**A Relational Frame Skills Training Intervention to Increase General**

**Intelligence and Scholastic Aptitude**

Sarah Cassidy[1], Bryan Roche[1], Dylan Colbert[1], Ian Stewart[2] & Ian M. Grey[3]

[1]Maynooth University, Ireland

[2]National University of Ireland, Galway

[3]Trinity College, Dublin

Running Head: Relational Skills Training

**Abstract**

This study aimed to replicate and extend the pilot findings of Cassidy et al. (2011) which found that teaching children to derive various relations among stimuli leads to increases in the full scale IQ scores of both typically developing children and those with educational and learning difficulties. In Experiment 1, fifteen 11-12 year old children were exposed over several months to an intensive training intervention to improve their understanding of the relations Same, Opposite and More and Less. Significant increases in full scale IQ of around one standard deviation were recorded for each child. In Experiment 2, the same intervention was delivered to thirty 15-17 year old children. Significant increases in verbal and numerical reasoning were recorded for almost every child. These findings corroborate the idea that relational skills may underlie many forms of general cognitive ability.

Keywords: Relational frame theory, Relational skills, Derived relational responding, Intelligence, Educational Aptitude.

## 1. Introduction

The idea that intelligence may be a malleable trait has a well-documented and contentious history. Arguably, the evidence base is shifting in favor of the idea that intelligence is not a stable trait, with leading researchers in the field arguing that an increasing role can be assigned to the environment in determining intelligence levels (e.g., Nisbett et al., 2012). Evidence for this perspective comes from educational, cognitive, neuroscientific and, most recently, behavior-analytic sources (the focus of the current study). For instance, educational studies have shown that the Intelligence Quotient (IQ) scores of children who miss a year of school drop several points and that such effects can even be measured over the summer months (see Ceci, 1991; Jencks et al., 1972). In one large-sample naturalistic and retrospective study (Brinch & Galloway, 2012), two extra years of compulsory schooling was associated with substantial IQ increases by age 19. Educational interventions for younger children, such as the Milwaukee Project (Garber, 1988), were reported as leading to an average IQ increase of 10 points over matched controls. Similarly, the Abecedarian pre-kindergarten program (Campbell, Ramey, Pungello, Sparling, & Miller- Johnson, 2002) was reported to have had an effect on IQ scores amounting to 4.5 points over matched controls by age 21. One study involving teaching reasoning skills to seventh-grade children improved IQ scores by 0.4 SD (Herrnstein, Nickerson, Sanchez, & Swets, 1986). More recently, a partial replication of that study reported intelligence gains across a range of measures (Sanz de Acedo Lizarraga, Ugarte, Iriarte, & Sanz de Acedo Baquedano, 2003). Several other studies have shown the impact of intensive intervention on scholastic performance (e.g., Campbell, Pungello,

Miller-Johnson, Burchinal, & Ramey, 2001; Nisbett, 2009; Schweinhart et al., 2005; Schweinhart & Weikart, 1980; Schweinhart & Weikart, 1993).

Research within the cognitive tradition has focused to a large extent on such features as working memory training and its impact on fluid intelligence, indexed using measures such as Raven's Progressive Matrices (RPM). For instance, Klingberg, Forssberg, and Westerberg (2002) reported improvements in RPM scores for both children with attention deficit/hyperactivity disorder (ADHD) and adults who had been provided with memory training. More recently, Mackey, Hill, Stone, and Bunge (2011) reported matrix reasoning improvements equating to 10 full scale IQ points in low SES children who had been exposed to memory training across a range of modalities. Similar effects have been reported for elderly populations in studies targeting memory training (Borella, Carretti, Riboldi, & De Beni, 2010) and executive functioning (Basak, Boot, Voss, & Kramer, 2008; Stephenson & Halpern, 2012).

