is a group-administered test of educational aptitude designed for use with secondary school students in Irish schools. The DRT permits computation of standardised scores for two subindices: Verbal Reasoning and Numerical Ability score, as well as an overall DRT score. The Verbal Reasoning subsection is a 40-item assessment of literacy and vocabulary, comprising four subtests: Synonyms, Classifications, Analogies and Antonyms. The Numerical Ability 40-item subsection assesses mathematical operations and general numeracy and consists of four further subtests: Operations with Numbers, Relations among numbers, Sequential Ordering and Numerical Abstractions.

The DRT is standardised against a sample of 6,000 Irish students and requires approximately 60 minutes to complete. The DRT can be delivered via five alternate but equivalent assessments, one of which is assigned randomly to each student at each administration.

6.2.2.1 Wechsler Abbreviated Scale of Intelligence.

The Wechsler Abbreviated Scale of Intelligence administration replicated the protocol employed in in Experiments 2 & 4.

5.2.2.3 Relational Abilities Index

The Relational Abilities Index administration replicated that assessment implemented previously by Cassidy et al. (2016) and in Experiments 1, 2 & 3.

6.2.2.4 Relational Training Protocol

The relational training intervention replicated the online SMART program implemented previously by Cassidy et al. (2016) and in Experiments 1, 2 & 3.

6.3 General Procedure

Following collection of consent forms, the sample was divided into two groups by school staff before SMART training and DRT assessments were administered. This was done as the school did not have a sufficient number of computers or personal tablets to provide online training to all 174 students at once. However, this facilitated the crossover design implemented with the IQ subsample, as it allowed the assessment of intellectual ability at baseline, following SMART training and following a control period. As the student cohort is divided into eight classes based on academic performance, school staff allocated group membership in such a way to ensure that each group was matched for academic ability (i.e. each group had an approximately even number of above average, average and below average students). Furthermore, online SMART training accounts were set up by a member of school staff, with each student being assigned a coded username. In addition, a subsample of the cohort (n = 39, 19 participants from each group) was selected by school staff to complete three WASI assessments across three time points in the study's duration (September/October, January/February & May). Group assignment was again controlled by school staff and based on students' academic performance level in order to ensure an ability-matched sample.

All participants were administered the DRT in a group setting at baseline. Upon signing on to the online DRT administration system, participants were randomly allocated one of its five equivalent forms. The results of this assessment were retained by the school and released to the experimenter following the completion of the experiment. Following the baseline DRT assessments, Group 1 was administered RAI assessments immediately and began SMART while Group 2 continued to attend their regular classroom activities. SMART sessions took place in bi-weekly sessions lasting an hour in a computer lab on school premises. All sessions were supervised by a member of school staff, and the experimenter did not attend any sessions. Following a period of approximately 4 months, Group 1 concluded their SMART sessions and completed a second RAI assessment. In addition, the IQ subsample was re-administered the WASI. In the second phase of the experiment, each group switched roles, as Group 2 completed baseline RAIs and began SMART, while Group 1 returned to regular classroom activities. Following a further period of four months, Group 2 ended their participation in SMART training and completed follow-up RAI assessments. The IQ subsample then completed a third and final WASI assessment. To conclude the experiment, all participants completed a follow-up DRT assessment. Participants were once again randomly allocated one of the five equivalent DRT forms by the online system. Once data collection had been completed, and all psychometric scores had been computed, the experimenter was then unblinded as to group membership, and given access to DRT scores for the sample.

6.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland**.** Parental consent and student assent were obtained for all participants, with separate consent forms provided to the general sample and the IQ-tested subsample. During the course of the current experiment, the experimenter was kept blind to any personally identifiable data, such as group allocation and training account usernames. As such, a member of school staff served as gatekeeper for the current study and was responsible for the collection of parental consent and student assent, the set-up of training accounts, group allocation, supervision of training sessions and selection of the IQ subsample. As the only direct contact between the experimenter and participants was during the IQ test administration, participants would provide the experimenter only with a further unique codename as a personal identifier (i.e. different to the training account codename). As such, the experimenter was completely blind to group allocation process, the selection of the IQ subsample and the administration of the DRT, and therefore could not link any IQ score, DRT score or online SMART training account to any student or to group membership. This could only be done by the gatekeeper of the data, who provided the experimenter with a document which allowed the linking of participants DRT, RAI and IQ scores, as well as provided access to relational training statistics.

6.3 Results and Discussion

6.3.1 Descriptive Statistics

Mean scores for Drumcondra Verbal Reasoning (*M* = 103.2, *SD* = 14.2), Numerical Ability (*M* = 104.7, *SD* = 14.6) and Overall score (*M* = 104.4, *SD* = 13.3) were all in the average range. For the IQ sample, mean scores for Full-Scale (*M* = 103.3, *SD* = 12.8), Verbal (*M* = 102.2, *SD* = 14.47) and Performance IQ (*M* = 104.2, *SD* = 11.92) were also in the average range. Paired-samples t-tests indicated that the two training groups did not differ significantly at baseline on any of the three DRT standardised scores (Overall, Verbal Reasoning, Numerical Ability) or the Relational Ability Index. Table 6.1 displays baseline descriptive statistics for the overall sample.

Table 6.1

Baseline Descriptive Statistics

Variable	M	SD	Range
RAI	37.3	7.75	19-53
Drumcondra Reasoning Test			
Verbal Reasoning	103.2	14.18	64-140
Numerical Ability	104.7	14.56	62-136
Overall Score	104.4	13.32	68-140
WASI			
Full Scale IQ	103.3	12.75	61-152
Verbal IQ	102.2	14.47	63-153
Performance IQ	104.2	11.92	66-138

In total, 98 participants (56%) completed all 55 relational training levels, with 147 participants (84%) completing at least the first block of training (coordination/opposition relations). In order to better understand the variables that may help explain the low level of training program completion, a correlational analysis was undertaken. Training level progress was predicted by baseline scores for Overall DRT (r =.5, p = .002), and both of its subindices: Verbal Reasoning (r = .42, p = .01) and

Numerical Reasoning (*r* = .42, *p* =.01). In addition, the number of training levels completed also correlated with FSIQ score (*r* =.42, *p* .01) and Performance IQ (*r* =.54, *p* = .001). Such positive correlations indicate that high ability participants were more likely to reach the latter stages of training, and conversely, those who presented with lower levels of ability were less likely to reach the latter stages and complete all 55 training levels. This trend is further underlined by the finding that those who completed training had significantly higher baseline FSIQ scores (*M* = 106.7, *SD* = 12.9) when compared to those who did not (*M* = 98.1, *SD* = 11.6) *t*(34) = 2.06, *p* = .047.

6.3.2 Correlational Analysis

RAI scores displayed moderate correlations with Verbal Reasoning standardised (*r* = .54, *p* <.001) and percentile scores (*r* = .52, *p* < .001), Numerical Ability standardised (*r* = .59, *p* < .001) and percentile scores (*r* = .57, *p* < .001), and overall DRT standardised (*r* = .61, *p* <.001) and percentile scores (*r* = .61, *p* < .001). RAI scores also correlated significantly with WASI Full Scale IQ (*rho* = .34, *p* = .04) and Performance IQ (*rho* = .42, *p* = .01), but not on this occasion with Verbal IQ (*rho* = .18, *p* = .3). RAI scores did not correlate significantly with any IQ subtest. Table 6.2 displays correlation coefficients and significance levels for RAI score and Drumcondra Reasoning Test scores.

Table 6.2		
Correlations between RAI accuracy score, IQ index scores and DRT Scores		
Measure	Correlation coefficient	Significance level
Drumcondra Reasoning Test		
Verbal Reasoning	.54**	<.001
Numerical Ability	.59**	<.001
Overall Score	.61**	<.001
WASI		
Full Scale IQ	.34*	.04
Verbal IQ	.18	.3
Performance IQ	.42**	.01
* Indicates correlation is significant at the 0.05 level (2-tailed) ** Indicates correlation is significant at the 0.01 level (2-tailed)		

6.3.3 Analysis of post-intervention changes in RAI and DRT scores

In order to analyse the effect of SMART on DRT and RAI test performance, a series of paired-samples t-tests were conducted. In line with Bonferroni procedures, an alpha level of .007 was set for these analyses.

RAI scores increased significantly following relational training, rising over 8.5 points from Time 1 (M = 40.4, SD = 7.6) to Time 2 (M = 48.9, SD = 6.63), $t(151)$ = -6.25, $p < .001$. The Cohen's d indicated that the effect size of this increase was very large (1.36).

In terms of Verbal Reasoning Scores, there was a significant increase in standardised scores from Time 1 (M = 103.4, SD = 14.2) to Time 2 (M = 106.7, SD = 15.01), $t(151)$ = -4.84, $p < .001$ (two-tailed). The Cohen's d indicated that the effect size of this increase was small (0.4). There was also a significant rise in percentile scores on this measure increasing from Time 1 (M = 56.3, SD = 27.3) to Time 2 (M = 62.3, SD = 26.92), $t(151)$ = -4.69, $p < .001$. The Cohen's d indicated that the effect size of this increase was small (0.38).

In terms of Numerical Ability Scores, **t**here was a significant increase in standardised scores from Time 1 (M = 104.7, SD = 14.95) to Time 2 (M = 107.6, SD =

14.56), *t*(151) = - 4.1, *p* < .001 (two-tailed). The Cohen's *d* indicated that the effect size of this increase was small (0.33). There was also a significant increase in percentile scores on this measure from Time 1 (*M* = 58.9, *SD* = 28.43) to Time 2 (*M* = 64.1, *SD* = 27.2), *t*(151) = - 3.84, *p* < .001. The Cohen's *d* indicated that the effect size of this increase was small (0.31).

In terms of Overall Drumcondra Reasoning Test scores, standardised scores, there was a significant increase of approximately 3.5 points from Time 1 (*M* = 104.5, *SD* = 14.6) to Time 2 (*M* = 107.96, *SD* = 14.75), *t*(151) = - 6.25, *p* <.001. The Cohen's *d* indicated that the effect size of this increase was medium (0.51). This increase was also reflected in a percentile score rise of almost 6 percentile ranks from Time 1 (*M* = 58.6, *SD* = 27.63) to Time 2 (*M* = 64.4, *SD* = 26.9), *t*(151) = -5.22, *p* < .001. The Cohen's *d* statistic indicated that the effect size of this increase was small (0.43).

6.3.4 Analysis of DRT score changes across ability levels

In order to further investigate the efficacy of relational skills training, individual paired samples t-tests were conducted to analyse changes in each of the three DRT indices across three ability levels based on initial DRT Overall Reasoning Scores (Below Average, Average, & Above Average: In line with Bonferroni procedures, an alpha level of .006 was set for these analyses. Figure 6.1 displays DRT percentile rank changes for all three groups, as well as the overall sample.

For the Below Average Group (n = 23, DRT Overall Score > 90), **t**here was a significant increase of 5.4 points in DRT Overall Reasoning standardised scores from Time 1 (*M* = 82, *SD* = 7) to Time 2 (*M* = 87.4, *SD* = 7.6), *t*(22) = -1.93, *p* = .001 (two-tailed). While mean Verbal Reasoning scores rose by 4 points from Time 1 (*M* = 84.7, *SD* = 9) to Time 2 (*M* = 88.7, *SD* = 11.2), this rise was not significant. The increase of

5.4 points in Numerical Ability from Time 1 (*M* = 83.1, *SD* = 8.6) to Time 2 (*M* = 88.5, *SD* = 9.2) was also not significant.

For the Average Group (n = 73, DRT Overall Score: 90 – 110), there was a significant increase of 3.8 points in DRT Overall Reasoning standardised scores from Time 1 (*M* = 100, *SD* = 4.7) to Time 2 (*M* = 103.8, *SD* = 7.6), $t(72) = -4.87$, $p < .001$ (two-tailed). Verbal Reasoning scores increased significantly by 3.2 points from Time 1 (*M* = 99.8, *SD* = 8.4) to Time 2 (*M* = 103, *SD* = 9.5), $t(72) = -3.47$, $p = .001$. Numerical Ability scores also rose significantly by 3.6 points from Time 1 (*M* = 100.1, *SD* = 6.98) to Time 2 (*M* = 103.7, *SD* = 8.7), $t(72) = 4.02$, $p < .001$.

For the Above Average Group (n = 56, DRT Overall Score > 110), there was an increase of 2.2 points in DRT Overall Reasoning standardised scores from Time 1 (*M* = 119.7, *SD* = 7.8) to Time 2 (*M* = 121.9, *SD* = 10.1), but this rise was not significant. Verbal Reasoning scores rose by 3.1 points from Time 1 (*M* = 115.7, *SD* = 10.7) to Time 2 (*M* = 118.8, *SD* = 12.2), but this rise was not significant. Finally, Numerical Reasoning Scores increased marginally by less than one point from Time 1 (*M* = 119.7, *SD* = 8.3) to Time 2 (*M* = 120.5, *SD* = 10.4), which did not reach statistical significance.

These results combine to suggest that the SMART system may be most effective in increasing academic performance for individuals within the average range of academic ability, at least for the current age cohort. While DRT score increases were highest for the below average group, there was slightly more variation in the magnitude of participants score changes for two of the three indices as the standard deviation in rises for Verbal Reasoning (9.7) and Numerical Ability (11.4) was slightly higher than that found for the Average group (SD = 7.9 & 7.7 respectively). This increased variance may perhaps account at least partially for the failure for theis score increase to reach statistical significance for the below average group. In comparison, no significant score increase was found for the Above Average group on any DRT index. This is perhaps unsurprising, given the finding that for each DRT index, post-intervention score rises showed weak inverse correlations with pre-intervention ability (Overall Reasoning: $r = -.2$, $p = .01$; Verbal Reasoning: $r = -.17$, $p = .04$; Numerical Reasoning: $r = -.33$, $p < .001$).

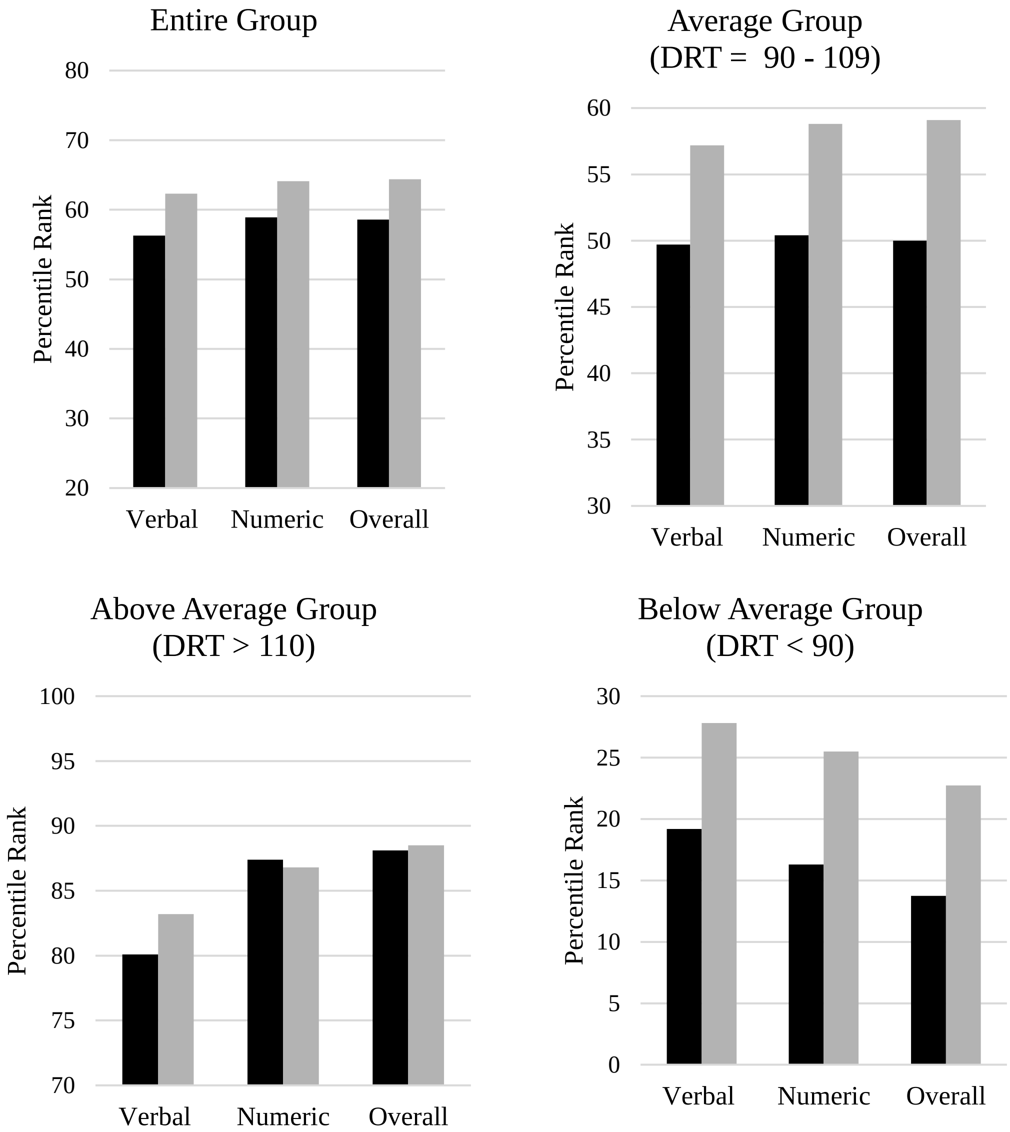


Figure 6.1. Bar chart depicting Drumcondra Reasoning Test Scores before (black) and after (grey) SMART Training.

6.3.5 Analysis of post-intervention changes in WASI IQ scores

For the IQ subsample, 22 of 38 participants completed all 55 training levels (58%), which was similar to the completion rate found for the entire sample (56%). In light of the results of Experiment 3, which clarified the importance of completion of all stages for significant IQ gains (see also Amd & Roche, 2018), only those who completed all 55 training levels will be included in subsequent analyses of IQ score changes. As only 6 participants in Group 2 completed all training levels the remaining samples were highly uneven across conditions and were small in size. In effect, the cross-over design was irreparably compromised. In addition, two participants who completed training failed to complete post-intervention IQ assessments. Therefore, simple pre- to post-intervention score changes for the three IQ indices and four IQ subtests will be studied using paired samples t-tests. For analyses of post-intervention IQ score changes, IQ tests completed immediately before and after the training intervention will be included. For Group 1, this consists of IQ scores at Time 1 and Time 2, while for Group 2, IQ scores at Times 2 and 3 will be analysed. Furthermore, pre- and post-control period IQ score changes will also be assessed, which will analyse IQ scores at Times 2 and 3 for Group 1, and IQ scores at Times 1 & 2 for Group 2 (i.e. before and after their period of no intervention). In line with Bonferroni procedures, an alpha level of .008 is set for these analyses.

Regarding Full Scale IQ, a significant increase of 3.8 points was found from pre-intervention (*M* = 107.6, *SD* = 12.1) to post-intervention (*M* = 111.5, *SD* = 13.9), *t*(19) = -3.1, *p* = .006. The Cohen's *d* effect size for this score change was small (0.3). Following control periods, there was no significant increase in Full Scale IQ with mean scores remaining virtually unchanged between pre-control (*M* = 109.4, *SD* = 14.8) and post-control period test administrations (*M* = 109.7, *SD* = 13.5), *t*(19) = -.24, *p* = .81.

Verbal IQ scores rose significantly by 4.7 points from pre-intervention (*M* = 103, *SD* = 14.5) to post-intervention (*M* = 107.7, *SD* = 15.8), *t*(19) = -3.37, *p* = .003. The Cohen's *d* statistic for this score rise was also small (0.31). Following control periods, Verbal IQ score did not change significantly from pre-control (*M* = 106.7, *SD* = 15.6) to post-control (*M* = 105.8, *SD* = 15.4), $t(19) = .5$, $p = .62$.

Performance IQ scores rose modestly from pre-intervention (*M* = 110.4, *SD* = 10.4) to post-intervention (*M*= 112.6, *SD* = 12), $t(19) = -1.54$, $p = .14$, but this rise did not reach statistical significance. There was an increase in Performance IQ scores following the control periods but this also did not reach statistical significance, with pre-control scores (*M* = 110, *SD* = 114.2) rising at post-control test administration (*M* = 114.2, *SD* = 11.9) $t(19) = - 2.4$, $p = .03$.

Regarding IQ subtests, Similarities was the only IQ subtest to show significant score increases at the current alpha level, rising 4.1 standardised points from pre-intervention (*M* = 49.8, *SD* = 9.3) to post-intervention (*M* = 53.9, *SD* = 9.1), $t(19) = -3.38$, $p = .003$. Scores for the other three IQ subtests, Vocabulary, Block Design & Matrix Reasoning did not change significantly following training.

The current investigation aimed to evaluate the utility of a relational skills training program (SMART) in improving academic ability, as assessed by the Drumcondra Reasoning Test, and intellectual performance, as assessed by the Wechsler Abbreviated Scale of Intelligence. In summary, relational skills training was found to be effective in significantly increasing Verbal Reasoning, Numerical Ability and overall academic ability, as assessed by the Drumcondra Reasoning Test. In addition, relational ability, as measured by the RAI, was found to show moderate-to-strong correlations with each of the three DRT indices: Overall Reasoning ($r = .61$, $p < .001$), Verbal Reasoning ($r = .59$, $p < .001$) and Numerical Ability ($r = .54$, $p < .001$). The SMART system displayed somewhat reduced efficacy in improving academic ability for those at the higher end of the ability spectrum, as significant score increases were found for both below average and average ability groups, but not for above average ability participants.

To complement this finding, post-intervention improvements for each of the DRT indices showed weak inverse correlations with baseline scores for each measure, further indicating that SMART may provide greater benefits for those with average or below average levels of ability. In terms of intellectual performance, small but significant score increases were found for Full Scale and Verbal IQ, but not for Performance IQ.

The current analysis demonstrated significant rises in both standardised and percentile scores for the Drumcondra Reasoning Test, as well as its two subscales (Verbal Reasoning & Numerical Ability) and therefore, represents the second analysis to report such increases on a standardised measure of academic aptitude. Indeed, the nature and intensity of the score increase found in the current analysis bear resemblance to those demonstrated by Cassidy et al. (2016), who reported significant post-SMART increases on a similar scholastic ability assessment, the Differential Aptitude Test (DAT; Bennett, Seashore, & Wesman, 1990) in a smaller (n = 30) and slightly older sample of secondary school students (Mean age = 16.4 years). As such, the current analysis broadly replicates this general effect of SMART on scholastic ability using a much larger sample of students (n = 174). In addition, the scholastic ability assessment in the current investigation exposed participants randomly to one of five alternate, equivalent forms at baseline and follow-up. This virtually eliminates the potential confounding effect of practice, which may have represented somewhat of a caveat regarding the validity of Cassidy et al.'s findings, as the DAT comprises a single assessment format. Furthermore, the academic ability measure utilised in the current study is the latest iteration of the Irish Department of Education's assessment of choice and was standardised using a sample of over 6,000 Irish children in 2016. As such, the modest but consistent improvements found on this scale, proposed to measure the skills deemed by the Department of Education as most essential for academic performance, is extremely encouraging.

The beneficial effect of SMART on DRT Verbal Reasoning scores is predicted by an expansive literature base which proposes that relational skills contribute heavily to the type of tasks being assessed. For example, numerous studies have proposed the foundational importance of relational responding and reasoning to various domains of literacy, such as vocabulary (Cassidy et al., 2010; Colbert et al., 2017; O'Hora et al., 2005; Stewart, et al., 2013), reading (de Rose et al., 1992; Farrington-Flint, Canobi, Wood, & Faulkner, 2007; Goswami, 1986; Mackay, 1985; Sidman, 1971), grammar (Hock, 1991, 2008) and spelling (Brown et al., 1996; Goswami, 1988; Mackay, 1985). The relevance of relational skills to performance on DRT Verbal Reasoning is rendered clearly evident by analysing the topography of the tasks included in this metric. Each of the four-component subtests on this scale can be understood entirely as assessments of relational skill. The Synonyms subtest assesses the participant's ability to select from an array a word that is semantically equivalent to a sample word. From an RFT perspective, this is an assessment of word-word coordination relations (i.e. X means the same as Y). The Antonym subtest is the direct inverse of the Synonym subtest, requiring the participant to identify a word which entails a definition that is the direct opposite of a target word, thereby assessing word-word opposition relations. The Analogy subtest assesses analogical reasoning, which is conceptualised as the derivation of relations between relational premises (e.g. the relation between A and B is the same/opposite as the relation between C and D). Finally, the Classification subtest can be considered an assessment of hierarchical relational responding, as trials consist of identifying which words do and do not belong in a given verbal categorisation (i.e. odd-one-out). Therefore, given the topography of the skills being assessed by the DRT and the generalised skills being trained as part of the intervention, it is unsurprising to discover such a finding. However, that being said, due to the clear relevance of relational skills to performance on DRT Verbal Reasoning subtest items, it may also follow that one may expect score increases of greater magnitude than those witnessed in the current analysis. While this may at least be partially explained by procedural issues, such as low completion rates (see below), this is an issue that perhaps requires further empirical investigation.

In terms of the Numerical Ability scale, two of its subcomponents, Relations among numbers and Sequential Ordering, may also ‘tap’ relational responding skills. The former subtest comprises a traditional number series task, whereby the participant must identify the next number in a sequence ordered in accordance to an unspecified rule (e.g., 1, 4, 7, 10, ?). In order to successfully respond to these tasks, the participant must derive the relation between each number in the sequence, and then apply that ‘rule” to find the missing number. In the above example, each successive number denotes a quantity which is three greater than its predecessor, a rule which can be applied to identify the missing number. In the Sequential Ordering task, participants must organise a collection of numerical quantities, depicted as percentages or fractions. As such, the participant must identify the value of each quantity and then order these quantities in a given sequence (e.g. from lowest to highest), which implicates comparison relations between each quantity. Once again, the similarity and overlap in the nature of the skills being trained by SMART and the skills being tested by the Numerical Ability scale would appear to precipitate the DRT score increases reported.

The potential utility of SMART in improving numerical skills is brought into sharp relief by the most recent PISA report (Shiel, Kelleher, McKeown, & Denner, 2016) which proposed that while Irish students are on average amongst the top performing in the domain of literacy amongst other OECD nations, scores on standardised numeracy tests are less impressive. In addition, this report highlighted that while male students tend to outperform female students on tests of numeracy, the ‘gender gap’ in standardised mathematics scores in Ireland is double the OECD average. As such, SMART may represent a potential resource in improving numerical skills in general, but may also harbour specific implications in facilitating the improvement of mathematical ability in female students in an attempt to reduce this inequality in scores.

While the significant post-SMART improvements on DRT scores show promise, it is important to underline that the magnitude of such rises was relatively small. Therefore, while an extensive literature base may predict that relational skills training may improve academic performance, future studies should investigate the impact of such training on actual school grades, rather than academic aptitude/ability assessments which may merely serve as proxies for grades. Indeed, there is considerable debate regarding the intimacy of the relationship between students’ actual grades and their scores on such academic ability assessments (Strauss et al., 2006).

Regarding the effect of SMART on intellectual performance, the post-intervention score increases found in the current sample are markedly more modest than those reported in previous analyses (e.g. Cassidy et al.,2011, 2016). The most readily accessible explanation for this would be the low level of program completion, as only 58% of the IQ subsample completed all 55 levels. There may be a number of contributing factors to this issue, most notably the finding that participants who presented with lower levels of ability at baseline were less likely to complete the training. Indeed, those who completed training had significantly higher baseline FSIQ scores than those who failed to complete training. This replicates previous findings, such as those reported in Experiment 3 which used a younger sample of 10-11-year olds. It appears, therefore, that even students 2-3 years older may struggle to complete training, at least within a 3-4 month window. While such a restrictive timeframe is necessitated by experimental designs such as the current one, this is not the case for actual non-experimental administrations of SMART within the school context. As such, it is important to elucidate whether the failure to complete training is due to an insufficient number of training sessions or may be in fact due to the rate or increment in task difficulty across stages being too high for younger students. If the latter proves to be true, additional modifications to the main program and the development of remedial programs (such as the SMART:R outlined in Chapter 5) may be necessary to allow students at the lower end of the ability spectrum to successfully complete all 55 training levels and access the benefits training completion can provide.

Another issue which may contribute to explaining this reduced effect may come in the form of procedural issues related to the administration of training. While previous analyses have included blind testing and group allocation, the current design is noteworthy due to the almost complete lack of experimenter engagement in the set-up of training accounts, day-to-day running of training and administration of RAI and DRT assessments. In fact, the only experimenter-student engagement took place during the WASI assessments for the IQ subsample, and even in this case, the experimenter was completely blind to participant names, group allocation and training statistics. While this was done deliberately in an attempt to minimise any potential experimenter bias, this endeavour may have left the experimenter blind to administrative, supervisory and/or procedural issues that may have affected the administration of training, and subsequent training effects. For instance, the member of school staff designated as supervisor to this process recounted issues regarding student's attendance for training sessions and the availability of computer labs for these sessions throughout the year. This issue was particularly prevalent during the second training period, which may have explained the relatively poor level of training completion for Group 2 compared to Group 1. As such, while there is a growing evidence base supporting the efficacy of

SMART in improving intellectual and academic performance, it appears that there remains much more to learn regarding the efficient integration of this program into the school curriculum and school environment.

In conclusion, the current study investigated the effect of SMART in increasing academic ability, as measured by the Drumcondra Reasoning Test, and intellectual performance, as assessed by the WASI. The current results suggest that SMART may offer benefits that extend beyond intellectual performance and into the academic domain, as small, but significant increases were found for DRT Overall Reasoning scores, and for both of its subindices: Verbal Reasoning and Numerical Ability. In addition, results from the IQ subsample showed that SMART was effective in increasing WASI Full Scale and Verbal IQ, but that the magnitude of these increases were considerably lower than those reported in previous analyses. In sum, the current study offers further suggestion of SMART's potential utility in fostering improvements in intellectual and academic performance using perhaps the most rigorous experimental designs and the largest sample to date. While the current results are far more understated than those reported previously, this progression to single-blind, large *n* studies represents an important progression in the current research stream and highlights a number of important issues that require consideration for future analyses.

Chapter 7

An investigation into the relationship between Intelligence Quotient scores and performance on the Multiple Relational Abilities Test

7.1 Introduction

Several authors have now argued that standardised IQ tests can be conceived as tests of DRR proficiency (e.g., Cassidy et al., 2010; Colbert et al., 2017; O'Hora et al., 2005). Indeed, in Experiments 2, 3, 4 and 5 significant correlations were found between relational responding proficiency and intellectual performance, as assessed by traditional IQ metrics. Previous analyses have also identified significant relationships between relational responding and various measurements of intelligence (Colbert et al., 2017; Hayes & Stewart, 2016; O'Hora et al., 2008). For instance, across two separate analyses, O'Hora et al. (2005) and O'Hora et al. (2008) reported significant correlations between performance on a temporal relations task and all three WAIS-III indices (Full-Scale, Verbal, and Performance IQ), two of four WAIS-III subindices (Verbal Comprehension and Perceptual Organisation) and two WAIS-III subtests (Vocabulary and Arithmetic). In addition, a recent study by Dixon, Belisle, Stanley, & Rowsey (2018) found that an assessment of DRR across numerous sensory modalities, the PEAK-E-PA (Promoting the Emergence of Advanced Knowledge Equivalence Pre-Assessment; Dixon et al., 2014) displayed a high level of correlation with performance on two IQ subtests, Vocabulary and Block Design.

The Relational Abilities Index (RAI), developed by Cassidy (2008), is a 55-item syllogistic reasoning assessment which measures proficiency in coordination, opposition and comparison relational responding, and is now regarded as an acceptable proxy measure of IQ (Colbert et al., 2017). Colbert et al. (2017) carried out the most in-depth analysis of the RAI to date, reporting medium-to-strong correlations between RAI scores and all three WAIS-III indices (Full-Scale, Verbal and Performance IQ), all four subscales (Verbal Comprehension, Working Memory, Perceptual Organisation and Processing Speed), as well as 10 of 13 IQ subtests. In addition, in the second of the Colbert et al.'s studies, RAI scores predicted performance various other measures of cognitive ability, including verbal ability (National Adult Reading Test; Nelson, 1982), visuospatial function (the Trail Making Test; Lezak, 1995) and memory (Rey Auditory Visual Learning Tests; Rey, 1958; English version: Taylor, 1959). While the RAI demonstrated considerable predictive validity across the test battery, closer investigation indicated the RAI's relatively limited utility in discriminating performance for high IQ participants due to a potential ceiling effect. As such, the authors concluded that the inclusion of a wider range of relational tasks, such as temporality, perspective-taking and analogy, may be beneficial in parsing out individual differences across a greater diversity of trial types and providing a more comprehensive account of relational ability and how its various aspects relate differentially to various aspects of intelligence.

Despite displaying considerable utility as a proxy measure of IQ, the Relational Abilities Index is somewhat limited in scope, due to the relatively narrow compendium of relational frames included. The three relational frames included in the RAI (coordination, opposition and comparison) were originally selected due to their apparent importance in standardised IQ tests (Cassidy et al., 2010; Stewart et al., 2013), along with their prominence in language acquisition and logistical reasoning (Barnes-Holmes, Barnes-Holmes, Smeets, Cullinan, & Leader, 2004; Hayes et al.,2001). However, a relational abilities index may benefit from comprehensive expansion in order to assess a wider range of relational skills, which have previously been associated with intellectual behaviour, namely distinction, temporality and analogy (Hayes & Stewart, 2016; O'Hora et al., 2005, 2008). Such a development should improve the utility of a relational abilities index in providing a more sensitive and nuanced differentiation of performances, particularly at the higher end of the performance spectrum (Colbert et al., 2017; Gore et al., 2010). In addition, in the event that positive correlations between this wider range of relational frames (e.g. distinction, temporality, analogy) and intellectual performance are found, such frames may be integrated into existing relational skills training interventions (such as the SMART program as administered in Experiments 1, 2, 3 and 5) as a means of building upon such protocol's utility in improving intellectual performance (see Cassidy et al., 2011; 2016).

In order to investigate the relationship between intellectual performance and a wider range of relational skills, scores on the Multiple Relational Assessment Task (MRAT) and the WASI were analysed in a sample of young adults. The MRAT extends upon the RAI by assessing 6 forms of relational responding (coordination, distinction, opposition, temporality, analogy, & perspective taking) across 5 blocks, comprising 78 trials in total. It is hoped that a correlational analysis of MRAT and WASI test scores will allow further elucidation of the contribution of these additional frames to intellectual performance, and potentially aid in identifying the relative influence of each of these frames on traditional assessments of intelligence. By doing this, it may be possible to integrate additional relational tasks into the current RAI format, as a means of providing a more comprehensive account of relational responding, and potentially, improve its utility as a proxy measure of intelligence.

7.2 Method

7.2.1 Participants

Thirty-six participants were selected at random convenience from a population of students at Maynooth University. Participants were recruited through the Department of Psychology's participant pool as well as through an anonymous online form shared on social media. Participants were informed they were not eligible for involvement if they had previously been tested using the Wechsler Abbreviated Scale of Intelligence, or if they had been diagnosed with any developmental disorders. Participants ranged from 18 to 38 years old (*M* = 21 years, 3 months).

7.2.2 Settings and Materials

All assessments took place in a private experimental room with no distractions. WASI administrations were delivered one-to-one, with the experimenter seated at a desk facing the participant. The MRAT was administered using a laptop computer.

7.2.2.1 Wechsler Abbreviated Scale of Intelligence.

The Wechsler Abbreviated Scale of Intelligence administration replicated the protocol employed in in Experiments 2, 4 & 5.

7.2.2.2 Multiple Relation Assessment Procedure.

The Multiple Relation Assessment Procedure (MRAT) is a 78-item assessment which measures an individual's relational responding proficiency in accordance with 6 relational frames across 5 testing blocks: coordination/distinction (Block 1), coordination/opposition (Block 2), temporality (Block 3), analogy (Block 4) and deixis (i.e. perspective-taking, Block 5). Blocks 1-4 comprised of 16 relational tasks, while Block 5 included 14 trials. Before commencing the assessment, instructions are displayed onscreen describing the test procedure and outlining example questions for each testing block. Once participants had read the instructions, they were directed to press a "Continue" button located onscreen in order to begin the assessment. Upon completion of a testing block, this process is repeated, providing participants with instructions and example trials before commencing each new testing block. No time limit was imposed for the MRAT. Sample trials for each of the 5 MRAT testing blocks can be found in Figure 7.1.

For Blocks 1 – 4, three letter (consonant-vowel-consonant) nonsense words are used as relata (e.g. CUG, TOF, FEG etc.). For each testing block, participants are presented with varying numbers of relational premises (e.g. "FEV is the same as TIV") which specify a network comprising a number of relations within a given relational frame (e.g. coordination, distinction, opposition etc.), followed by a relational question (e.g. “Is TIV opposite to FEV?”) which probes for their understanding of the specified relational network. Participants were required to register their answer by selecting one of the binary response options “YES” or “NO” located in the bottom corners of the screen. In order to control for positional responding, the placement of these response options is switched repeatedly throughout the assessment.

Block 1 assesses the participants ability to respond in accordance to coordination and distinction relations by presenting them with two relational premises, specifying “same” and/or “different” relations between three nonsense stimuli (e.g. “CUG is different to TOF, TOF is the same as FEG”), followed by a relational question (e.g. “Is CUG different to FEG?”). Blocks 2 and 3 repeat this general format but specify relations within different relational frames. Block 2 included relational premises which specify relations in accordance to coordination and opposition relations (e.g. “LEW is same as RIF, RIF is opposite to NOQ”), while Block 3 includes temporal relations (e.g. “RUQ is before POY, TOK is after POY”).

Block 4 assesses analogical reasoning proficiency, which refers to the participant’s ability to derive the relations between multiple relational premises. As such, for this testing block, participants were presented with two relational premises, followed by a relational question which probed for the relation between the relationships specified in each premise. These premises included a variety of relational frames: coordination, distinction, comparison (more than/less than) and temporality. For example, “FUD is more than DET, FUJ is less than BIV. Is FUD to DET the same as BIV to FUJ?”. All relational questions asked participants whether the relationship specified in the one premise were “the same as” or “different to” the relationship specified in the other premise.

Finally, Block 5 assessed deictic (perspective-taking) relations, by requiring participants to switch perspectives and to respond in accordance with their new relative 'position'. An example trial would be: "I am sitting here on the blue chair. You are sitting there on the black chair. If I were you and you were me, where would you be sitting?". In addition, this block involved trials which asked to participants to view a situation from a particular temporal position. For example, "Yesterday I was swimming. Today I am running. If now was then and then was now, what would I be doing now?". In all cases, the relational premises involved sitting/standing in a particular location (e.g. on the blue/black/red chair, at the yellow/green/black door, etc.), or performing a particular activity (e.g. swimming, reading, running) and the relational question involved changing either physical ("if here was there" and vice-versa) or temporal positions ("if now was then" and vice-versa). Upon completion, the MRAT provides data on overall response accuracy, accuracy for each individual testing block and response latencies for each of the 78 trials.