Perhaps no single paradigm for increasing intelligence has received as much scientific and media attention as that involving extended dual n-back training, a task designed to target working memory. While not without its critics, research reported by Susan Jaeggi and colleagues indicates increases in fluid intelligence ($g_f$) as measured by RPM across short term and long-term (3-month) follow-up measurement periods (see Jäeggi, Buschkuehl, Jonides & Perrig, 2008; Jäeggi, Studer-Luethi, Buschkuehl, Su, Jonides & Perrig, 2010; Jaeggi, Buschkuehl, Jonides & Shah, 2011).

Evidence from the field of neuroscience also points to neural changes that might be expected to accompany improved brain functioning following interventions designed to enhance reasoning skills. For instance, Mackey, Miller-Singley and

Bunge (2013) reported that intensive reasoning training in the form of preparation for the Law School Admissions Test (LSAT) resulted in tighter coupling among regions in the lateral fronto-parietal network, as measured with resting-state fMRI (rs-fMRI). A demonstration of such plasticity is important because the regions affected by the training are associated with complex cognition (see also Mackey, Whitaker & Bunge, 2012).

Psychologists in the behavior-analytic tradition have always adopted a purely functional approach to understanding intelligence and general cognitive ability (e.g., Schlinger, 2003). Instead of viewing intelligence as an invariant trait (encapsulated by the concept of "g"), they see it as a malleable skill set, while viewing IQ measures as mere indices of the fluency of the skills involved (e.g., Cassidy, Roche & O'Hora, 2010; Cassidy, Roche & Hayes, 2011). The late O. Ivar Lovaas (1987) reported IQ gains up to 30 points (roughly two standard deviations) following a three-year applied behavior analysis (ABA) intervention for autism. Nearly half of the autistic children in that study were not noticeably intellectually different from normally functioning children after the three-year program (Reed, Osbourne & Corness, 2005). In an independent replication of the Lovaas study, Sallows and Graupner (2005) recorded similar significant IQ rises among autistic children. In a further study, Smith, Eikeseth, Klevstrang and Lovaas (1997) used an ABA treatment program to improve expressive speech and adaptive behavior among severely mentally retarded children with autistic features. They also measured IQs at follow-up. While behavioral problems diminished in both groups, children in the treatment condition displayed a higher mean IQ at follow-up and evinced more expressive speech than those in the comparison group.

A more recent behavioral approach to intellectual functioning is provided by *Relational Frame Theory* (RFT; Hayes, Barnes-Holmes, & Roche, 2001; Dymond & Roche, 2013) that attempts to codify a wide range of cognitive skills in terms of a smaller range of underlying, teachable skills. These are known as relational framing skills (or relational skills for short) and might be considered loosely as the functional counterpart of the more widely used concept of relational reasoning skills (e.g., Halford, Wilson & Philips, 2010).

RFT draws together several decades of research focused on a key repertoire referred to as *derived relational responding*, the most widely studied form of which is *stimulus equivalence*. The latter refers to the emergence of stimulus relations that are untrained but which are nevertheless predictable and controllable. A typical procedure for studying this behavioral phenomenon involves training participants to make stimulus selections from a pair (or more) of stimuli, in the presence of a sample stimulus (i.e., a conditional discrimination made in a matching-to-sample format). All stimuli are unrelated to each other along any physical continuum. For example, given sample stimulus A, selecting comparison stimulus B is reinforced (i.e., A-B). On other trials selecting comparison stimulus C given sample stimulus A is reinforced (i.e., A-C). For most verbally-able individuals, B-A, C-A (i.e., symmetry), B-C and C-B (i.e., combined symmetry and transitivity) relations emerge without further instruction or reinforcement, and if they do stimulus equivalence has been observed (Sidman, 1971).

Researchers have been fast to capitalize upon the obvious relevance of this and related phenomena to a wide range of cognitive skills, including language, reasoning, and problem solving (see Dymond & Roche, 2013). Galizio, Stewart and Pilgrim (2001), for example, described how derived relational responding can be used

as a new paradigm for understanding reasoning insofar as the emergence of stimulus equivalence classes is not unlike the process of category clustering. Galizio et al. argued that traditionally the study of categorization has been the province of cognitive psychology and psycho-linguistics (e.g., Margolis & Laurence, 1999) but that a behavior-analytic approach to categorization can complement the cognitive approach to understanding such phenomena as organizational processes in memory (e.g., Bousfield, 1953). Indeed, it is a widely held view that derived relational responding and language processes are in fact synonymous (see Dymond & Roche, 2013 for a book-length review).

RFT extends the analysis of derived relational responding to include relations other than equivalence, such as comparison (more than/less than), time (before/after), opposition and difference. All of the deriving skills relevant to these frames are thought to be learned through multiple exemplars delivered through extended social and educational interactions (see, e.g., Stewart, Tarbox, Roche & O'Hora, 2013) and each is characterized by a pattern of derivation (mutual and combinatorial entailment and transformation of response functions of each stimulus by the others) that confirms the emergence of a particular type of frame. The important point in the current context, however, is that a relatively narrow range of taught and derived stimulus relation types (i.e., frames) can be used to describe a very wide range of intellectual processes (e.g., problem solving, vocabulary, etc). For instance, vocabulary test items can be viewed as tests for taught or derived frames of coordination (e.g., "What does brave mean?"). Some IQ test items require definitively derived, rather than taught relational responding in accordance with a frame of coordination. As an example, the *Picture Concepts* subtest on the Wechsler Intelligence Scale for Children fourth edition, UK (WISC-IV[UK]), requires a child to

look at several rows of images, and choose one image from each row to form a cluster of images that have a meaningful relationship to each other. The stimuli may be novel in the given arrangement, but for children who have been exposed to a typical education, the skill of deriving relations among an arbitrary array of stimuli is not itself novel. Frames of opposition are also specifically probed for in vocabulary items asking the respondent to identify a word with the opposite meaning to a sample (e.g., the AH4). The reader is referred to Cassidy, Roche & O'Hora (2010) and Roche, Cassidy and Stewart (2013) for an extensive analysis of popular IQ tests along these lines.

The relational frame approach to intelligence is somewhat commensurate with several mainstream cognitive approaches to understanding intellectual skills. The most obvious of these is the concept of relational reasoning or knowledge. Specifically, relational knowledge is thought to integrate heuristic and analytic cognition and to be important for symbolic processes. As it happens, the regions of the brain activated by relational reasoning are in the prefrontal cortex, which further corroborates the view that relational reasoning is central to many higher cognitive processes (see Halford et. al., 2010). Indeed, behavior-analytic theorists agree with the cognitive position that mathematics can be characterized as a set of relational networks or concepts (Ninness et al., 2006; 2009; 2015; Marr, 2015).

It has not gone unnoticed by cognitive psychologists that if relational reasoning skills are teachable, that it should be possible to enhance general intelligence by enhancing relational reasoning (e.g., Mackey et al., 2013). A similar line of logic has been followed within RFT. More specifically, the development of framing appears to be correlated with the development of language, itself seen as a crucial aspect of intellectual development and ability (Smith, Smith, Taylor & Hobby,

2005). Framing has been shown to emerge at the same time as language (Lipkens, Hayes & Hayes, 1993) and to be absent in language disabled individuals (Barnes, McCullagh, & Keenan, 1990; Devaney, Hayes & Nelson, 1986). In addition, numerous empirical and conceptual research papers have presented evidence that the ability to derive relations is associated with, and possibly even underpins language ability (Moran, Stewart, McElwee & Ming, 2010, 2014; O'Connor, Rafferty, Barnes-Holmes & Barnes-Holmes, 2009).