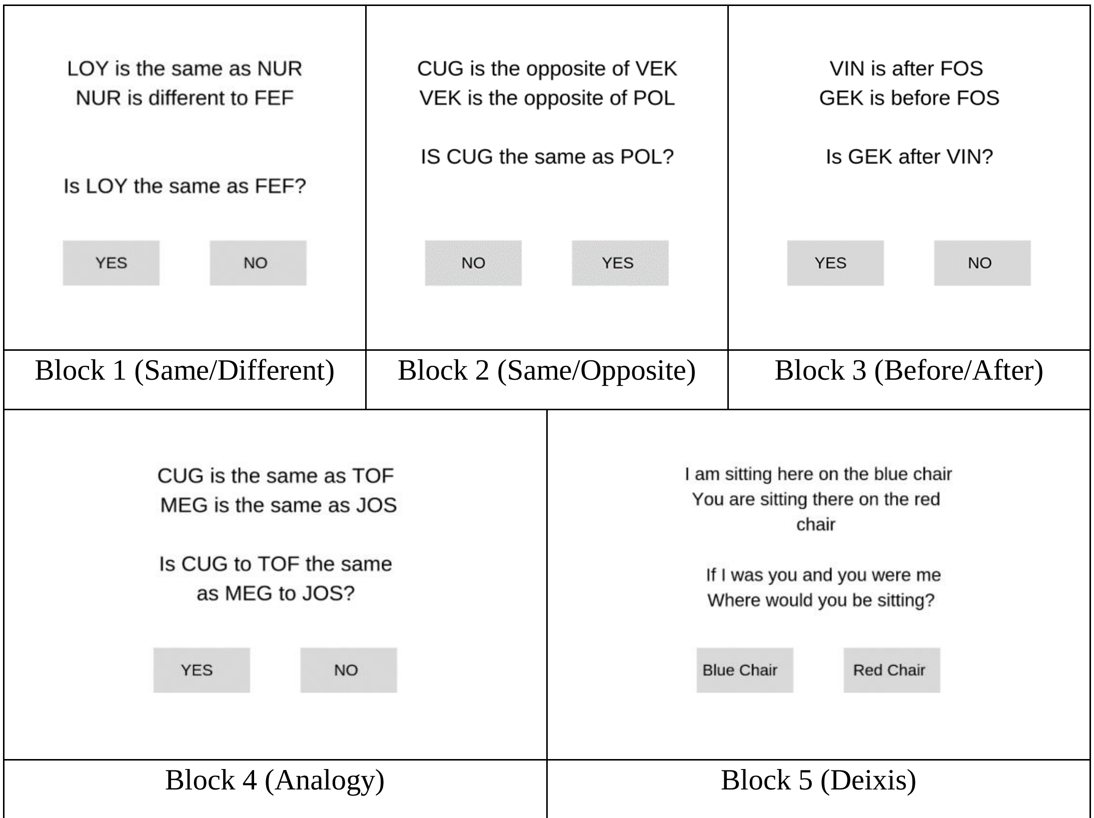


Figure 7.1. Sample relational tasks for each of the 5 MRAT modules.

7.2.3 General Procedure

Once each participant had been briefed regarding the nature of the study, they were asked to sign a consent form indicating that they understood what the study involved and gave permission for the data they provided to be included in this study. All participants were administered the full four subtest battery of the WASI. Following completion of the WASI, each participant was then administered the MRAT using a laptop, which took 20-25 minutes to complete

7.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland. As a study originally designed as an undergraduate research project, it was not required to undergo explicit committee approval but conformed to a checklist of considerations and was supervised throughout by the author.

7.3 Results and Discussion

7.3.1 Descriptive Statistics

Table 7.1 below displays baseline descriptive statistics for WASI IQ scores. Mean scores for Full Scale (M = 113.5, SD = 9.6), Verbal (M = 113. 2, SD = 110.9) and Performance IQ (M = 110.9, SD = 9.8) were all in the high average range. Test of normality indicated that scores for Full-Scale IQ, Verbal IQ, Performance IQ and MRAT were all normally distributed.

Table 7.1

Baseline Descriptive Statistics for WASI IQ scores

Variable	Mean (SD)	Range
Full-Scale IQ	113.5 (9.6)	86 - 131
Verbal IQ	113.2 (9.7)	94 - 134
Vocabulary	57.5 (7.6)	44 - 70
Similarities	57.9 (6.2)	46 - 68
Performance IQ	110.9 (9.8)	79 - 125
Block Design	56.7 (6.4)	39 - 66
Matrix Reasoning	55.6 (8.1)	21 - 66

Mean MRAT score for the current sample was 84.6% (*SD* = 8). Correct responding among the MRAT testing blocks was highest for Same/Different (*M* = 89.6%, *SD* = 13.4) and Same/Opposite (*M* = 88.9%, *SD* = 10.9), closely followed by Analogical (*M* = 87.5%, *SD* = 10.1) and Before/After (*M* = 85.8%, *SD* = 13.8). The only MRAT module in which mean correct responding was below 85% was the deictic block (*M* = 71%, *SD* = 15.9), in which participants were considerably less accurate, in comparison to each of the other MRAT blocks. Full descriptive statistics for the MRAT and its assessment blocks are displayed in Figure 7.2.

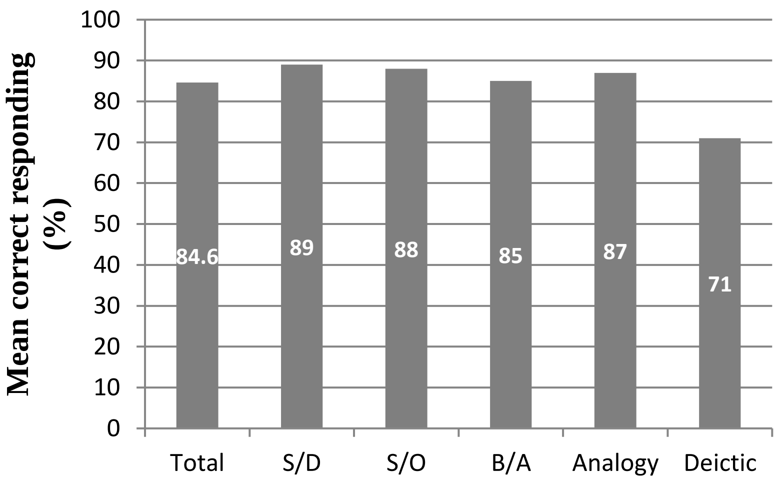


Figure 7.2. Bar charts displaying mean correct responses for total MRAT score and each of the 5 individual testing blocks.

7.3.2 Correlational Analysis

Regarding intercorrelations between overall MRAT score and each of the 5 individual assessment blocks, each relational frame demonstrated moderate-to-strong significant correlations with overall MRAT score. Of the five frames, before/after (r = .67, p < .001) and same/opposite (r = .6, p < .001) showed the closest relationship with overall relational performance, followed by the deictic (r = .56, p < .001), same/different (r = .54, p = .001) and analogical blocks (r = .53, p < .001). Internal consistency, as indicated by Cronbach's alpha statistic (.54), was quite low.

7.3.3 WASI IQ and MRAT Performance

Table 7.2 displays correlations between total MRAT score and each of the WASI's indices and subtests. MRAT scores displayed moderate-to-strong significant correlations with Full-Scale IQ (r = .63, p < .001) and Performance IQ (r = .67, p < .001), and a moderate relationship with Verbal IQ (r = .44, p = .007). In addition, MRAT score predicted scores on both Performance IQ subtests, Block Design (r = .46, p < .001) and Matrix Reasoning (r = .53, p = .001). In terms of Verbal IQ subtests, MRAT scores also correlated significantly with scores for the Vocabulary subtest (r = .4, p = .02), but not the Similarities subtest.

Table 7.2

Correlations between baseline MRAT scores and IQ indices and subtests

Variable	Correlation coefficient	Significance level
Full Scale IQ	.63**	<.001
Verbal IQ	.44**	.007
Vocabulary	.4*	.02
Similarities	.31	.07
Performance IQ	.67**	<.001
Block Design	.56*	.<.001
Matrix Reasoning	.53**	.001

* Indicates correlation is significant at the 0.05 level (2-tailed)

** Indicates correlation is significant at the 0.01 level (2-tailed)

In order to glean a better understanding of the relevance of individual relational frames to the various domains of intelligence as assessed by the WASI, the degree of correlation between each of the MRAT's 5 testing blocks and IQ index and subtest scores was analysed. Correlational analyses indicated that the Same/Opposite block showed the closest relationship to Full-Scale IQ (r = .63, p < .001), followed by blocks assessing Same/Different (r = .47, p = .004), Before/After (r = .39, p = .018) and Analogical responding (r = .34, p = .046). Scores for the Deictic block did not correlate with Full-Scale IQ. Verbal IQ showed significant correlations with Same/Opposite (r = .43, p = .008) and Same/Different (r = .38, p = .02), but not with the other three relational blocks. Performance IQ showed significant correlations with Same/Opposite (r = .63, p < .001), Before/After (r = .49, p = .002) and Same/Different trials (r = .47, p = .004), but not with Analogy or Deictic trials. As such, performance on Same/Opposite trials not only correlated with all three WASI IQ indices, but also showed the strongest degree of correlation in each case. The Same/Different block also correlated with each index, albeit with weaker effect sizes.

In terms of Verbal IQ subtests, Vocabulary scores were found to display significant correlations with Same/Different trials (r = .36, p = .03), but none of the other 4 testing blocks. Similarities subtests scores correlated significantly with scores for Same/Opposite trials (r = .41, p = .01), but none of the other testing blocks. In terms of Performance IQ subtests, Block Design scores were predicted by Same/Opposite (r = .66, p < .001) and Before/After trials (r = .37, p = .03), but not Same/Different, Analogy or Deictic trial performance. Matrix Reasoning scores correlated with scores for Same/Different (r = .59, p < .001), Same/Opposite (**r** = .37, p = .03) and Before/After trials (r = .44, p = .008).

Of note is the finding that participant's scores for Deictic responding did not correlate with any IQ index or subtest score in the current analysis. In fact, if performance on Deictic trials was removed from overall MRAT score, the resulting 4-block MRAT score (referred to henceforth as MRAT-4; M = 56.3 out of a possible 64, SD = .51) demonstrated greater utility in predicting Full Scale IQ (r = .69, p <.001), Verbal IQ (r = .48, p = .003) and Performance IQ (r = .71, p < .001) when compared to the original 5-block MRAT Score. Correlations between this MRAT-4 score also retained the original significant correlations for three IQ subtests, with similar effect sizes for Vocabulary (r = .39, p =.02) and Block Design (r = .52, p = .001), but greater effect size for the Matrix Reasoning subtest (r = .64, p < .001). Finally, unlike the original MRAT score, by removing Deictic trials, MRAT-4 score correlated significantly with the Similarities subtest (r = .41, p = .01). As such, the MRAT-4 score correlated, at least to a moderate extent to all 7 WASI IQ index and subtest scores. Table 7.3 displays the comparative efficacy of the full MRAT and MRAT-4 in predicting IQ index and subtest scores.

Table 7.3

Correlations between baseline MRAT and MRAT-4 scores and IQ indices and subtests

	MRAT		MRAT-4	
Variable	Correlation coefficient	Significance	Correlation coefficient	Significance
Full Scale IQ	.63*	<.001	.69**	<001
Verbal IQ	.44**	.007	.48**	.003
Vocabulary	.4*	.02	.39*	.02
Similarities	.31	.07	.41*	.01
Performance IQ	.67**	<.001	.71**	<.001
Block Design	.56**	<.001	.52**	.001
Matrix Reasoning	.53**	.001	.64**	<.001

* Indicates correlation is significant at the 0.05 level (2-tailed)

** Indicates correlation is significant at the 0.01 level (2-tailed)

7.3.4 Analysis of predictive validity of MRAT and MRAT-4 scores for high IQ participants

Due to the presence of a potential ceiling effect, and its entailed effect on the efficacy of the MRAT to predict IQ at the upper end of ability, an additional correlational analysis was conducted to investigate the relationship between MRAT score and IQ index and subtest score for high IQ participants (FSIQ = 110+, n = 25). Table 7.4 below displays correlation coefficients found between MRAT and MRAT-4 scores and each IQ index and subtest score for high IQ participants.

Results suggest a general reduction in the predictive validity of the MRAT (and MRAT-4 subindex) for high IQ participants, as the strength of the relationship between FSIQ, VIQ and PIQ is significantly reduced for both MRAT and MRAT-4 scores. MRAT scores showed significant correlations for FSIQ (r = .46, p =.02), PIQ (r = .51, p = .009) and Matrix Reasoning scores (r = .48, p = .01), but the strength of these correlations are considerably weaker than those reported for the overall sample.

Similarly, MRAT-4 predicted FSIQ (r = .3, p = .007), PIQ (r = .48, p = .02), Vocabulary (r = .41, p = .04) and Matrix Reasoning (r = .46, p = .02). However, as was the case with total MRAT score, with the exception of Vocabulary scores, the strength of the relationships between MRAT-4 and each of these IQ metrics was diminished for the high IQ group. Such results combine to indicate that the MRAT may be a less effective proxy for general intellectual performance for those at the higher end of ability

Table 7.4

Correlations between baseline MRAT and MRAT-4 scores and IQ scores for high IQ participants

	MRAT		MRAT-4	
Variable	Correlation coefficient	Significance	Correlation coefficient	Significance
Full Scale IQ	.46*	.02	.53**	.007
Verbal IQ	.21	.3	.34	.1
Vocabulary	.33	.11	.41*	.04
Similarities	-.05	.8	.05	.81
Performance IQ	.51**	.009	.48*	.02
Block Design	.27	.19	.24	.26
Matrix Reasoning	.48*	.01	.46*	.02

* Indicates correlation is significant at the 0.05 level (2-tailed)

** Indicates correlation is significant at the 0.01 level (2-tailed)

The purpose of the current study was to investigate the utility of an extended relational skills assessment, the MRAT, in predicting performance on a widely-administered assessment of intellectual function, the Wechsler Abbreviated Scale of Intelligence. In summary, the current investigation found that total MRAT score showed moderate-to-strong correlations with each of the three WASI IQ indices (Full Scale, Verbal & Performance IQ), and moderate strength correlations with scaled scores for three of four IQ subtests (Vocabulary, Block Design & Matrix Reasoning).

Further investigation analysed the relationship between domains of intellectual performance and proficiency for each of the relational frames assessed by the MRAT. Such analyses indicated that of the 5 relational blocks included, performance on Same/Opposite (Coordination/Opposition) and Same/Different (Coordination/Distinction) were the best predictors of general intellectual performance, with significant correlations for each of the three WASI IQ index scores. Before/After scores were found to correlate significantly with Full-Scale IQ and Performance IQ, but not Verbal IQ. Analogy scores correlated significantly with Full-Scale IQ only. In contrast, Deictic trial scores were found to show no significant correlation with any IQ index or subtest score. In light of such a finding, the removal of deictic trials from overall MRAT score (i.e. MRAT-4 score) increased the strength of correlations between relational task performance and IQ index scores.

Given that the Colbert et al., (2017) correlational analysis of the Relational Abilities Index is perhaps the most comprehensive psychometric investigation into the relationship between relational skills and intelligence, it is prudent to place the current results into the context of their analysis. The Relational Abilities Index, as administered in Colbert and colleagues study, is a 55-item assessment of relational responding in accordance with the frames of coordination/opposition and comparison. While the MRAT administered in the current analysis assesses a wider range of relational frames, across a greater number of trials (78 trials), it is illuminating to note that the RAI demonstrated a greater utility in predicting intellectual performance as assessed by the WAIS-III. For example, the correlations reported for the RAI with Full Scale (r = .74) and Verbal IQ (r = .78) are considerably stronger than those found for MRAT in the current analysis (r = .63 & .44 respectively).

There are numerous possible explanations for this reduced predictive validity. Firstly, it may be the case that the differing levels of correlation between IQ and each relational skills assessment may be due to differences in the type of relational frames included in each assessment and the relative relevance of such frames to intellectual performance. While the MRAT assesses proficiency in a wider range of relational frames, it may be the case that the RAI assesses the relational frames which are more (or most) relevant to intellectual performance, at least as assessed by traditional IQ metrics. This point is strengthened by the finding that of the MRAT's 5 relational blocks, coordination/opposition tasks, which are assessed in detail in the RAI, showed the closest relationship to overall intelligence. In addition, despite its extended battery, trials that specifically assess More/Less responding are not included in the MRAT (although some Analogy block tasks include more/less relational premises), unlike the RAI. More than/Less than trials, considered an assessment of the frame of comparison (a categorisation shared with Before/After trials), have been shown to be of fundamental importance to a range of intellectual tasks (see Colbert et al., 2017), and therefore their absence may at least partially explain the reduced utility of the MRAT in this regard. While the MRAT's Before/After trials can also be considered as assessments of comparison relational responding, it may be the case that proficiency in responding to such trials is a form of comparison responding that is less relevant to overall intellectual performance. Additionally, it may be the case that the content of the WASI 'taps' more/less responding to a greater extent than before/after responding. This question requires empirical investigation, which will be conducted in Experiment 7.

Secondly, it may be proposed that while the MRAT offers a greater breadth of relational frame trials, this may entail a reduced level of depth in terms of the complexity of the trials included, as well as the nuance with which relational responding can be measured. For instance, while the MRAT includes five testing blocks in comparison to the RAI's two, the length of such blocks (14-16 trials) is much shorter than that of the RAI (26 & 29 trials). This extended length and complexity may feasibly offer a more sensitive assessment of individual performance, and therefore provide a more comprehensive account of individual differences. Indeed, extended block length may have been of benefit in the current analysis, as there was a clear clustering of MRAT scores towards the scale's upper limit, thereby reducing overall score variance and the likelihood of identifying significant correlations (see Goodwin & Leech, 2006).

However, the MRAT demonstrated greater utility in predicting Performance IQ (r = .67) when compared to the RAI (r = .55). This indicates that due to the wider range of relational frames (i.e. distinction, temporality, analogy and deictic) assessed, the MRAT was better able to provide a metric of skills which may be relevant to Performance IQ task performance. Correlational analyses indicate that of these additional frames, Before/After scores (i.e. temporal relations) were significantly correlated with Performance IQ (r = .47, p = .004) and both of its subtests: Block Design (r = .37, p = .03) and Matrix Reasoning (r = .44, p < .008). In addition, Same/Different scores (coordination/distinction relations) predicted scores for Performance IQ (r = .47, p = .004) and one Performance IQ subtest: Matrix Reasoning (r = .59, p < .000). These results would, therefore, implicate the potential relevance of temporal and distinction relations to Performance IQ test items in particular, and more generally, the domain of intellectual skills such test items are intended to assess, e.g. non-verbal fluid reasoning, perceptual organisation and abstract conceptualisation (Wechsler, 1999). As such, given the RAI's tendency to correlate more closely with Verbal rather than Performance IQ scores (see Colbert et al., 2017), the inclusion of such relational frames into a modified RAI may provide a more comprehensive proxy measure of intellectual performance.

A key limitation of the RAI is that while it displayed a close relationship with IQ test performance for the Colbert et al. (2017) sample as a whole, it did not predict performance for high IQ (FSIQ = 110+) participants, most likely due to a clear ceiling effect in RAI scores. In the current analysis, while the MRAT score distribution was indeed skewed to the left and ceiling effects appeared to reduce the strength of correlations between MRAT scores and FSIQ scores, these correlations remained significant for participants in this high-ability bracket. While the MRAT's reduced number of trials per relational frame would appear to make ceiling effects more likely, it may be the case that the extended number of frames assessed aided in delineating scores at higher levels of ability. This preliminary finding offers a potential solution to the ceiling effects previously found when administering the RAI to high-ability participants, and as such, the integration of a greater number of relational frames into the current RAI will be investigated in the following chapter.

Upon closer investigation of the individual correlations between intellectual performance and each of the relational frames assessed, it was found that performance on the Same/Opposite block was the best predictor of Full-Scale, Verbal and Performance IQ scores. Such a finding is predicted by theoretical (Cassidy et al., 2010) as well as empirical work (Colbert et al., 2017), as the relational frames being assessed by this block, coordination and opposition, are deemed to be among the most basic, foundational and important frames in regard to intellectual and linguistic performance (Stewart et al., 2013). Indeed, the relevance of such frames is clearly evident by analysing the type of questions included in the WASI, as well as many other gold-standard psychometric measures of intelligence. For example, the Vocabulary subtest can also be viewed as an assessment of coordination relations, as participants must simply express the meaning of a given word, by re-expressing it in different, but synonymous terms (i.e. probing for word-word coordination relations). As these WASI Verbal IQ subtests clearly implicate relational responding repertoires, it is perhaps unsurprising to find significant correlations between such test items and MRAT scores.

Upon analysis of the types of tasks included in the WASI assessment, it comes as a surprise that of the four WASI subtests, the only subtest which did not show a significant correlation with overall MRAT performance was the Similarities subtest. The Similarities subtest is essentially an assessment of word-word and object-object coordination relational responding and/or hierarchical relational responding, as participants are required to identify how one word/concept/object is similar to another or which verbal category both can be considered members of (e.g. "how is a cow similar to a bear?"). While performance on coordination/opposition relational tasks correlated with moderate strength with Similarities scores, a significant correlation between performance on this subtest and overall MRAT performance could be expected. In explaining this lack of effect, it may be the case that only coordination relational responding specifically is relevant to performance on this subtest and that the additional frames exert no influence on this task. Therefore, only coordination relational task performance specifically correlated with Similarities subtest scores, but this effect washed out in the context of overall performance on all 5 relational blocks. In addition, it may be the case that hierarchical responding (which was not assessed by the MRAT), may be of greater importance than coordination relations to this type of task, as many Similarities test items can be correctly answered by providing a categorisation common to both stimuli. For example, a Similarities test item such as "how are a plane and a bus alike?" can be answered correctly by stating that both are types of transport. Indeed, in order to score highly on this subtest, a participant must identify an overarching, more general similarity between two items (e.g. group categorisation) rather than surface-level, physical similarities (e.g. both a plane and a bus have wheels), which may implicate hierarchical responding moreso than coordination responding.

Another noteworthy result of the current analysis is the finding that one of the five testing blocks, deictic relations, did not correlate significantly with any IQ index or subtest. In fact, the mean MRAT-4 score, which summed total performance across the four frames which correlated with Full-Scale IQ (i.e. Same/Opposite, Same/Different, Before/After & Analogy) thereby removing deictic tasks from participants' overall score, improved upon the utility of the original 5-block MRAT score in predicting intellectual performance for both average and above-average IQ participants. There are a number of potential implications of this result, chiefly that perspective-taking relational responding may not be of great relevance to intellectual behaviour, at least as assessed by traditional IQ assessments. Previous research would appear to contradict such an assertion, as a small number of studies have found that intelligence and the ability to embody different spatial, temporal and personal positions may be related (Gore et al., 2010; Rehfeldt, Dillen, Ziomek, & Kowalchuk, 2007; Tarshis & Shore, 1991). However, it may be the case that this relationship is mediated by general relational responding proficiency, rather than a direct influence of deictic relational responding *per se*. Indeed, theoretically, such perspective-taking tasks bear little resemblance topographically to most IQ test items. Future analyses should aim to investigate this relationship more closely, by assessing the relationship between deictic relational responding and intellectual performance, while controlling for the influence of proficiency in other forms of relational responding.

In contrast to the above, it may be suggested that, given the remit of the specific IQ assessment administered, it may have been unlikely to find a correlation between IQ and deictic relational responding proficiency. Whether this simply reflects deictic relational responding's irrelevance to the specific IQ test administered, or to intellectual performance more generally, is open to debate. However, it must be noted that perspective-taking does not appear, at least in any great capacity, to be within the remit of many traditional assessments of intelligence, such as the WAIS, WASI or WISC batteries. For example, even in the case of the more comprehensive WAIS-III, perspective-taking could not be considered a core skill required for any of its 14 subtests (although a very limited number of questions in the commonly-included Comprehension subtest may involve empathetic perspective-taking – e.g. "Why should you apologise if you have hurt someone?" or "Why should you keep a promise?").

The Wechsler Block Design subtest would appear to be the task which is most likely to 'tap' deictic relational responding, as it may require participants to engage in object-based spatial transformations, in which participants reorient and reorganise a number of coloured blocks to match to a sample arrangement. However, this form of mental reasoning has been suggested as a process related to, but clearly distinct from, ego-centric perspective transformation (as employed in most deictic responding tasks; Bryant & Tversky, 1999; Hegarty & Waller, 2004; Kozhevnikov, Motes, Rasch, & Blajenkova, 2006). This distinction has received support from neuroimaging studies, which propose that egocentric spatial transformation is underpinned by activation in the left parietal-temporal-occipital junction whereas object-based transformation is associated by activity in inferior and posterior parietal areas (Zacks, Vettel, & Michelon, 2003).

In conclusion, the current analysis found that scores on the MRAT were significantly correlated with all three WASI indices, indicating the relevance of a wider range of relational responding to general intellectual performance. In terms of the individual relational frames, it was found that Full-Scale IQ was predicted by performance on the Same/Opposite, Same/Different, Before/After and Analogy testing blocks, but not Deictic trials. Verbal IQ showed significant correlations with Same/Opposite and Same/Different block scores, with Performance IQ score showing a significant relationship with these two relational frames, alongside Before/After scores. In comparison to the two-frame RAI, the MRAT demonstrated an increased utility in predicting Full-Scale and Performance IQ scores for high-ability participants. As such, the current investigation proposes that the inclusion of additional relational task types, specifically Same/Different, Before/After and Analogy, may extend the utility of the RAI as a proxy measure of intellectual performance. Therefore, Experiment 7 will assess the validity of a modified version of the RAI, the RAI+, which includes these three additional frames, in predicting intellectual performance. In addition, in order to build upon the findings of Experiment 5, further analysis of the relationship between relational skill and the academic domains of literacy and numeracy will be investigated using this newly-developed assessment.

Chapter 8

The Relational Abilities Index+: An investigation into the relationship between relational skills and measures of intellectual performance, academic attainment and numeracy

8.1 Introduction

In the previous chapter, the influence of the wider range of relational frames on intellectual performance was assessed by administering the MRAT, an assessment of coordination, opposition, distinction, temporal and deictic relational responding, and the WASI, a gold-standard IQ measure. This was done in order to investigate whether the addition of a greater range of relational frames may improve the utility of the Relational Abilities Index as a proxy measure of intelligence. The results of this study indicated that while coordination and opposition responding proficiency were the strongest predictors of performance on many IQ subindices and subtests (and are currently assessed by the RAI), distinction, temporal and analogical relational also displayed significant correlations with such IQ metrics. As such, in light of such findings, the Relational Abilities Index+ has been devised to expand the remit of the original two-block RAI (assessing coordination/opposition and comparison) by including an additional three relational frames: distinction, temporality, and analogy. The primary aim of this analysis, therefore, is to assess the relationship between this new measure of relational responding and intellectual performance as assessed by the WASI. Secondly, the relevance of relational responding to academic attainment is relatively unelucidated, and therefore, the current analysis will also investigate the degree of correlation between the RAI+ and a measure of academic attainment, the WIAT-T.

Indeed, several studies have revealed high levels of correlations between measures of relational responding and various tests of verbal ability (Barnes, McCullagh, & Keenan, 1990; Colbert et al., 2017; Dugdale & Lowe, 2000). In the domain of academic attainment specifically, relational responding has been shown to be of key importance to reading (de Rose et al., 1992; Farrington-Flint & Wood, 2007; Goswami, 1986; Mackay, 1985; Sidman, 1971), vocabulary (Edwards et al., 2011; McHugh et al., 2004; Nippold & Sullivan, 1987), grammar (Hock, 1991, 2008) and even spelling (Brown et al., 1996; Goswami, 1988; Mackay, 1985). Such findings would appear to highlight the importance of an established proficiency in relational responding as a key contributor to literacy. Critically, RFT has produced dozens of studies which have shown that DRR interventions usher in language ability (e.g., Cowley, Green, & Braunuing-Mcmorrow, 1992; Cullinan, Barnes, Hampson, & Lyddy, 1994; de Rose et al., 1992; Matos & d'Oliveira, 1992; Murphy & Barnes-holmes, 2009; Murphy & Barnes-Holmes, 2010; Murphy et al., 2005; Wulfert & Hayes, 1988), and so it is widely argued that the former produces the latter rather than vice versa (Barnes-Holmes, Finn, McEnteggart, & Barnes-Holmes, 2018).

Sophistication in relational responding may also comprise a key facet of numeracy and mathematical fluency (Carpenter et al., 2003). Molina et al. (2005) found that the encouragement of "relational thinking" (i.e., analysing the relationships specified in mathematical problems before engaging in mathematical computation) afforded a meaningful and comprehensive learning of arithmetic and provided a foundational basis for the study of algebra in a sample of primary school children. Indeed, the mathematical symbols which receive such focus in these relational thinking interventions are conceived of as contextual cues from an RFT perspective. In addition, several RFT studies conducted by Ninness and colleagues (McGinty et al., 2012; Ninness et al., 2006, 2009; Ninness, Rumph, McCuller, Harrison, et al., 2005; Ninness, Rumph, McCuller, Vasquez, III, et al., 2005) have demonstrated the utility of training students to derive relations as a means of improving advanced mathematical skills.

The Relational Abilities Index+ (RAI+), which assesses performance across five modules of relational responding (Same/Opposite, More/Less, Same/Different, Before/After and Analogy), was developed for the purpose of this study. The current study aims to investigate the validity and utility of the RAI+ by assessing its degree of correlation with well-established assessments of intelligence (Wechsler Abbreviated

Scale of Intelligence; WASI), numeracy (Wechsler Adult Intelligence Scale: Arithmetic subtest), literacy and educational attainment (Wechsler Individual Achievement Test: Teacher Edition). In addition, the relative contribution of each of the five relational frame test modules will be considered.

8.2. Method

8.2.1 Participants

A total of 97 individuals (50 female) participated in this study. Participants' ages ranged from 18 to 45 (M = 25.4 years). All participants were fluent English speakers with no incidence of any cognitive disorders or impairments which could have impacted the current results. The vast majority of participants (n = 85) were attending third-level education across a range of disciplines at the time of participation.

8.2.2 Materials

8.2.2.1 Relational Abilities Index+.

A revised version (termed the RAI+) of the Relational Abilities Index employed in Colbert et al. (2017) and Cassidy et al. (2016) was administered through the website proprofs.org to assess participants' relational abilities. The RAI+ consists of a battery of 67 syllogistic relational puzzles, assessing proficiency in responding in accordance with Same/Opposite (15 trials), Same/Different (14 trials), More/Less (13 trials), Before/After (13 trials) and Analogy (12 trials) frames in that order. The RAI+ required approximately 20 to 25 minutes to complete. Figure 8.1 displays sample tasks for each RAI+ module.

The general format of trials utilized in the RAI+ mirrored that of the original RAI. Each task consisted of between one and three relational premise(s) in which relations between nonsense words were stated (e.g. "CUG is the same as TOF"), followed by a question based on the relationship(s) specified in the premise(s) (e.g., “Is TOF the same as CUG?”). A total of 227 stimuli comprised of three-letter nonsense words (e.g. ‘CUG’, ‘TOF’, ‘JOS’) in the format “consonant-vowel-consonant” (to ensure pronounceability) were presented with no stimulus being repeated throughout the assessment. Participants indicated their response by using the computer mouse to click on either a ‘YES’ or ‘NO’ button onscreen. Positional responding was controlled for by switching the positions of the response options throughout the assessment. A countdown timer was also visible on the page at all times, imposing a limit of 34 minutes to complete the assessment (i.e. approx. 30 seconds per question).

LUZ is the same as TIV Is LUZ opposite to TIV?	SAJ is different to LIR LIR is the same as VUS Is VUS the same as SAJ?
Block 1: Same/Opposite	Block 2: Same/Different
NOG is less than HAV HAV is less than WUQ WUQ is less than VUN Is HAV more than VUN?	LOF is after FEH WUC is after LOF LON is after WUC Is FEH before LON?
Block 3: More/Less	Block 4: Before/After
QUD is less than KON JOL is more than JIT Is KON to QUD the same as JIT to JOL?	REG is same as CAS WIX is opposite to DOM Is CAS to REG different from DOM to WIX?
Block 5: Analogy	

Figure 8.1. Sample relational tasks for each of the 5 RAI+ modules.

Task complexity was therefore controlled by modifying; 1) the number of relational premises (1-3); 2) the order of relational premises (sequential or random); 3) the directionality of the relational question (i.e., whether or not the relational question probes for first term-last term relations, or last term-first term relations as specified in the premises); 4) the number of relation types presented in each trial (e.g., only "same" relations, or a combination of "same" and "opposite"); and 5) the presence/absence of the relational cue used in the question in the relational premise(s), (e.g., "CUG is same as LER, is CUG same as LER?").

With the exception of Analogy trials, the first trial for each relational frame included a single premise, followed by a relational question and as such assessed the participant's ability to derive mutually-entailed relations. This involved either changing the directionality of the relational statement or switching the relational frame to its inverse in the relational question. Each block then progressed to 10 two-premise trials which included three relata, in which every possible derived relation within this network was probed for. Finally, each block then included a number of three-premise trials (4-6) which specified a relational network across four relata. Any additional relations entailed by the presentation of the fourth relational premise (e.g., between stimulus A/B/C and stimulus D, and vice versa) were assessed during these trials. For the Analogy block, 12 two-premise trials were included. Each premise stated the relation between two stimuli in accordance with same/opposite (four trials), before/after (four trials) and more/less (four trials), followed by a "same/different" relational question which probed for relationship between each of relational premises specified (e.g. "is FEG to TID the same as VER to RUF?").

8.2.2.1 Wechsler Abbreviated Scale of Intelligence.

The Wechsler Abbreviated Scale of Intelligence administration replicated the protocol employed in in Experiments 2, 4, 5, & 6.

8.2.2.3 Wechsler Individual Achievement Test

The Wechsler Individual Achievement Test Second UK Edition for Teachers (WIAT-II-T; Pearson Education, 2006a), an adaptation of the traditional WIAT-II battery, is an individually-administered standardised assessment of educational attainment and is comprised of three subtests of Reading, Spelling and Reading Comprehension. The WIAT-II has satisfactory construct, content and criterion validity as well as test-retest reliability for an adult population (Pearson Education, 2006b).

8.2.2.4 WAIS-III: Arithmetic

The Wechsler Adult Intelligence Scale (WAIS-III: UK; Wechsler, 1998) is an individually administered assessment of intellectual ability. It is one of the most popular IQ measures and is often considered a "gold standard" of intelligence testing (Butler, Retzlaff, & Vanderploeg, 1991; Ivnik et al., 1992; Strauss, Sherman, & Spreen, 2006). The Arithmetic subtest of the WAIS-III comprises one part of the Working Memory subindex of Verbal IQ and consists of 20 arithmetic questions which successively increase in difficulty and are subject to a time limit. Normed tables for this subtest are available for the computation of a standardised score.

8.2.3 General Procedure

The study was conducted in a private experimental room, free from noise and other distracting stimuli. Participants were seated at a desk directly opposite the researcher and were required to provide valid consent before participation. Each participant was engaged in the task individually, on a one-to-one basis with the researcher. Participants were briefed on the general nature of the study and signed a consent form at this first stage. While all participants completed the RAI+, a section of the sample completed this alongside the WASI (n=60) and another subsample completed this alongside the WIAT-T (n=37).

8.2.4 Ethics

The current study was conducted in adherence to guidelines specified by Maynooth University's Social Research Ethics Committee and the Psychological Society of Ireland. As a study originally designed as an undergraduate research project, it was not required to undergo explicit committee approval but conformed to a checklist of considerations.

8.3. Results & Discussion

8.3.1 Descriptive Statistics

Mean RAI+ scores for the current sample was 59.82 out of 67 (89.3%), with scores ranging from 41 to 67. Table 8.1 details full descriptive statistics for individual RAI+ scores, WASI subindices, WIAT-T-II scores and WAIS-III Arithmetic scores.

Table 8.1

Descriptive statistics for RAI+, WASI, WIAT-T II and WASI-III Arithmetic scores

Measure	M	SD	Range
RAI+	89.3%	9.2	61.2 - 100%
Same/Opposite	88.9%	12.3	40-100%
Same/Different	88.2%	8.8	40-100%
More/Less	80%	11.8	46.2 – 100%
Before/After	76%	14	38.5 – 100%
Analogy	65.4%	15.1	16.7 – 100%
WASI			
Full Scale IQ	109.6	9.2	89-128
Verbal IQ	109.6	10.7	88-126
Performance IQ	107.2	9.4	88-126
WIAT	108.8	6.1	96-119
Reading Comprehension	102.9	7.9	78-115
Reading	113.2	4.5	99-119
Spelling	110.3	9.4	87-125
WAIS Arithmetic	126.3	17.5	85-145

In relation to the WASI, scores for Full-Scale IQ (*M* = 109.6, *SD* = 9.2), Verbal IQ (*M* = 109.6, *SD* = 10.7) and Performance IQ (*M* = 107.2, *SD* = 9.37) were all towards the upper limit of the average range. For the WIAT-II-T, scores for WIAT Overall standardised (*M* = 108.8, *SD* = 6.1), Reading (*M* = 102.9, *SD* = 7.9), Reading Comprehension (*M* = 113.2, *SD* = 4.5) and Spelling (*M* = 110.3, *SD* = 9.4) were all in the average to above average range. For the WAIS Arithmetic subtest, standardised score estimates (converted from scaled scores) ranged from 85 to 145 (*M* = 126.32, *SD* = 17.5).

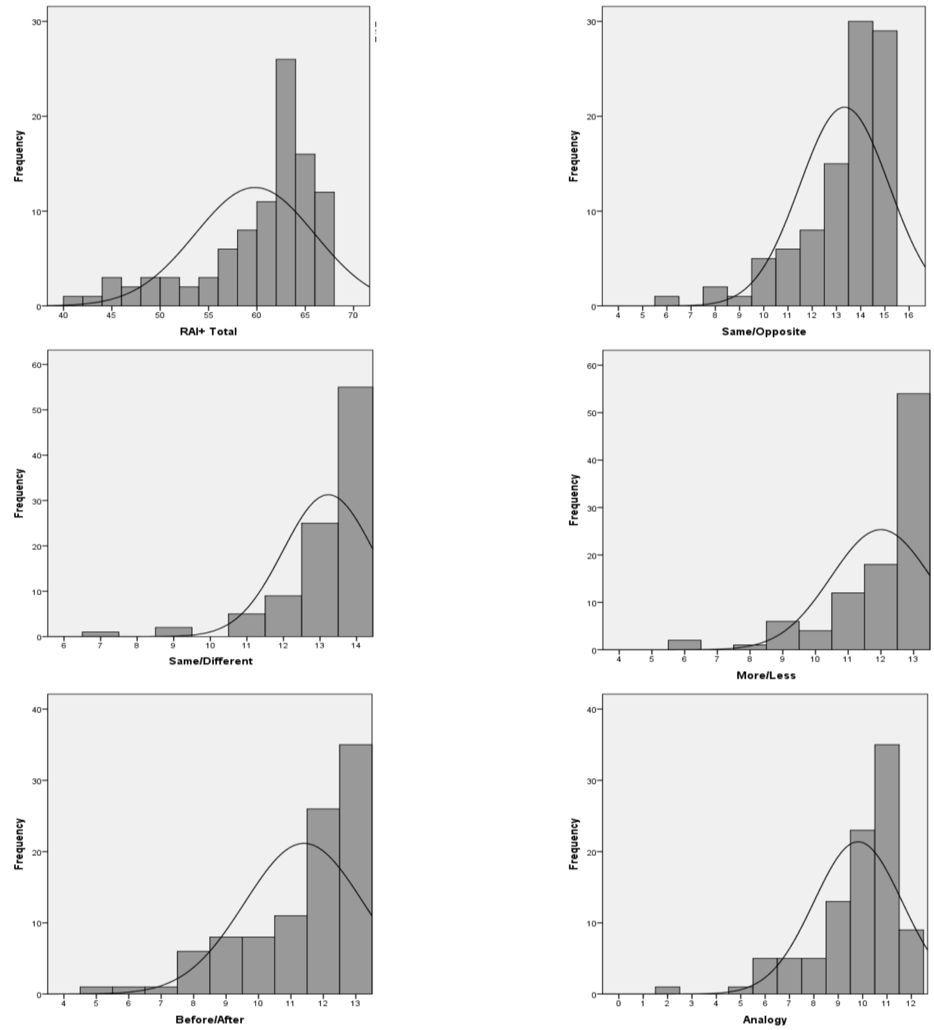


Figure 8.2. Histograms displaying the distribution of scores for overall RAI+ and each of the 5 testing blocks. Plots were calculated using the total number of correct trials per block.