In addition, several correlational studies have concluded that there are important overlaps between the ability to derive relations and intellectual skills established in the school setting. For example, O'Hora and colleagues showed correlations between college students' ability on a test requiring derivation of temporal, sameness and distinction relations and their performance on a number of different subscales of the Wechsler Adult Intelligence Scale (WAIS-III) including vocabulary, arithmetic, digit-symbol encoding, verbal comprehension, and perceptual organization (O'Hora, Pelaez and Barnes-Holmes, 2005; O'Hora, Pelaez, Barnes-Holmes, Rae, Robinson and Chaudhary, 2008). O'Toole and Barnes-Holmes (2009) measured the fluency of sameness, distinction and temporal (before/after) relational skills using a procedure known as the Implicit Relational Assessment Procedure. (IRAP; e.g., Barnes-Holmes, Barnes-Holmes, Stewart & Boles, 2010) and found significant moderate correlations between scores on the Kaufman Brief Intelligence Test (K-BIT) and relational skill fluency. Gore, Barnes-Holmes and Murphy (2010) reported significant correlations between perspective-taking relations (e.g., here and there) and Full Scale ($r = .43$), Verbal ($r = .45$) and Performance IQ ($r = .45$) as measured by the Wechsler Abbreviated Scale of Intelligence (WASI). Finally, most recently, a number of studies have demonstrated relationships between relational skill

9

and IQ in children with autism and other forms of intellectual disability (Dixon, Whiting, Rowsey, & Belisle, 2014; Moran, Stewart, McElwee & Ming, 2010, 2014).

Given these converging strands of research, behavioral psychologists have argued that increasing an individual's proficiency in relational framing should lead to improvements in a range of mental aptitudes relevant to education, as well as in general intelligence (Cassidy, Roche & O'Hora, 2010; Roche, Cassidy & Stewart, 2013). Relational framing skills are usually taught using a multiple exemplar training (MET) method. For example, we might teach a child that, given statements of the form *A is more than B*, and *B is more than C*, then it can be derived that C is less than A. Patterns of derivation such as this are taught across a very large number of exemplars, using many different stimulus sets, but always reinforcing the same contextually correct response pattern. Responses to relational probe questions (e.g., Is A more or less than C?) are usually followed by feedback, but testing blocks in which no feedback is given are also interspersed with training trials. In this way, a child is taught across multiple examples that the forms of the various stimuli are irrelevant to deriving novel relations between them; instead, it is the contextually determined relational pattern that determines the correct answer. According to RFT, it is as a result of MET in relational framing occurring naturalistically in everyday social and educational contexts that understanding of complex relations emerges, rather than vice versa, a conclusion supported by studies showing that MET can be used to train relational framing skills even before they naturally emerge through such interactions (e.g., Vitale, Barnes-Holmes, Barnes-Holmes & Campbell, 2008).

Only one study to date had tested an intervention to increase general intelligence using a relational frame training intervention. In Experiment 1 of Cassidy, Roche and Hayes (2011), four 10-12 year old children were given MET in

10

the relations *Same*, *Opposite* and *More/Less,* across several weekly sessions spanning several months. Baseline and post-intervention IQ scores were assessed using the Wechsler Intelligence Scale for Children (WISC-III[UK]). A further four matched participants received no training but also underwent the two IQ assessments. The training group showed significant improvements in full scale IQ, with no increase lower than one standard deviation. None of the control participant IQs rose by more than four points; far short of one standard deviation. The mean IQ of the training group rose from 105.5 to 132.75 points, while that of the controls actually fell from 106.5 to 104.25. In Experiment 2, a further eight 11-12 year old children with below average IQs were exposed to a slightly modified version of the training. Full scale IQ measured using the WISC-IV[UK], rose by at least one standard deviation for seven of the eight children and this increase was significant. An anecdotal four-year follow-up measure recently reported by Roche, Cassidy and Stewart (2013) found that for all seven of the eight participants revisited, IQ rises were maintained and were still significantly higher than the baseline measures taken four years previously.

Aside from requiring replication, there are several shortcomings of the foregoing exploratory research that should be addressed using more robust research designs. The first issue relates to the small sample sizes employed. The second relates to the need to measure the influence of relational frame training not just on IQ but on critically important outcomes such as scholastic ability. The third relates to the need for a single blind procedure to be employed whereby those administering the pre and post intervention assessments are unconnected with the study. These issues are all addressed in various ways in the current research.

The current study reports two separate experiments, which assessed the impact of a relational frame training intervention on general intelligence (Experiment

1) and scholastic ability (Experiment 2). Both experiments involved the administration of the computer-based training by school teachers in the school setting over several months.  Experiment 1 involved the use of the WISC-IV$^{UK}$ as both the baseline and follow-up dependent measure with a sample of fifteen 11-12 year old children. Experiment 2 involved a larger sample (n=30) of 15-17 year old children who were exposed to the same intervention but who received the Differential Aptitude Test as both the baseline and post-intervention measure.

## Experiment 1

## 2. Method

### 2.1 Participants

Participants consisted of an entire class cohort of one primary level school in Co. Meath, Ireland. As such, a convenience non-random sampling method was employed. All participants were native English speakers with IQ scores distributed as might be expected for a sample of this size taken from a mainstream school (range = 84-110, M = 97, SD = 10.21). The sample consisted of 9 males and 6 females ranging between the ages of 10 and 12 years (M = 11.1), none of whom had any general intellectual deficits (although three of the participants were diagnosed with Dyslexia and a further one was diagnosed with a Disruptive Behavior Disorder.  One of the children with Dyslexia also presented with a specific speech and language impairment).  Written consent for participation was provided by legal guardians in all cases and ongoing verbal assent was sought from each individual child.

### 2.2 Materials

**2.2.1 WISC IV $^{(UK).}$** The Wechsler Intelligence Scale for Children IV $^{(UK)}$ is an individually administered, comprehensive clinical instrument for measuring the

intelligence of children under 17 years of age (Wechsler, 2004). It provides composite scores that index intellectual ability in a series of specific domains; Verbal Comprehension Index, Perceptual Reasoning Index, Working Memory Index, and Processing Speed Index. It also provides a composite score that represents a child's full-scale intelligence quotient (IQ), which was the dependent measure for this experiment. During the test re-test standardization process, the children were administered the WISC-IV twice, and the average re-test interval was 32 days. Composite indices reliability coefficients ranged from .86 to .93, while coefficients for the subtests ranged from .72 to .92 on average (Wechsler et al., 2004).

**2.2.2 Relational Abilities Index (RAI).** The Relational Abilities Index assessment (see Cassidy, Roche & Hayes, 2011) was delivered online via the website raiseyouriq.com[1]. The relational skills training that followed the RAI assessment, mirrored it precisely, insofar as it taught the participants to correctly answer each of the 55 RAI question types in separate training blocks. Thus, it would be worthwhile to outline the RAI in detail here, as a springboard for outlining the relational skills intervention phase to follow (but see next section for procedural details).

The RAI consists of a battery of 55 relational tasks presented across two contiguous blocks of 29 Same/Opposite relational probes and 26 More/Less probes (see Table 1). For example, one Same/Opposite probe trial might be; B is the same as A, B is Opposite to C, Is C the same as A? (where the A, B and C labels represent randomly generated nonsense words; see Figure 1 for examples).

RAI tasks increase in complexity as the test proceeds, always beginning in task 1 with a single premise task, accompanied by a question regarding the relata (e.g., A is the same as B, is B opposite to A?). As outlined in Table 1, for each stage

---

[1] While this is now a commercial website run by the first two authors, at the time of this research this site was not commercially active and the software described here was online for research purposes only. The authors have declared a potential conflict of interest to the Editor of this journal.

of testing (and later for each stage of relational skills training and testing) several possible relational tasks are possible within the defining parameters.  These are always selected randomly (within the parameters) by the computer software. Thus, no two participants are given precisely the same tasks within a given stage, but all are presented with the same 55 task *types*.  In addition, it should be noted that no nonsense word is ever used twice for a single participant, across RAI assessment and subsequent training and testing combined, such that each and every task presented on screen is unique within participants.  There is a 30s time limit on responding to each task.  RAI scores are calculated simply as the total number of correct responses out of 55.