Mean accuracy scores were highest for the Same/Opposite (*M* = 88.9%, *SD* = 12.3) and Same/Different (*M* = 88.2%, *SD* = 8.3) modules, followed by More/Less (*M* = 80%, *SD* = 10.2) and Before/After (*M* = 76%, *SD* = 12.2). Performance on the Analogy module was significantly lower, with mean accuracy at 65.4% (*SD* = 9.2). Figure 8.2 displays the distribution of scores for total RAI+ score, as well as the distribution of scores for each testing module.

8.3.2 Correlational Analyses

Each individual relational skills module demonstrated strong, significant correlations with overall RAI+ score, suggesting respectable internal consistency. Of the five frames, Same/Opposite (*rho* = .79, $p < .001$), Before/After (*rho* = .78, $p < .001$) and More/Less (*rho* = .75, $p <.001$) tasks displayed the closest relationship, followed by Analogy (*rho* = .67, $p < .001$) and Same/Different (*rho* = .52, $p < .001$). The Cronbach's Alpha statistic for the RAI+ was 0.79. Results from the correlational analysis of RAI+ performance and WASI, WIAT and WAIS Arithmetic scores are shown in Table 8.2.

Table 8.2
Correlations between RAI+ accuracy scores, and WASI IQ and its subindex scores.

Measure	Correlation coefficient	Significance level
WASI		
Full Scale IQ	0.54**	<.001
Verbal IQ	0.42**	.001
Vocabulary	0.37**	.003
Similarities	0.37**	.003
Performance IQ	0.48**	<.001
Block Design	0.42**	.001
Matrix Reasoning	0.42**	.001
WIAT-T	0.27	.1
Reading	0.14	.416
Reading Comprehension	0.08	.673
Spelling	0.29	.29
WAIS-III Arithmetic	0.43**	.009

** Indicates correlation is significant at the 0.01 level (2-tailed)

8.3.2.1 WASI scores.

Overall RAI+ scores correlated significantly with FSIQ (*rho* = .54, $p < .001$), as well as both VIQ (*rho* = .42, $p < .001$) and PIQ (*rho* = .48, $p < .001$). Performance on the RAI+ also correlated significantly with all four IQ subtests: Vocabulary (*rho* = .37, $p =.003$), Similarities (*rho* = .37, $p = .003$), Block Design (*rho* = .42, $p = .001$) and Matrix Reasoning (*rho* = .42, $p = .001$). Figure 8.3 represents scatterplots outlining the relationship between RAI+ scores and each of the WASI's three IQ indices.

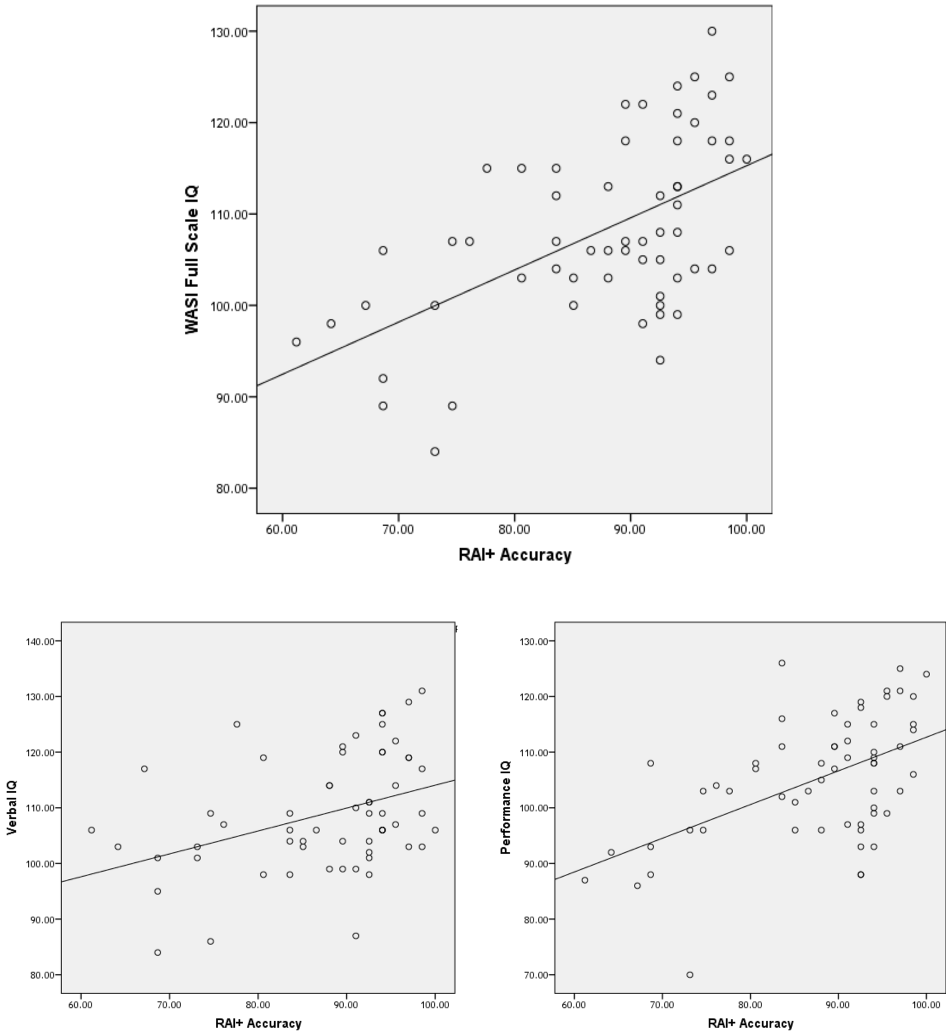


Figure 8.3. Scatterplots of the relationship between RAI+ and each of the three WASI IQ indices.

Additional exploratory analyses revealed that out of the five relational task blocks which comprise the RAI+, More/Less tasks exhibited the strongest correlation with WASI FSIQ (*rho* = .49, *p* < .001), closely followed by Same/Different (*rho* = .48, *p* < .001), Same/Opposite (*rho* = .44, *p* < .001), and Before/After (*rho* = .42, *p* = .001) tasks. Each of these four relational skillsets also displayed a significant relationship with VIQ (Before/After, *rho* = .43, *p* = .001; More/Less, *rho* = .42, *p* = .001; Same/Different, *rho* = .31, *p* = .015; Same/Opposite, *rho* = .26, *p* = .045). Similarly, scores for the relational skills Same/Different (*rho* = .48, *p* < .001), Same/Opposite (*rho* = .44, *p* < .001), More/Less (*rho* = .44, *p* = .001) and Before/After (*rho* = .33, *p* = .01) were correlated moderately with PIQ. Analogy tasks, surprisingly, displayed a significant correlation with PIQ (*rho* = .3, *p* =.02), but not with FSIQ or VIQ. Removal of Analogy module scores from the aggregate RAI+ score, increased the overall RAI+ correlation with FSIQ (*rho* = .55, $p < .001$) and VIQ (*rho* = .46, $p < .001$), but the correlation with PIQ was unaffected (*rho* = .48, $p < .001$).

The relationship between RAI+ scores and FSIQ for high IQ individuals (FSIQ: 110+, *rho* = .44, *p* = .03) was statistically significant. In addition, significant correlations were found between RAI+ scores and both VIQ (*rho* = .64, *p* = .001) and PIQ (*rho* = .47, *p* = .02), for this group.

8.3.2.2 WIAT-T.

RAI+ performance did not show a significant level of correlation with WIAT-T Standardised Score or any of the three WIAT-T subtests: Reading, Reading Comprehension and Spelling. Further analyses revealed that performance on the Same/Opposite block showed a moderate significant relationship with WIAT-T Standardised Score (*rho* = .35, *p* = .04). However, of the four other relational modules administered, no significant correlations were found with WIAT-T Standardised Score, Reading, Reading Comprehension or Spelling.

8.3.2.3 WAIS Arithmetic.

Overall scores for the RAI+ showed a moderate positive correlation with WAIS-III arithmetic scores (*rho* = .43, *p* = .009). WAIS-III arithmetic scores also correlated significantly with Same/Opposite (*rho* = .6, $p < .001$) and Before/After scores (*rho* = .4, *p* = .01), but not with other RAI+ subtest scores.

In summary, the RAI+ aggregate score displayed significant levels of correlation with all seven WASI IQ indices and subtests, as well as WAIS-III Arithmetic. Upon investigation, it was found that scores for the Same/Opposite, Same/Different, More/Less and Before/After test modules all correlated with the three WASI IQ indices, while the Analogy module only correlated with one IQ index (Performance IQ). RAI+ total and module scores generally did not correlate with any of the WIAT-T metrics, with the sole exception of the Same/Opposite module.

The purpose of the current study was to investigate and evaluate the utility of the RAI+ as a potential proxy measure of intellectual and scholastic ability, through assessing its degree of correlation with measures of intellectual performance (WASI IQ), educational/verbal attainment (WIAT-T-II) and numeracy (WAIS-III Arithmetic). Consistent with our expectations, the results from a correlational analysis revealed the presence of a significant relationship between scores of relational responding on the RAI+ and Full Scale, Verbal and Performance IQ on a standardised measure of intelligence, a finding which is highly consistent with previous studies (Colbert et al., 2017; Dixon et al., 2014; Gore et al., 2010; O'Hora et al., 2008; O'Toole & Barnes-Holmes, 2009). In addition, significant correlations were also observed between RAI+ scores and each of the four WASI IQ subtests. Broadly speaking, these results appear to support the assertion that relational responding may play an influential role in intellectual behaviour (Andrews & Halford, 1998; Barnes-Holmes et al., 2010; Colbert et al., 2017; Gentner & Loewenstein, 2002; Halford, Wilson, & Phillips, 2010; Moran et al., 2014; O'Hora et al., 2005; Roche et al., 2013; Stewart et al., 2013).

A significant correlation between RAI+ scores and WAIS-III Arithmetic scores further underline the relevance of relational skill proficiency to numeracy (Carpenter et al., 2003; Koehler, 2004; Molina, 2005; Molina et al., 2005; Ninness et al., 2006). In terms of academic attainment, the Same/Opposite test module displayed a significant relationship with the WIAT-T index score, a result predicted by the relevance of coordination relations to language acquisition (Hayes et al., 2001). Critically, however, the current study failed to identify a correlation between WIAT scores and any other RAI+ metric. This is inconsistent with the RFT perspective that relational abilities are functionally associated with academic attainment. However, there are a number of factors and limitations that may have affected this outcome. Firstly, the WIAT-II may not have been the most appropriate assessment for the current sample, as it may be less sensitive in assessing high levels of performance and has somewhat limited utility in predicting actual school grades (Spreen & Strauss, 2006). In effect, the question of how educational achievement relates to relational abilities perhaps cannot be solved psychometrically, but may instead require correlational analyses between relational abilities and actual academic attainment (itself a highly variable metric, bringing with it further challenges). While this issue renders the possibility of finding significant correlations between relational ability and academic achievement less likely, this does not alter our finding that the RAI+ did not predict performance on a well-validated and widely-administered academic achievement test.

While two distinct, but closely related Wechsler measures of intelligence were administered in the Colbert et al. (2017) study and in the current study (WAIS-III and WASI, respectively), there is considerable overlap in terms of the outcomes these studies report. While the pattern of significant relationships is similar, the strength of correlations varies between these two analyses. For example, the correlation coefficients reported for relational ability and Full-Scale (.54), Verbal (.42) and Performance IQ (.48) in the current study are considerably lower than that reported in the Colbert et al. analysis (.74, .78 & .55 respectively). These studies also differ in terms of correlations between relational skills measures and the four IQ subtests shared by each IQ measure (i.e., Vocabulary; .63 and .38, respectively, Similarities; .58 and .37, respectively, Block Design; .6 and .42, respectively, and Matrix Reasoning; .48 and .42, respectively). As such, we must conclude that the addition of further relational frames into the assessment was not beneficial in improving the predictive utility of the RAI and in fact increased variance along dimensions perhaps not as strongly related to IQ as the Same, Opposite, More than and Less than relational skills proficiencies. However, decreased sensitivity to the relevant relational skills by the WASI compared to the WAIS cannot be ruled out. Furthermore, it is still crucial to understand that the relative contributions of each relational skill to IQ and their inter-relationships with each other as part of a larger effort to elaborate a different perspective on the nature of human intelligence. In this regard, the current exercise has been informative.

In terms of what we have learned about the inter-relationships between various relational skill repertoires, perhaps the most illuminating have been, firstly, the confirmation of More/Less, Same/Different, Same/Opposite and Before/After as perhaps the most strongly related to IQ (Berens & Hayes, 2007; O'Hora et al., 2008; O'Toole & Barnes-Holmes, 2009; O'Toole et al., 2009; Stewart et al., 2013). Secondly, we have learned much from the surprising lack of correlation between analogical skills and IQ, as well as WIAT scores and WAIS arithmetic. This is highly unexpected because analogical reasoning is consistently associated with many higher cognitive skills such as abstract reasoning (Gentner, Holyoak, & Kokinov, 2001; Richland & Simms, 2015), problem solving (Gentner & Smith, 2013), creative endeavours such as writing poetry or prose (Shen & Lai, 2014) and more generally, is considered a ubiquitous aspect of everyday human communication (Stewart, Barnes-Holmes, & Roche, 2004).

We may make sense of the latter outcome in several ways. Firstly, it may be suggested that the limited number or type of analogical reasoning trials included in the RAI+ may not sufficiently assess subtle individual differences in this skill repertoire. However, it may also be that the WASI and WIAT have a poor representation of such tasks in their battery. In the case of the WASI, while Matrix Reasoning can be considered an assessment of visual-spatial analogical reasoning (Carpenter et al., 1990), none of the four WASI subtests directly assess verbal analogical reasoning. In addition, despite the fact that analogical reasoning is pertinent to a number of important verbal competencies, such as reading (Farrington-Flint et al., 2007; Farrington-Flint & Wood, 2007), vocabulary (Edwards et al., 2011; Nippold & Sullivan, 1987), grammar (Edwards et al., 2011; Hock, 1991, 2008), and spelling (Brown et al., 1996; Goswami, 1988), none of the WIAT-T's three subtests of Reading, Reading Comprehension and Spelling explicitly employ analogical tasks as a means of measuring verbal attainment.

At this point, it is important to note that the aim of the current research stream is not to compose a "better" measure of IQ but to provide a functional account of intellectual performance and an accompanying assessment tool. As such, dissimilarities in the remit of measurement and/or failures to find significant correlations do not necessarily represent a psychometric failure of the RAI+, but may, in fact, reflect a theoretical divergence in terms of what constitutes intellectual performance. The strength of the correlations reported in the current analysis suggests that while these repertoires may be related, they are not equivalent or synonymous, at least as assessed by the testing battery administered. That global issue notwithstanding, the RFT literature would propose that due to the advanced level of complexity inherent in analogical reasoning, for example, its proficiency levels should predict IQ (see Carpentier, Smeets, & Barnes-Holmes, 2003; McHugh et al., 2007; McLoughlin et al., 2018), particularly for high ability individuals. The fact that this was not the case, may point to construct validity issues for either the RAI+ or the WASI, depending on what *a priori* definition of intelligence one begins with.

Secondly, it may be the case that analogical responding is not as relevant to intellectual performance as other "core" relational skills (e.g., same/opposite, more than/less than), a suggestion supported by the reduced level of inter-correlation between analogical trials and scores on the other RAI+ modules, as well as the finding that the removal of analogy test trials actually increased the internal consistency and predictive validity of the RAI+. The apparent distinction between "core" relational skills and analogical reasoning may be related to its unfolding in the developmental process. Specifically, there is some modest evidence that relational skills of coordination and comparison (more/less) emerge first and appear to be well-established prior to the development of many higher-level relational skills (Carpentier et al., 2003). In this sense, perhaps analogy comprises part of a higher-level skill set that is still unfolding in adults, insofar as it depends upon proficiency in each of the other relations and involves learning to relate relations to each other. Future research should aim to investigate to what extent some of these skills precede or functionally overlap with each other and should attempt to map out the developmental trajectory of analogical reasoning, which may extend well into adulthood.

In assessing the distribution of RAI+ test scores, the reduced variance of scores for the current sample is noteworthy. A large proportion of our sample (29%) achieved an overall RAI+ score of 95% or above. In contrast, only one participant displayed a Full-Scale IQ above the 95th percentile, and none scored more than 95% on the WASI Similarities, Vocabulary or Matrix Reasoning subtests. The skewed distribution of RAI+ scores would, therefore, reduce the likelihood of significant correlation with IQ metrics, and diminish its utility as a proxy measurement on intelligence more generally.

One possible mechanism for enhancing the predictive validity of the RAI+ is to ensure a wider range of scores and therefore improve the sensitivity of the test. This could perhaps be achieved most readily by reducing the time limit at either a global or per trial basis. While previous research suggests that response fluency, in general, may not correlate with Full-Scale IQ (Binder, 1996), from an RFT and behaviour-analytic perspective, the fluency with which responding occurs is an important component of intelligent behaviour (Cassidy et al., 2016; Thorndike, Bregman, Cobb, & Woodyard, 1926). Secondly, more difficult tasks could be included in the RAI+ in an effort to increase its sensitivity at the top end of the scale. This could involve increasing the nodal distance of relations tested, or the number of nodes which link any two stimuli in a set of conditional relations (Sidman, 2009). For instance, most tasks in the RAI+ assess two nodal (e.g., A is the same as B, B is opposite to C) or three nodal (e.g., A is the same as B, B is the same as C, C is the same as D) relational reasoning. The addition of further nodes could be integrated into the current RAI+ and would potentially allow us to ascertain a more comprehensive profile of individual strengths and weaknesses, as well as more balanced data distributions. However, it is critical to state at this point that there is no conceptual requirement that RAI+ scores be distributed normally across the population, precisely because they index a malleable skill set that is considered to be continually in flux and therefore at varying levels in various environmental contexts across various times.

The aim of this present study was to test a prototype extended relational abilities index which built upon the RAI in terms of the range of relational frames it assessed. Our results indicate that while the RAI+ exhibited significant correlations with a range of IQ indices and subtests, its inclusion of additional relational frames did not improve upon the predictive validity demonstrated by the original RAI. This may not be surprising given both the previously reported high correlations between the shorter RAI and Full-Scale IQ (.74; Colbert et al., 2017), and the currently reported high

Cronbach's alpha for the RAI+ (.79). In other words, any subset of relational tasks may hold the potential to function as a useful proxy of both overall relational skills and IQ.

Interestingly, performance on the RAI+ displayed a general failure to predict educational and verbal attainment as measured by the WIAT-II-T, despite a wealth of previous theoretical and empirical work which would anticipate such relationships. The work did, however, reveal important inter-correlations between relational skills repertoires, and found a respectable level of internal consistency for the RAI+. Overall, the study confirms that relational skill indices may represent useful proxies of full-scale intelligence and potentially numeracy, but that such indices bear a more complex relationship to academic aptitude.

Future studies may endeavour to provide a more comprehensive examination of the relationship between academic attainment and relational skill fluency. Interestingly, the respectable correlations obtained between RAI+ scores and the standardised measures of numeracy and Full-Scale IQ, suggest that it may well be the standardised nature of the WASI and WAIS indices that facilitated such correlations. The WIAT, in contrast, is not a very good predictor of school grades (Spreen & Strauss, 2006) and so its construct validity may be in question, rather than that of the RAI+. In addition, the most important measure of academic attainment is actual scholastic performance, and it is more fitting for a behavioural science to validate a proxy measure for academic attainment against real school performance, than against further proxies for the same. In the meantime, the RAI+ is not ready for use as a proxy measure of academic ability but would appear to hold promise as a functionally understood, behaviour-analytically acceptable proxy for assessing intellectual function. If this is so, we have moved some way forward in developing a progressive behaviour-analytic, functionally understood assessment of the broad skill set widely referred to as "intelligence".

Chapter 9

General Discussion

In this chapter, a review of the essential findings of the current thesis will first be described, and then placed into the context of the previous theoretical and empirical research base. Subsequently, more general, overarching themes concerning the efficacy of relational skills training interventions and the relationship between AARR and intelligence will be extracted from the current research findings and discussed in detail. Such a discussion will, therefore, attempt to identify the general results and implications of this research, and highlight what has been learned, and was remains to be investigated, in regard to the current research stream.

Experiment 1 entailed the first large-scale investigation into the efficacy of an online training intervention, SMART (Cassidy et al., 2011, 2016) in improving derived relational responding in accordance with the frames of coordination, opposition and comparison. As derived relational responding has been conceptualized as a generalized operant (Hayes et al., 2001), this repertoire is therefore inherently flexible and open to manipulation by environmental contingencies. Due to the proposed relevance of such a repertoire to intellectual performance (see Colbert et al., 2017), a growing body of empirical investigations (Amd & Roche, 2018; Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; McLoughlin et al., 2018; Parra & Ruiz, 2016; Thirus et al., 2016; Vizcaíno-Torres et al., 2015) have investigated the application of SMART as a means of improving intellectual performance. As an investigation into SMART's efficacy in increasing intellectual ability is one of the central themes of the current thesis, Experiment 1 first endeavoured to establish the utility of SMART in increasing sophistication in the skillset that it targets, relational responding, in a sample of secondary school students (n = 169), to serve as a manipulation check for the intervention and a basis for future explanation of SMART's potential ramifications regarding intelligence.

Results of this analysis indicate SMART's clear and pronounced success in improving relational skill for the current sample, with RAI scores increasing significantly from 38.8 to 46.8 out of a possible 55. RAI Fluency scores, which took into account both speed and accuracy of response, also increased significantly as the average time taken to complete the RAI dropped from 10 minutes 24 seconds, to 9 minutes 18 seconds. Finally, a linear regression equation based on published RAI and IQ data allowed the calculation of a Full-Scale IQ estimate, which was also found to increase significantly following intervention from 102.1 to 113.8. These results combine to suggest that SMART is effective in improving both the accuracy and fluency of the relational skills it targets.

Upon closer inspection of the impact of relational skills training at different baseline ability levels, a trend emerged suggesting an inverse relationship between post-intervention RAI score gains and baseline IQ. Correlational analysis revealed a moderate, negative correlation between baseline ability and subsequent RAI score gain (r = -.45). In addition, those who scored in the top 20% of RAI scores at baseline displayed post-intervention RAI score increases (M = 4.6) considerably lower than the average rise for the entire sample (M = 8). The most apparent explanation of this relationship would be that participants who recorded pre-intervention RAI scores at the higher end of the scale already displayed proficiency in most forms of relational responding that would be subsequently trained in the training program. To illustrate this point, as discussed in Chapter 2, a participant with an RAI score of 50 out of 55 stands far less to 'gain' than a participant with an RAI score of 25, as the former individual has demonstrated a relational responding repertoire in which approximately 50 of the 55 forms of relational responding has already been established. As a relatively reduced number of trials represent skills not yet acquired, SMART offers such an individual much more restricted scope for learning and therefore, performance improvement when compared to an individual with a lower RAI. The issue of ceiling effects will be discussed further in Section 9.3.2.1.

To complement this trend, further analysis found that those who completed all 55 training levels (n = 135) had a significantly higher baseline RAI (M = 38.9) than those that did not (M = 31, n = 34). While it may appear superfluous to propose that those with weaker relational skills were less likely to complete a relational skills training program, this finding requires further attention due to the clear relationship between this repertoire and intellectual performance. As such, it may be the case that those at the lower end of the IQ spectrum (at least for this age group) may have difficulty in completing the training program. Given the substantial efficacy of SMART in increasing relational skill for this cohort in particular, it is extremely important to ensure that every opportunity and resource is provided to such participants to allow them to access to the benefits that SMART can provide. Relational skills training for those at the lower end of the ability spectrum is addressed by Experiment 4 and will be further discussed in Section 9.4.4.

In sum, Experiment 1 provides considerable support for the utility of SMART in successfully targeting and subsequently improving proficiency in derived relational responding in accordance with the frames of coordination, opposition and comparison. Subsequent studies will address the application of this program in improving a range of intellectual and academic skills, as well as the relationship between post-intervention improvements in relational, intellectual and academic skills.

Experiment 2 aimed to build upon the findings of the previous experiment by investigating the efficacy of SMART in increasing intellectual performance as assessed by a gold-standard IQ test, the Wechsler Abbreviated Scale of Intelligence (WASI; Wechsler, 2013). While a small number of published analyses have previously highlighted the potential application of relational skills training as a means of ameliorating intellectual performance (Amd & Roche, 2018; Cassidy et al., 2011, 2016; Hayes & Stewart, 2016; McLoughlin et al., 2018; Parra & Ruiz, 2016; Thirus et al., 2016; Vizcaíno-Torres et al., 2015), the promising effects reported have, to some extent, been undermined by a number of concerns regarding the methodological rigour of the experimental designs employed. Such studies have failed to include control groups (e.g. Amd & Roche, 2018; Cassidy et al., 2011; Hayes & Stewart, 2016; McLoughlin et al., 2018; Vizcaíno-Torres et al., 2015) and blinded testers (e.g. Amd & Roche, 2018; Cassidy et al., 2011; Hayes & Stewart, 2016; McLoughlin et al., 2018; Parra & Ruiz, 2016; Vizcaíno-Torres et al., 2015). In addition, a number of studies have included insufficient training periods and test-retest intervals (Amd & Roche, 2018; McLoughlin et al., 2018), and/or utilised small-n or single case designs (Cassidy et al., 2011; Luciano et al., 2007; McLoughlin et al., 2018; Vizcaíno-Torres et al., 2015). As such, the design implemented in Experiment 2 aimed to investigate the effect of SMART on intellectual performance using a single-blind randomised controlled design over a period deemed sufficient to complete training and reduce the potential for IQ test practice effects, using a sample of Irish secondary school students.

Correlational analyses uncovered significant, moderate-to-strong correlations between scores on the relational responding metric administered, the RAI, and all three WASI IQ indices (Full-Scale, Verbal and Performance IQ), and three of four IQ subtests (Vocabulary, Similarities & Matrix Reasoning). The relevance of relational responding to intellectual performance has received support from a small number of published investigations (e.g. Colbert et al., 2017; O'Hora et al., 2005, 2008). In light of a further replication of this effect, the current analysis adds support to the potential of utilising relational responding assessments (specifically the RAI) as a potential functional alternative to more traditional assessments of intelligence (Colbert et al., 2017). The merits and limitations of this approach to assessing intelligence will be further addressed in Section 9.2.

In line with the results of previous intervention studies (e.g. Cassidy et al., 2011; 2016; Hayes & Stewart, 2016), SMART was found to lead to significant improvements in scores in Full-Scale IQ (*M* = 18.4), Verbal IQ (*M* = 19.7) and Performance IQ (*M* = 13.5), while scores on these indices remain virtually unchanged for control participants. The reported increase for Full-Scale IQ bears resemblance to the rise of 23 points found in the first of the Cassidy et al. (2016) studies, further underlining the magnitude of intellectual performance improvement following relational skills training. To put this rise into perspective, if an individual presenting with an IQ in the 50^{th} percentile (i.e. FSIQ = 100, the average level of intellectual performance for his/her age group) was to demonstrate an increment of 18 IQ points following intervention, that individual would rise 32 percentile ranks, thus scoring above 82% of his/her peers. Given the importance of IQ to an extensive variety of socially-desirable outcomes such as academic attainment (Deary, 2012; Jensen, 1998; S. B. Kaufman, Reynolds, Liu, Kaufman, & McGrew, 2012; Rindermann, 2007), income (Irwing & Lynn, 2006; Lynn, 2010; Lynn & Vanhanen, 2011; Meisenberg, 2012) and even self-reported happiness (Ali et al., 2013; Isaacowitz & Smith, 2003; Kanazawa, 2014; Siedlecki, Tucker-Drob, Oishi, & Salthouse, 2008), the intellectual gains being reported here reflect more than just an inconsequential improvement on an arbitrary, irrelevant metric, but rather hold genuine implications for social, academic and occupational success.

In terms of Verbal IQ, previous analyses have also reported increases in verbal intelligence following relational skills training interventions, a finding which is replicated in the current analysis. The Verbal IQ rises reported in the current study (*M* = 19.6 points) are similar to the 18-point increase reported in Cassidy et al. (2011) which used a small sample of 8- to 12-year old children. Hayes and Stewart (2016) also report significant rises in a number of verbal indices, such as WIAT Spelling, WIAT Reading, WISC Letter-Number Sequencing and WISC Digit Span following SMART training, indicating that improvements in IQ scores following SMART training may further extend to increments in scholastic aptitude. In addition to the score rises found for Verbal IQ, scores for both Verbal subtests (Vocabulary & Similarities) were found to increase significantly following training. The efficacy of the SMART program in improving aspects of verbal intelligence is predicted by an extensive theoretical and empirical literature base proposing the importance of relational skill to language development and proficiency (Colbert et al., 2017; Gore et al., 2010; Hayes et al., 2001; O'Connor, Rafferty, Barnes-Holmes, & Barnes-Holmes, 2009; Roche et al., 2013; Sidman, 1971; Stewart & Roche, 2013). From an RFT perspective, word-word and object-word relations underpin language development and serve as the basis for linguistic reference (Stewart & Roche, 2013). As such, relational responding proficiency would appear to facilitate verbal intellectual performance as assessed by Verbal IQ subtests. As outlined by Cassidy et al., (2010) many of the Verbal subtests can be understood as tests of relational responding to a greater or lesser degree (see Section 9.1.1 for further detail), and therefore predict gains for this IQ index. Indeed, numerous studies have reported significant correlations between measures of relational responding and Verbal IQ items (D. Colbert et al., 2017; Gore et al., 2010; O'Hora et al., 2008). In a correlational analysis of relational responding and scores on the WAIS-III, Colbert et al. (2017) reported moderate-to-strong statistically significant correlations between RAI scores and Verbal IQ, both Verbal IQ subindices (Working Memory & Verbal Comprehension) and 6 of 7 Verbal IQ subtests, indicating a wide-ranging relationship between relational responding and virtually all aspects of verbal intelligence as assessed by the WAIS-III.

Regarding Performance IQ, while the ANOVA did not uncover a significant between-groups effect on this metric, the finding that experimental participants showed a significant increase of 13.5 points, whereas scores for the control group barely changed (< 2 points), is telling. Furthermore, scores for both Performance IQ subtests (Block Design & Matrix Reasoning) were found to increase significantly after training. Such results further acknowledge the far-reaching influence of relational responding to domains that bear little topographical resemblance, as well as the wide-ranging benefits an improved relational responding proficiency can catalyse. Further discussion of the relevance of relational responding to Performance IQ test items can be found in Section 9.1.2.

In contrast to the findings of Experiment 1, the current data do not suggest a relationship between baseline levels of ability and post-intervention improvements in intellectual performance. In the previous experiment, it was found that those with lower baseline levels of relational ability were less likely to complete the training, and therefore less likely to demonstrate the greater score improvements predicted by completion of the training program. Perhaps the most readily available explanation of this inconsistency may be found in the difference between baseline levels of relational responding in each group. Mean baseline RAI scores for the current sample ($M = 39.3$) were 10% higher than the mean scores found for the sample used in Experiment 1, who were also 3 years younger. Due to the relatively sophisticated relational skills displayed by the current sample at baseline, it may be the case the vast majority of participants demonstrated the prerequisite level of relational skills required from the outset. As such, individual differences in baseline relational skills may not have predicted training progress, as most students began training with a responding repertoire that was a sufficient foundation to allow the establishment of all 55 relational responding task types. On the other hand, the participants used in Experiment 1 may have shown greater variation in the establishment of these pre-requisite skills, and therefore, individual differences in baseline relational skills (i.e. the presence or absence of the pre-requisite relational skills) may have exerted a greater impact on training progress. Alternatively, given the lower level of ability for Experiment 1 participants, it may also be the case that the progressive increases in training task difficulty occurred at a rate that was too fast or in increments too great.

Experiment 2 represents an important extension of previous similar studies, insofar as it was the first to employ blind testers and participant allocation by a third party, as well as third-party management of the training intervention. However, there are a number of potential limitations of the current study's methodology. Perhaps foremost among these was the failure to implement an active control measure. Specifically, it could be suggested that the IQ gains observed following the intervention are not due to the relational skills intervention *per se*, but are instead due to general factors related to engagement in any form of intensive training (Melby-Lervag et al., 2013). While this possibility cannot be directly contested, it should be remembered that the Hayes and Stewart (2016) study did use an active control group and found similar effects to those observed here. Moreover, while no study can ever serve as the elusive *experimentum crucis* on SMART, it can help to triangulate in on the SMART effect using varying methodologies and in so doing also produce another replication of an increasingly reported intervention outcome. This reticulated approach to theory development is a key feature of the scientific approach with which RFT is associated (Hayes, Barnes-Holmes, & Wilson, 2012). As such, the emergence of the SMART effect under varying conditions can be viewed not as an inconvenient inconsistency across studies, but as a support for the idea that the SMART effect may be a real and robust effect that can be observed across contexts and situations. That said, it would of course be prudent for future studies to examine the non-specific effects of study participation on IQ gain.

Another limitation of the Experiment 2 may be the lack of a manipulation check of the variable being trained (i.e., relational skills). Studies generally administer some form of relational skills assessment, such as the recently developed Relational Abilities Index (RAI; Colbert et al., 2017) at baseline and at follow-up in order to see if skill improvements have been made on a direct measure of the very skill being trained. However, because the training was administered by school authorities and not the researchers, this data was not obtained and therefore, this type of analysis was not possible. Having access to such measures would allow for more complex statistical analyses of the relationship between IQ gains and relational skills improvements and should be a feature of all future studies. However, following the results of Experiment 1, which provided a clear indication that SMART is effective in targeting and improving relational responding proficiency, it is reasonably safe to assume that relational skills were improved in the current sample and that this improved performance likely contributed to the intellectual gains witnessed.

In summary, the current analysis represents an important progression in investigations into relational skills training programs as a means of improving intellectual function. The results of the current study lend further support to the burgeoning research stream which promotes the efficacy of the SMART training program in increasing IQ scores, and importantly, under more controlled and methodologically rigorous conditions.

In light of the complex relationship that baseline ability and SMART training progress appear to share, Experiment 3 administered this intervention to a sample of 10- and 11-year-old children in order to investigate the feasibility and utility of administering SMART to younger children with assumedly weaker relational responding repertoires. While baseline ability may have precluded the possibility of some 12-14-year-olds of completing training (as witnessed in Experiment 1), this was not a factor for the 16-17-year-old sample studied in Experiment 2. As such, the motivation behind Experiment 3 was threefold: (1) to shed further light on the impact of baseline ability on training progress and potential IQ improvement; (2) to assess the applicability and utility of the current SMART protocol using the youngest normally-developing sample of children to date and (3) to further increase the degree of experimental rigour of studies analysing the potential impact of SMART by implementing a crossover design.

While ANOVA results did not identify a clear between-group effect over the course the three WISC-IV IQ administrations, thereby indicating a lack of effect of SMART, post-hoc t-tests indicated significant Full-Scale IQ score increases of over 6 points following training, but not following control periods. In addition, one of the four IQ subindices, Verbal Comprehension showed a similar trend, with mean scores for the sample rising by 7.2 points following training, but not changing significantly following control periods. Regarding the other three IQ subindices, the effect of SMART is less clear. While Group 1 displayed significant increases in Perceptual Organisation scores following training and not following the control period, this pattern was reversed for Group 2. In addition, neither group demonstrated post-intervention performance improvements for the Processing Speed subindex, but Group 2 scores rose significantly following the no-intervention period. Finally, SMART was conclusively shown to exert no significant influence on the fourth WISC subindex: Working Memory.

Results from a correlational analysis indicate that, as predicted, baseline RAI scores displayed a strong correlation with Full-Scale IQ (r = .69). To complement this, medium-strength correlations were also found for three of four IQ subindices (Verbal Comprehension, Perceptual Reasoning and Working Memory) and 7 of 10 IQ subtests (Similarities, Vocabulary, Block Design, Picture Concepts, Matrix Reasoning, Digit Span, Arithmetic). Such widespread correlation to various topographically dissimilar domains of intellectual performance further underlines the relevance of relational skill to intellectual performance.

However, given the somewhat conflicting result of a reduction (or even absence) of a ‘SMART effect’ despite a considerable level of correlation between IQ and RAI scores, further analysis was undertaken to gain a better understanding of the current results. It appears that the diminished efficacy of SMART in increasing IQ scores may be due to the finding that most participants (23 of 28) failed to complete all 55 training levels within their 3-4 month training period. As post-intervention Full-Scale IQ score increases were predicted by number of training levels completed (even after controlling for baseline IQ), it appears that completion of the entire training program is essential to produce IQ score improvements at the level seen in Experiment 2, regardless of pre-intervention intellectual ability. To further strengthen this point, it was found that the small number of participants who completed all training levels (n = 5) showed FSIQ rises (M = 13 points) that were 2.5 times greater than those who did not (M = 4.9). This pattern was replicated for IQ subindices, as rises found for those who completed training far outweighed those that did not: Verbal Comprehension (11 points to 6.4 points), Perceptual Reasoning (11 points to 3.3 points), Working Memory (7 points to 3.2 points) and Processing Speed (9 points to -0.3 points). As such, the current analysis concludes that the most important contributory factor towards post-intervention IQ score increases is training progress/completion, irrespective of the individual’s pre-intervention intellectual ability.

As the mean baseline RAI score was only slightly lower (32.9) than the average number of training levels completed (37), it can be suggested that, on average, participants were only exposed to training levels for 4 forms of relational responding that were not already present at pre-intervention. Due to the fact that the RAI is not a Guttmann-style scale (an issue addressed in Section 9.4.3), this is far from a watertight estimation. However, the test-training discrepancy score (i.e. training levels completed minus pre-intervention RAI score) may give some tentative indication of how many novel or additional forms of relational responding were successfully established following training. The current analysis found that a positive test-training discrepancy score (i.e., training levels completed > pre-intervention RAI score) resulted in FSIQ rises that were double those reported for individuals that showed a negative score on this metric.

Such results, alongside the finding that only a small proportion of a sample of normally-developing 10- and 11-year-olds were able to complete the SMART program (albeit within a 3-4 month period), call for the development of further resources to allow younger children and/or those with lower levels of relational skill and intellectual performance to access the benefits SMART has demonstrated the ability to provide. Upon supervision of training sessions, it became apparent that, for many students, one of the most delimiting stumbling blocks was the arbitrary nature of the stimuli included as relata in the program. This difficulty was all the more evident for those at the lower end of the ability spectrum, who, given their baseline IQ, were found to be less likely to reach the later stages of training. During sessions, it was found that the substitution of these arbitrary relata for physical aids (such as pencils and counters) was extremely effective in overcoming these difficulties and allowing students to progress further than they may have done otherwise. Therefore, it appears that for the average 10- or 11-year-old child, the SMART program necessitates a level of sophistication in AARR that typically may be beyond his/her current capabilities. The SMART:R program, designed specifically to address this issue, was assessed in Experiment 4, and will be discussed subsequently.

Experiment 4 represented the first investigation of the SMART:Remedial program, a relational skills training intervention which aims to establish arbitrarily-applicable coordination, opposition and comparison responding by presenting participants with non-arbitrary relational tasks and slowly phasing-in arbitrary relata as stimuli. This program was developed as a direct response to the findings of Experiment 3, which proposed that post-SMART improvements in intellectual performance are heavily dependent on training progress and that individuals who are younger and/or display weaker relational responding repertoires may not be able to achieve the level of progress necessary to facilitate these performance improvements. Experiment 3 further identified that much of the difficulty in completing training stages was attributable to an inability to reason effectively with the abstract relata used in relational tasks. This issue was found to be more prevalent for those at the lower end of the ability spectrum. Given that SMART treats such a competency as a prerequisite, the SMART:R system was designed to train more basic relational responding as a means of establishing the relational skills necessary to complete the main SMART program. Therefore, this system was administered over the course of 4 months to a sample of 9- and 10-year-old students with below-average levels of intellectual performance (Mean FSIQ = 87.7) using an ability-matched control group.

Results from an ANOVA indicate that SMART:R was extremely effective in improving relational responding proficiency (as measured by the SMART:R assessment) with mean scores rising from 62.2% to 90.7% following training when compared to a group who continued remedial support classes but received no additional intervention. Furthermore, individual ANOVAs indicated that the SMART:R training protocol resulted in large gains in both coordination/opposition and comparison relational responding. In terms of AARR specifically, scores on such trials were found to increase significantly from 51.9% (approx. chance level) to 81.8% following training, while control group scores were relatively unchanged and remained at approximately chance levels. Therefore, these results combine to suggest that the SMART:R program is an efficacious means of building upon non-arbitrary relational responding repertoires as a means of establishing/improving AARR.

Correlational analyses revealed significant, medium-strength correlations between SMART:R assessment scores and Full-Scale IQ, Performance IQ and one of four IQ subtests, Block Design. The failure of SMART:R assessment scores to predict Verbal IQ is somewhat contradictory to an array of both theoretical and empirical work proposing such a relationship (see Colbert et al., 2017 for an overview). This finding will be discussed in more detail in Section 9.2. Of note is the finding that post-training score changes for Full-Scale, Verbal and Performance IQ were not predicted by baseline Full-Scale IQ or SMART:R assessment scores, indicating that pre-intervention ability is not a significant determinant of post-intervention gains, at least with regard to this ability range. Of additional importance is the result that all participants completed training within 17 weeks of biweekly 45-minute sessions (excluding one student who was removed from analysis due to an extended school absence), with a mean completion time of 14 weeks.

As a secondary aim, this analysis investigated the impact of the SMART:R program in increasing intellectual performance, as assessed by the WASI. While the nature of the skills being trained by the main SMART program appear to be more germane to intellectual performance (a point exemplified by the relatively weaker correlations between SMART:R assessment scores and IQ metrics), the current analysis aim to explore whether post-training improvements in more basic relational skills would be accompanied by improvements in general intelligence. Results from further ANOVAs identified a significant between-group effect of training on Full-Scale and Performance IQ, with the experimental group recording post-training score increases of 6.2 and 3.4 points, respectively. While no between-group effect was found for Verbal IQ, experimental participants displayed large rises of 8.6 points, compared to a mean control group increase of 1.2 points, which may indicate the presence of an effect. Such results indicate that the benefits of the SMART:R protocol may extend beyond the improvement of relational responding proficiency, and may, in a similar vein to the main SMART program, convey implications of overall intellectual performance.

As Experiments 2, 3 & 4 added further support to the suggestion that relational skills training may represent an effective means of ameliorating intellectual performance, Experiment 5 explored the impact of SMART on a related domain, academic performance. While intelligence and academic performance share a close relationship (Bourneville, 1895; Deary et al., 2007; Jensen, 1998; Laidra et al., 2007; Neisser et al., 1996; Roth et al., 2015), there is a relative dearth of published research concerning the relevance of relational responding to scholastic ability, and specifically, the potential application of relational skills training in producing demonstrable improvements on this skill set. Currently, only two published studies have investigated such an effect (Cassidy et al., 2016; Hayes & Stewart, 2016), finding significant improvements on measures of academic aptitude in two small samples of children. This effect, however, has not yet been validated using a large sample. Therefore, Experiment 5 aimed to investigate the utility of SMART in improving academic performance on the Drumcondra Reasoning Test, the Irish Department of Education's academic assessment of choice, in an entire year cohort of 12- to 14-year-old children (n = 174). As a secondary analysis, a subsample of this group (n = 38) completed WASI IQ assessments in order to explore any potential concurrency in relational, academic and intellectual performance increments following training.

At baseline, correlational analyses indicated that RAI scores correlated significantly with both DRT subindices, Verbal Reasoning (r = .54) and Numerical Ability (r = .59), as well as the overall DRT composite score (r = .61), indicating a medium-strength relationship between relational skill and academic performance. Such positive correlations are predicted by an established research base proposing the relevance of relational responding to various verbal (Brown et al., 1996; Cassidy, Roche, & O'Hora, 2010; Colbert et al., 2017; de Rose et al., 1992; Edwards et al., 2011; Farrington-Flint et al., 2007; Goswami, 1986; Mackay, 1985; McHugh et al., 2004; Nippold & Sullivan, 1987; Sidman, 1971) and numerical skills (Carpenter et al., 2003, 2005; Cassidy et al., 2016; Colbert et al., 2017; Molina et al., 2005, 2008; Stephens, 2007).

In terms of post-intervention score increases, scores for DRT Overall, Verbal Reasoning and Numerical Reasoning increased significantly by 3.5, 3.3 and 2.9 points respectively. While these score increases are modest, the finding of a significant effect of SMART on this particular measure of academic ability is promising for a number of reasons. Firstly, the DRT is the assessment of choice for the Department of Education, and therefore can be considered to assess the academic skills deemed most important to scholastic performance. Secondly, this study administered a brand-new iteration of the DRT, which randomly administers one of five equivalent, but topographically distinct forms to each participant. As such, unlike previous studies in this area (i.e. Cassidy et al., 2016; Hayes & Stewart, 2016), the potential that follow-up score increases are due to mere practice effects is rendered extremely unlikely because of the variety in the assessment forms implemented. Thirdly, the discovery an effect of SMART using such a large sample allows a much more solid and empirically-validated argument to be made concerning the utility of this program in providing real benefits to student's scholastic performance.

As previously reported in Experiments 1 & 3, there was an inverse relationship between post-intervention score gains and baseline ability found with the current sample. After dividing up the sample based on baseline DRT standardised scores into below average (DRT < 90), average (DRT = 90 – 110) and above average (DRT > 110) cohorts, score increases, while still significant for all groups, diminished as a function of baseline ability (5.4, 3.8 & 2.2 points respectively). A similar pattern was replicated for both Verbal Reasoning and Numerical Ability. Such results converge to propose that participants with lower baseline ability may stand more to 'gain' from SMART than those with higher levels of ability, due to the more expansive and proficient relational responding repertoires that higher intellectual and academic ability likely entails. Therefore, a greater proportion of the skills being targeted may already be established for high-ability individuals, and as such, training these skills results in lesser increments in performance. This trend will be further discussed in Section 9.3.2.1. However, the recurrent finding that SMART displays reduced efficacy in improving performance for high-ability individuals calls for amendments to be made to this protocol in order to include more complex trials and/or elevated task difficulty, in order to establish more advanced forms of relational responding for those who present with relatively sophisticated relational skills at the outset. Experiment 6 will endeavour to assess the relationship of a wider range of relational frames with intelligence, with a view of potentially integrating training protocols for these additional frames into the SMART program.

As such, Experiment 6 comprised a correlational analysis of the relationship between intellectual performance and scores on an extended relational skill assessment, the Multiple Relational Assessment Test, which measures proficiency in accordance with the frames of coordination, distinction, opposition, temporality, analogy, & deixis. This was done to glean a better understanding of the relative contribution of a wider range of relational skills to intellectual performance, and potentially, to confirm or disconfirm the potential utility of the MRAT as a proxy measure of intelligence.

Results indicated a moderate-to-strong correlation between overall MRAT score and WASI Full-Scale IQ ($r = .63$), Verbal ($r = .44$) and Performance IQ ($r = .67$). In addition, overall MRAT score was found to correlate moderately with three of four IQ subtests (Vocabulary, Block Design, and Matrix Reasoning). Full-Scale IQ was also shown to be predicted by Coordination/Opposition ($r = .63$), Coordination/Distinction ($r = .47$), Temporality ($r = .39$) and Analogical Reasoning ($r = .34$). Verbal IQ correlated significantly with Coordination/Opposition ($r = .43$) and Coordination/Distinction ($r = .38$). Finally, Performance IQ was found to correlate significantly with Coordination/Opposition ($r = .47$), Temporality ($r = .49$) and Coordination/Distinction ($r = .47$).

In terms of Verbal IQ subtests, Vocabulary and Similarities each correlated significantly with performance on Coordination/Distinction ($r = .36$) and Coordination/Opposition ($r = .41$) respectively, but with no other relational skills modules. Regarding Performance IQ measures, Block Design correlated with Coordination/Opposition ($r = .66$) and Temporality ($r = .37$), while Matrix Reasoning scores correlated significantly with Coordination/Distinction ($r = .59$), Coordination/Opposition ($r = .37$) and Temporality ($r = .44$). As expected, this analysis somewhat replicated previous AARR-IQ correlational analyses, as performance on the MRAT's Coordination/Opposition block shared the closest relationship with overall intellectual function, correlating significantly with all three IQ indices and 3 of 4 IQ subtests. In contrast, Deictic relational responding did not correlate significantly with any IQ index or subtest score, indicating an apparent irrelevance to intelligence. Such is the extent of this lack of correlation, removal of Deictic block scores from overall MRAT scores (i.e. MRAT-4) actually increased the strength of the MRAT's relationship to Full-Scale (r = .69), Verbal IQ (r = .48) and Performance IQ (r = .71). Therefore, the current analysis recommends the removal of deictic score blocks and the adoption of the MRAT-4 score as a more reliable estimate of overall intelligence. Further discussion of the relative contribution of each relational frame to intellectual performance will be discussed further in Section 9.1.3.

A further aim of Experiment 6 was to explore the efficacy of the MRAT in predicting IQ scores for high-ability participants, as Colbert et al. (2017) previously identified a potential ceiling effect for the RAI, and therefore its reduced utility as a proxy measure of IQ for those at the higher end of the ability spectrum. While the strength of the relationship between MRAT-4 and IQ measures was generally weaker for high IQ (110+) participants, the correlation between MRAT-4 and FSIQ was still significant (r = .53). While this effect size is insufficient in supporting the MRAT's use as a proxy measure of IQ, it represents an advantage over the RAI in this regard. As such, the inclusion of additional relational frames, specifically those found to correlate significantly with Full-Scale IQ (i.e. distinction, temporality and analogy), may improve the RAI's ability to approximate IQ for high ability individuals. This will be further explored in Experiment 7, and a comparison of the variety of relational skill assessments administered in the current thesis will be discussed in Section 9.2.

Experiment 7 represented the first analysis of a novel iteration of the RAI, the RAI+, which, based on the results of Experiment 6, integrated trials assessing distinction, temporal and analogical relational responding. As such, the two-block RAI which included Coordination/Opposition and Comparison trials was modified into a 5-block RAI via the addition of blocks assessing Coordination/Distinction, Temporal and Analogy trials. In order to assess the utility of the RAI+ in predicting intellectual ability, in comparison to the original RAI, an investigation into its covariance with IQ scores was conducted using an adult sample. In addition, in order to build upon the findings of Experiment 5, a second group of participants completed well-established metrics of numeracy (Wechsler Adult Intelligence Scale: Arithmetic subtest), literacy and educational attainment (Wechsler Individual Achievement Test: Teacher Edition) in order to assess the relationship between RAI+ scores and academic performance.

Regarding IQ indices, RAI+ showed a medium-strength correlation with Full-Scale (*rho* = .54), Verbal (*rho* = .48) and Performance IQ (*rho* = .48). In addition, RAI+ scores correlated significantly with all four IQ subtests: Vocabulary (*rho* = .37), Similarities (*rho* = .37), Block Design (*rho* = .42) and Matrix Reasoning (*rho* = .42). As such, in comparison to the effect sizes reported by Colbert et al., (2017), the strength of the correlations between the RAI+ and IQ metrics are somewhat underwhelming, despite a widespread pattern of significance. The relative merits of both the RAI+ and MRAT in comparison to the RAI will be elaborated upon in Section 9.2. That point withstanding, while the inclusion of additional frames did not provide a more accurate IQ proxy, it did reduce, to a certain extent, the ceiling effect of the original RAI, as RAI+-FSIQ correlation was significant for high IQ individuals (*rho* = .44). Furthermore, RAI+ scores also predicted Verbal IQ (*rho* = .64) and Performance IQ (*rho* = .47). In light of this finding, despite the diminished utility of the RAI+ in predicting IQ in general, this exploration has been informative insofar as providing insight into a potential remedy to the RAI's ceiling effect. The issue of ceiling effects will be addressed in more detail in Section 9.3.2.1.

In terms of verbal academic ability, RAI+ scores did not correlate with WIAT-T standardized score, or any of its three component subtests (Reading, Reading Comprehension and Spelling) despite extensive research predicting such a relationship (e.g. Barnes-Holmes et al., 2004; Brown et al., 1996; Cowley et al., 1992; Cullinan et al., 1994; de Rose et al., 1992; Farrington-Flint & Wood, 2007; Hayes et al., 2001; Goswami, 1986; Mackay, 1985; Sidman, 1971; Stewart et al., 2004; Wulfert & Hayes, 1988). There a number of potential explanatory factors for this lack of correlation, perhaps most notably the potential unsuitability of the testing battery utilised in high-ability samples.

Correlational analyses also revealed that Full-Scale IQ shared a significant relationship with scores for Comparison (*rho* = .49), Coordination/Distinction (*rho* = .48), Coordination/Opposition (*rho* = .44) and Temporal trials (*rho* = .42). Verbal IQ was also correlated significantly with the following four relational trial blocks: Temporal (*rho* = .43), Comparison (*rho* = .42), Coordination/Distinction (*rho* = .31) and Coordination/Opposition (*rho* = .26). This trend continued, with Performance IQ showing a significant correlation with Coordination/Distinction (*rho* = .48), Coordination/Opposition (*rho* = .44), Comparison (*rho* = .44) and Temporal relations (*rho* = .33). Scores on Analogical task, surprisingly, correlated only with Performance IQ (*rho* = .3), but not Full-Scale or Verbal IQ. In fact, the removal of Analogy scores actually improved the correlation between overall RAI+ and both Full-Scale IQ (*rho* = .55) and Verbal IQ (*rho* = .46), with no effect on Performance IQ. The relevance of individual relational frames to overall intelligence will be further discussed in Section 9.1.3.

9.1 Reconsidering the Relationship between AARR and Intelligence

The current thesis provides substantial support for the assertion that relational skill is functionally related to intellectual performance. In each experiment which assessed both relational responding and IQ (Experiments 2-7), the various assessments of relational responding administered (Relational Abilities Index, SMART:R Assessment, Multiple Relational Assessment Procedure) showed moderate-to-strong significant correlations with Full-Scale IQ as assessed by gold-standard, widely-administered IQ assessments (WASI & WISC). Of great interest to these correlations between relational ability and specific IQ metrics is the point that previous literature predicts such relationships (Cassidy et al., 2010; Colbert et al., 2017; O'Hora et al., 2005, 2008; O'Toole et al., 2009), and furthermore, can provide a functional account of performance on these subtests. Such accounts can contribute to demystifying individual differences in IQ test performance by highlighting the centrality of clearly demonstrable skills, rather than innate, mentalistic or inaccessible faculties. As outlined in Section 1.4, a wide array of tasks commonly employed by traditional IQ tests can be understood as metrics of generalised relational skills, whose establishment and application are functionally understood. As such, a key strength of the current research stream is the clarity and accuracy with which it can identify the crucial underlying skills that facilitate intellectual performance, as assessed by traditional IQ assessments. Indeed, a growing literature base proposes that these functional accounts are both theoretically-grounded (Cassidy et al., 2010; Hayes et al., 2001; Roche et al., 2013) as well as empirically-supported (Colbert et al., 2017; Moran et al., 2015; O'Hora et al., 2005, 2008; O'Toole & Barnes-Holmes, 2009). The key implication of such a proposition is that if intelligence can be defined as a set of skills, the improvement of these skills can foster intellectual growth.

9.1.1 Relational Responding & Verbal IQ

In addition to the widespread correlations reported between relational skill and Full-Scale IQ, Verbal IQ scores showed moderate strength significant correlations with measures of relational skill in 3 of the 5 studies which computed this IQ subindex (the WISC-IV administered in Experiment 3 does not provide Performance or Verbal IQ subindex scores). Indeed, an extensive research base proposes the importance of relational responding to a Verbal IQ test items and general verbal competency (e.g. Cassidy et al., 2010; Colbert et al., 2017; Dymond & Roche, 2013; Hayes et al., 2001), as well as a wide range of language specific skills, such as language acquisition (Barnes-Holmes et al., 2004; Cowley et al., 1992; Cullinan et al., 1994; Hayes et al., 2001; Murphy & Barnes-Holmes, 2010; Stewart et al., 2004; Wulfert & Hayes, 1988), reading (de Rose et al., 1992; Farrington-Flint & Wood, 2007; Goswami, 1986; Mackay, 1985; Sidman, 1971), vocabulary (Edwards et al., 2011; McHugh et al., 2004; Nippold & Sullivan, 1987), grammar (Hock, 1991, 2008) and spelling (Brown et al., 1996; Goswami, 1988; Mackay, 1985). Therefore, the finding of a significant relationship between various assessments of relational skill and Verbal IQ and its subtests is unsurprising.

As outlined in Section 1.4.1, many of the commonly-administered tests of Verbal IQ clearly invoke various relational skills, perhaps most notably the Wechsler Vocabulary and Similarities subtests. As both the WASI and WISC-IV include these two subtests, both subtests were administered in each of the 6 current studies which employed an IQ assessment. Vocabulary is deemed to be one of the strongest predictors of overall intellectual functioning (Marchman & Fernald, 2008; Smith et al., 2005; Vetterli & Furedy, 1997; Wechsler, 1949, 1955, 1974, 1991, 2011), and therefore is included in an extensive array of traditional IQ assessments (e.g. WAIS, WISC, WASI, Stanford-Binet). The Vocabulary subtest presents the participant with a series of words progressively increasing in complexity and requires him/her to define each in turn. As such, this subtest can be considered as an assessment of word-word and word-object coordination relational responding, as the correct answer in any trial requires the participant to provide a synonym or series of words equivalent in meaning to the word to be defined. In line with previous analyses (Colbert et al., 2017; O'Hora et al., 2005, 2008), a significant correlation between Vocabulary and relational responding scores were found for 4 of 5 studies, serving to further underline the relevance of relational responding in vocabulary acquisition. As discussed in Section 1.4.3.1, the relational frame of coordination, in particular, can be viewed as the basis of linguistic reference (Stewart & Roche, 2013) as it is involved in the mapping of words to their physical referent and establishing word-word equivalence relations that facilitates vocabulary expansion. In a similar way, the frame of opposition may facilitate word knowledge, as a once a trained or untrained opposition relation has been established between a novel and a known word (e.g. tiny means the opposite of gargantuan), the latter word's meaning can thus be derived. Therefore, a correlation between relational skills and vocabulary tests is predicted by such functional accounts.

In addition, three of the current analyses found significant correlations between relational responding and the Similarities subtest, thereby adding to previous studies reporting such a relationship (Colbert et al., 2017; O'Hora et al., 2008). The Similarities subtest presents participants with a pair of words (e.g. pen and pencil) and requires the participant to identify in what way these two words are the same or similar. Correct answers usually involve the identification of a functional (e.g. they both write) or categorical classification (e.g. they are both stationary) to which both words are party to. In this way, Similarities invokes both coordination and hierarchical relational responding. Similarities can perhaps be viewed as one of the most obvious examples of

IQ test items which 'tap' relational responding, a topographical similarity that predicts the significant correlations currently reported.

9.1.2 Relational Responding & Performance IQ

In addition to findings in regard to Verbal IQ test items, each of the five experiments which computed a Performance IQ score also reported moderate-to-strong significant correlations between this metric and relational skill. Furthermore, moderate-to-strong significant correlations were found for both Performance IQ subtests administered, Matrix Reasoning (in 4 of 5 experiments) and Block Design (in 4 of 5 experiments), and also one of the two indices, Perceptual Reasoning, which replaced Performance IQ in the WISC-IV testing battery administered in Experiment 3. This is a particularly interesting result, as traditionally, research on relational responding and intelligence has emphasised the clear relevance of the former repertoire to verbally-based tasks (e.g. Cassidy et al., 2010; Colbert et al., 2017; Dymond & Roche, 2013; Hayes et al., 2001). Performance IQ, on the other hand, is intended as a relatively non-verbal measure, being defined as “a measure of fluid reasoning, spatial processing, attentiveness to details, and visual-motor integration” (Lange, 2011, p.1). Therefore, the relevance and application of the currently proposed collection of relational skills to Performance IQ is less readily observable, and thus requires further elucidation.

With the exception of Experiment 3 which administered the WISC-IV, Performance IQ was computed for all other studies on the basis of performance on two subtests: Matrix Reasoning and Block Design. In the case of the former subtest, four of the current studies found significant correlations with relational responding, a result which thus far has only been reported by one previous analysis (Colbert et al., 2017). In this subtest, participants are presented with an array of geometric designs with a section or shape missing. The participant must then select from a selection of sample designs which ‘fits’ into the model displayed, in accordance with some rule of inclusion (e.g. completes some visual pattern) or as the next shape in a given progression (e.g. geometric shapes changing in size/number of sides/orientation etc.). As such, Matrix Reasoning (and other similar assessments, such as Raven’s Matrices; Raven & Court, 2000) can be considered tests of visuospatial analogy and therefore entail a relational component. By analysing the relationship between each member of a given series analogically, the participant can correctly respond to each trial by applying this relational ‘rule’ to the missing shape or section and selecting the appropriate response from the options provided.

While understanding Matrix Reasoning in these terms may contribute to explaining the relevance of relational responding to performance on this task, analogical reasoning was only assessed in 2 of the studies included in the current thesis, Experiments 6 & 7, yet significant correlations were also found between relational responding and this subtest in Experiments 2 and 3. As discussed briefly in Experiment 7, the relationship between analogical responding, viewed as a more advanced form of relational responding as it involves the relating of relations, and more basic forms of relational responding requires further delineation. It may be the case that more foundational relational skills, such as coordination and opposition, may provide the basis for analogical reasoning. Indeed, a small number of studies provided some evidence that sophistication in basic coordination and opposition relational responding is prerequisite for the establishment of analogical reasoning (e.g. Carpentier et al., 2003). Therefore, while the types of relational responding assessed by the RAI as administered in Experiments 2 & 3 (i.e. coordination, opposition, & comparison), do not appear to bear direct relevance to Matrix Reasoning trials, the correlations found may in part be explained by the foundational importance of these more basic relational skills to analogical relational responding. The correlational analyses of Experiments 6 and 7 would appear to support the centrality of coordination, opposition and comparison relational responding, as it was performance on these tasks that were the most closely associated with overall relational responding as assessed by two separate multi-frame assessments of relational skill. Therefore, as analogical relational responding may 'rest' upon one's acquired proficiency in more basic relational frames, the relationship between performance on the RAI and Matrix Reasoning may be mediated by analogical reasoning proficiency.

The second of the Performance IQ subtests, Block Design involves the composition of a geometric design using red and white blocks based on a 2D target design ranging from 1 x 2 to 3 x 3 block arrangements. Block Design is intended to be a measure of visuospatial processing and nonverbal problem solving, as well as fine motor skill (Soto & Kraper, 2013). As was the case with Matrix Reasoning, 4 of 5 relevant studies reported a significant correlation between relational responding and Block Design test scores, a relationship that has been previously reported by O'Hora et al., (2008) and Colbert et al. (2017). As outlined by Hayes et al., 2001, performance on this subtest implicates a form of relational responding termed pragmatic verbal analysis. Pragmatic verbal analysis refers to arbitrarily-applicable relational responding under the control of nonarbitrary physical world relations. When applied to the Block Design subtest specifically, pragmatic verbal analysis refers to the participants' continuous awareness and interpretation of the relation between a progressively evolving current state (i.e. an arrangement of blocks that does not match the target arrangement) and a goal state (i.e. the target arrangement of blocks). In order to achieve the goal state and complete each trial, the participant must analyse the physical relation between the current and desired arrangement of blocks (e.g., this red block should be further to the left, the right side of my design does not match the target arrangement, the two white blocks should be placed above the two red blocks etc.). Therefore, for a given 3 x 3 block design, the participant must construct a design in which each of the 9 blocks included match those in the target design in terms of orientation, colour and placement. To do this, as evinced above, the participant must ensure that relation between his/her blocks and that of the target is one of equivalence/coordination, but also must ensure that spatial relations between each of his/her 9 individual blocks are analogous to the individual spatial relations displayed in that target design. While the relevance of relational responding to performance on this type of test has been expounded theoretically, the current analysis substantially bolsters the relatively limited empirical research which has investigated such a relationship.

9.1.3 Individual Relational Frames and Intelligence

A collection of studies included in the current thesis highlight the general relevance of relational responding to intelligence, most notably in the case of coordination, opposition and comparison relations. While the relevance of these relational frames has been established by previously published analyses (Colbert et al., 2017; Hayes & Stewart, 2016), the results of Experiments 2, 3 & 5 further highlight the important role of these forms of relational responding with regard to general intelligence. That being said, the pertinence of the wider range of relational skills (e.g. distinction, deixis, analogy) has not received a comparable degree of attention and as such, the role such relational skills play with regard to intelligence is poorly understood. Therefore, one of the primary aims of the current thesis was to investigate the relative contribution and relevance of this extended relational skill repertoire to intellectual function. Experiments 6 & 7 endeavoured to pursue this aim by administering extended relational skills testing batteries (MRAT & RAI+ respectively) and assessing the degree of correlation between specific relational frames and various IQ indices and subtests.

For both Experiments 6 & 7, coordination/opposition and comparison relational responding were found to be among the strongest predictors of not only intellectual performance on multiple metrics, but overall relational skill. Such a finding would propose that these more basic relational frames may play a fundamental and pervasive role with regard to the wider relational responding repertoire and general intellectual skills. This relative centrality, in the context of other relational frames, provides an important insight into why the RAI has shown such promise as a functional alternative to traditional IQ assessments. Indeed, before this delineation of the individual relationships of specific relational frames to intellectual function, it was entirely unclear as to whether the strong RAI-IQ correlations may be explained by the relevance of the specific forms of relational responding being assessed, or whether this relationship was due to the fact that the testing battery merely tapped a more general, overarching repertoire that was functionally related to IQ. It appears that due to the weaker levels of correlation found between IQ and other relational frames, that coordination, opposition and comparison relations may be the relational frames most closely associated with intellectual skill.

In terms of the additional relational frames that were assessed, correlations between such frames and various IQ metrics were generally found to be either non-significant, or significant with diminished effect. One of the additional frames assessed by both the MRAT and the RAI+ was Analogy, proposed to be of importance to an array of higher order cognitive skills (Gentner et al., 2001; Gentner & Smith, 2013; Richland, Holyoak, & Stigler, 2004; Stewart et al., 2004). Due to its status as a more advanced form of relational responding (necessitating the relating between relations), it was hoped that the inclusion of such complex trials would aid in reducing the ceiling effects commonly witnessed in the original RAI (Colbert et al., 2017). The lack of correlation between performance on analogical reasoning trials and most IQ metrics, therefore, is striking. For example, in our analyses of RAI+ and WASI IQ score distributions in Experiment 7, Analogy was the only module out of the five relational frames assessed that did not show a significant correlation with Full-Scale IQ, as each of the other frames displayed very similar, moderate-strength correlations. Furthermore, while each of the 4 other frames correlated significantly with Verbal IQ and Performance IQ, Analogy only correlated significantly with the latter, albeit with the lowest level of covariance amongst other relational frames. To complement these findings, the Analogy block included in the MRAT as administered in Experiment 6 was similarly outperformed by coordination/opposition, coordination/distinction and comparison relational responding as predictors of WASI index, subindex and subtest scores. While Analogy scores did correlate with Full-Scale IQ, the strength of this relationship was considerably lower than that of coordination/opposition and coordination/distinction block scores and marginally lower than comparison block scores. In addition, unlike these three blocks, Analogy failed to correlate significantly with Verbal IQ, Performance IQ or any WASI subtest.

The lack of correlation between analogical reasoning performance and IQ can be interpreted in a number of ways. Perhaps the most obvious interpretation is to take the results *prima facie* and acknowledge that this repertoire is not closely related (or related to any significant degree) to general intellectual performance. However, given that this relationship has been propagated by extensive theoretical accounts within behaviour analysis (Hayes et al., 2001; McLoughlin et al., 2018; Roche et al., 2013; Stewart et al., 2004; Stewart et al., 2002), and additional evidence from external research streams (Geake & Hansen, 2005; Gentner et al., 2001; Hofstadter, 2001; Mulholland, Pellegrino, & Glaser, 1980; Spearman, 1946), it is perhaps premature to accept such a conclusion. Instead, it may be the case that, in light of such an extensive research base proposing the contrary, that the two current studies failing to find a relationship may be underpowered (combined n = 133) for example. Furthermore, it may be argued that while analogical reasoning does contribute to intellectual performance, the IQ testing battery administered (the WASI) does not adequately or accurately assess this specific facet of intelligence. While Matrix Reasoning may be considered a test of visuospatial analogy, the WASI does not include any explicit assessment of verbal analogical reasoning, unlike a number of other IQ testing batteries. Therefore, it may be the case that analogical reasoning proficiency was taken into consideration in the demarcation of individual differences in relational responding, but not intelligence. Had the Stanford-Binet or Woodcock-Johnson IQ assessments ben administered instead, both of which include subtests with a defined emphasis on analogical reasoning, perhaps a significant relationship between IQ test scores and analogy block scores may have emerged.

That being said, the current research endeavours to improve the validity of relational skills as reliable alternatives to traditional IQ tests. Therefore, if the relationship between intelligence a given relational skill is rendered wholly insignificant based on an arbitrary procedural detail such as the selection of one gold-standard IQ testing battery over another, perhaps that specific relational skill does not warrant inclusion in such a relational skills battery. Currently, the relational skill assessment of choice, the RAI, has shown considerable efficacy across multiple analyses in predicting IQ based on an individual's proficiency in coordination, opposition and comparison relational responding. If the RAI's format is to be altered, therefore, the inclusion of additional relational frames must only be catalysed by a succession of empirical investigations reporting a strong and significant relationship between that proposed frame and intellectual performance. As the current evidence base is not sufficient in the case of analogical reasoning, there is no reason to expect that the inclusion of analogical tasks would improve the utility of the RAI. However, given the swathe of theoretical accounts proposing the relevance of this form of responding to intelligence, further research should build upon the findings of the current thesis by further investigating the empirical relationship between these two repertoires.

Alongside analogy, deictic or perspective-taking relational responding were analysed regarding its relevance to intellectual performance using the MRAT in Experiment 6. Unlike in the case of analogical reasoning, whereby a positive correlation with IQ was predicted by a considerable literature base, only a very limited number of studies (Gore et al., 2010; Rehfeldt et al., 2007; Tarshis & Shore, 1991) proposed an association between deixis and intellectual performance. Furthermore, unlike many other forms of relational responding, few IQ tests include items that would appear to rely primarily on the ability to embody different temporal, spatial or personal perspectives. A small number of individual test items included in the Wechsler Comprehension subtest may possibly involve this form of responding with an empathetic context (e.g. "Why should you keep a promise?"). However, examples such as these are minimal. Perhaps the IQ subtest which most obviously requires perspective-taking is the Wechsler Block Design subtest, but emerging evidence has suggested that object-centred reorientations are distinct to ego-centric perspective transformation (as discussed in Chapter 7). In light of such points, this analysis of deixis was far more exploratory in nature.

The results of this investigation were relatively clear-cut, as performance on deictic relational trials was not found to correlate significantly with any WASI IQ index or subtest score. In fact, the removal of deictic scores from the MRAT's testing battery (i.e. the MRAT-4 score) actually led to a slightly increased efficacy in predicting Full-Scale, Verbal and Performance IQ. Therefore, on the basis of current results, and the relative dearth of opposing evidence, it may be tentatively concluded that deictic relational responding may not share a close relationship with general intellectual performance. Any significant relationship found between these two repertoires may, therefore, most likely be a statistical artefact of a mediating mutual relationship to general relational skill, rather than a direct relationship *per se*. In sum, while an array of research highlights the importance of perspective-taking to a number of important behaviours, such as empathy (e.g. Decety & Lamm, 2006; Lamm, Batson, & Decety, 2007) and cooperation (e.g. Galinsky & Mussweiler, 2001; Johnson, 1975), it appears that the influence of this form of responding does not extend into the intellectual domain.

At this point, it is important to highlight that, in pursuit of a behaviour-analytic, functional account of intelligence, a degree of divergence in what is considered to constitute intelligence is to be expected. Indeed, the question of whether this lack of correlation between deictic relational responding and a more traditional conceptualisation of intellectual performance serves as a basis to conclude that the former is not relevant to the latter, or alternatively, that the former represents an aspect of intellectual performance not adequately assessed or described by the latter, is a matter of debate heavily reliant on an individual's pre-existing theoretical position. Furthermore, as was discussed in regard to analogical relational responding, in the event that a different IQ assessment was administered, specifically one which included trials that bore closer resemblance to deictic relational responding and found a significant correlation – would that be sufficient evidence to support claims of the relevance of deixis to intelligence? Indeed, given the lack of theoretical and psychometric consensus on what intelligence constitutes, such a question may be answered more easily via appeal to theoretical, rather than empirical arguments. While the current thesis provides empirical support for theoretical accounts proposing a relationship between intellectual performance and a wide range of relational frames, it appears that this support is lacking in the case of deictic relational responding.

9.2 Comparison of Relational Skill Measures

The current thesis comprised, in part, of the study of a small collection of relational skill assessments, each of which was evaluated regarding their degree of correlation to general intellectual performance. While the context and rationale for the administration of each relational skills assessment varied across studies, it is nonetheless illuminating to compare their relative efficacy in predicting IQ, as a means of understanding what kind of relational skills and assessment formats best lend themselves to this purpose. In total, four different relational skills assessments were administered, the Relational Abilities Index (55-trials assessing abstract coordination, opposition and comparison relations), the SMART:R assessment (56 trials assessing basic coordination, opposition and comparison relations), the Multiple Relational Assessment Test (78 trials assessing coordination, opposition, comparison/temporal, distinction, analogical and deictic relations) and the Relational Abilities Index + (67 trials assessing coordination, opposition, comparison, distinction, and analogical relations). For clarity purposes, it may be prudent to remove the SMART:R from this analysis, due to the clear functional and topographical distinction between this basic relational skills assessment designed to test NARR as well as AARR, and the other three assessments, each of which employed abstract relata exclusively in assessing higher level AARR and DRR. The reduced level of covariance found between scores for the SMART:R assessment and Full-Scale IQ (r = .45) is expected, due to the assertion that the relationship between relational responding and general intelligence is due to the generalised application of DRR. The SMART:R, on the other hand, is designed to establish this form of responding (i.e. the ability to arbitrarily apply relational responding to novel relata), and therefore these more basic relational skills may not be as germane to domains beyond the specific task topography, such as intellectual performance.

In comparing the degree of correlation reported between Full-Scale IQ and each of the remaining three relational skill assessments (RAI, MRAT & RAI+), it appears that the original RAI is still superior in predicting intellectual performance. While the strength of the correlations between RAI and FSIQ were not quite as strong as those reported by Colbert et al. (2017), the effect sizes for this relationship were approaching this level in Experiment 2 ($r = .64$) and Experiment 3 ($r = .69$). The strength of the correlation between FSIQ and MRAT score was also of similar intensity ($r = .63$), as was the MRAT-4 score ($r = .69$). The correlation between scores on the RAI+, a test developed as a potential improvement upon the original RAI format, and FSIQ ($r = .54$) was lower than both the original RAI and the MRAT.

In line with previous comments regarding the centrality of coordination, opposition and comparison relations, the superior performance of the RAI in predicting intellectual performance is perhaps unsurprising. While both the MRAT and the RAI+ assessed a wider range of relational frames, analyses of the additional relational frames found that they did not show as close a relationship to IQ metrics as coordination, opposition and comparison. It appears, therefore, at least based on the current series of investigations, that it is these three foundational frames that may be most closely related to intellectual performance, and therefore, future assessment should retain, and possibly extend, such trials as a means of predicting IQ. As the MRAT and RAI+ were both studied to investigate the potential of improving upon the RAI's utility as a proxy measure of intelligence, it would appear that the current thesis indicates a failure in this regard. However, these analyses provided extremely important insights into the relevance of the wider collection of relational frames to intelligence and helped rule out (at least tentatively) additional forms of responding that do not appear to exert as influential an impact on general intelligence.

Interestingly, however, upon analysis of the relative correlation between the RAI, RAI+ and MRAT and Performance IQ, it appears that the MRAT displays a considerably stronger correlation ($r = .67$) than either of the other two assessments. This result harbours considerable promise, as it may offer the potential of increasing the relevance of the original RAI to Performance IQ test items. It appears that this result may most likely be due to the inclusion of temporal and/or distinction relations, as of the five frames assessed by the MRAT, only coordination/opposition, coordination/distinction and temporal relations were significantly correlated with Performance IQ. As coordination and opposition relations are already assessed by the RAI, the inclusion of temporal and distinction tasks may have contributed to the MRAT's increased efficacy in this regard.

It is also possible that this effect may be due to more general procedural details of the MRAT, the most obvious of which is the extended length of the assessment (78 trials compared to the 55-trial RAI). As such, it may be the case that a longer assessment provides a greater opportunity for individual differences in relational responding proficiency to emerge, thus resulting in a more accurate and nuanced approximation of relational skill. The potential of this simple difference explaining this improved effect is rendered somewhat less likely due to the failure of the 67-trial RAI+ to improve upon the original RAI's degree of correlation with IQ. Nonetheless, this is a potential contributory factor and requires further empirical investigation. In line with the recommendations to reduce the RAI's ceiling effect (see Section 9.3.2.1), the addition of a small number of more complex 4- or 5- node trials onto each of the RAI's two existing blocks may at least partially satisfy this query.