Stimuli for all trials consist of three-letter or four-letter nonsense words (e.g., GUH, KIR, TIGA) always of the form *consonant-vowel-consonant, consonant-vowel-vowel-consonant, consonant-vowel-consonant-consonant, or consonant-consonant-vowel-consonant*. These are generated randomly on the fly by the test software.  English language words and known slang words are omitted from the stimulus database. A total of 248 stimuli are used in the RAI assessment, with no stimulus appearing on more than one trial.

The increasing complexity of the task types tested in the RAI (and trained in the intervention phase) are controlled in terms of;

1) The number of premise statements presented to the participant before a relational question is posed (one, two or three)

2) The order in which the premises are presented

3) The nodal distance between the relata referred to in the question. As an example, given the premises; A is the same as B, B is the same as C, and C is the same as D, the question; Is B the same as A? represents a zero-node derived relation

probe. The question; Is C the same as A? represents a one-node derived relation

probe, and the question; Is D the same and A? represents a two-node derived relations

probe. Generally, the greater the nodal distance between the two relata referred to in

the relational question, the more complex the relational task.

     4) The directionality of the relational question. For instance, given A is more

than B, B is more than C, and C is more than D, correctly answering the question; Is

D more than A? (i.e., reverse direction to that of the linear logic of the premise

sequence) involves a more highly developed relational skills repertoire than correctly

answering the question; Is A more than D?

     5) Whether or not more than one relation type was specified across the

premises.

     6) Whether or not the relational term employed in the question was employed

in any of the premises.

     The reader is referred to Table 1 and Figure 1 for details and examples of all

of the relational tasks presented during both the RAI assessment and relational skills

training stages.

     The RAI was not intended to function as a standardized test of cognitive

ability but as a baseline measure of the precise skill domains that were targeted in the

online relational frame training intervention employed. It therefore functioned as a

manipulation check insofar as for theoretical reasons, RAI scores must increase, if

full-scale IQ scores are to increase.   The full details of the trial format are outlined in

the Procedure section.

<div align="center">***Insert Figure 1 Around Here***</div>

## 2.3 Procedure

Before commencing the study, ethical approval was provided by the relevant school principal. Detailed information was provided to each participant's parents and written consent was provided from these for each child. Children were provided with a full IQ assessment approximately one month before commencing the training, which took on average 90.47 days to complete (Range 37-222 days; SD=59.96 days). The follow up IQ test was administered approximately one month following the completion of the training for each child. The RAI was first administered on the first day of the intervention and again on the final day of the intervention when participants had completed all stages of the training and testing.

During the RAI assessment and all training and testing stages that followed, participants were seated approximately 4 feet apart in a small room (approx. 8m by 4m), each facing a 15" computer monitor. Each participant wore headphones so that auditory feedback provided by the computer software was not audible to other participants. Participants could not see each other's computer screens.

**2.3.1 IQ testing.** Baseline and follow-up IQ testing took place in a quiet room in the school setting, with only the student and experimenter present. Administrations of the test took approximately 90 minutes.

**2.3.2 Relational Abilities Index Assessment.** After the general nature of the task was briefly explained, the participant was asked to read the instructions presented on the screen. These stated that on each trial a number of statements would be presented that would involve made-up words as well as a question about a specific relationship between these words that could be worked out on the basis of the statements.

The testing proceeded across 55 trials as outlined above and described in detail in Table 1. Each trial consisted of one to three statements presented in black

bold font, a question in blue writing, a green YES button and a red NO button (see Figure 1). Questions were answered by clicking on either the YES or NO buttons, which were located on the bottom left and right of the computer screen (with left-right location randomized across trials to control for positional responding). A 30 second countdown clock was presented on the right hand side of the screen and turned from black to red after 20 seconds. A 30 second time limit applied to each trial, whereby failure to produce a response was recorded as an incorrect response. As soon as a response was made within the 30s response window, the screen was cleared and the next task was immediately presented without feedback.