The current thesis has proven extremely informative in delineating the relevance of the wider range of relational frames to intellectual performance. Previously, assertions of the contribution of relational responding to intelligence have rested mainly on theoretical, rather than empirical accounts. As such, the correlational analyses of the current thesis have added considerably to our understanding of which relational frames are most closely related to intelligence. A secondary motivation behind the expansion of the RAI test battery was to shed light on which relational frames may be most relevant to intelligence, with a view of subsequently integrating these frames into the SMART program for training. Indeed, the correlational analyses included in the current thesis contribute significantly to further delineating the relationship between relational skill and intelligence, which may yet extend the efficacy of relational skills training in facilitating performance improvements of greater intensity and variety moving forward.

9.3 Evaluating Intervention Efficacy in Improving Intellectual Performance

The current analysis adds considerably to the emerging line of research proposing that intellectual performance can be improved via relational skills intervention. Most of the more notable investigations contributing to this research stream have focussed on the application of the SMART program, and as such, several of the current studies analysed the efficacy of this program across various domains and populations. In general, the results of the three analyses which assessed the effectiveness of SMART in improving IQ scores combine to support previous reports of such an efficacy in improving intellectual function (e.g. Amd & Roche, 2018; Cassidy et al., 2011; 2016; Hayes & Stewart, 2016). In particular, the results of Experiment 2, which reported a mean WASI Full-Scale IQ increase of 18.4 points, bear resemblance to the results of perhaps the most noteworthy analysis in the field currently, Cassidy et al. (2016), which found post-intervention increases of 23 points. The results of this experiment also support Cassidy et al.'s findings of significant score increases for Verbal and Performance IQ, with Experiment 2 finding increases of 19.7 and 13.5 points, respectively. Significant score increases were also found for Full-Scale IQ in

Experiments 3 & 5. However, the scale of these score increases was substantially smaller (6 points & 3.8 points respectively). In the case of Experiment 3, perhaps the most obvious explanation for this reduced effect would be the generally lower training completion rate for this sample, as only 5 of 28 participants completed all 55 training levels. The issue of completion rates and their impact on post-intervention outcomes will be discussed in detail in Section 9.3.2.2.

9.3.1 Comparative efficacy of other "cognitive enhancement" training methods

Of particular interest in the current discussion is the finding that while there are a number of intervention studies reporting success in improving intellectual function, very few have been able to produce improvements as large and widespread as the SMART training program. There have been numerous training programs that have been proposed to improve intellectual function by targeting performance in very specific cognitive domains, such as working memory (e.g. Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011; Klingberg et al., 2005), attention (e.g. Rueda et al., 2004), mental planning and strategy (Basak et al., 2008) and general problem solving and creativity (Tranter & Koutstaal, 2008). However, none of these studies have demonstrated reliable rises using a full-scale IQ assessment, an effect which has now been found repeatedly using SMART in the current thesis and in previously published analyses (e.g., Cassidy et al., 2011; 2016; Hayes & Stewart, 2016). The subsequent sections will address some of the most noteworthy issues regarding the comparative efficacy of other training interventions in improving intellectual function.

9.3.1.1 Far Transfer.

Some of the most noteworthy research on intellectual enhancement in recent times has focused on improving levels of working memory, using what is called the dual *n*-back procedure (e.g. Buschkuehl & Jaeggi, 2010; Jaeggi et al., 2008, 2011). However, despite the acclaim bestowed upon such reports (see Sternberg, 2008), the efficacy of working memory interventions in improving intellectual performance has also come under severe scrutiny due to various methodological issues (see Section 1.5). In addition, doubts have been raised over the generalizability of training benefits beyond working memory and into more general intellectual domains (Ackerman et al., 2005; Colom et al., 2008; Kane et al., 2005; Moody, 2009) as such studies have shown insufficient evidence of far transfer (Lampit, Hallock, & Valenzuela, 2014; Melby-Lervag et al., 2013; Rapport, Orban, Kofler, & Friedman, 2013). Indeed, in a systematic meta-analysis of working memory interventions designed to improve general cognitive ability and/or intellectual performance, Melby-Lervag et al. (2013) report that not one of the 23 studies included administered a full-scale IQ assessment, which, given the conclusions drawn from many of these studies, the authors contest should have been one of the primary outcome measures for this research stream. In comparison to the current research stream, working memory research reports more modest gains, typically of just a few standardized points, on a specific domain of intellectual performance (fluid intelligence) as assessed almost exclusively by matrix reasoning tasks. Furthermore, a number of attempts at replicating these results have not been successful (Chooi & Thompson, 2012; Lawlor-Savage & Goghari, 2016; Owen et al., 2010; Redick, Shipstead, Wiemers, Melby-Lervåg, & Hulme, 2015).

Indeed, it may seem that the most prevalent criticism of working memory interventions, as well as other ‘brain training’ interventions, concerns an apparent failure to demonstrate far-transfer. Based on such a finding, several meta-analyses (Melby-Lervag et al., 2013; Moreau et al., 2019; Simons et al., 2016) and open-letters (e.g. Stanford Center on Longevity, 2014) have now asserted such a conclusion based on the current research literature. Interestingly, however, to the author’s knowledge at least, no such analysis has analysed empirical investigations into the efficacy of relational skills training in increasing IQ before reaching such a conclusion, as none of the above-mentioned meta-analyses included such studies in their analysis.

The recurrent finding that relational skills training results in increments in performance on tasks that bear little topographical similarity to those being trained (i.e. far transfer), may therefore represent an advantage of relational skills training over other 'brain training' protocols. Of the two IQ assessments utilised by the current thesis (WISC and WASI), one would be hard-pressed to find even one subtest that could be considered topographically identical or even similar to the skills being targeted by the relational skills training interventions administered. This is perhaps most evident in the case of improved scores for Performance IQ test items such as Block Design and Matrix Reasoning (as found in Experiment 2, for example). While competency on these subtests and many others can be understood theoretically as a reflection of relational skill sophistication, the task themselves are in no way similar to the verbal syllogisms being trained as part of the SMART program. Indeed, few would argue that improving skill in deriving untrained relations in two- and three-premise logical syllogisms and witnessing resultant increments in an individual's ability to arrange physical blocks to match a target arrangement represents near- or even moderate-transfer, such is the discrepancy in task topography. Beyond this point, the sheer variety of IQ tasks that show score increases following relational skills training would seem to undermine claims that SMART results exclusively in limited, domain-specific effects, rather than generalised, far-transfer effects. Based on the current evidence, it therefore appears that SMART achieves far-transfer in ways seldom witnessed for other 'cognitive enhancement' interventions.

.3.1.2 Mechanisms of Intervention.

Another common criticism of interventions designed to improve intellectual and/or cognitive function is that the mechanisms by which such interventions exert their effects are poorly understood and/or vaguely described, as outlined by Simons et al. (2016). A clear advantage of the current intervention is that it is underpinned by a clear and well-documented theoretical account of how and why relational skills contribute to intellectual performance, as an extensive range of intellectual, verbal, logical and mathematical skills can be understood from a relational responding perspective (see Section 1.4.3). In addition, the training protocol utilised, Multiple Exemplar Training, is well-established within the behavioural literature as an efficacious means of establishing various forms of generalised responding by 'shaping' responses across a large number of stimuli by means of corrective feedback (see Holth, 2017 for an overview of the history of MET interventions). Furthermore, due to the conceptualisation of derived relational responding as a generalised operant (see Healy et al., 2000), the process by which an improved relational skill repertoire can exert far-reaching impact into topographically distinct, but related, domains (such as IQ test performance) is both explained and predicted. As such, the relevance of relational skills to intelligence, the means by which such skills can be improved, and the mechanism underlying the effects of an enhanced relational skill repertoire into other domains are all well understood and elucidated by a considerable body of empirical and theoretical accounts.

9.3.1.3 Efficacy across Age and Ability Levels.

Another of the recommendations proposed by Simons et al., (2016) concerns the need to explore the effects of training interventions across a range of potential samples. Simons et al. stress that the claims of numerous interventions' efficacy in benefiting cognitive function in normally-developed, healthy adults are actually founded upon research using individuals with cognitive deficits. In the case of relational skills interventions, the evidence base has displayed considerable growth in recent times, as numerous analyses (including those within the current thesis) have confirmed the positive effect of training across the age and ability spectrum, reporting benefits for normally-developing young children (Parra & Ruiz, 2016; Vizcaíno-Torres et al., 2015), pre-adolescent (Cassidy et al., 2011; 2016), early adolescent (Experiment 5) and late adolescent children (Experiment 2; Cassidy et al, 2016), as well as those with learning difficulties and/or below average ability (Experiment 4; Hayes & Stewart, 2016; Ruiz, Suarez, & Lopes, 2012).

9.3.1.4 Opportunity Costs.

In an open letter released by the Max Planck Institute for Human Development and the Stanford Center on Longevity entitled “A Consensus on the Brain Training Industry From the Scientific Community” (2014), the issue of opportunity costs are also highlighted, as the authors warn that time spent engaging in computerised training protocols may drain time spent engaging in other more useful activities. The relative value of a given training intervention therefore may be best established via comparison with active control measures believed to produce benefits for cognitive function. Indeed, the current evidence base supporting SMART requires further work of this kind, as few studies have implemented active control measures. Hayes & Stewart (2016) is one such exception, as SMART was found to result in intellectual and academic ability improvements that were not replicated in an ability-matched group who completed a computer-coding training program. In addition, Experiment 4 found significantly greater IQ score improvements in a SMART-trained group of individuals with additional educational needs, in comparison to an ability-matched group who did not receive the training intervention but continued with their regular remedial classes. While such studies provide some evidence that SMART provides benefits beyond what could be expected of a non-specific effect of intervention, further research designs utilising active control measures are required to confirm or disconfirm this point.

In sum, upon reflection of the wider range of intervention protocols currently proposed to exert a positive effect on intellectual function, it appears that many of the caveats that burden other programs' claims with heavy suspicion are not applicable (or at the very least, are significantly less applicable) to relational skills interventions. Indeed, the results of the current thesis add considerably to establishing the efficacy of relational skills training as amongst the most promising and effective means of producing demonstrable improvements in intellectual performance, producing a number of investigations that appear to satisfy many of the concerns commonly levelled against 'cognitive enhancement' interventions.

9.3.2 Factors influencing IQ score rises

9.3.2.1 Ceiling effect.

One of the recurrent findings of the current thesis is the SMART program's reduced efficacy in improving intellectual and academic performance for those at the higher end of the ability spectrum. Across 4 of the 5 intervention studies included in the current thesis, statistical analyses revealed that those with higher levels of ability at baseline demonstrated score increases of lower magnitude than those with average or below average ability on a given metric. This effect, therefore, was apparent across samples and across performance domains (i.e. relational responding, intellectual performance and academic ability).

For example, Experiment 1 revealed an inverse correlation between baseline RAI scores and post-intervention RAI score increases. In addition, those who recorded scores within the top 20% of the sample showed RAI score increases significantly lower (M = 4.6) than the rest of the sample (M = 7.5). In the case of relational responding at least, this ceiling effect may be explained simply by the fact that those who score highly on the RAI have already displayed proficiency in the majority of the tasks that will subsequently be targeted and trained by the training intervention. As such, an individual who records an RAI score of 50 out of 55 stands ‘less to gain’ from the SMART program than an individual who registers a score of 25, as the former individual has responded correctly to 50 of the 55 types of relational skill that will then be trained. In light of this point, future analyses should attempt to extend both the RAI and the SMART program to include additional trials of greater complexity in order to allow high-ability participants to benefit in a manner commensurate to lower ability individuals. This point will be further discussed in Section 9.3.2.1.

A similar trend was discovered for academic ability, as there appeared to be a clear ceiling effect with regard to post-intervention Drumcondra Reasoning Test score rises. Following training, significant score increases were found for DRT Overall Reasoning Test scores for those within the below average and average categorisations but not for participants in the above average range. There are a number of interpretations of this finding, perhaps the most obvious of which is that due to the correlation between RAI scores and DRT scores, the participants in the above average DRT range likely presented with higher baseline RAIs, and therefore acquired a lower number of ‘new’ relational skills, resulting in a lower level of DRT performance improvement. Upon analysis of the post-intervention RAI score changes across ability groups, this potential interpretation is not supported, as RAI score increases were significant, and remarkably similar, for each of the ability groups: below average (M = 9), average (M = 8.3) and above average (M = 8.8). However, there was a far stronger inverse correlation between baseline RAI and post-intervention RAI change for the above-average group (r = -.86) when compared to the overall sample (r = -.56). This effect somewhat carried over into DRT outcomes, as pre-intervention RAI scores were inversely correlated with post-intervention increases in DRT Overall Reasoning scores (r = -.28). This effect was not apparent in either of the two other ability groupings.

Based on this pattern of results, a second potential explanation would be that the complexity of the tasks currently trained by the SMART program do not contribute as strongly to academic performance (at least as assessed by the DRT) at the upper end of the ability spectrum (i.e. the establishment of more advanced or more varied relational responding repertoires may be necessary to elevate already high levels of performance). Similarly, it may be the case that the most advanced DRT items may not invoke proficiency in coordination, opposition and comparison relational responding, as this test involves a number of items which can be considered assessments of analogical and/or hierarchical relational responding. The question of if and how a more sophisticated repertoire of coordination, opposition and comparison relational responding may transfer into performance on other types of relational skill, such as analogical and hierarchical responding is yet to be elucidated empirically. As such, it may be the case that the current program was insufficient in improving the skills required to record DRT scores at the extreme upper end of performance. In sum, however, the current data do not provide a clear-cut explanation for this reduced effect of SMART for high-ability individuals, and as such, future analyses should aim to further investigate this trend.

In terms of intellectual performance, evidence for a ceiling effect is less clear. Experiments 2 and 4 report no correlation between baseline Full-Scale IQ score and post-intervention IQ gains. While Experiment 3 found a strong correlation between these two variables (r = .68), this effect disappeared after controlling for number of training levels completed. Indeed, there may be a complex relationship between baseline ability, training level completion and post-intervention outcomes for this age group (10-11-year-olds), as it was nonetheless found that baseline IQ significantly predicted number of training levels completed, which then predicted subsequent IQ gains. As such, our analyses indicate that while higher IQ individuals are more likely to complete the training program, it appears that it is the number of training levels completed, in isolation of baseline ability that predicts post-training improvements in intellectual performance. It may be the case that higher IQ individuals were simply more likely to complete training within the relatively restrictive window of 3-4 months, and that, if lower IQ individuals had more time, they too may have completed training and therefore displayed greater performance improvements. This 'dosage effect' would therefore heavily emphasise the importance of completing the training program. Such an effect will be discussed subsequently in Section 9.3.2.2.

9.3.2.2 Dosage effect.

Another factor closely associated with post-intervention outcomes was the number of training levels completed, with a dosage effect of this kind reported in Experiments 3 and 6. In Experiment 3, number of training levels completed correlated significantly with post-intervention IQ score increases. In addition, those that completed training displayed IQ rises multiple times greater than those who did not. To complement these findings, Experiment 6 reported significant Full-Scale and Verbal IQ score rises for those who completed all 55 training levels, but not for those who failed to do so. Such results would appear to indicate that training level completion is essential to produce significant increments in intellectual performance. These results complement the findings of Amd & Roche (2018) who reported such an effect in a small sample of non-English speaking children. An analysis of dosage effect was not possible for Experiments 2 and 4, as the training data was lost for the former analysis and all participants were required to complete training in the latter. In addition, as the second RAI was only administered following training completion in Experiment 1, it was not possible to calculate post-intervention score changes in RAI for uncompleted participants. Therefore, in both experiments that facilitated such an analysis, the number of training levels completed was found to be a significant contributor to post-intervention score increases.

9.4 Future research

While the current thesis represents a substantial progression for the current research stream and has catalysed a wide range of novel insights into relational skills training and the relationship between relational responding and intelligence, the results of the current analyses have also identified a number of issues that require further investigation to continue to move this research agenda forward both in terms of validity and utility.

9.4.1 Experimental control

In assessing the potential effect of a given training intervention, double-blind, placebo-controlled randomised control trials are considered the ‘gold-standard’ means of inferring causality (Simons et al., 2016). While the current thesis represents a clear advance in the experimental rigour and stringency employed in assessing the SMART program, an experimental design of this type is yet to be implemented. While Experiments 2 and 5, for example, utilised a single-blind randomised control trial, they did not include a placebo intervention that would facilitate a double-blinded study. Future studies should therefore further improve the experimental control of intervention studies by utilising active control groups who complete placebo interventions or other interventions intended to improve intellectual/cognitive/academic performance in order to assess the proposed ‘SMART effect’. While the body of evidence collated thus far would seem to reliably propose such an effect, it is essential to base such assertions on research investigations of the highest quality, particularly given the issues regarding the publicising of spurious and/or exaggerated effects of many other ‘brain training’ programs. In consideration of such a point, it is important that relational skill interventions extricate themselves from other programs that profess similar applications and benefits, based on marketing aspirations moreso than empirical evidence. As such, future investigations should aim to continue to explore the efficacy of the SMART program by conducting analyses representing best practice in the analysis of intervention effects.

9.4.2 Mediation analysis

While the current thesis provides a theoretical account to explain the mechanisms explaining how an improved relational responding repertoire may improve intellectual and academic performance, the validity of such an account may be investigated statistically using mediation analysis. This form of statistical analysis allows the investigation of whether performance improvements on a given metric can be explained by concurrent improvements on another. Mediation analysis, in the current context, would thereby aid in discriminating whether post-intervention score increases can be attributed to the effects of improving the repertoire being trained (i.e. relational skills), rather than other extraneous factors. Indeed, in a meta-analysis of working memory interventions, Melby-Lervag et al. (2016) implemented this form of analysis and proposed that any far-transfer effects reported by such interventions are not explained by improvements in working memory capacity (i.e. that the training did not cause these far-transfer effects). While the current literature base would strongly assert that relational skill training exerts a positive effect on intellectual performance in particular, the implementation of mediation analysis may provide a greater degree of certainty in making such claims, in the event enhanced relational skills were found to explain improved intellectual performance following training.

9.4.3 Use of Guttmann-style scales

While assessments of relational responding, such as the RAI, RAI+ and MRAT have shown reasonable promise as potential functional alternatives to more traditional IQ assessments, one clear advantage of more mainstream intelligence assessments is their use of Guttman-style (also known as cumulative) scales. Such scales rank-order individual test items in levels of difficulty so that if an individual is capable of responding correctly to a given test item, he/she can be expected to respond correctly to all previous test items if required to (Guttman, 1950, 1954). Traditional IQ subtests tend to adopt this scaling to facilitate different starting points for various age groups under the assumption that a participant will most likely respond correctly to earlier trials, and therefore can 'skip' such trials. For example, the WASI is an IQ test designed for both children and adults, and as such necessitates set starting points for individuals based on age (and entailed expected ability). For example, for the Similarities subtest, 9-year-old children begin on item 5 (i.e. "How are red and blue alike?"), whereas 12-year-old begin on item 7 (i.e. "How are Grapes and Strawberries alike?"). In a similar fashion, various endpoints designating the final test item to be administered to members of a given age group are also demarcated (e.g. item 20 for 6-8-year-olds, item 24 for 9-11 –year-olds). In the case of IQ subtests, this is done to minimise administration time and any potential adverse effect of repeatedly exposing individuals to trials deemed beyond their capabilities.

While relational skill assessments may adopt the Guttman-style scaling for these reasons also, such a scale would serve a function of even greater importance: reducing the likelihood of score inflation due to random responding. Unlike most IQ test items, assessments such as the RAI, RAI+ and MRAT involve binary response options (i.e. "Yes" or "No"), thereby providing participants with a 50% chance of responding correctly to any trial regardless of their actual ability. As such, by presenting participants to a series of tasks which require relational skills beyond their current repertoires, scores will be artificially improved by purely chance responding. For instance, if a given child does not possess the level of derived relational responding to respond correct to even the final 10 RAI tasks, he/she will, on average, respond correctly to 5 of these final 10 trials. This, therefore, reduces the utility of the RAI and other similar scales in providing an accurate measurement of an individual's relational skills.

In line with this modification, relational skills assessments should also utilise discontinue rules. As these assessments are already formatted to incrementally increase task difficulty as an individual progresses through testing blocks, such rules will also aid in avoiding the redundant presentation of a series of trials that the participant is not capable of responding correctly to. Traditionally, IQ subtests specify the number of incorrect responses that must be emitted before the test is prematurely ended. For example, in the Wechsler Vocabulary subtest, if a participant provides 4 incorrect responses in a row, it is deemed unnecessary to continue the test. In a similar vein to the addition of specified endpoints, this modification will tailor the test administration not only to the level of ability expected of a participant's age group, but to his/her specific level of ability by ending the test once the participant registers a specified number of incorrect responses. As such, the addition of discontinue rules will help reduce administration time and improve the accuracy of the assessment in estimating relational skill by further reducing the number of trials an individual is likely to resort to randomly responding to.

9.4.4 Modifications to SMART:Remedial program

While the SMART:R program was found to be effective in achieving its primary aim (i.e. to establish arbitrarily-applicable relational responding), future administrations may modify the program in a number of ways. For one, archaic and/or complex words were utilised as a means of reducing participants reliance on word knowledge in deriving relations in Levels 14-18, thereby requiring them to infer word meaning based on the relational premises presented. While many of these words could be assumed to be beyond the participants' comprehension given their age and ability level, this was not confirmed by experimental means. Future administrations of the program should aim to identify whether the "novel" words introduced from Level 14 are already established in participants' vocabularies, in order to ensure that individuals can only respond correctly by truly deriving the definition of such words.

Furthermore, this research stream may benefit from a more extensive experimental design, that administered both the SMART:R and traditional SMART programs consecutively to a sample which has not demonstrated AARR proficiency at the outset. As the SMART:R program was intended to remediate the skills prerequisite for the main SMART program, future investigations should administer the SMART:R program in a sample deficient in regards to these prerequisite skills and then explore whether the main SMART program can be successfully completed following administration of the SMART:R intervention. As the main SMART program provides a greater opportunity to improve for those with lower levels of relational skills proficiency, the implementation of a SMART:R plus SMART design in low-ability samples harbours the potential for considerable improvement in intellectual performance, perhaps beyond what has been thus far achieved in published studies.

9.4.5 Longitudinal research

Most of the currently published interventions proposing the efficacy of SMART have analysed simple pre- to post-intervention score changes, with assessments taking place in close proximity to the start and finish of training. As such, the issue of whether post-SMART score increases are transient has been left relatively unexplored. While most analyses of the SMART program have administered follow-up assessments well beyond the recommended test-retest interval for their given testing batteries (thereby eliminating practice effects as a significant factor), the question of whether the performance improvements witnessed following training sustain over the longer term following the cessation of training remains unanswered. Indeed, rather than expecting post-intervention improvements to 'wash-out' over a period of time, it could equally be suggested that due to the nature of the skills being trained, specifically the characteristic generalisability of such skills, that intellectual and/or academic performance may continue to rise following training completion. In essence, the skills being trained facilitate future learning, and as such, the benefits of an enhanced repertoire of relational skills may foster improvement far beyond the intervention period. It may be the case that an improved ability to construct relational networks may lead to exponential intellectual growth moving forward, as such networks may serve as the basis for an accelerated learning process. However, such hypotheses require further empirical investigation to assess the stability of post-intervention gains, and to identify the long-term implications of relational skills training, as well as procedures to maintain performance increments beyond the training period (if required).

9.5 Conclusion

The current thesis offers validation of the efficacy of relational skills training interventions in producing significant improvements in both intellectual and academic performance across a range of age and ability levels. In addition, the results of the current collection of investigations highlight a number of important variables moderating this effect, as well as identifying several avenues to explore in an effort to further improve the efficacy of such protocols. Furthermore, by analysing the relationship between the wider range of relational frames and intellectual performance, the current thesis offers considerable value in elucidating the nature of this association. Due to the insights gained from such an analysis, it is hoped that such knowledge may be invoked to increase the accuracy and utility of relational skills assessments in serving as functional alternatives to traditional IQ tests. In conclusion, it is proposed that the current thesis strongly contributes to a research stream that now shows genuine promise in not only fostering potentially life-changing improvements in intellectual performance, but revolutionising how we understand intelligence itself.

Reference List

Abramson, C. (2013). Problems of Teaching the Behaviorist Perspective in the Cognitive Revolution. *Behavioral Sciences*, *3*(1), 55-71. https://doi.org/10.3390/bs3010055

Aciego, R., García, L., & Betancort, M. (2011). Los beneficios de la práctica del ajedrez en el enriquecimiento intelectual y socioafectivo en escolares. *Universidad de La Laguna.* …. Retrieved from http://es.chessbase.com/portals/0/files/images/2012/Leontxo/TBAH/El trastorno por déficit de atención con hiperactividad.pdf

Ackerman, P. L., Beier, M. E., & Boyle, M. O. (2005). Working memory and intelligence: The same or different constructs? *Psychological Bulletin*, *131*(1), 30–60. https://doi.org/10.1037/0033-2909.131.1.30

Ahmann, J. S. (1985). Otis Lennon School Ability Test. *Measurement & Evaluation in Counseling & Development*, *17*(4), 226-229.

Ali, A., Hassiotis, A., Dein, S., McManus, S., Rai, D., Cooper, C., … Meltzer, H. (2012). The relationship between happiness and intelligent quotient: the contribution of socio-economic and clinical factors. *Psychological Medicine*, *43*(06), 1303–1312. https://doi.org/10.1017/s0033291712002139

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

Amd, M., & Roche, B. (2018). Assessing the Effects of a Relational Training Intervention on Fluid Intelligence Among a Sample of Socially Disadvantaged Children in Bangladesh. *Psychological Record*, *68*(2), 141–149. https://doi.org/10.1007/s40732-018-0273-4

Andrews, G., & Halford, G. S. (1998). Children’s ability to make transitive inferences: The importance of premise integration and structural complexity. *Cognitive Development*, *13*(4), 479–513. https://doi.org/10.1016/S0885-2014(98)90004-1

Arréat, L. (2012). *Alfred Binet*. *Monist* (Vol. 22). Chicago, Illinois: Chicago University Press. https://doi.org/10.5840/monist191222115

Au, J., Sheehan, E., Tsai, N., Duncan, G. J., Buschkuehl, M., & Jaeggi, S. M. (2015). Improving fluid intelligence with training on working memory: a meta-analysis. *Psychonomic Bulletin and Review*, *22*(2), 366–377. https://doi.org/10.3758/s13423-014-0699-x

B., H., & Binet, A. (2006). *Les idées modernes sur les enfantsLes idees modernes sur les enfants*. *Population (French Edition)* (Vol. 29). Paris, France: Flammarion. https://doi.org/10.2307/1530945

Backstrand, B., Svartengren, M., Nilsson, M., Johnsson, T., Cooper-Kuhn, C. M., Aberg, N. D., … Kuhn, H. G. (2009). Cardiovascular fitness is associated with cognition in young adulthood. *Proceedings of the National Academy of Sciences*, *106*(49), 20906–20911. https://doi.org/10.1073/pnas.0905307106

Ball, K., Edwards, J. D., Ross, L. A., & McGwin, G. (2010). Cognitive training decreases motor vehicle collision involvement of older drivers. *Journal of the American Geriatrics Society*, *58*(11), 2107–2113. https://doi.org/10.1111/j.1532-5415.2010.03138.x

Ballesteros, S., Mayas, J., Prieto, A., Barrera, P. T., González, C. P., De León, L. P., … 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., … 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

Barkham, M. (2011). *Single Case Experimental Design : 1*. *Psychological Services*. New York, New York, New York: Pergamon.

Barnes-Holmes, D., & Barnes-Holmes, Y. (2000). Explaining complex behavior: Two perspectives on the concept of generalized operant classes. *Psychological Record*, *50*(2), 251-165. https://doi.org/10.1007/BF03395355

Barnes-Holmes, D., Barnes-Holmes, Y., & Cullinan, V. (2000). Relational frame theory and Skinner's verbal behavior: A possible synthesis. *Behavior Analyst*, *23*(1), 69–84. https://doi.org/10.1007/BF03392000

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the implicit relational assessment procedure (IRAP) and the relational elaboration and coherence (REC) model. *Psychological Record*, *60*(3), 527–542. https://doi.org/10.1007/BF03395726

Barnes-Holmes, D., Finn, M., McEnteggart, C., & Barnes-Holmes, Y. (2018). Derived Stimulus Relations and Their Role in a Behavior-Analytic Account of Human Language and Cognition. *Perspectives on Behavior Science*, *41*(1), 155–173. https://doi.org/10.1007/s40614-017-0124-7

Barnes-Holmes, D., Ninness, S. K., Ninness, C., Payne, R., Ward, T. A., Elliott, M. P., … McCuller, G. (2007). Transformations of Mathematical and Stimulus Functions. *Journal of Applied Behavior Analysis*, *39*(3), 299–321. https://doi.org/10.1901/jaba.2006.139-05

Barnes-holmes, D., Smeets, Y. B. P. M., Cullinan, V., & Leader, G. (2004). Relational Frame Theory and Stimulus Equivalence : Conceptual and Procedural Issues. *International Journal of Psychology and Psychological Therapy*, *4*, 181–214. Retrieved from http://redalyc.uaemex.mx/redalyc/html/560/56040202/56040202.html

Barnes-holmes, Y., Barnes-holmes, D., Healy, O., Lyddy, F., Cullinan, V., & Hayes, S. C. (1992). Psychological development. In *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 157–158).

Barnes-Holmes, Y., Barnes-Holmes, D., Roche, B., & Smeets, P. M. (2001). 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. *Establecimiento En Ninos Pequenos de Comportamiento Relacional de Acuerdo a M√°s-Qu√© y Menos-Qu√© Como Conducta Operante Generalizada*, *4*(3), 531–558.

Barnes, D., McCullagh, P. D., & Keenan, M. (2017). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior*, *8*(1), 19–30. https://doi.org/10.1007/bf03392844

Basak, C., Boot, W. R., Voss, M. W., & Kramer, A. F. (2008). Can Training in a Real-Time Strategy Video Game Attenuate Cognitive Decline in Older Adults? *Psychology and Aging*, *23*(4), 765–777. https://doi.org/10.1037/a0013494

Basso, M. R., Bornstein, R. A., & Lang, J. M. (2003). Practice Effects on Commonly Used Measures of Executive Function Across Twelve Months. *The Clinical Neuropsychologist*, *13*(3), 283–292. https://doi.org/10.1076/clin.13.3.283.1743

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

Bell, R. C., & Bell, R. C. (1981). *The Wechsler Intelligence Scale For Children*. New York, New York: Psychological Corporation. https://doi.org/10.1177/014662168100500117

Bennett, S. L., Gaudiani, J. L., Brinton, J. T., & Mehler, P. S. (2015). Motivated to Survive: High Cooperativeness in Severe Anorexia Nervosa. In *Eating Disorders* (Vol. 23, pp. 430–438). https://doi.org/10.1080/10640266.2015.1027118

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

Beran, M. J., & Washburn, D. A. (2006). Chimpanzee responding during matching to sample: control by exclusion. *Journal of the Experimental Analysis of Behavior*, *78*(3), 497–508. https://doi.org/10.1901/jeab.2002.78-497

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

Binder, C. (1996). Behavioral fluency: Evolution of a new paradigm. *The Behavior Analyst*, *19*(2), 163–197. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/22478257

Binet, A. (n.d.). Nos commissions de travail. *Bulletin de La Société Libre Pour l'Etude Psychologique de l'Enfant*, *4*(14), 337–346.

Binet, A. (2007). Sommaire des travaux en cours à la Société de psychologie de l'enfant. *L'année Psychologique*, *10*(1), 116–130. https://doi.org/10.3406/psy.1903.3542

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. (2006). New methods for the diagnosis of the intellectual level of subnormals. (L'Année Psych., 1905, pp. 191-244). In *The development of intelligence in children (The Binet-Simon Scale).* (pp. 37–90). https://doi.org/10.1037/11069-002

Blackledge, J. T. (2014). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today*, *3*(4), 421–433. https://doi.org/10.1037/h0099997

Block, N. (2006). Psychologism and Behaviorism. *The Philosophical Review, 90*(1), 5. https://doi.org/10.2307/2184371

Boag, S. (2011). Explanation in personality psychology: "Verbal magic" and the five-factor model. *Philosophical Psychology, 24*(2), 223–243. https://doi.org/10.1080/09515089.2010.548319

Bodily, K. D., Katz, J. S., & Wright, A. A. (2008). Matching-to-Sample Abstract-Concept Learning by Pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *34*(1), 178–184. https://doi.org/10.1037/0097-7403.34.1.178

Boring, E. G. (2007). Intelligence as the Tests Test It. *Studies in Individual Differences: The Search for Intelligence.*, 210–214. https://doi.org/10.1037/11491-017

Bourneville, D. M. (1895). *Assistance, traitement et éducation des enfants idiots et dégénérés.* Paris, France: Alcan.

Bourneville, D. M. (1898). *Création de classes spéciales pour les enfants arriérés*. Paris, France: Alcan.

Bourneville, D. M. (1899). *Lettre à M. Charles Dupuy, président du Conseil, Ministre de l'Intérieur, sur la création de classes spéciales pour les enfants arriérés.* Paris, France: Alcan.

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

Brinch, C. N., & Galloway, T. A. (2011). Schooling in adolescence raises IQ scores. *Proceedings of the National Academy of Sciences, 109*(2), 425–430. https://doi.org/10.1073/pnas.1106077109

Brown, K. J., Sinatra, G. M., & Wagstaff, J. M. (2005). 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

Brown, M., Askew, M., Baker, D., Denvir, H., & Millett, A. (1998). Is the national numeracy strategy research-based? *British Journal of Educational Studies, 46*(4), 362–385. https://doi.org/10.1111/1467-8527.00090

Bryant, D. J., & Tversky, B. (1999). Mental Representations of Perspective and Spatial Relations from Diagrams and Models. *Journal of Experimental Psychology: Learning Memory and Cognition*, *25*(1), 137–156. https://doi.org/10.1037/0278-

7393.25.1.137

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: A literature review. *Swiss Medical Weekly*, *140*(19-20), 266-272. https://doi.org/smw-12852

Butler, A. C., Chapman, J. E., Forman, E. M., & Beck, A. T. (2006). The empirical status of cognitive-behavioral therapy: A review of meta-analyses. *Clinical Psychology Review*, *26*(1), 17-31. https://doi.org/10.1016/j.cpr.2005.07.003

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

Campbell, J. A., & Willis, J. (1978). Modifying Components of "Creative Behavior" in the Natural Environment. *Behavior Modification*, *2*(4), 549–564. https://doi.org/10.1177/014544557824007

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven progressive matrices test. *Psychological Review*, *97*(3), 404–431. https://doi.org/10.1037/0033-295X.97.3.404

Carpenter, T. P., Franke, M. L., & Levi, L. (2003). *Thinking Mathematically: Integrating Arithmetic and Algebra in Elementary School*. Toronto, Canada: Pearson Education Canada.

Carpenter, T. P., Levi, L., Franke, M. L., & Zeringue, J. K. (2005). Algebra in Elementary School: Developing Relational Thinking. *ZDM - International Journal on Mathematics Education*, *37*(1), 53–59. https://doi.org/10.1007/BF02655897

Carpentier, F., Smeets, P. M., & Barnes-Holmes, D. (2003). Equivalence-equivalence as a model of analogy: Further analyses. *The Psychological Record*, *53*(3), 349–371.

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

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

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

Cassidy, S., Roche, B., & O'Hora, D. (2017). Relational Frame Theory and Human Intelligence. *European Journal of Behavior Analysis*, *11*(1), 37–51. https://doi.org/10.1080/15021149.2010.11434333

Catania, C. A. (1992). *Learning*. Englewood Cliffs, New Jersey: Prentice Hall.

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

Chand, I., & Runco, M. A. (1993). Problem finding skills as components in the creative process. *Personality and Individual Differences*, *14*(1), 155–162. https://doi.org/10.1016/0191-8869(93)90185-6

Charen, S. (2006). *Psychological Appraisal of Children with Cerebral Defects*. *The Journal of Nervous and Mental Disease* (Vol. 131). Cambridge, Massachsetts: Harvard University Press. https://doi.org/10.1097/00005053-196011000-00020

Chooi, W. T., & Thompson, L. A. (2012). Working memory training does not improve intelligence in healthy young adults. *Intelligence*, *40*(6), 531–542. https://doi.org/10.1016/j.intell.2012.07.004

Clarke, G. (2003). *Consumer behaviour analysis: Critical perspectives on business and management*. *Journal of Consumer Behaviour* (Vol. 3). Retrieved from http://search.ebscohost.com/login.aspx?direct=true&db=buh&AN=11661616&site=ehost-live

Cohen, H., Amerine-Dickens, M., & Smith, T. (2006). Early Intensive Behavioral Treatment. *Journal of Developmental & Behavioral Pediatrics*, *27*(Supplement 2), S145–S155. https://doi.org/10.1097/00004703-200604002-00013

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

Colom, R., Abad, F. J., Quiroga, M. Á., Shih, P. C., & Flores-Mendoza, C. (2008). Working memory and intelligence are highly related constructs, but why? *Intelligence*, *36*(6), 584–606. https://doi.org/10.1016/j.intell.2008.01.002

Compas, B. E., Haaga, D. A. F., Williams, D. A., Keefe, F. J., & Leitenberg, H. (2005). Sampling of empirically supported psychological treatments from health psychology: Smoking, chronic pain, cancer, and bulimia nervosa. *Journal of Consulting and Clinical Psychology, 66*(1), 89-112. https://doi.org/10.1037/0022-006x.66.1.89

Cooper, J. O. (1982). Applied Behavior Analysis in Education. *Theory Into Practice, 21*(2), 114–118. https://doi.org/10.1080/00405848209542992

Cowley, B. J., Green, G., & Braunling-McMorrow, D. (2006). Using stimulus equivalence procedures to teach name-face matching to adults with brain injuries. *Journal of Applied Behavior Analysis*, *25*(2), 461–475. https://doi.org/10.1901/jaba.1992.25-461

Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, *66*, 93–99. https://doi.org/10.1016/j.jesp.2015.10.002

Cuijpers, P., Hollon, S. D., Van Straten, A., Bockting, C., Berking, M., & Andersson, G. (2013). Does cognitive behaviour therapy have an enduring effect that is superior to keeping patients on continuation pharmacotherapy? A meta-analysis. *BMJ Open*, *3*(4). https://doi.org/10.1136/bmjopen-2012-002542

Cullinan, V. A., Barnes, D., Hampson, P. J., & Lyddy, F. (2017). A Transfer of Explicitly and Nonexplicitly Trained Sequence Responses Through Equivalence Relations: An Experimental Demonstration and Connectionist Model. *The Psychological Record*, *44*(4), 559–585. https://doi.org/10.1007/bf03395144

Cumming, W. W., & Berryman, R. (2006). Some data on matching behavior in the pigeon1. *Journal of the Experimental Analysis of Behavior*, *4*(3), 281–284. https://doi.org/10.1901/jeab.1961.4-281

D'Amato, M. R., & Colombo, M. (1985). Auditory matching-to-sample in monkeys (Cebus apella). *Animal Learning & Behavior*, *13*(4), 375–382. https://doi.org/10.3758/BF03208013

D'Amato, M. R., & O'Neill, W. (2006). Effect of delay-interval illumination on matching behavior in the capuchin monkey1. *Journal of the Experimental Analysis of Behavior*, *15*(3), 327–333. https://doi.org/10.1901/jeab.1971.15-327

D'Amato, M. R., & Worsham, R. W. (1972). Delayed matching in the capuchin monkey with brief sample durations. *Learning and Motivation*, *3*(3), 304–312. https://doi.org/10.1016/0023-9690(72)90026-4

D'Amato, M. R., & Worsham, R. W. (1974). Retrieval cues and short-term memory in capuchin monkeys. *Journal of Comparative and Physiological Psychology*, *86*(2), 274–282. https://doi.org/10.1037/h0035950

Dahlin, E., Nyberg, L., Bäckman, L., & Neely, A. S. (2008). Plasticity of Executive Functioning in Young and Older Adults: Immediate Training Gains, Transfer, and Long-Term Maintenance. *Psychology and Aging*, *23*(4), 720–730. https://doi.org/10.1037/a0014296

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

Davachi, L., & Goldman-Rakic, P. S. (2017). Primate Rhinal Cortex Participates in Both Visual Recognition and Working Memory Tasks: Functional Mapping With 2-DG. *Journal of Neurophysiology*, *85*(6), 2590–2601. https://doi.org/10.1152/jn.2001.85.6.2590

de Houwer, J., Barnes-Holmes, D., & Moors, A. (2013). What is learning? On the nature and merits of a functional definition of learning. *Psychonomic Bulletin and Review*, *20*(4), 631–642. https://doi.org/10.3758/s13423-013-0386-3

De Rose, J., & De Souza, D. (2006). Teaching Reading And Spelling: Exclusion And Stimulus Equivalence. *Journal of Applied Behavior Analysis*, *29*(4), 451–469. https://doi.org/10.1901/jaba.1996.29-451

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

Decety, J., & Lamm, C. (2006). Human empathy through the lens of social neuroscience. *TheScientificWorldJournal*, *6*, 1146-1163. https://doi.org/10.1100/tsw.2006.221

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

DiClemente, D. F., & Hantula, D. A. (2003). Applied behavioral economics and consumer choice. *Journal of Economic Psychology*, *24*(5), 589-602. https://doi.org/10.1016/S0167-4870(03)00003-5

Dixon, M., Ninness, S. K., Barnes-Holmes, D., Rumph, R., McGinty, J., Rehfeldt, R. A., … Holland, J. (2009). Constructing and Deriving Reciprocal Trigonometric Relations: a Functional Analytic Approach. *Journal of Applied Behavior Analysis*, *42*(2), 191–208. https://doi.org/10.1901/jaba.2009.42-191

Dixon, M. R., Belisle, J., Stanley, C. R., & Rowsey, K. (2018). Student outcomes after 1 year of front line staff implementation of the PEAK curriculum. *Behavioral Interventions*, *33*(2), 185–195. https://doi.org/10.1002/bin.1516

Dixon, M. R., Whiting, S. W., Rowsey, K., & Belisly, J. (2014). Assessing the relationship between intelligence and the PEAK relational training system. *Research in Autism Spectrum Disorders*, *8*(9), 1208–1213. https://doi.org/10.1016/j.rasd.2014.05.005

Dodrill, C. B., & Troupin, A. S. (1975). Effects of repeated administrations of a comprehensive neuropsychological battery among chronic epileptics. *Journal of Nervous and Mental Disease*, *161*(3), 185–190. https://doi.org/10.1097/00005053-197509000-00006

Doll, E. A. (2006). *The Measurement of Intelligence. An Explanation of and a Complete Guide for the Use of the Stanford Revision and Extension of the Binet-Simon Intelligence Scale. Journal of Educational Psychology* (Vol. 8). Boston: Houghton Mifflin. https://doi.org/10.1037/h0068271

Dugdale, N., & Lowe, C. F. (2006). 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

DUNCAN, C. P. (2010). Attempts To Influence Performance on an Insight Problem. *Psychological Reports*, *9*(5), 35. https://doi.org/10.2466/pr0.9.5.35-42

Dunn, L. M., Dunn, L. M., Bulheller, S., & Häcker, H. (1965). *Peabody Picture Vocabulary Test PPVT. Peabody picture vocabulary test*. Circle Pines, MN: American Guidance Service.

Dunnett, S. B., Martel, F. L., & Iversen, S. D. (1990). Proactive interference effects on short-term memory in rats: II. Effects in young and aged rats. *Behavioral Neuroscience, 104*(5), 666-670.

Dutton, E., & Lynn, R. (2014). A negative flynn effect in Finland, 1997-2009. *Intelligence*, *41*(6), 817–820. https://doi.org/10.1016/j.intell.2013.05.008

Dutton, E., & Lynn, R. (2015). A negative Flynn Effect in France, 1999 to 2008-9. *Intelligence*, *51*, 67–70. https://doi.org/10.1016/j.intell.2015.05.005

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

https://doi.org/10.1016/j.intell.2016.10.002

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. (2006). 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 Behavior*, *64*(2), 163–184. https://doi.org/10.1901/jeab.1995.64-163

Dymond, S., & Roche, B. (2013). *Advances in Relational Frame Theory*. Oakland, CA: Context Press.

Education Research Centre. (2006). *Drumcondra Primary Mathematics Test - Revised.* Dublin, Ireland: Educational Research Centre, St. Patrick's Collge.

Educational Research Centre. (2007). Drumcondra primary reading test—revised. Levels 3–6 administra- tion manual and technical manual. Dublin, Ireland: Educational Research Centre, St. Patrick's Collge.

Educational Research Centre. (2016). *Drumcondra Reasoning Test*. Dublin, Ireland: Educational Research Centre, St. Patrick's Collge.

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

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2003). Intensive Behavioral Treatment at School for 4- to 7-Year-Old Children with Autism. *Behavior Modification*, *26*(1), 49–68. https://doi.org/10.1177/0145445502026001004

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2007). Outcome for children with autism who began intensive behavioral treatment between ages 4 and 7: A comparison controlled study. *Behavior Modification*, *31*(3), 264-278. https://doi.org/10.1177/0145445506291396

Elliott, C. D. (2007). *Differential Ability Scales - Second Edition (DAS-II): Introductory and technical handbook*. San Antonio, Texas: Psychological Corporation.

Engle, R. W., Laughlin, J. E., Tuholski, S. W., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*(3), 309–331. https://doi.org/10.1037/0096-3445.128.3.309

Epstein, R. (1987). The spontaneous interconnection of four repertoires of behavior in a pigeon (Columba livia). *Journal of Comparative Psychology (Washington, D.C. : 1983)*, *101*(2), 197–201. https://doi.org/10.1037/0735-7036.101.2.197

Epstein, R. (1990). Generativity theory and creativity Runco. In M. A. Runco & R. S. Albert (Eds.), *Sage focus editions*, Vol. 115. Theories of creativity (pp. 116-140). Thousand Oaks, CA, US: Sage Publications, Inc.

Epstein, R., Kirshnit, C. E., Lanza, R. P., & Rubin, L. C. (1984). "Insight" in the pigeon: Antecedents and determinants of an intelligent performance. *Nature*, *308*(5954), 61–62. https://doi.org/10.1038/308061a0

Etkin, M., & D'Amato, M. R. (1969). Delayed matching-to-sample and short-term memory in the capuchin monkey. *Journal of Comparative and Physiological Psychology, 69*(3), 544–549. https://doi.org/10.1037/h0028209

Fancher, R. E. (1985). *The Intelligence Men: Makers of the IQ controversy*. New York, New York, New York: Norton.

Fantino, E. (1998). Judgment and decision making: Behavioral approaches. *The Behavior Analyst*.

Fantino, E. (2004). Behavior-analytic approaches to decision making. In *Behavioural Processes* (Vol. 66, pp. 279–288). https://doi.org/10.1016/j.beproc.2004.03.009

Fantino, E., & Stolarz-Fantino, S. (2005). Decision-making: Context matters. In *Behavioural Processes* (Vol. 69, pp. 165–171). https://doi.org/10.1016/j.beproc.2005.02.002

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

Farrington-Flint, L., & Wood, C. (2007). The role of lexical analogies in beginning reading: Insights from children's self-reports. *Journal of Educational Psychology*, *99*(2), 326–338. https://doi.org/10.1037/0022-0663.99.2.326

Finch, G. (1942). Delayed matching-from-sample and non-spatial delayed response in champanzees. *Journal of Comparative Psychology*, *34*(3), 315–319. https://doi.org/10.1037/h0062821

Flemming, T. M., & Kennedy, E. H. (2011). Chimpanzee (Pan troglodytes) Relational Matching: Playing by Their Own (Analogical) Rules. *Journal of Comparative Psychology*, *125*(2), 207–215. https://doi.org/10.1037/a0022098

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*. New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511605253

Foxall, G. R. (1994). Behavior analysis and consumer psychology. *Journal of Economic Psychology, 15*(1), 5–91. https://doi.org/10.1016/0167-4870(94)90032-9

Foxall, G. R. (2001). Foundations of consumer behaviour analysis. *Marketing Theory, 1*(2), 165-199. https://doi.org/10.1177/147059310100100202

Foxall, G. R. (2015). Operant Behavioral Economics. *Managerial and Decision Economics*, *37*, 4–5.

Foxall, G. R. (2017). Behavioral Economics in Consumer Behavior Analysis. *Behavior Analyst*, *40*(2), 309–313. https://doi.org/10.1007/s40614-017-0127-4

Furrebøe, E. F., & Sandaker, I. (2017). Contributions of Behavior Analysis to Behavioral Economics. *Behavior Analyst*, *40*(2), 315–327. https://doi.org/10.1007/s40614-017-0110-0

Gale. H. Roid. (2003). *Stanford Binet Intelligence Scale, Fifth edition*. Itasca, IL:Riverside Publishing.

Galinsky, A. D., & Mussweiler, T. (2001). First offers as anchors: The role of perspective-taking and negotiator focus. *Journal of Personality and Social Psychology*, *81*(4), 657-669. https://doi.org/10.1101/gad.535209

Garofoli, D., Natale, F., Truppa, V., Castorina, G., Piano Mortari, E., & Visalberghi, E. (2010). Identity concept learning in matching-to-sample tasks by tufted capuchin monkeys (Cebus apella). *Animal Cognition*, *13*(6), 835–848. https://doi.org/10.1007/s10071-010-0332-y

Gateaux-Mennecier, J. (1989). Bourneville et l'enfance aliennee. Paris, France: Centurion.

Gateaux-Mennecier, J. (2002). Bourneville, humaniste et réformateur. *Rev Prat*, *52*, 1517–1521.

Geake, J. G., & Hansen, P. C. (2005). Neural correlates of intelligence as revealed by fMRI of fluid analogies. *NeuroImage*, *26*(2), 555–564. https://doi.org/10.1016/j.neuroimage.2005.01.035

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

Gentner, D., Holyoak, K. J., & Kokinov, B. N. (2001). *The Analogical Mind: Perpectives from Cognitive Science*.Cambrideg, Massachusetts: MIT Press. https://doi.org/10.1017/CBO9781107415324.004

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.

Gentner, D., & Smith, L. A. (2013). Analogical learning and reasoning. In *The Oxford Handbook of Cognitive Psychology*. https://doi.org/10.1093/oxfordhb/9780195376746.001.0001

Gisquet-Verrier, P., & Delatour, B. (2006). The role of the rat prelimbic/infralimbic cortex in working memory: Not involved in the short-term maintenance but in monitoring and processing functions. *Neuroscience*, *141*(2), 585–596. https://doi.org/10.1016/j.neuroscience.2006.04.009

Glanz, K., Rimer, B. K., & Viswanath, K. (2008). *Health Behaviour and Health Education*. *Health Education.*, *63*, 105-107. https://doi.org/10.1016/S0033-3506(49)81524-1

Gómez, S., Barnes-Holmes, D., Barnes-Holmes, Y., López, F., & Martín, C. B. (2017).

Exemplar Training and A Derived Transformation of Functions in Accordance with Symmetry and Equivalence. *The Psychological Record*, *57*(2), 273–293. https://doi.org/10.1007/bf03395577

Goodwin, A., Kelly, E., Kelso, G., Ninness, C., Rumph, R., McCuller, G., … McGinty, J. (2017). Training and Deriving Precalculus Relations: A Small-Group, Web-Interactive Approach. *The Psychological Record*, *62*(2), 225–242. https://doi.org/10.1007/bf03395799

Goodwin, L. D., & Leech, N. L. (2006). Understanding correlation: Factors that affect the size of r. *Journal of Experimental Education*, *74*(3), 249–266. 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 and Psychological Therapy*, *10*(1), 1–17.

Goswami, U. (1986). Children's use of analogy in learning to read: A developmental study. *Journal of Experimental Child Psychology*, *42*(1), 73–83. https://doi.org/10.1016/0022-0965(86)90016-0

Goswami, U. (2011). Children's use of analogy in learning to spell. *British Journal of Developmental Psychology*, *6*(1), 21–33. https://doi.org/10.1111/j.2044-835x.1988.tb01077.x

Goto, K., & Watanabe, S. (2009). Visual working memory of jungle crows (Corvus macrorhynchos) in operant delayed matching-to-sample. *Japanese Psychological Research*, *51*(3), 122–131. https://doi.org/10.1111/j.1468-5884.2009.00400.x

Goto, K., & Watanabe, S. (2012). Large-billed crows (Corvus macrorhynchos) have retrospective but not prospective metamemory. *Animal Cognition, 15*(1), 27-35. https://doi.org/10.1007/s10071-011-0428-z

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

Gottfredson, L. S. (2003). g, Jobs and Life. In *The Scientific Study of General Intelligence: Tribute to Arthur Jensen* (pp. 293–342). Oxford, UK: Pergamon. https://doi.org/10.1016/B978-008043793-4/50053-2

Grant, D. S. (1976). Effect of sample presentation time on long-delay matching in the pigeon. *Learning and Motivation*, *7*(4), 580–590. https://doi.org/10.1016/0023-9690(76)90008-4

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

Griffee, K., & Dougher, M. J. (2006). Contextual control of stimulus generalization and stimulus equivalence in hierarchical categorization. *Journal of the Experimental Analysis of Behavior, 78*(3), 433–447. https://doi.org/10.1901/jeab.2002.78-433

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

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149–161. https://doi.org/10.1007/BF02289162

Guttman, L. L. (2019). *The Basis for Scalogram Analysis*. *Scaling*. Indianapolis Ind.: Bobbs-Merrill College Division. https://doi.org/10.4324/9781315128948-16

Hahn, U. (2011). The problem of circularity in evidence, argument, and explanation. *Perspectives on Psychological Science, 6(2), 172-182*. https://doi.org/10.1177/1745691611400240

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

Hale, S., & Jansen, J. (1994). Global Processing-Time Coefficients Characterize Individual and Group Differences in Cognitive Speed. *Psychological Science*, *5*(6), 384–389. https://doi.org/10.1111/j.1467-9280.1994.tb00290.x

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

Hambrick, D. Z., Redick, T. S., Engle, R. W., Harrison, T. L., Shipstead, Z., & Hicks, K. L. (2013). Working Memory Training May Increase Working Memory Capacity but Not Fluid Intelligence. *Psychological Science*, *24*(12), 2409–2419. https://doi.org/10.1177/0956797613492984

Hammill, D. D., & Pearson, N. (2017). Comprehensive test of nonverbal intelligence: Second edition. *Jounral of Psychoeducational Assessment, 23*(2), 161-172. https://doi.org/10.1007/978-3-319-50604-3_10

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

Harrison, C., Ninness, S. K., Ford, A. M., Rumph, R., McCuller, G., & Ninness, C. (2006). A Functional Analytic Approach To Computer-Interactive Mathematics. *Journal of Applied Behavior Analysis*, *38*(1), 1–22. https://doi.org/10.1901/jaba.2005.2-04

Hashiya, K., & Kojima, S. (2001). Acquisition of auditory-visual intermodal matching-to-sample by a chimpanzee (Pan troglodytes): Comparison with visual-visual intramodal matching. *Animal Cognition*, *4*(3–4), 231–239. https://doi.org/10.1007/s10071-001-0118-3

Haun, D. B. M., & Call, J. (2009). Great apes' capacities to recognize relational similarity. *Cognition*, *110*(2), 147–159. https://doi.org/10.1016/j.cognition.2008.10.012

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., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1*(1-2), 1-16. https://doi.org/10.1016/j.jcbs.2012.09.004

Hayes, S. C., & Brownstein, A. J. (2016). Mentalism, behavior-behavior relations, and a behavior-analytic view of the purposes of science. In S. C. Hayes (Ed.) *The Act in Context: The Canonical Papers of Steven C. Hayes*. New York: Routledge. https://doi.org/10.4324/9781315745138

Hayes, S. C., Gifford, E. V., Townsend, R. C., & Barnes-Holmes, D. (2005). Thinking, Problem-solving, and Pragmatic Verbal Analysis. In *Relational Frame Theory* (pp. 87–101). https://doi.org/10.1007/0-306-47638-x_5

Healy, O., Barnes-Holmes, D., & Smeets, P. M. (2006). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior*, *74*(2), 207–227. https://doi.org/10.1901/jeab.2000.74-207

Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, *32*(2), 175–191. https://doi.org/10.1016/j.intell.2003.12.001

Heim, A. W., Watts, K. P., & Simmonds, V. (1968). *AH4 Question Book*. UK: NFER-Nelson Publishing Company Ltd.

Heitz, R. P., Unsworth, N., & Engle, R. W. (2005). Working memory capacity, attention control, and fluid intelligence. In *Handbook of Understanding and Measuring Intelligence* (pp. 61–78). Thousand Oaks, CA, US: Sage Publications, Inc. https://doi.org/10.4135/9781452233529.n5

Herman, L. M., & Gordon, J. A. (2006). Auditory delayed matching in the bottlenose dolphin1. *Journal of the Experimental Analysis of Behavior, 21*(1), 19–26. https://doi.org/10.1901/jeab.1974.21-19

Herman, L. M., Hovancik, J. R., Gory, J. D., & Bradshaw, G. L. (1989). Generalization of Visual Matching by a Bottlenosed Dolphin (Tursiops truncatus): Evidence for Invariance of Cognitive Performance with Visual and Auditory Materials. *Journal of Experimental Psychology: Animal Behavior Processes*, *15*(2), 124–136. https://doi.org/10.1037/0097-7403.15.2.124

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

Hobbs, S. A., Moguin, L. E., Tyroler, M., & Lahey, B. B. (1980). Cognitive behavior therapy with children: Has clinical utility been demonstrated? *Psychological Bulletin*, *87*(1), 147–165. https://doi.org/10.1037/0033-2909.87.1.147

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science*, *12*(4). https://doi.org/10.1111/j.1467-7687.2009.00848.x

Holmes, J., Gathercole, S. E., Place, M., Dunning, D. L., Hilton, K. A., & Elliott, J. G. (2010). Working memory deficits can be overcome: Impacts of training and medication on working memory in children with ADHD. *Applied Cognitive Psychology*, *24*(6), 827–836. https://doi.org/10.1002/acp.1589

Holth, P. (2014). Psychological behaviorism: A path to the grand reunification of psychology and behavior analysis? *The Behavior Analyst Today*, *4*(3), 306–317. https://doi.org/10.1037/h0100019

Holth, P. (2017). Multiple Exemplar Training: Some Strengths and Limitations. *Behavior Analyst*, *40*(1), 225–241. https://doi.org/10.1007/s40614-017-0083-z

Hopkins, W. D., & Washburn, D. A. (2002). Matching visual stimuli on the basis of global and local features by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatta). *Animal Cognition*, *5*(1), 27–31. https://doi.org/10.1007/s10071-001-0121-8

Hossiep, R., Turck, D., & Hasella, M. (1999). *Bochumer Matrizentest: BOMAT Advanced-Short Version*. Gottingen: Hogrefe.

Howe, M. J. A. (1990). Does intelligence exist? *The Psychologist*, *Novemeber*, 490–493.

Howlin, P., Goode, S., Hutton, J., & Rutter, M. (2004). Adult outcome for children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *45*(2), 212–229. https://doi.org/10.1111/j.1469-7610.2004.00215.x

Hribar, A., Haun, D., & Call, J. (2011). Great apes' strategies to map spatial relations. *Animal Cognition*, *14*(4), 511–523. https://doi.org/10.1007/s10071-011-0385-6

Hudson, J. A., & Mayhew, E. M. Y. (2008). The development of memory for recurring events. In *The Development of Memory in Infancy and Childhood: Second Edition* (pp. 69–91). https://doi.org/10.4324/9780203934654

Illinois State Board of Education. (2013). Illinois Early Learning and Development Standards: For Preschool 3 years old to Kindergarten Enrollment Age. Retrieved from http://www.isbe.state.il.us/earlychi/pdf/early_learning_standards.pdf.

Irwin, K. C., & Britt, M. S. (2005). The algebraic nature of students' numerical manipulation in the New Zealand Numeracy Project. *Educational Studies in Mathematics*, *58*(2), 169–188. https://doi.org/10.1007/s10649-005-2755-y

Irwing, P., & Lynn, R. (2006). *The Relation Between Childhood IQ and Income in Middle Age* (Vol. 31). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.683.1926&rep=rep1&type=pdf

Isaacowitz, D. M., & Smith, J. (2003). Positive and negative affect in very old age. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *58*(3). https://doi.org/10.1093/geronb/58.3.P143

Jacobson, J. W., Mulick, J. A., & Green, G. (1998). Cost-benefit estimates for early intensive behavioral intervention for young children with autism - General model and single state case. *Behavioral Interventions*, *13*(4), 201–226. https://doi.org/10.1002/(SICI)1099-078X(199811)13:4<201::AID-BIN17>3.0.CO;2-R

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

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences*, *108*(25), 10081–10086. https://doi.org/10.1073/pnas.1103228108

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

Jahr, E., Brown, T., Lemaic, M., degli Espinosa, F., Kovshoff, H., Alsford, P., … Remington, B. (2007). Early Intensive Behavioral Intervention: Outcomes for Children With Autism and Their Parents After Two Years. *American Journal on Mental Retardation*, *112*(6), 418. https://doi.org/10.1352/0895-8017(2007)112[418:eibiof]2.0.co;2

Jencks, S. (1972). *Inequality*. London, UK: Penguin.

Jensen, A. R. (1980). Chronometric analysis of intelligence. *Journal of Social and Biological Systems, 3*(2), 103–122. https://doi.org/10.1016/0140-1750(80)90003-2

Jensen, A. R. (1998). *The g factor: The science of mental ability*. *Westport, CT: Praeger/Greenwood.* https://doi.org/10.1007/s13398-014-0173-7.2

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

Johnson, D. W. (1975). Cooperativeness and social perspective taking. *Journal of Personality and Social Psychology, 31*(2), 241–244. https://doi.org/10.1037/h0076285

Johnston, J. M., & Pennypacker, H. (1980). *Strategies and tactics of human behavior research*. Hillsdale, New Jersey, New Jersey: Erlbaum.

Jones, F. N., & Skinner, B. F. (2006). *The Behavior of Organisms: An Experimental Analysis*. *The American Journal of Psychology* (Vol. 52). Acton, Massachusetts, Massachusetts: Copley. https://doi.org/10.2307/1416495

Jonides, J., Yip, C.-Y., Franklin, M. S., Rattray, K., Moher, J., & Sledge Moore, K. (2008). The effects of musical training on verbal memory. *Psychology of Music, 36*(3), 353-365. https://doi.org/10.1177/0305735607086044

Joseph, B. D., & Hock, H. H. (2006). *Principles of Historical Linguistics*. *Language* (2nd ed., Vol. 65). Berlin, Germany: Mouton de Gruyter. https://doi.org/10.2307/414853

Juliano, J. M., Haddad, F. A., & Carroll, J. L. (1988). Three-year stability of WISC-R factor scores for black and white, female and male children classified as learning-disabled. *Journal of School Psychology*, *26*(4), 317–325. https://doi.org/10.1016/0022-4405(88)90040-4

Kanazawa, S. (2014). Intelligence and childlessness. *Social Science Research, 48,* 157–170. https://doi.org/10.1016/j.ssresearch.2014.06.003

Kane, M. J., Hambrick, D. Z., & Conway, A. R. A. (2005). Working memory capacity and fluid intelligence are strongly related constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 66-71. https://doi.org/10.1037/0033-2909.131.1.66

Kane, M. J., Harrison, T. L., Hicks, K. L., Engle, R. W., Redick, T. S., Shipstead, Z., … Fried, D. E. (2012). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General*, *142*(2), 359–379. https://doi.org/10.1037/a0029082

Kaufman, A. S., & Kaufman, N. L. (2004a). *Kaufman Brief Intelligence Test, Second Edition.* Bloomington, Minnesota: Pearson, Inc.

Kaufman, A. S., & Kaufman, N. L. (2004b). *Manual for Kaufman Brief Intelligence Test*. Circle Pines, Minnesota: American Guidance Service.

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

Kazdin, A. E. (1980). Acceptability of time out from reinforcement procedures for disruptive child behavior. *Behavior Therapy*, *11*(3), 329–344. https://doi.org/10.1016/S0005-7894(80)80050-5

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

Kerr, B., & Lohman, D. F. (2012). Cognitive Abilities Test. In Kerr, B. (Ed) *Encyclopedia of Giftedness, Creativity, and Talent*. Thousand Oaks, CT: SAGE Publications, Inchttps://doi.org/10.4135/9781412971959.n71

Kieran, C. (1981). Concepts associated with the equality symbol. *Educational Studies in Mathematics*, *12*(3), 317–326. https://doi.org/10.1007/BF00311062

Kilian, A., Von Fersen, L., & Güntürkün, O. (2005). Left hemispheric advantage for numerical abilities in the bottlenose dolphin. *Behavioural Processes*, *68*(2), 179–184. https://doi.org/10.1016/j.beproc.2004.11.003

Kızıltoprak, A., & Köse, N. Y. (2017). Relational thinking: The bridge between arithmetic and algebra. *International Electronic Journal of Elementary Education, 10*(1), 131–145. https://doi.org/10.26822/iejee.2017131893

Klingberg, T., Fernell, E., Olesen, P. J., Johnson, M., Gustafsson, P., Dahlström, K., … Westerberg, H. (2005). Computerized training of working memory in children with ADHD - A randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 177–186. https://doi.org/10.1097/00004583-200502000-00010

Koehler, O. (1956). The Ability of Birds to Count. *The World of Mathematics. Volume One, 9*, 489–496.

Kozhevnikov, M., Motes, M. A., Rasch, B., & Blajenkova, O. (2006). Perspective-taking vs. mental rotation transformations and how they predict spatial navigation performance. *Applied Cognitive Psychology*, *20*(3), 397–417. https://doi.org/10.1002/acp.1192

Kuśmierek, P., & Kowalska, D. M. (2002). Effect of sound source position on learning and performance of auditory delayed matching-to-sample task in dogs. *Acta Neurobiologiae Experimentalis*, *62*(4), 251–262.

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

Laidra, K., Pullmann, H., & Allik, J. (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. *Personality and Individual Differences*, *42*(3), 441–451. https://doi.org/10.1016/j.paid.2006.08.001

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*(1), 42-58. https://doi.org/10.1162/jocn.2007.19.1.42

Lampit, A., Hallock, H., & Valenzuela, M. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLoS Medicine*, *11*(11). https://doi.org/10.1371/journal.pmed.1001756

Lange, R. T. (2010). Performance IQ. In *Encyclopedia of Clinical Neuropsychology* (pp. 1907–1908). New York, New York: Springer Publishing Co. https://doi.org/10.1007/978-0-387-79948-3_1066

Lawlor-Savage, L., & Goghari, V. M. (2016). Dual N-back working memory training in healthy adults: A randomized comparison to processing speed training. *PLoS ONE*, *11*(4). https://doi.org/10.1371/journal.pone.0151817

Leader, G., & Barnes-Holmes, D. (2001). Establishing fraction-decimal equivalence using a respondent-type training procedure. *Psychological Record*, *51*(1), 151–165. https://doi.org/10.1007/BF03395391

Lezak, M. D., & Lezak, M. D. (2004). *Neuropsychological assessment*. New York: Oxford University Press

Lind, J., Enquist, M., & Ghirlanda, S. (2015). Animal memory: A review of delayed matching-to-sample data. *Behavioural Processes*, *117*, 52-58. https://doi.org/10.1016/j.beproc.2014.11.019

Locurto, C. (1991). Beyond IQ in preschool programs? *Intelligence*, *15*(3), 295–312. https://doi.org/10.1016/0160-2896(91)90038-F

Longevity, S. C. on. (2014). A Consensus on the Brain Training Industry from the Scientific Community. Retrieved from http://longevity3.stanford.edu/blog/2014/10/15/the-consensus-on-the-brain-training-industry-from-the-scientific-community/

Louis, T. A., Lavori, P. W., Bailar, J. C., & Polansky, M. (2010). Crossover and Self-Controlled Designs in Clinical Research. *New England Journal of Medicine*, *310*(1), 24–31. https://doi.org/10.1056/nejm198401053100106

Lovaas, O. I. (2005). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology*, *55*(1), 3–9. https://doi.org/10.1037/0022-006x.55.1.3

Luciano, C., Becerra, I. G., & Valverde, M. R. (2007). The Role of Multiple-Exemplar Training and Naming in Establishing Derived Equivalence in an Infant. *Journal of the Experimental Analysis of Behavior*, *87*(3), 349–365. https://doi.org/10.1901/jeab.2007.08-06

Lynch, D. C., & Cuvo, A. J. (2006). Stimulus equivalence instruction of fraction—decimal relations. *Journal of Applied Behavior Analysis*, *28*(2), 115–126. https://doi.org/10.1901/jaba.1995.28-115

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

Lynn, R. (2010a). Differential rates of secular increase of five major primary abilities. *Biodemography and Social Biology*, *37*(1–2), 137–141. https://doi.org/10.1080/19485565.1990.9988753

Lynn, R. (2010b). In Italy, north-south differences in IQ predict differences in income, education, infant mortality, stature, and literacy. *Intelligence, 38*(1), 93–100. https://doi.org/10.1016/j.intell.2009.07.004

Lynn, R., & Vanhanen, T. (2012). National IQs: A review of their educational, cognitive, economic, political, demographic, sociological, epidemiological, geographic and climatic correlates. *Intelligence, 40*(2), 226–234. https://doi.org/10.1016/j.intell.2011.11.004

Maccorquodale, K., & Meehl, P. E. (1948). On a distinction between hypothetical constructs and intervening variables. *Psychological Review, 55*(2), 95–107. https://doi.org/10.1037/h0056029

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

Mackey, A. P., Hill, S. S., Stone, S. I., & Bunge, S. A. (2011). Differential effects of reasoning and speed training in children. *Developmental Science, 14*(3), 582–590. https://doi.org/10.1111/j.1467-7687.2010.01005.x

Mackey, A. P., Miller Singley, A. T., & Bunge, S. A. (2013). Intensive Reasoning Training Alters Patterns of Brain Connectivity at Rest. *Journal of Neuroscience*, *33*(11), 4796–4803. https://doi.org/10.1523/jneurosci.4141-12.2013

Mackey, A. P., Miller Singley, A. T., Wendelken, C., & Bunge, S. A. (2015). Characterizing behavioral and brain changes associated with practicing reasoning skills. *PLoS ONE*, *10*(9). https://doi.org/10.1371/journal.pone.0137627

Makrygianni, M. K., Gena, A., Katoudi, S., & Galanis, P. (2018). The effectiveness of applied behavior analytic interventions for children with Autism Spectrum Disorder: A meta-analytic study. *Research in Autism Spectrum Disorders, 51, 18-*

31. https://doi.org/10.1016/j.rasd.2018.03.006

Maltzman, I. (1955). Thinking: from a behavioristic point of view. *Psychological Review*, *62*(4), 275–286. https://doi.org/10.1037/h0041818

Maltzman, I. (1962). Motivation and the direction of thinking. *Psychological Bulletin*, *59*(6), 457–467. https://doi.org/10.1037/h0046851

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

Marteau, T., Dieppe, P., Foy, R., Kinmonth, A. L., & Schneiderman, N. (2006). Behavioural medicine: Changing our behaviour. *British Medical Journal*, *332*(7539), 437–438. https://doi.org/10.1136/bmj.332.7539.437

Matos, M. A., & d'Oliveira, M. M. H. (1992). Equivalence relations and reading. In *Understanding verbal relations* (pp. 83–94).

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

McCarthy, D. (1988). *Escalas McCarthy de aptitudes y psicomotricidad para niños*. Madrid, Spain: TEA.

McCuller, G., Ford, A. M., Bradfield, A., Harrison, C., Rumph, R., Vasquez, E., … Ninness, C. (2017). A Relational Frame and Artificial Neural Network Approach to Computer-Interactive Mathematics. *The Psychological Record*, *55*(1), 135–153. https://doi.org/10.1007/bf03395503

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

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

McHugh, L., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & Dymond, S. (2007). Deictic relational complexity and the development of deception. *Psychological Record*, *57*(4), 517–531. https://doi.org/10.1007/BF03395592

McIlvane, W. J., Bevilacqua, M. C., Huziwara, E. M., Alves, C. O., Lopes, J., de Souza, D. G., … Almeida-Verdu, A. C. (2008). Relational Learning in Children with Deafness and Cochlear Implants. *Journal of the Experimental Analysis of Behavior*, *89*(3), 407–424. https://doi.org/10.1901/jeab.2008-89-407

McIlvane, W. J., Dube, W. V, & Callahan, T. D. (1996). Attention: A behavior analytical perspective. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, memory, and executive function* (pp. 97–117). Baltimore: Brookes.

McIlvane, W. J., Dube, W. V, Kledaras, J. B., Iennaco, F. M., & Stoddard, L. T. (1990). Teaching relational discrimination to individuals with mental retardation: some problems and possible solutions. *American Journal of Mental Retardation*, *95*(3), 283-296.

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 Behavior Analysis*, *19*(2), 228–246. https://doi.org/10.1080/15021149.2018.1507087

Meisenberg, G. (2012). National IQ and economic outcomes. *Personality and Individual Differences*, *53*(2), 103–107. https://doi.org/10.1016/j.paid.2011.06.022

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

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

Melchiori, L. E. (2006). Reading, equivalence, and recombination of units: a replication with students with different learning histories. *Journal of Applied Behavior Analysis*, *33*(1), 97–100. https://doi.org/10.1901/jaba.2000.33-97

Mello, N. K. (1971). Alcohol effects on delayed matching to sample performance by rhesus monkey. *Physiology and Behavior*, *7*(1), 77–101. https://doi.org/10.1016/0031-9384(71)90239-3

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

Molina, M., Castro, E., & Ambrose, R. (2018). Enriching Arithmetic Learning by Promoting Relational Thinking. *The International Journal of Learning: Annual Review*, *12*(5), 265–270. https://doi.org/10.18848/1447-9494/cgp/v12i05/47502

Molina, M., Castro, E., & Castro, E. (2008). Third Graders´ Strategies and Use of Relational Thinking When Solving Number Sentences. *Proceedings of the Joint Meeting of PME 32 and PME-NA XXX*, *3*, 399–406.

Moll, F. W., & Nieder, A. (2015). Cross-modal associative mnemonic signals in crow endbrain neurons. *Current Biology*, *25*(16), 2196–2201. https://doi.org/10.1016/j.cub.2015.07.013

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/s10803-010-0968-0

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational Ability and Language Performance in Children With Autism Spectrum Disorders and Typically Developing Children: A Further Test of the TARPA Protocol. *Psychological Record*, *64*(2), 233–251. https://doi.org/10.1007/s40732-014-0032-0

Moreau, D., Macnamara, B. N., & Hambrick, D. Z. (2019). Overstating the Role of Environmental Factors in Success: A Cautionary Note. *Current Directions in Psychological Science*, *28*(1), 28–33. https://doi.org/10.1177/0963721418797300

Mulholland, T. M., Pellegrino, J. W., & Glaser, R. (1980). Components of geometric analogy solution. *Cognitive Psychology*, *12*(2), 252–284. https://doi.org/10.1016/0010-0285(80)90011-0

Murphy, C., & Barnes-Holmes, D. (2011). Establishing Five Derived Mands in Three Adolescent Boys With Autism. *Journal of Applied Behavior Analysis*, *43*(3), 537–541. https://doi.org/10.1901/jaba.2010.43-537

Murphy, C., & Barnes-Holmes, D. (2017a). Establishing Complex Derived Manding with Children with and Without A Diagnosis of Autism. *The Psychological Record*, *60*(3), 489–503. https://doi.org/10.1007/bf03395723

Murphy, C., & Barnes-Holmes, D. (2017b). Establishing Derived Manding for Specific Amounts with Three Children: An Attempt at Synthesizing Skinner's Verbal Behavior with Relational Frame Theory. *The Psychological Record*, *59*(1), 75–91. https://doi.org/10.1007/bf03395650

Murphy, C., Barnes-Holmes, D., & Barnes-Holmes, Y. (2006). Derived Manding in Children with Autism: Synthesizing Skinner's Verbal Behavior with Relational Frame Theory. *Journal of Applied Behavior Analysis*, *38*(4), 445–462. https://doi.org/10.1901/jaba.2005.97-04

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. (2008). Traditional IQ: 100 years of misconception and its relationship to minority representation in gifted programs. In J. Van Tassel-Basks (Ed.), *Critical issues in equity and excellence in gifted education series, alternative assessment of gifted learners* (pp. 67–88). Waco, TX: Prufrock Press.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. 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*. *1982*.

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. (2014). Verbal and Perceptual Analogical Reasoning and Proportional Metaphor Comprehension in Young Children. *Journal of Speech, Language, and 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

No authorship indicated. (2006). *Differential Aptitude Tests. Journal of Consulting Psychology* (Vol. 12). New York, New York: The Psychological Corporation. https://doi.org/10.1037/h0049421

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’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’Hora, D., Roche, B., Barnes-Holmes, D., & Smeets, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of relational frame theory. *Psychological Record*, *52*(1), 51–75. https://doi.org/10.1007/BF03395414

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.

Oberauer, K., Wilhelm, O., Schulze, R., & Süß, H. M. (2005). Working memory and intelligence -Their correlation and their relation: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin, 131*(1), 61-65. https://doi.org/10.1037/0033-2909.131.1.61

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Personality at work: Raising awareness and correcting misconceptions. *Human Performance*, 18(4), 389-404. https://doi.org/10.1207/s15327043hup1804_5

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

Parish, L., Terman, L. M., & Merrill, M. A. (2006). *Stanford-Binet Intelligence Scale. Third Revision. Form LM. British Journal of Educational Studies* (Vol. 11). Boston: Houghton Mifflin. https://doi.org/10.2307/3118634

Parr, L. A., Winslow, J. T., Hopkins, W. D., & De Waal, F. B. M. (2000). Recognizing facial cues: Individual discrimination by chimpanzees (Pan troglodytes) and rhesus monkeys (Macaca mulatto). *Journal of Comparative Psychology*, *114*(1), 47–60. https://doi.org/10.1037/0735-7036.114.1.47

Parra, I., & Ruiz, F. J. (2016). The effect on intelligence quotient of training fluency in relational frames of coordination. *International Journal of Psychology and Psychological Therapy*, *16*(1), 1–12.

Perrig, W. J., Breil, F., Hoppeler, H., Buschkuehl, M., Jaeggi, S. M., Hutchison, S., … Perrig-Chiello, P. (2009). Impact of working memory training on memory performance in old-old adults. *Psychology and Aging*, *23*(4), 743–753. https://doi.org/10.1037/a0014342

Peters-Scheffer, N., Didden, R., Korzilius, H., & Sturmey, P. (2011). A meta-analytic study on the effectiveness of comprehensive ABA-based early intervention programs for children with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders, 5(1), 60-69*. https://doi.org/10.1016/j.rasd.2010.03.011

Peterson, L., Homer, A. L., & Wonderlich, S. A. (1982). The integrity of independent variables in behavior analysis. *Journal of Applied Behavior Analysis*, *15*(4), 477-492. https://doi.org/10.1901/jaba.1982.15-477

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*. https://doi.org/10.1177/1745691615577701

Porritt, M., & Poling, A. (2008). Scopolamine effects under a titrating-delayed-nonmatching-to-position procedure. *Psychological Record*, *58*(1), 37–49. https://doi.org/10.1007/BF03395601

Premack, D. (2004). Is Language the Key to Human Intelligence? *Science*, *303*(5656), 318-320. https://doi.org/10.1126/science.1093993

Privitera, S., Visalberghi, E., Piano Mortari, E., Truppa, V., & Garofoli, D. (2011). Same/Different Concept Learning by Capuchin Monkeys in Matching-to-Sample Tasks. *PLoS ONE*, *6*(8), e23809. https://doi.org/10.1371/journal.pone.0023809

Prost, A. (2010). 1968 : Mort Et Naissance De L’Université Française. *Vingtième Siècle, Revue d’histoire*, *23*(1), 59–70. https://doi.org/10.3406/xxs.1989.2835

Putnam, H. (1983). *Mind, Language and Reality, Philosophical Papers, volume 2*. Cambridge, Massachsetts: Cambridge University Press.

Ramsden, S., Richardson, F. M., Josse, G., Thomas, M. S. C., Ellis, C., Shakeshaft, C., … Price, C. J. (2011). Verbal and non-verbal intelligence changes in the teenage brain. *Nature*, *479*(7371), 113–116. https://doi.org/10.1038/nature10514

Rapport, L. J., Brooke Brines, D., Axelrod, B. N., & Theisen, M. E. (1997). Full scale IQ as mediator of practice effects: The rich get richer. *Clinical Neuropsychologist*, *11*(4), 375–380. https://doi.org/10.1080/13854049708400466

Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review*, *33*(8), 1237-1252. https://doi.org/10.1016/j.cpr.2013.08.005

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's Standard Progressive Matrices and vocabulary scales*. San Antonio, Texas: Pearson, Inc.

Redick, T. S., Shipstead, Z., Wiemers, E. A., Melby-Lervåg, M., & Hulme, C. (2015). What's Working in Working Memory Training? An Educational Perspective. *Educational Psychology Review, 27*(4), 617–633. https://doi.org/10.1007/s10648-015-9314-6

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence*, *16*(3–4), 259–272. https://doi.org/10.1016/0160-2896(92)90009-G

Reese, H. W., & Parnes, S. J. (1970). Programming creative behavior. *Child Development*, *41*(2), 413-423. https://doi.org/10.1111/j.1467-8624.1970.tb01002.x

Reese, H. W., Parnes, S. J., Treffinger, D. J., & Kaltsounis, G. (1976). Effects of a creative studies program on structure-of-intellect factors. *Journal of Educational Psychology*, *68*(4), 401–410. https://doi.org/10.1037/0022-0663.68.4.401

Rehfeldt, R. A., Dillen, J. E., Ziomek, M. M., & Kowalchuk, R. K. (2007). Assessing relational learning deficits in perspective-taking in children with high-functioning autism spectrum disorder. *Psychological Record*, *57*(1), 23–47. https://doi.org/10.1007/BF03395563

Rey, A. (1958). *L'examen clinique en psychologie. [The clinical examination in psychology.]*. *L'examen clinique en psychologie.* Paris, France: Presses Universitaires de France.

Reynolds, C. R., Gutkin, T. B., Dappen, L., & Wright, D. (2011). Differential Validity of the Wisc-R for Boys and Girls Referred for Psychological Services. *Perceptual and Motor Skills*, *48*(3), 868–870. https://doi.org/10.2466/pms.1979.48.3.868

Richardson, K. (2002). What IQ Tests Test. *Theory & Psychology*, *12*(3), 283–314. https://doi.org/10.1177/0959354302012003012

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

Richland, L. E., Holyoak, K. J., & Stigler, J. W. (2004). Analogy use in eighth-grade mathematics classrooms. *Cognition and Instruction, 22*(1), 37-60. https://doi.org/10.1207/s1532690Xci2201_2

Richland, L. E., & Simms, N. (2015). Analogy, higher order thinking, and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(17), 1213-1216. https://doi.org/10.1002/wcs.1336

Rindermann, H. (2007). The g-factor of international cognitive ability comparisons: The homogeneity of results in PISA, TIMSS, PIRLS and IQ-tests across nations. *European Journal of Personality, 21*(5), 667-706. https://doi.org/10.1002/per.634

Roche, B., Barnes-Holmes, Y., Barnes-Holmes, D., Stewart, I., & O'Hora, D. (2002). Relational frame theory: A new paradigm for the analysis of social behavior. *The Behavior Analyst*, *25*(1), 75-91.

Roche, B., & Barnes, D. (2006). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, *67*(3), 275–301. https://doi.org/10.1901/jeab.1997.67-275

Roche, B., Cassidy, S., & Stewart, I. (2013). Nurturing genius: Using relational frame theory to address a foundational aim of psychology. In Kashdan, T. B., & Ciarrochi, J. (Eds.) *Mindfulness, acceptance, and positive psychology: The seven foundations of well-being.* (pp. 267–302). Oakland, CA, US: Context Press/New Harbinger Publications.

Roitblat, H. L., Penner, R. H., & Nachtigall, P. E. (1990). Matching-to-Sample by an Echolocating Dolphin (Tursiops truncatus). *Journal of Experimental Psychology: Animal Behavior Processes*, *16*(1), 85–95. https://doi.org/10.1037/0097-7403.16.1.85

Rosales, R., Rehfeldt, R. A., & Lovett, S. (2017). Effects of Multiple Exemplar Training on the Emergence of Derived Relations in Preschool Children Learning a Second Language. *The Analysis of Verbal Behavior*, *27*(1), 61–74. https://doi.org/10.1007/bf03393092

Rose, J. C. De, Rossito, A. L., Rose, T. M. S. De, Peder, U., & Sao, D. (1985b). Chapter 4 Stimulus Equivalence and Generalizatíon in Reading After Matching to Sample by Exclusión. *Reading*, (1984), 69–82.

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

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P., & Posner, M. I. (2004). Development of attentional networks in childhood. *Neuropsychologia*, *42*(8), 1029–1040. https://doi.org/10.1016/j.neuropsychologia.2003.12.012

Ruiz, F. J., Suárez, J. J., & López-López, J. C. (2012). Tratamiento de un caso de autismo a través del entrenamiento en comportamiento relacional derivado. In R. Quavedo-Blasco & V. J. Quavedo-Blasco (Eds.), *ResearchGate* (pp. 609–612). Granada, Spain: Asociación Española de Psicología Conductual. Retrieved from https://www.researchgate.net/profile/Francisco_Ruiz8/publication/234002185_Tratamiento_de_un_caso_de_autismo_a_travs_del_entrenamiento_en_comportamiento_relacional_derivado/links/02bfe50e0bd3511eca000000.pdf

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

Sallows, G. O., & Graupner, T. D. (2005). Intensive Behavioral Treatment for Children

With Autism: Four-Year Outcome and Predictors. *American Journal on Mental Retardation*, *110*(6), 417. https://doi.org/10.1352/0895-8017(2005)110[417:ibtfcw]2.0.co;2

Salthouse, T. A. (2014). Quantity and structure of word knowledge across adulthood. *Intelligence*, *46*(1), 122–130. https://doi.org/10.1016/j.intell.2014.05.009

Salthouse, T. A., & Pink, J. E. (2008). Why is working memory related to fluid intelligence? *Psychonomic Bulletin and Review*, *15*(2), 364–371. https://doi.org/10.3758/PBR.15.2.364

Sattler, J. M. (2001). *Assessment of children: Cognitive applications.* San Diego, CA: Jerome M. Sattler Publisher Inc.

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

Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Schrank, F. A., McGrew, K. S., Mather, N., & Woodcock, R. W. (2014). *Woodcock-Johnson IV.* Rolling Meadows, Illinois: Riverside Publishing.

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

Schulz, K. F., Chalmers, L., Hayes, R. J., & Altman, D. G. (1995). Empirical evidence of bias: Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association, 273*(5), 8-12. https://doi.org/10.1001/jama.273.5.408

Sedgwick, P. (2014). Spearman’s rank correlation coefficient. *BMJ (Online), 349*. https://doi.org/10.1136/bmj.g7327

Selnes, O. A. (2012). A Compendium of Neuropsychological Tests: Administration, Norms, and Commentary. *Neurology*, *41*(11), 1856–1856. https://doi.org/10.1212/wnl.41.11.1856-a

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*(3), 409–418. https://doi.org/10.1348/978185408X383123

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*. https://doi.org/10.1348/000709906X96987

Sheinkopf, S. J., & Siegel, B. (1998). Home-based behavioral treatment of young children with autism. *Journal of Autism and Developmental Disorders*, *28*(1), 15–23. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9546298

Shen, T. (2014). Exploring the Relationship between Creative Test of ATTA and the Thinking of Creative Works. *The European Journal of Social & Behavioural Sciences*, *7*(4), 1159–1167. https://doi.org/10.15405/ejsbs.95

Sheppard, L. D., & Vernon, P. A. (2008). Intelligence and speed of information-processing: A review of 50 years of research. *Personality and Individual Differences, 44(3), 535-551*. https://doi.org/10.1016/j.paid.2007.09.015

Shiel, G., Kelleher, C., McKeown, C., & Denner, S. (2015). *Future Ready? The Performance of 15-year-olds in Ireland on Science, Reading Literacy and Mathematics in PISA 2015*. *Educational Research Centre*. Dublin, Ireland, Ireland: Educational Research Centre, St. Patrick's Collge. https://doi.org/10.1002/9781119206613

Shipstead, Z., Hicks, K. L., & Engle, R. W. (2012). Working memory training remains a work in progress. *Journal of Applied Research in Memory and Cognition*, *1*(3), 217–219. https://doi.org/10.1016/j.jarmac.2012.07.009

Shipstead, Z., Redick, T. S., & Engle, R. W. (2012). Is working memory training effective? *Psychological Bulletin*, *138*(4), 628–654. https://doi.org/10.1037/a0027473

Sidman, M. (1960). *Tactics of scientific research*. *Tactics of scientific research.* New York, New York, New York: Basic Books. Retrieved from http://jproxy.lib.ecu.edu/login?url=http://search.ebscohost.com/login.aspx?direct=true&db=psyh&AN=1961-01453-000&site=ehost-live

Sidman, M. (2014). Reading and Auditory-Visual Equivalences. *Journal of Speech and Hearing Research*, *14*(1), 5–13. https://doi.org/10.1044/jshr.1401.05

Sidman, M. (2017). Equivalence Relations and Behavior: An Introductory Tutorial. *The Analysis of Verbal Behavior*, *25*(1), 5–17. https://doi.org/10.1007/bf03393066

Siedlecki, K. L., Tucker-Drob, E. M., Oishi, S., & Salthouse, T. A. (2008). Life satisfaction across adulthood: Different determinants at different ages? *Journal of Positive Psychology*, *3*(3), 153–164. https://doi.org/10.1080/17439760701834602

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

Siemsen, H., Testelin, J., Martin-Hansen, L., Siemsen, K. H., Andrieu, B., & Fevre, J. M. (2017). *Transforming IQ into "Orthopédie Mentale"*. Paris, France: Springer International.

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. (1953). *Science and behavior*. New York, New York, New York: The Free Press.

Skinner, B. F. (1974). *About behaviorism.* Oxford, England: Alfred A. Knopf.

Skinner, B. F. (2010). An operant analysis of problem solving. *Behavioral and Brain Sciences*, *7*(04), 583. https://doi.org/10.1017/s0140525x00027412

Skov-Rackette, S. I., Miller, N. Y., & Shettleworth, S. J. (2006). What-where-when memory in pigeons. *Journal of Experimental Psychology: Animal Behavior Processes*, *32*(4), 345–358. https://doi.org/10.1037/0097-7403.32.4.345

Slattery, B., Stewart, I., & O'Hora, D. (2011). Testing for transitive class containment as a feature of hierarchical classification. *Journal of the Experimental Analysis of Behavior*, *96*(2), 243-260. https://doi.org/10.1901/jeab.2011.96-243

Sliwa, J., Duhamel, J.-R., Pascalis, O., & Wirth, S. (2011). Spontaneous voice-face identity matching by rhesus monkeys for familiar conspecifics and humans. *Proceedings of the National Academy of Sciences*, *108*(4), 1735–1740. https://doi.org/10.1073/pnas.1008169108

Smirnova, A. A., Lazareva, O. F., & Zorina, Z. A. (2003). Prototype symbolization in hooded crows. *Neuroscience and Behavioral Physiology*, *33*(4), 335–348. https://doi.org/10.1023/A:1022839405933

Smirnova, A. A., Lazareva, O. F., & Zorina, Z. A. (2006). Use of number by crows: investigation by matching and oddity learning. *Journal of the Experimental Analysis of Behavior*, *73*(2), 163–176. https://doi.org/10.1901/jeab.2000.73-163

Smirnova, A., Zorina, Z., Obozova, T., & Wasserman, E. (2015). Crows spontaneously exhibit analogical reasoning. *Current Biology*, *25*(2), 256–260. https://doi.org/10.1016/j.cub.2014.11.063

Smith, B. L., Smith, T. D., Taylor, L., & Hobby, M. (2005). Relationship between Intelligence and Vocabulary. *Perceptual and Motor Skills*, *100*(1), 101–108. https://doi.org/10.2466/pms.100.1.101-108

Smith, T., Groen, A. D., & Wynn, J. W. (2008). Randomized Trial of Intensive Early Intervention for Children with Pervasive Developmental Disorder. *Early Intervention: The Essential Readings*, *105*(4), 153–182. https://doi.org/10.1002/9780470755778.ch6

Soto, T., & Kraper, C. (2013). Block Design Subtest. In F. R. Volkmar (Ed.), *Encyclopedia of Autism Spectrum Disorders*. New York, New York: Springer Publishing Co.

Soveri, A., Antfolk, J., Karlsson, L., Salo, B., & Laine, M. (2017). Working memory training revisited: A multi-level meta-analysis of n-back training studies. *Psychonomic Bulletin and Review*, *24*(4), 1077–1096. https://doi.org/10.3758/s13423-016-1217-0

Sparling, J., Miller-Johnson, S., Pungello, E., Campbell, F. A., & Ramey, C. T. (2004). Early Childhood Education: Young Adult Outcomes From the Abecedarian Project. *Applied Developmental Science*, *6*(1), 42–57. https://doi.org/10.1207/s1532480xads0601_05

Sparrow, S. S., & Cicchetti, D. V. (1989). The Vineland Adaptive Behavior Scales. In *Major psychological assessment instruments* (Vol. 2, pp. 199–231).

Spearman, C. (1927). *The abilities of a man*. New York, New York, New York: Macmillan Co. https://doi.org/10.1054/ijom.2003.0429

Spearman, C. (1946). Theory of General Factor. *British Journal of Psychology. General Section*, *36*(3), 117–131. https://doi.org/10.1111/j.2044-8295.1946.tb01114.x

Spearman, C. (2006). "General Intelligence," Objectively Determined and Measured. *The American Journal of Psychology*, *15*(2), 201. https://doi.org/10.2307/1412107

Staats, A. W. (1968). *Learning, Language, and Cognition.* New York, New York: Holt, Rinehart & Winston.

Staats, A. W. (1971). *Child learning, intelligence and personality*. New York, New York: Harper & Row.

Staats, A. W. (1996). *Behavior and personality: Psychological behaviorism.* (New York,). Springer Publishing Co.

Staats, A. W. (1996). *Behavior and Personality: Psychological Behaviorism.* (New York,). New York, New York, New York: Springer Publishing Co.

Staats, A. W., & Burns, G. L. (1981). Intelligence and child development: What intelligence is and how it is learned and functions. *Genetic Psychology Monographs*, *104*(2), 237–301. Retrieved from http://login.proxy.library.vanderbilt.edu/login?url=http://search.proquest.com/docview/616605665?accountid=14816%5Cnhttp://sfx.library.vanderbilt.edu/vu?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:journal&genre=article&sid=ProQ:ProQ%3Apsycinfo&at

Stanhope, K. J., McLenachan, A. P., & Dourish, C. T. (1995). Dissociation between cognitive and motor/motivational deficits in the delayed matching to position test: effects of scopolamine, 8-OH-DPAT and EAA antagonists. *Psychopharmacology*, *122*(3), 268–280. https://doi.org/10.1007/BF02246548

Steele, D., & Hayes, S. C. (2006). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, *56*(3), 519–555. https://doi.org/10.1901/jeab.1991.56-519

Stephens, A. C. (2008). What "counts" as algebra in the eyes of preservice elementary teachers? *Journal of Mathematical Behavior*, *27*(1), 33–47. https://doi.org/10.1016/j.jmathb.2007.12.002

Stephens, M. (2006). Describing And Exploring The Power Of Relational Thinking. In *Group* (pp. 1370–1374).

Stephens, M. (2007). Students ' Emerging Algebraic Thinking in the Middle School Years. *Mathematics: Essential Research, Essential Practice: Proceedings of the 30th Annual Conference of the Mathematic Education Research Group of Australasia*, 678–687.

Stephens, M., & Ribeiro, A. (2012). Working towards algebra: The importance of relational thinking. *Revista Latinoamericana de Investigacion En Matematica Educativa*, *15*(3), 373–402.

Stephenson, C. L., & Halpern, D. F. (2013). Improved matrix reasoning is limited to training on tasks with a visuospatial component. *Intelligence*, *41*(5), 341–357. https://doi.org/10.1016/j.intell.2013.05.006

Sternberg, R. J. (2008). Increasing fluid intelligence is possible after all. *Proceedings of the National Academy of Sciences*, *105*(19), 6791–6792. https://doi.org/10.1073/pnas.0803396105

Stewart, I., & Barnes-Holmes, D. (2001). Understanding metaphor: A relational frame perspective. *Behavior Analyst*, *24*(2), 191–199. https://doi.org/10.1007/BF03392030

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. (2006). A functional-analytic model of analogy: a relational frame analysis. *Journal of the Experimental Analysis of Behavior*, *78*(3), 375–396. https://doi.org/10.1901/jeab.2002.78-375

Stewart, I., Ming, S., Walsh, L., Moran, L., & McElwee, J. (2015). Correlating derived relational responding with linguistic and cognitive ability in children with Autism Spectrum Disorders. *Research in Autism Spectrum Disorders*, *19*, 32–43. https://doi.org/10.1016/j.rasd.2014.12.015

Stewart, I., & Roche, B. (2013). Relational frame theory: An overview. In *Advances in relational frame theory: Research and application.* (pp. 51–71). Oakland, CA, US: Context Press/New Harbinger Publications.

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). Oakland, CA: New Harbinger.

Stoney, C. M., & Lee Johnson, L. (2012). Design of Clinical Studies and Trials. In *Principles and Practice of Clinical Research* (pp. 225–242). https://doi.org/10.1016/b978-0-12-382167-6.00019-9

Sulzer-Azaroff, B., & Gillat, A. (2006). *Trends in behavior analysis in education*. *Journal of Applied Behavior Analysis* (Vol. 23). https://doi.org/10.1901/jaba.1990.23-491

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*(4), 349–362. https://doi.org/10.1016/j.intell.2004.06.004

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

Symonds, P. M., & Spearman, C. (2006). The Abilities of Man; Their Nature and Measurement. *The Journal of Philosophy*, *25*(1), 20. https://doi.org/10.2307/2015168

Tarshis, E., & Shore, B. M. (1991). Perspective taking in high and above average IQ preschool children. *European Journal of High Ability*, *2*(2), 201–211. https://doi.org/10.1080/0937445910020209

Teasdale, T. W., & Owen, D. R. (2005). A long-term rise and recent decline in intelligence test performance: The Flynn Effect in reverse. *Personality and Individual Differences*, *39*(4), 837-843. https://doi.org/10.1016/j.paid.2005.01.029

Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* Boston: Houghton Mifflin.

Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology and Psychological Therapy*, *16*(2), 141–155.

Thompson, T. W., Waskom, M. L., Garel, K. L. A., Cardenas-Iniguez, C., Reynolds, G. O., Winter, R., … Gabrieli, J. D. E. (2013). Failure of Working Memory Training to Enhance Cognition or Intelligence. *PLoS ONE*, *30*(4), 306-312. https://doi.org/10.1080/08824096.2013.837042

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Stanford-Binet Intelligence Scale: Fourth Edition*. Itasca, Illinois: Riverside Publishing.

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

Tranter, L. J., & Koutstaal, W. (2008). Age and flexible thinking: An experimental demonstration of the beneficial effects of increased cognitively stimulating activity on fluid intelligence in healthy older adults. *Aging, Neuropsychology, and Cognition*, *15*(2), 184–207. https://doi.org/10.1080/13825580701322163

Trask, P. C., Schwartz, S. M., Deaner, S. L., Paterson, A. G., Johnson, T., Rubenfire, M., & Pomerleau, O. F. (2002). Behavioral medicine: the challenge of integrating psychological and behavioral approaches into primary care. *Effective Clinical Practice: ECP*, *5*(2), 75-83.

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

Unsworth, N., & Engle, R. W. (2006). Simple and complex memory spans and their relation to fluid abilities: Evidence from list-length effects. *Journal of Memory and Language*, *54*(1), 68–80. https://doi.org/10.1016/j.jml.2005.06.003

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*(3), 498–505. https://doi.org/10.3758/BF03192720

van der Maas, H., Kan, K.-J., & Borsboom, D. (2014). Intelligence Is What the Intelligence Test Measures. Seriously. *Journal of Intelligence*, *2*(1), 12–15. https://doi.org/10.3390/jintelligence2010012

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research*, *54*(5), 433–447. https://doi.org/10.1111/j.1365-

2788.2010.01285.x

Van Etten, M. L., & Taylor, S. (1998). Comparative efficacy of treatments for PTSD: A meta-analysis. *Clinical Psychology and Psychotherapy*, *5*(3), 126–145. https://doi.org/10.1002/(SICI)1099-0879(199809)5:3<126::AID-CPP153>3.0.CO;2-H

Vernon, P. A., & Jensen, A. R. (2002). Individual and group differences in intelligence and speed of information processing. *Personality and Individual Differences, 5*(4), 411–423. https://doi.org/10.1016/0191-8869(84)90006-0

Vetterli, C. F., & Furedy, J. 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. https://doi.org/10.1016/S0191-8869(96)00265-6

Virués-Ortega, J. (2010). Applied behavior analytic intervention for autism in early childhood: Meta-analysis, meta-regression and dose-response meta-analysis of multiple outcomes. *Clinical Psychology Review*. https://doi.org/10.1016/j.cpr.2010.01.008

Vizcaíno-Torres, R. M., Ruiz, F. J., Luciano, C., López-López, J. C., Barbero-Rubio, A., & Gil, E. (2015). The effect of relational training on intelligence quotient: A case study. *Psicothema*, *27*, 120–127. https://doi.org/10.7334/psicothema2014.149

Vonk, J. (2003). Gorilla (Gorilla gorilla gorilla) and orangutan (Pongo abelii) understanding of first- and second-order relations. *Animal Cognition*, *6*(2), 77–86. https://doi.org/10.1007/s10071-003-0159-x

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

Waris, O., Soveri, A., & Laine, M. (2015). Transfer after working memory updating training. *PLoS ONE, 10*(9). https://doi.org/10.1371/journal.pone.0138734

Wechsler, D. (n.d.). *Wechsler Intelligence Scale for Children, Third Edition, manual.* San Antonio, Texas: Psychological Corporation.

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York, New York: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale manual*. New York, New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale-Revised manual*. New York, New York: Psychological Corporation.

Wechsler, D. (1991). Wechsler intelligence scale for children–Third edition. *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (1997a). *WAIS-III WMS-III technical manual.* San Antonio, Texas: Psychological Corporation.

Wechsler, D. (1997b). *WAIS-III administration and scoring manual*. *The Psychological Corporation, San Antonio, TX*. https://doi.org/10.1177/1073191102009001003

Wechsler, D. (1999). Wechsler Abbreviated Scale of Intelligence manual. San Antonio, Texas: Psychological Corporation.

Wechsler, D. (2002). Wechsler Preschool and Primary Scale of Intelligence. *Wechsler Preschool and Primary Scale of Intelligence*. https://doi.org/10.1007/978-1-4419-1698-3_866

Wechsler, D. (2003). Wechsler intelligence scale for children–Fourth Edition (WISC-IV). *San Antonio, TX: The Psychological Corporation*.

Wechsler, D. (2005). *Wechsler Individual Achievement Test*. London, UK: The Psychological Corporation.

Wechsler, D. (2008). Wechsler Adult Intelligence Scale - Fourth Edition. *Statistics Solutions*.

Wechsler, D. (2011). *Wechsler Abbreviated Scale of Intelligence-Second Edition, manual. Wechsler Abbreviated Scale of Intelligence--Second Edition (WASI-II)*. San Antonio, Texas: Psychological Corporation. https://doi.org/citeulike-article-id:6135820

Wechsler, D. (2013). WASI -II: Wechsler abbreviated scale of intelligence - second edition. *Journal of Psychoeducational Assessment*. https://doi.org/10.1177/0734282912467756

Westerberg, H., & Klingberg, T. (2007). Changes in cortical activity after training of working memory - a single-subject analysis. *Physiology and Behavior*, *92*(1–2), 186–192. https://doi.org/10.1016/j.physbeh.2007.05.041

Wikipedia, F., & Herrnstein, R. (2005). *The Bell Curve*. New York, New York: Free Press. https://doi.org/10.5408/0022-1368-43.1.2

Willis, S. L., Tennstedt, S. L., Marsiske, M., Ball, K., Elias, J., Koepke, K. M., … Wright, E. (2006). Long-term effects of cognitive training on everyday functional outcomes in older adults. *Journal of the American Medical Association*, *296*(23), 2805–2814. https://doi.org/10.1001/jama.296.23.2805

Woodcock, R. W., & Johnson, M. B. (1977). *Woodcock-Johnson Psycho- Educational Battery*. Allen, Texas: DLM.

Woodcock, R. W., & Johnson, M. B. (1989). *Woodcock-Johnson Psycho- Educational Ba ery–Revised*. Rolling Meadows, Illinois: Riverside Publishing.

Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III Tests of Achievement*. *Test*.

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences*, *54*(8), 871–876. https://doi.org/10.1016/j.paid.2012.12.022

Worsham, R. W., & D'Amato, M. R. (1973). Ambient light, white noise, and monkey vocalization as sources of interference in visual short-term memory of monkeys. *Journal of Experimental Psychology*, *99*(1), 99–105. https://doi.org/10.1037/h0034770

Wright, A. A., Cook, R. G., Rivera, J. J., Sands, S. F., & Delius, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning & Behavior*, *16*(4), 436–444. https://doi.org/10.3758/BF03209384

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

Wright, A. A., & Delius, J. D. (2005). Learning processes in matching and oddity: The oddity preference effect and sample reinforcement. *Journal of Experimental Psychology: Animal Behavior Processes*, *31*(4), 425–432. https://doi.org/10.1037/0097-7403.31.4.425

Wulfert, E., & Hayes, S. C. (2006). Transfer of a conditional ordering response through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior*, *50*(2), 125–144. https://doi.org/10.1901/jeab.1988.50-125

Yang, S., Paynter, J. M., & Gilmore, L. (2016). Vineland Adaptive Behavior Scales: II Profile of Young Children with Autism Spectrum Disorder. *Journal of Autism and Developmental Disorders*, *46*(1), 64–73. https://doi.org/10.1007/s10803-015-2543-1

Yousafzai, S. Y., Foxall, G. R., & Pallister, J. G. (2010). Explaining internet banking behavior: Theory of reasoned action, theory of planned behavior, or technology acceptance model? *Journal of Applied Social Psychology*, *40*(5), 1172–1202. https://doi.org/10.1111/j.1559-1816.2010.00615.x

Zachariah, M., & Gould, S. J. (2006). *The Mismeasure of Man*. *Canadian Journal of Education / Revue canadienne de l'éducation* (Vol. 9). New York, New York, New York: Norton. https://doi.org/10.2307/1494302

Zacks, J. M., Vettel, J. M., & Michelon, P. (2003). Imagined Viewer and Object Rotations Dissociated with Event-Related fMRI. *Journal of Cognitive Neuroscience*, *15*(7), 1002–1018. https://doi.org/10.1162/089892903770007399

Zheng, Y., Myerson, J., & Hale, S. (2000). Age and individual differences in visuospatial processing speed: Testing the magnification hypothesis. *Psychonomic Bulletin and Review*, *7*(1), 113–120. https://doi.org/10.3758/BF03210729

Zimmerberg, B., Glick, S. D., & Jarvik, M. E. (1971). Impairment of recent memory by marihuana and THC in rhesus monkeys. *Nature*, *233*(5318), 343–345. https://doi.org/10.1038/233343a0

Zorina, Z. A., & Smirnova, A. A. (1996). Quantitative evaluations in gray crows: Generalization of the relative attribute "larger set." *Neuroscience and Behavioral Physiology*, *26*(4), 357–364. https://doi.org/10.1007/BF02359040

Zürcher, N. R., Rodriguez, J. S., Jenkins, S. L., Keenan, K., Bartlett, T. Q., McDonald, T. J., … Nijland, M. J. (2010). Performance of juvenile baboons on neuropsychological tests assessing associative learning, motivation and attention. *Journal of Neuroscience Methods*, *188*(2), 219–225. https://doi.org/10.1016/j.jneumeth.2010.02.011

Appendix A

Experiment 1: Information Sheet for Parents & Guardians

The purpose of this psychological research is to examine the benefits of the SMART program previously completed by your son. The data required for this study has already been collected in the 2015-2016 academic year, and as such, your son will not be required to complete any further assessment or training. All that is being sought is permission to access previously collected records which detail my son's progress and performance during the training program.

As part of the SMART programme, your son completed two Relational Abilities Index assessments (RAI; a measure of problem-solving ability) and a series of relational skills training tasks. There are two objectives of this study: (1) an analysis of the general performance of secondary school students on the Relational Abilities Index assessment and (2) an investigation into the effectiveness of the SMART program programme in increasing these RAI scores.

These investigations are being conducted in order to help develop the RAI into an even more practical and illustrative measure of ability. For most educational or intellectual assessments, it is very important to be able to give participants scores which reflect their relative level of performance when compared to their peers (e.g. your score was higher than 25/50/75% of other students in your age group). In order to be able to do this, we must gather a large amount of data, which will tell us how students in each age group generally perform on the RAI assessment. That is the primary aim of this study. In addition, we would also like to ascertain how effective this program is in increasing relational skills (i.e. how good is this program at doing what it aims to do?). In order to maximise the potential benefits of this program, we must further investigate how effective it currently is, in order to identify areas to improve upon.

This research is being conducted as part of PhD. level research by Mr. Dylan Colbert under the supervision of Dr. Bryan Roche of Maynooth University. Dylan Colbert has been fully Garda vetted. Dylan Colbert (the main researcher) is currently being supervised by a qualified educational psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing and that data is processed and stored correctly and in accordance with normal data protection procedures. All RAI scores and other statistics will be passed on by Summerhill College to Mr Colbert in a anonymised form so that confidentiality is assured. The only information being sought are individual scores on the RAI assessment (pre- and post-training), number of training levels completed, and number and duration of training sessions completed. As such, the researcher will not receive any personal or identifiable data (names, address etc.).

All students participating in this study will remain completely anonymous and will not be referred to by name in any publication or document. Furthermore, as this information will be anonymised before being sent to the primary experimenter, it will not be possible to link any score or training statistic to any specific student. The data collected will be used only by the researchers. This data will be available to each participant's parents/guardians.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. Declining to participate in this research will in no way affect your child's education or access to normal teaching services.

You can contact Dr. Bryan Roche at Bryan.t.roche@mu.ie of Maynooth University as supervisor of this research if you have any concerns or queries. The study has been approved by the school's Board of Management and the Principal is happy for it to go ahead for those parents who give their consent. The study has also been approved by the Research Ethics Committee of Maynooth University and it adheres to their Child Protection Policy.

Finally, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at DYLAN.COLBERT.2011@mumail.ie. Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@mu.ie or by telephone at (01) 7086026. Dr. Sarah Cassidy can be reached at smithsfieldclinic@gmail.com.

Appendix B

Experiment 1: Consent Form for Parent or Guardian

In agreeing to allow my son to participate in the research project I understand the following:

- Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator, a psychology graduate currently gathering data for his PhD at Maynooth University.

- The purpose of this psychological research is to measure the intellectual benefits of the SMART program previously completed by my son at Summerhill College. The data required for this study has already been collected in the 2015-2016 academic year, and as such, my son will not be required to complete any further assessment or training. What is being sought is permission to access previously collected records which detail my son's progress and performance during the training program and to use these in scientific research publications.

- All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This data will be available to each participant and his parents/guardians should they request it.

- The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and the Ethics Committee of Maynooth University. The study design adheres to Maynooth University's Child Protection Policy.

- I understand that I may withdraw my son's data from the study at any stage up to but not following publication of the data.

- I understand that I may contact Dr. Bryan Roche at Bryan.T.Roche@mu.ie of Maynooth University as supervisor of this research programme.

- I acknowledge that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Signed in duplicate:

________________________________ Parent/Guardian

________________________________ Researcher

________________________________ Date

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

Appendix C

Experiment 2: Information Sheet for Parents/Guardians

This project is designed to test the effectiveness of a newly developed technique for raising general intellectual ability. The particular method used in this study is based on a scientific theory of cognitive development, known as Relational Frame Theory, which was partly developed by one of the consultants on this project (Dr. Bryan Roche), at Maynooth University. The method is called SMART training and the online tool used to deliver this training was developed at Maynooth University and is used in several Irish schools as part of the normal curriculum. The name given to this online tool is SMART (Strengthening Mental Abilities with Relational Training).

What is SMART?

SMART training is based on the finding that most of what psychologists consider intelligent behaviour involves relating things to each other in a variety of ways (i.e. seeing the connections and links between things). Intelligent people have a good understanding of simple concepts such as before/after, more/less, opposite/same, here/there and so on. These are called "relational skills". When we teach these skills, intelligence, as measured by IQ tests, appears to rise. This project, being run by County Carlow Development Project, based in Banglestown, is being targeted at students in the Carlow area and is hoped to boost the general cognitive ability (i.e, intelligence) of the children chosen to take part in this first trial in the region.

While early results have been promising, and several published scientific studies have shown that IQ gains result from using this online intellectual skills training tool, there is no guarantee that volunteers who undergo the online training will experience increases in their IQ following participation.

This delivery of the SMART programme is being overseen by Dr. Bryan Roche of Maynooth University who is a Psychologist and who has many years of experience running trials of this type in Irish school settings.

What does SMART involve?

SMART training involves a system of online puzzles or exercises on a personal computer or other internet connected device. During the training period, your child will be asked to complete three training sessions per week during school hours. Each one of these sessions will take around 45 minutes to complete. The training usually requires approximately three months completing and involves a quick (10 minutes) relational skills assessment at the outset, which is then repeated at the end of the training. All of this is automated and will be delivered entirely online. While your child will be taken out of class for these sessions, along with the other children participating in the programme, the skills which will be trained in these sessions are considered integral in establishing an improved ability to perform scholastically.

Training consists of solving a number of logical puzzles, followed by feedback from the computer in some cases but not in others. For example, users may be asked;

"If A is more than B and C is less than B, is A more than C?"

Users indicate their answer in all cases by clicking on the words "Yes" or "No" on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer, which delivers feedback to help the user improve their accuracy at these types of tasks. This technique allows us to train the relational abilities that are so important to intelligence.

IQ and standardised school testing

As part of this research it is also necessary to administer a shortened intelligence (IQ) test two times; in January 2017, and again upon completion of training in May 2017. The test used will be the WASI IQ test, a test designed for both children and adults. The test will take approximately 30-40 minutes, and the participant will be able to take breaks whenever needed. These IQ scores will NOT be delivered to your child but will be made available to you at the end of the study in a confidential letter delivered via the school. That document will explain what IQ is and how to interpret the IQ scores recorded for your child. You can decide at that point if you think it is appropriate to tell your child their IQ or not.

A detailed psychological report on each child will ***not*** be provided because IQ is being measured in this project for research purposes only (i.e., not for diagnostic purposes). We simply want to see if your child's general cognitive ability improves. If users have any concerns about their scores, however, they will be referred to Dr. Bryan Roche of Maynooth University.

As part of the project we are asking that all parents allow the CCDP to use the standardised mathematics and verbal ability scores recorded yearly by your child's school to be used to assess improvements. These will be used only in anonymous form using code names, so that the project managers can assess whether or not improvements in school aptitudes result from the SMART training.

Use of and access to your child's data

The CCDP and school teachers will be able to remotely monitor the progress and login times for all students on the project and will have access to all test results. These, however, will be known to Dr. Roche at Maynooth University only using pseudonyms, and will not be known to anyone other than staff at CCDP who are working on this project and your child's teachers who are involved in delivering the project at their school. This information will be kept strictly confidential at all times.

All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only in any subsequent publication or conference presentation. Data will only be discussed in terms of the group, no individual participant will be singled out and analysed. The data collected will be used only by the researchers. This data will be available to each participant's parents/guardians. No personal data will be recorded except your child's name and date of birth, and these are recorded solely so that the recorded IQ scores can be returned to each child's parents.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. Declining to participate in this research or withdrawing from it before it is completed will in no way affect your child's education or access to normal teaching services. You can contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if you have any concerns or queries.

Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@nuim.ie or by telephone at (01) 7086026.
Dylan Thomas at County Carlow Development project can be reached at dthomas@carlowdevelopment.ie

Appendix D

Experiment 2: Consent Form for Parent or Guardian

In agreeing to allow my son to participate in the research project I understand the following:

County Carlow Development Project, in collaboration with Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this programme in my child's school. The purpose of this project is to examine the effectiveness of a form of online intellectual skills for increasing intellectual ability or IQ.

I understand that my child will be asked to complete a short-form IQ test in January 2017 and again following the conclusion of the programme in May 2017.This assessment will be conducted by Dr. Bryan Roche of Maynooth University, in a suitable room in my child's school under the supervision of a school teacher. The wishes and comfort of my child will be treated with the utmost sensitivity during the assessment and during every training session which will be overseen by a teacher from my child's school. My child's training will take place online, three times weekly in one of the school resource rooms and these training sessions will last around 45 minutes.

All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the project managers at CCDP and Dr. Bryan Roche at Maynooth University. This data will be available to each participant and his parents/guardians at the end of the training period.

I understand that I may withdraw my child from the study at any stage even after giving my consent. My child may also leave the study at any time. I may also withdraw his data at the conclusion of my participation if I still have concerns.

I understand that I may contact Dr. Bryan Roche at Bryan.T.Roche@nuim.ie of Maynooth University or Dylan Thomas (dthomas@carlowdevelopment.ie) of CCDP if I have any concerns about the project.

I understand that increases in IQ are not guaranteed as a result of participating in this project. I understand that the project is experimental and not clinical in nature and that my child will not receive a full psychological report although I will receive their IQ scores at the end of the study.

I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed by CCDP or Br. Bryan Roche.

Signed in duplicate:

________________________________ Parent/Guardian

________________________________ Project representative

________________________________ Date

Appendix E

Experiment 3: Information Sheet for Participants

Drimnagh Castle has agreed to take part in this study, which is trying to see if a new type of brain training can improve intelligence. The brain training program that is being used is actually available online where it is called SMART training.

This program has been created by a group of psychologists – people who study how people think and behave. Psychologists measure intelligence by using a test known as an IQ test. You will take three IQ tests as part of this study, to check the effect the brain training has. Each IQ test takes around an hour to complete.

An IQ test measures how good you are at solving different types of problems – for example, maths problems and word problems. For example, we will ask you what certain words mean, we will test your memory and we will ask you some mental math problems. Your score will be kept totally private but we will pass on the IQ test results to your parents at the end of the study and they can choose what to do with them. You will be asked many different types of questions.

Psychologists have discovered that really intelligent people are very good at understanding the relationships between things. For example, they can easily figure out if two things are the same or the opposite, or they can work out if one thing is bigger or smaller than another – even when it may not be so obvious.

SMART training teaches people how to see these relationships more easily.

The training is very simple – all you have to do is solve some mental puzzles online. As you train, these puzzles will get harder and will help you to understand those important relationships. The training is a bit like some brain training games you might have played on the Nintendo DS or on a tablet.

You will be asked to train twice a week in school, and once at home, each time for around 30 minutes. Half of your class will train before Christmas, and the other half will do the same training after Christmas.

You are allowed to stop taking part at any time if you would rather not do the training or take the IQ tests.

The study will be run by Dylan Colbert from Maynooth University. Mr. Colbert was once a student of Drimnagh Castle.

Appendix F

Experiment 3: Information Sheet for Parents & Guardians

The current research project is designed to test the effectiveness of a newly developed technique for raising intellectual ability. The particular method used in this study is based on a theory known as Relational Frame Theory and is called SMART training (Strengthening Mental Abilities with Relational Training).

SMART training is based on the finding that most of what psychologists consider intelligent behaviour involves relating things to each other in a variety of ways (i.e. seeing the connections and links between things). Intelligent people have a good understanding of simple concepts such as before/after, more/less, opposite/same, here/there and so on. These are called “relational skills”. When we teach these skills, intelligence, as measured by IQ tests, appears to rise. However, more research is needed to confirm that this is the case.

The tool we are using to train “relational skills” in this study has already been developed in previous research at Maynooth University. The main findings of that research have been published, and a web-based tool developed within Maynooth University has also now been made publicly available. Basically, we are trying to assess how effective this training can be. While early results have been promising, there is no absolute guarantee that volunteers who undergo the online training will experience increases in their IQ following participation.

This research is being conducted as part of PhD. level research by Mr. Dylan Colbert under the supervision of Dr. Bryan Roche of NUI Maynooth. Dylan Colbert has been fully Garda vetted and is a former student of Drimnagh Castle CBS.

SMART training involves a system of online puzzles or exercises on a personal computer or other internet connected device. During the training period, your son will be asked to complete three training sessions per week – twice in school, and once as part of his homework. Each one of these sessions will take around 30 minutes to complete. The training usually requires approximately three months to complete and involves a quick (10 minutes) intellectual assessment at the outset, which is then repeated at the end of the training. All of this is automated and will be delivered entirely online. While your son will be taken out of class for these sessions, along with the other children participating in the study, the skills which will be trained in these sessions are considered integral in establishing an improved ability to perform scholastically.

Training consists of solving a number of logical puzzles, followed by feedback from the computer in some cases but not in others. For example, users may be asked;

“If A is more than B and C is less than B, is A more than C?”

Users indicate their answer in all cases by clicking on the words “Yes” or “No” on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer, which delivers feedback to help the user improve their accuracy at these types of tasks. This technique allows us to train the relational abilities that are so important to intelligence.

All volunteers will be given the online “brain training” as part of the study. Half of the volunteers will begin training in September and should finish by December. The other half of the volunteers will train during the second training period (January to March/April). All participants will receive the same training, for the same period of time. The researcher, his supervisor and your son’s teacher will be able to see the frequency of logins by the user as well as their progress.

As part of this research it is also necessary to administer a full intelligence (IQ) test three times; in September 2016, December 2016/January 2017 and then finally in April/May 2017. The test used will be the WISC IQ test, a test specially designed for children. The test will take approximately 60-90 minutes, and the participant will be able to take breaks whenever needed. These IQ scores will NOT be delivered to your child but will be made available to you at the end of the study in a confidential letter delivered via the school. That document will explain what IQ is and how to interpret the IQ scores recorded for your child throughout the study.

A detailed psychological report on each child will not be provided because IQ is being measured in this study for research purposes only (i.e., not for diagnostic purposes). If users have any concerns about their scores, however, they will be referred to the research team's educational Psychologist Dr. Sarah Cassidy free of charge. She will address any concerns participants have about their IQ score, and will help them to understand the meaning of their IQ test performance. They may also contact the research supervisor Dr. Bryan Roche (details below).

Dylan Colbert (the main researcher) is currently being supervised by a qualified educational psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing that these tests are administered appropriately and that data is processed and stored correctly and in accordance with normal data protection procedures. All IQ scores will be associated with user names using an encryption technique so that confidentiality is assured.

It is advised that you do not volunteer your child for the study if he has at any point attended a school of special education outside of the mainstream school system, or experiences any intellectual problems that you know or feel constitute an intellectual disability. Additionally, any child who is currently waiting for an educational assessment of his intellectual ability is advised not to take part in the study.

All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only in any subsequent publication or conference presentation. Data will only be discussed in terms of the group, no individual participant will be singled out and analysed. The data collected will be used only by the researchers. This data will be available to each participant's parents/guardians. No personal data will be recorded except your child's name and date of birth, and these are recorded solely so that the recorded IQ scores can be returned to each child's parents. Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. Declining to participate in this research, or withdrawing from it before it is completed will in no way affect your child's education or access to normal teaching services. You can contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if you have any concerns or queries.

The study has been approved by the school's Board of Management and the Principal is happy for it to go ahead for those parents who give their consent. The study has also been approved by the Research Ethics Committee of Maynooth University and it adheres to the University's Child Protection Policy. Finally, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at DYLAN.COLBERT.2011@nuim.ie.
Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@nuim.ie or by telephone at (01) 7086026.
Dr. Sarah Cassidy can be reached at soconnor00@hotmail.com
If during your participation in this study you feel the information and guidelines that you were given have been neglected or disregarded in any way, or if you are unhappy about the process, please contact the Secretary of the Maynooth University Ethics Committee at esearch.ethics@nuim.ie or +353 (0)1 708 6019. Please be assured that your concerns will be dealt with in a sensitive manner.

Appendix G

Experiment 3: Consent Form for Parent or Guardian

In agreeing to allow my son to participate in the research project I understand the following:

- Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator, psychology graduate and former Drimnagh Castle student currently gathering data for his PhD at Maynooth University.

- The purpose of this psychological research is to examine the effectiveness of a form of "brain training" for increasing intellectual ability or IQ. Each student will be randomly allocated to one of two groups, with the first group receiving this training from September to December, and the second group training from January to March/April. All participants will receive the same training, for the same period of time.

- I understand that my son will be asked to complete a standard full scale IQ test in September 2016, Dec/Jan 2016/17 and April/May 2017. This IQ test is known as the WISC and has been specifically designed for use with children. This assessment will take place in Drimnagh Castle CBS and requires around 60-90 minutes to complete. The wishes and comfort of the child will be treated with the utmost sensitivity during the assessment and every session will be overseen by a teacher from Drimnagh Castle primary school. Each child will be free to take breaks whenever he wishes. He will also complete a short assessment of his "relational skills" before and after training which will take the form of a test for logical reasoning, not unlike an algebra test (around 10 minutes).

- My son's training will take place online, twice weekly in Drimnagh Castle's computer room and once a week as part of his homework. Training sessions will last around 30 minutes

- All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This data will be available to each participant and his parents/guardians should they request it.

- The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and the Ethics Committee of Maynooth University. The study design adheres to Maynooth University's Child Protection Policy.

- I understand that I may withdraw my son from the study at any stage even after giving my consent. My son may also leave the study at any time. I may also withdraw his data at the conclusion of my participation if I still have concerns.

- I understand that I may contact Dr. Bryan Roche at Bryan.T.Roche@nuim.ie of Maynooth University as supervisor of this research if I have any concerns and can access a private clinical consultation with educational psychologist Dr. Sarah Cassidy (private practice and Maynooth University) if I have concerns about my son's IQ score.

- I understand that increases in IQ are not guaranteed as a result of participating in this study.

- I understand that the study is experimental and not clinical in nature and that my son will not receive a full psychological report although I will receive his IQ scores at the end of the study.

- I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed.

- I acknowledge that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Signed in duplicate:

____

Parent/Guardian

____

Researcher

____

Date

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

Appendix H

Experiment 4: Information Sheet for Participants

Drimnagh Castle has agreed to take part in this study, which is trying to see if a new type of brain training can improve intelligence. A version of this brain training program that is being used is actually available online where it is called SMART training.

This program has been created by a group of psychologists – people who study how people think and behave. Psychologists measure intelligence by using a test known as an IQ test. You will take three IQ tests as part of this study, to check the effect the brain training has. Each IQ test takes around 30 minutes to complete.

An IQ test measures how good you are at solving different types of problems – for example, maths problems and word problems. For example, we will ask you what certain words mean or to make some designs using blocks. Your score will be kept totally private but we will pass on the IQ test results to your parents at the end of the study and they can choose what to do with them.

Psychologists have discovered that really intelligent people are very good at understanding the relationships between things. For example, they can easily figure out if two things are the same or the opposite, or they can work out if one thing is bigger or smaller than another – even when it may not be so obvious.

SMART training teaches people how to see these relationships more easily.

The training is very simple – all you have to do is solve some mental puzzles online. As you train, these puzzles will get harder and will help you to understand those important relationships. The training is a bit like some brain training games you might have played on the Nintendo DS or on a tablet. You will be asked to train twice a week in school, each time for around 45 minutes.

You are allowed to stop taking part at any time if you would rather not do the training or take the IQ tests.

The study will be run by Dylan Colbert from Maynooth University. Mr. Colbert was once a student of Drimnagh Castle.

Appendix I

Experiment 4: Information Sheet for Parents/Guardians

The current research project is designed to test the effectiveness of a newly developed technique for raising intellectual ability. The particular method used in this study is based on a theory known as Relational Frame Theory and is called SMARTr training (Strengthening Mental Abilities with Relational Training: Remedial).

SMARTr training is based on the finding that most of what psychologists consider intelligent behaviour involves relating things to each other in a variety of ways (i.e. seeing the connections and links between things). Intelligent people have a good understanding of simple concepts such as before/after, more/less, opposite/same, here/there and so on. These are called "relational skills". When we teach these skills, intelligence, as measured by IQ tests, appears to rise. However, more research is needed to confirm that this is the case.

The tool we are using to train "relational skills" in this study is a modification of a program that has already been developed in previous research at Maynooth University (SMART). The main findings of that research have been published, and a web-based tool developed within Maynooth University has also now been made publicly available. Basically, we are trying to assess how effective this training can be. While early results have been promising, there is no absolute guarantee that volunteers who undergo the online training will experience increases in their IQ following participation. Dylan Colbert recently completed a year long study of the SMART program with a group of Drimnagh Castle's 4th class students. Following the results of this study, the SMARTr program has been developed to allow a wider range of students to access the potential benefits of relational skills training.

This research is being conducted as part of PhD. level research by Mr. Dylan Colbert under the supervision of Dr. Bryan Roche of Maynooth University. Dylan Colbert has been fully Garda vetted and is a former student of Drimnagh Castle CBS.

SMARTr training involves a system of online puzzles or exercises on a personal computer or other internet connected device. During the training period, your son will be asked to complete two training sessions per week during school hours. Each one of these sessions will take around 45 minutes to complete. The training usually requires approximately three months completing and involves a quick (10 minutes) relational skills assessment at the outset, which is then repeated at the end of the training. All of this is automated and will be delivered entirely online. While your son will be taken out of class for these sessions, along with the other children participating in the study, the skills which will be trained in these sessions are considered integral in establishing an improved ability to perform scholastically.

Training consists of solving a number of logical puzzles, followed by feedback from the computer in some cases but not in others. For example, users may be asked;

"If A is more than B and C is less than B, is A more than C?"

Users indicate their answer in all cases by clicking on the words "Yes" or "No" on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer, which delivers feedback to help the user improve their accuracy at these types of tasks. This technique allows us to train the relational abilities that are so important to intelligence.

As part of this study, Drimnagh Castle teaching staff have indentified your son as being suitable for participation in this study . If you feel that participation may in any way interfere with your son's classroom and/or learning support activities, we recommend that you decline to participate. All volunteers will be given the online "brain training" as part of the study. All participants will receive the same training, for the same period of time. The researcher, his supervisor and your son's teacher will be able to see the frequency of logins by the user as well as their progress.

As part of this research it is also necessary to administer a shortened intelligence (IQ) test three times; in October/November 2016, halfway through the training program then finally upon completion of training in approximately three months time.

The test used will be the WASI IQ test, a test designed for both children and adults. The test will take approximately 30 minutes, and the participant will be able to take breaks whenever needed. These IQ scores will NOT be delivered to your child but will be made available to you at the end of the study in a confidential letter delivered via the school. That document will explain what IQ is and how to interpret the IQ scores recorded for your child throughout the study. Participants will also be asked to complete two short "relational skills" assessments after each IQ test, which take approximately 10-15 minutes to complete.

A detailed psychological report on each child will not be provided because IQ is being measured in this study for research purposes only (i.e., not for diagnostic purposes). If users have any concerns about their scores, however, they will be referred to the research team's educational Psychologist Dr. Sarah Cassidy free of charge. She will address any concerns participants have about their IQ score, and will help them to understand the meaning of their IQ test performance. They may also contact the research supervisor Dr. Bryan Roche (details below).

Dylan Colbert (the main researcher) is currently being supervised by a qualified educational psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing that these tests are administered appropriately and that data is processed and stored correctly and in accordance with normal data protection procedures. All IQ scores will be associated with user names using an encryption technique so that confidentiality is assured.

All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only in any subsequent publication or conference presentation. Data will only be discussed in terms of the group, no individual participant will be singled out and analysed. The data collected will be used only by the researchers. This data will be available to each participant's parents/guardians. No personal data will be recorded except your child's name and date of birth, and these are recorded solely so that the recorded IQ scores can be returned to each child's parents.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns Declining to participate in this research or withdrawing from it before it is completed will in no way affect your child's education or access to normal teaching services. You can contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if you have any concerns or queries.

The study has been approved by the school's Board of Management and the Principal is happy for it to go ahead for those parents who give their consent. The study has also been approved by the Research Ethics Committee of Maynooth University and it adheres to the University's Child Protection Policy. Finally, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at DYLAN.COLBERT.2011@nuim.ie.
Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@nuim.ie or by telephone at (01) 7086026.
Dr. Sarah Cassidy can be reached at soconnor00@hotmail.com

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

Appendix J

Experiment 4: Consent Form for Parent or Guardian

In agreeing to allow my son to participate in the research project I understand the following:

- Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator, psychology graduate and former Drimnagh Castle student currently gathering data for his PhD at Maynooth University. In the last school year, Mr. Colbert conducted a similar study using one of the school's 4th class groups.
- The purpose of this psychological research is to examine the effectiveness of a form of "brain training" for increasing intellectual ability or IQ.
- I understand that my son will be asked to complete a short-form IQ test in October/November 2016, halfway through the training program and then following the conclusion of the study in approximately 3 months time. This IQ test is known as the WASI and has been specifically designed for use with both adults and children. This assessment will take place in Drimnagh Castle CBS and requires around 30 minutes to complete. The wishes and comfort of the child will be treated with the utmost sensitivity during the assessment and every session will be overseen by a teacher from Drimnagh Castle CBS. He will also complete a short assessment of his "relational skills" before and after training which will take the form of a test for logical reasoning, not unlike an algebra test (around 20 minutes).
- My son's training will take place online, twice weekly in one of Drimnagh Castle's classrooms Training sessions will last around 45 minutes.
- All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This data will be available to each participant and his parents/guardians should they request it.
- The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and the Ethics Committee of Maynooth University. The study design adheres to Maynooth University's Child Protection Policy.
- I understand that I may withdraw my son from the study at any stage even after giving my consent. My son may also leave the study at any time. I may also withdraw his data at the conclusion of my participation if I still have concerns.
- I understand that I may contact Dr. Bryan Roche at Bryan.T.Roche@nuim.ie of Maynooth University as supervisor of this research if I have any concerns and can access a private clinical consultation with educational psychologist Dr. Sarah Cassidy (private practice and Maynooth University) if I have concerns about my son's IQ score.
  - I understand that increases in IQ are not guaranteed as a result of participating in this study. I understand that the study is experimental and not clinical or therapeutic in nature and that my son will not receive a full psychological report although I will receive his IQ scores at the end of the study.
- I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed.
- I acknowledge that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Signed in duplicate:

________________________________ Parent/Guardian

________________________________ Researcher

________________________________ Date

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

Appendix K

Experiment 5: Information Sheet for Participants

Summerhill College has agreed to take part in this study, which is trying to see if a new type of brain training can improve intelligence and academic performance. A version of this brain training program that is being used is actually available online where it is called SMART training.

This program has been created by a group of psychologists – people who study how people think and behave. Psychologists measure intelligence by using a test known as an IQ test. A small number of students will be selected randomly to complete three IQ tests throughout the year. An IQ test measures how good you are at solving different types of problems – for example, maths problems and word problems. For example, we will ask you what certain words mean or to make some designs using blocks.

Every student will be asked to take assessments of your language skills, maths skills, problem solving and reading at three times throughout the year. These tests are used in schools across the country and are administered on a computer. All of your scores will be kept totally private but we will pass on your test results to your parents at the end of the study and they can choose what to do with them.

Psychologists have discovered that really intelligent people are very good at understanding the relationships between things. For example, they can easily figure out if two things are the same or the opposite, or they can work out if one thing is bigger or smaller than another – even when it may not be so obvious.

SMART training teaches people how to see these relationships more easily.

The training is very simple – all you have to do is solve some mental puzzles online. As you train, these puzzles will get harder and will help you to understand those important relationships. The training is a bit like some problem-solving you might have played on the Nintendo DS or on a tablet. You will be asked to train twice a week in school, each time for around 30 minutes.

You are allowed to stop taking part at any time if you would rather not do the training or take the tests.

The study will be run by Dylan Colbert, a postgraduate studying conducting research to complete his PhD. in Maynooth University.

Appendix L

Experiment 5: Information Sheet for Parents & Guardians

The current research project is designed to test the effectiveness of a newly developed technique for raising intellectual and/or scholastic ability. The particular method used in this study is based on a theory known as Relational Frame Theory and is called SMART training (Strengthening Mental Abilities with Relational Training).

SMART training is based on the finding that most of what psychologists consider intelligent behaviour involves relating things to each other in a variety of ways (i.e. seeing the connections and links between things). Intelligent people have a good understanding of simple concepts such as before/after, more/less, opposite/same, here/there and so on. These are called “relational skills”. When we teach these skills, intelligence, as measured by IQ tests, appears to rise. However, more research is needed to confirm that this is the case.

The tool we are using to train “relational skills” in this study has already been developed in previous research at Maynooth University and has been implemented in Summerhill College in past academic years. The main findings of that research have been published, and a web-based tool developed within Maynooth University has also now been made publicly available. Basically, we are trying to assess how effective this training can be. While early results have been hugely promising, there is no absolute guarantee that volunteers who undergo the online training will experience increases in their IQ following participation. This research is being conducted as part of PhD. level research by Mr. Dylan Colbert under the supervision of Dr. Bryan Roche of NUI Maynooth. Dylan Colbert has been fully Garda vetted and is a former student of Drimnagh Castle CBS.

SMART training involves a system of online puzzles or exercises on a personal computer or other internet connected device. During the training period, your son will be asked to complete 2 of training sessions per week. Each one of these sessions will take around 45 minutes to complete. The training usually requires approximately three months to complete and involves a quick (10 minutes) intellectual assessment at the outset, which is then repeated at the end of the training. All of this is automated and will be delivered entirely online. While your son will be taken out of class for these sessions, along with the other students participating in the study, the skills which will be trained in these sessions are considered integral in establishing an improved ability to perform scholastically. As such, undergoing this training is expected to have long-term benefits for your child’s school performance, although an increase in general intelligence (IQ) cannot be guaranteed.

Training consists of solving a number of logical puzzles, followed by feedback from the computer in some cases but not in others. For example, users may be asked;

“If A is more than B and C is less than B, is A more than C?”

Users indicate their answer in all cases by clicking on the words “Yes” or “No” on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer, which delivers feedback to help the user improve their accuracy at these types of tasks. This technique allows us to train the relational abilities that are so important to intelligence.

All volunteers will be given the online “brain training” as part of the study. Half of the volunteers will begin training in September and should finish by December. The other half of the volunteers will train during the second training period (January to March/April). All participants will receive the same training, for the same period of time. The researcher and his supervisor will be able to see the frequency of logins by the user as well as their progress.
As part of this research it is also necessary to administer a short IQ test three times; in September/October 2017, December 2017/January 2018 and then finally in April/May 2018. The test used will be the WASi IQ test, a test specially designed for children and adults. The test will take approximately 30 minutes, and the participant will be able to take breaks whenever needed.

These IQ scores will NOT be delivered to your child but will be made available to all parents at the end of the study in a confidential letter delivered via the school. That document will explain what IQ is and how to interpret the IQ scores recorded for your child throughout the study. In addition, your son will be asked to complete three standardised assessments of his academic ability, in September/October 2017, December 2017/January 2018 and then finally in April/May 2018. These assessments will consist of the Drumcondra Primary Reading and Mathematics Tests, and will require approximately 50 minutes to complete. Your son will also be administered a school-administered group reading assessment during these testing periods. If, for any reason, you would prefer for your son not to complete one or more of these assessments, you will be able to indicate this on the consent form attached.

A detailed psychological report on each child will not be provided because IQ and scholastic ability are being measured in this study for research purposes only (i.e., not for diagnostic purposes). If users have any concerns about their scores, however, they will be referred to the research team's educational Psychologist Dr. Sarah Cassidy free of charge. She will address any concerns participants have about their IQ score, and will help them to understand the meaning of their IQ test performance. They may also contact the research supervisor Dr. Bryan Roche (details below).

Dylan Colbert (the main researcher) is currently being supervised by a qualified educational psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing that these tests are administered appropriately and that data is processed and stored correctly and in accordance with normal data protection procedures. All IQ and scholastic ability scores will be associated with user names using an encryption technique so that confidentiality is assured.

All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This data will be available to each participant's parents/guardians. No personal data will be recorded except your child's name and date of birth, and these are recorded solely so that the recorded IQ scores can be returned to each child's parents.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. Declining to participate in this research or withdrawing from it before it is completed will in no way affect your child's education or access to normal teaching services.

You can contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if you have any concerns or queries.

The study has been approved by the school's Board of Management and the Principal is happy for it to go ahead for those parents who give their consent. The study has also been approved by the Research Ethics Committee of Maynooth University and it adheres to their Child Protection Policy.
Finally, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at DYLAN.COLBERT.2011@nuim.ie. Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@mu.ie or by telephone at (01) 7086026. Dr. Sarah Cassidy can be reached at soconnor00@hotmail.com

Appendix M

Experiment 5: Consent Form for Parent or Guardian

In agreeing to allow my son to participate in the research project I understand the following:

- Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator, a psychology graduate student currently gathering data for his PhD at Maynooth University.

- The purpose of this psychological research is to examine the effectiveness of SMART, a form of "brain training" for increasing intellectual and/or scholastic ability. Each student will be randomly allocated to one of two groups, with the first group receiving this training from September to December, and the second group training from January to March/April. All participants will receive the same training for the same period of time.

- I understand that my son will be asked to complete a short-form IQ test in September/October 2017, Dec/Jan 2017/18 and April/May 2018. This IQ test is known as the Wechsler Abbreviated Scale of Intelligence (WASi) and has been specifically designed for use with both children and adults. This assessment will take place during school hours and on school premises and requires around 30 minutes to complete. The wishes and comfort of each student will be treated with the utmost sensitivity during the assessment and every session will be overseen by a teacher from Summerhill College. Each student will be free to take breaks whenever he wishes. He will also complete a short assessment of his "relational skills" before and after training which will take the form of a test for logical reasoning, not unlike an algebra test (around 10 minutes).

- I understand that my son will also be asked to complete standardised assessments of scholastic ability in September/October 2017, Dec/Jan 2017/18 and April/May 2018. These assessments will consist of the Drumcondra Primary Reading and Mathematics Tests, which will require approximately 50 minutes to complete in total.

- My son's training will take place online during school hours. Training sessions will last around 45 minutes.

- All students participating in this study will remain anonymous and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This data will be available to each participant and his parents/guardians should they request it.

- The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and the Ethics Committee of Maynooth University. This study design adheres to Maynooth University's Child Protection Policy.

- I understand that I may withdraw my son from the study at any stage even after giving my consent. My son may also leave the study at any time. I may also withdraw his data at the conclusion of my participation if I still have concerns.

- I understand that I may contact Dr. Bryan Roche at Bryan.T.Roche@mu.ie of Maynooth University as supervisor of this research if I have any concerns and can access a private clinical consultation with educational psychologist Dr. Sarah Cassidy (private practice and Maynooth University) if I have concerns about my son's IQ score.

- I understand that increases in IQ or scholastic ability are not guaranteed as a result of participating in this study.

- I understand that the study is experimental and not clinical in nature and that my son will not receive a full psychological report although I will receive his IQ scores at the end of the study.

- I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed.

- I acknowledge that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

- I hereby give my consent for the following assessments to be administered to my son as part of this study (tick where appropriate):

  - ○ Wechsler Abbreviated Scale of Intelligence
  - ○ Drumcondra Primary Reading Scale
  - ○ Drumcondra Primary Mathematics Scale
  - ○ All of the above

Signed in duplicate:

______________________________ Parent/Guardian

______________________________ Researcher

______________________________ Date

Appendix N

Experiment 6: Information Sheet for Potential Research Participants

The current research project is designed to test the effectiveness of a newly developed technique for assessing intellectual ability. The particular method of interest in this study is based on a psychological theory known as Relational Frame Theory and this new test is called the Multiple Relational Assessment Test (MRAT).

This research is being conducted as part of a project led Mr. Dylan Colbert, under the supervision of Dr. Bryan Roche of Maynooth University.The MRAT is a test of your ability to relate things to each other in a variety of ways. Research suggests that intelligent reasoning involves the use of "relational" concepts such as before, after, more, less, opposite, different, same, here, there, and so on. This study is investigating how important these skills are to overall intellectual performance. This will be done by assessing your "relational skills" ability, using the MRAT, and then assessing your IQ, using a widely used IQ test called the Wechsler Adult Intelligence Scale.

The tool we are using to assess your "relational skills" in this study (the MRAT) has been developed in previous research at Maynooth University and other academic institutions. The MRAT assessment consists of solving a number of logical puzzles. For example, users may be asked "If A is more than B and C is less than B, is A more than C?". Users indicate their answer in all cases by clicking on the words "Yes" or "No" on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer.

As part of this research, it is also necessary to administer a brief intelligence (IQ) test. The standard test used will be the WASi IQ test. This test consists of a wide range of verbal, mathematical and other reasoning tasks and will take approximately 30 minutes to complete. The IQ assessments are being overseen by a qualified psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing that these tests are administered appropriately. The researcher supervisor, Dr. Bryan Roche will ensure that data is processed and stored correctly and in accordance with current data protection procedures. All IQ scores will be associated with your name using an encryption technique so that confidentiality is assured. Your name will be linked to your IQ score by a code that will be stored separately to your IQ score record.

While the IQ scores from the assessment will be provided to the user at the end of their participation, a detailed psychological report will not be provided because IQ is being measured for research purposes only. If users have any concerns about their scores, however, they will be referred to the research team's educational Psychologist Dr. Sarah Cassidy free of charge. She will address any concerns participants have about their IQ score, and will help them to understand the meaning of their IQ test performance.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. If at any point you have attended a school of special education outside of the mainstream school system due to learning difficulties, or if you suffer with any intellectual problems that you know or feel constitute an intellectual disability then you may not be of use to us in this study and you should not volunteer to participate.

Before you volunteer to participate, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at dylan.colbert.2011@mumail.ie. Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@nuim.ie or by telephone at (01) 7086026. Dr. Sarah Cassidy can be reached at soconnor00@hotmail.com

Appendix O

Experiment 6: Consent Form for Participants

In agreeing to participate in the research project I understand the following:

• Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator and is a psychology graduate, currently gathering data for his postgraduate studies at Maynooth University.

• The purpose of this psychological research is to examine the effectiveness of a new form of assessment for measuring intellectual ability or "IQ".

• I understand that I will be asked to complete a full-scale IQ test and brief new IQ assessment called the MRAT.

• I understand that I will not be given access to my IQ test results until after I have completed my participation in the study.

I understand that the study involves two assessment sessions, the first of approximately 90 minutes, and the second 15 minutes.

• All persons participating in this study will remain anonymous from each other and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This study data will be available to each participant should they request it.

• The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and in line with the Research Ethics guidelines of Maynooth University.

• I understand that I may withdraw from the study at any stage even after giving my consent. I may also withdraw my data at the conclusion of my participation if I still have concerns.

• I understand that I may contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if I have any concerns and can access a private clinical consultation with Dr. Sarah Cassidy (private practice and Maynooth University) if I have concerns about my IQ score.

• I have been informed as to the general nature of the study. I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed.

• I acknowledge that in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of an 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.

• I am over 18 years of age.

Signed in duplicate:

______________________________ Participant

______________________________ Researcher

______________________________ Date

Appendix P

Experiment 7: Information Sheet for Potential Research Participants

The current research project is designed to test the effectiveness of a newly developed technique for assessing intellectual ability. The particular method of interest in this study is based on a psychological theory known as Relational Frame Theory and this new test is called the Multiple Relational Assessment Test (MRAT).

This research is being conducted as part of a project led Mr. Dylan Colbert, under the supervision of Dr. Bryan Roche of Maynooth University. The RAI+ is a test of your ability to relate things to each other in a variety of ways. Research suggests that intelligent reasoning involves the use of "relational" concepts such as before, after, more, less, opposite, different, same, here, there, and so on. This study is investigating how important these skills are to overall intellectual performance. This will be done by assessing your "relational skills" ability, using the RAI+, and then assessing your IQ, using a widely used IQ test called the Wechsler Adult Intelligence Scale.

The tool we are using to assess your "relational skills" in this study (the RAI+) has been developed in previous research at Maynooth University and other academic institutions. The RAI+ assessment consists of solving a number of logical puzzles. For example, users may be asked "If A is more than B and C is less than B, is A more than C?". Users indicate their answer in all cases by clicking on the words "Yes" or "No" on the computer screen. Puzzles get increasingly more difficult but progression is guided by the computer.

As part of this research, it is also necessary to administer a brief intelligence (IQ) test. The standard test used will be the WASi IQ test. This test consists of a wide range of verbal, mathematical and other reasoning tasks and will take approximately 30 minutes to complete. The IQ assessments are being overseen by a qualified psychologist (Dr. Sarah Cassidy, Maynooth University) who will also take responsibility for seeing that these tests are administered appropriately. The researcher supervisor, Dr. Bryan Roche will ensure that data is processed and stored correctly and in accordance with current data protection procedures. All IQ scores will be associated with your name using an encryption technique so that confidentiality is assured. Your name will be linked to your IQ score by a code that will be stored separately to your IQ score record.

While the IQ scores from the assessment will be provided to the user at the end of their participation, a detailed psychological report will not be provided because IQ is being measured for research purposes only. If users have any concerns about their scores, however, they will be referred to the research team's educational Psychologist Dr. Sarah Cassidy free of charge. She will address any concerns participants have about their IQ score, and will help them to understand the meaning of their IQ test performance.

Volunteers can withdraw from the study at any stage even after giving consent, and may also withdraw their data at the conclusion of the study if they still have concerns. If at any point you have attended a school of special education outside of the mainstream school system due to learning difficulties, or if you suffer with any intellectual problems that you know or feel constitute an intellectual disability then you may not be of use to us in this study and you should not volunteer to participate.

Before you volunteer to participate, you should be made aware that in some circumstances, 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 law to ensure that confidentiality is maintained to the greatest possible extent.

Mr. Dylan Colbert can be reached by email at dylan.colbert.2011@mumail.ie. Dr. Bryan Roche can be reached by email at the Department of Psychology, NUI Maynooth at Bryan.T.Roche@nuim.ie or by telephone at (01) 7086026. Dr. Sarah Cassidy can be reached at soconnor00@hotmail.com

Appendix Q

Experiment 7: Consent Form for Participants

In agreeing to participate in the research project I understand the following:

• Mr. Dylan Colbert and Dr. Bryan Roche, of the Department of Psychology, Maynooth University are conducting this research. Mr. Colbert is the principal investigator and is a psychology graduate, currently gathering data for his postgraduate studies at Maynooth University.

• The purpose of this psychological research is to examine the effectiveness of a new form of assessment for measuring intellectual ability or "IQ".

• I understand that I will be asked to complete a full-scale IQ test and brief new IQ assessment called the RAI+.

• I understand that I will not be given access to my IQ test results until after I have completed my participation in the study.

I understand that the study involves two assessment sessions, the first of approximately 90 minutes, and the second 15 minutes.

• All persons participating in this study will remain anonymous from each other and will not be referred to by name in any publication or document. The data will remain confidential at all times and will be referred to by code names only. The data collected will be used only by the researchers. This study data will be available to each participant should they request it.

• The researchers will conduct all parts of this study in line with the ethical code of conduct laid down by the Psychological Society of Ireland and in line with the Research Ethics guidelines of Maynooth University.

• I understand that I may withdraw from the study at any stage even after giving my consent. I may also withdraw my data at the conclusion of my participation if I still have concerns.

• I understand that I may contact Dr. Bryan Roche at Bryan.t.roche@nuim.ie of Maynooth University as supervisor of this research if I have any concerns and can access a private clinical consultation with Dr. Sarah Cassidy (private practice and Maynooth University) if I have concerns about my IQ score.

• I have been informed as to the general nature of the study. I have read the research information sheet. I understand that at the conclusion of my participation, any further questions or concerns I have will be fully addressed.

• I acknowledge that in some circumstances, confidentiality of research data and records may be overridden by courts in the event of litigation or in the course of an 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.

• I am over 18 years of age.

Signed in duplicate:

_______________________________ Participant

_______________________________ Researcher

_______________________________ Date

Appendix R

Experiment 7: Table displaying format of RAI+

The format and sequence of all 67 RAI+ trials.

Block	No	Premise 1			Premise 2			Premise 3			Question		
1	1	a	same as	b							a	opposite to	b
	2	a	same as	b	b	same as	c				a	same as	c
	3	a	same as	b	b	same as	c				a	same as	c
	4	a	opposite to	b	b	opposite to	c				c	same as	b
	5	a	same as	b	b	same as	c				c	same as	a
	6	a	same as	b	c	same as	a				b	opposite to	c
	7	a	opposite to	b	b	opposite to	c				c	same as	a
	8	a	opposite to	b	c	opposite to	a				a	same as	c
	9	a	opposite to	b	c	opposite to	a				b	same as	a
	10	a	opposite to	b	b	same as	c				c	opposite to	a
	11	a	opposite to	b	c	same as	a				a	opposite to	c
	12	a	same as	b	b	same as	c	c	same as	d	d	opposite to	b
	13	a	opposite to	b	b	opposite to	c	c	opposite to	d	d	opposite to	a
	14	a	same as	b	b	opposite to	c	c	opposite to	d	d	same as	a
	15	a	opposite to	b	b	opposite to	c	c	same as	d	b	opposite to	c
2	16	a	different to	b							b	same as	a
	17	a	same as	b	b	same as	c				c	same as	a
	18	a	different to	b	b	different to	c				b	same as	c
	19	a	different to	b	b	different to	c				a	same as	c
	20	a	same as	b	c	same as	a				b	same as	c
	21	a	same as	b	c	same as	a				b	same as	c
	22	a	different to	b	b	different to	c				a	same as	b
	23	a	different to	b	c	different to	a				b	different to	a
	24	a	different to	b	b	same as	c				c	same as	a
	25	a	same as	b	b	different to	c				a	different to	b
	26	a	same as	b	b	same as	c	c	same as	d	b	same as	c
	27	a	different to	b	b	different to	c	c	different to	d	b	different to	a
	28	a	same as	b	b	same as	c	c	different to	d	b	same as	c
	29	a	different to	b	b	same as	c	c	same as	d	a	different to	c
3	30	a	more than	b							b	more than	a
	31	a	more than	b	b	more than	c				a	more than	b
	32	a	more than	b	b	more than	c				b	less than	c
	33	a	less than	b	b	less than	c				a	less than	c
	34	a	less than	b	b	less than	c				b	more than	c
	35	a	more than	b	c	more than	a				a	more than	c
	36	a	more than	b	c	more than	a				c	more than	b
	37	a	less than	b	c	less than	A				c	more than	a
	38	a	more than	b	b	more than	c	c	more than	d	a	more than	c
	39	a	less than	b	b	less than	c	c	less than	d	c	less than	b
	40	a	more than	b	c	more than	b	d	more than	c	d	less than	a

	41	a	more than	b	c	more than	a	d	more than	c	d	less than	b
	42	a	less than	b	c	less than	a	d	less than	b	d	less than	b
4	43	a	after	b							a	after	b
	44	a	before	b	b	before	c				c	before	b
	45	a	before	b	b	before	c				c	before	a
	46	a	after	b	b	after	c				b	after	a
	47	a	after	b	b	after	c				c	before	a
	48	a	before	b	c	before	a				c	after	b
	49	a	after	b	c	after	a				c	after	a
	50	a	after	b	c	after	a				b	before	c
	51	a	before	b	b	before	c	c	before	d	d	before	b
	52	a	after	b	b	after	c	c	after	d	b	before	d
	53	a	before	b	c	before	b	d	before	c	b	after	d
	54	a	after	b	c	after	a	d	after	c	a	before	d
	55	a	after	b	c	after	a	d	after	c	b	before	d
5	56	a	same as	b	c	same as	d				a/b	same as	c/d
	57	a	opposite to	b	c	opposite to	d				b/a	different to	c/d
	58	a	before	b	c	before	d				a/b	same as	c/d
	59	a	before	b	c	before	d				b/a	different to	c/d
	60	a	same as	b	c	opposite to	d				a/b	same as	c/d
	61	a	opposite to	b	c	same as	d				b/a	different to	c/d
	62	a	after	b	c	before	d				a/b	same as	c/d
	63	a	after	b	c	before	d				b/a	different to	c/d
	64	a	more than	b	c	more than	d				a/b	same as	c/d
	65	a	less than	b	c	less than	d				b/a	same as	c/d
	66	a	more than	b	c	less than	d				a/b	same as	c/d
	67	a	less than	b	c	more than	d				a/b	same as	c/d