***Insert Table 1 About Here***

**2.3.3 Relational Frame Training**.

The training was conducted via the website raiseyouriq.com[1]. Training proceeded across 55 trials precisely mirroring those presented in the RAI assessment. The key difference was that for each task presented in the assessment, an entire stage of training, containing a potentially infinite number of trials (and therefore stimuli), was presented. The purpose of training was to establish correct and rapid responding to a single 16-trial sequence of tasks of each type (see Table 1) presented during the RAI assessment (i.e., 55 stages of training to reflect the 55 trial types). This was achieved by presenting a potentially infinite sequence of relational framing tasks (i.e., of the same complexity level but using different stimuli on every single trial) in a single stage of training and requiring 16 correct consecutive responses in order to complete the stage.

During training stages, all responses were followed by corrective feedback (i.e., the words Correct or Wrong appearing on the Screen for 1s with a short beep accompanying the word Correct). Different nonsense word stimuli were employed

17

on every task of the training such that no nonsense word was ever seen twice by the same participant across the entire 55 stages of training (and testing; see below).

Once 16 successive correct responses were made the relevant training stage was completed. The participant was then informed by the computer that a test would be presented at that same level of complexity. During the test, 16 trials of the same kind were presented but in the absence of feedback and the test could only be passed by errorless responding across all 16. The 30s response time limit applied on each trial of both training and testing.

If a participant failed to pass the test, they were cycled back to the completed training stage until they once again satisfied the training criterion (16 consecutively correct responses). Thus, on each stage participants were cycled between training and testing blocks until they completed training and then passed a testing block for that stage. Once that was achieved the participant advanced to the next stage and this continued until all 55 stages were completed.

Training took place in sessions of around 45-60 minutes, 2-3 times per week in a classroom setting under teacher supervision. Students could stop training at any point and the computer software kept track of each participant's progress, which followed its own unique trajectory. A limit of five passed stages per session was imposed to make sure that training was dispersed over a sufficiently long period to allow for consolidation. While participants found it easy to complete five stages during early sessions, progress became increasingly slow as the sessions progressed and the complexity of the training and testing tasks increased. The number of days required to complete training, ranged from 37-222 days (Mean=90.47, SD=59.96). No students dropped out of the training or failed to complete the assessments.

### 3. Results and Discussion

18

A Kolmogorov-Smirnov test for normality indicated that the IQ and RAI scores did not deviate significantly from a normal distribution at either pre or post-intervention ($p > .05$ in all cases). Thus a planned comparison t-test was employed to test the hypothesis that IQ scores would rise as a result of the relational frame training intervention. A t-test was also used to examine increases in relational ability from pre- to post-intervention. Bonferroni correction was applied to the .05 alpha level.

***Insert Figure 2 Around Here***

**3.1 RAI scores.** The mean RAI score at baseline was 33.8 (SD=8.11), while at post-intervention it was 48.53 (SD=5.54). This increase in RAI score was significant, $t(14) = 7.23$, $p < .001$. The Cohen's $d$ effect size was large at 2.127 (Cohen, 1988). This suggests that the training successfully targeted the relational skills of interest. Figure 2 illustrates graphically the increase in RAI scores for each participant from pre to post-intervention.

***Insert Table 2 About Here***

**3.2 Full Scale IQ Scores (FSIQ).** Table 2 shows that the average full scale IQ at baseline was in the average range at 97 (SD=10.21). At post-intervention full scale IQ was in the superior range at 120 (SD=9.97). This increase in full scale IQ from pre to post-intervention was significant, $t (14) = 19.183$, $p < .001$. The Cohen's $d$ effect size was large at 2.329 (Cohen, 1988). The lowest IQ rise (for P12) was 14 full-scale IQ points, which is just 1 point shy of a full standard deviation. The remaining participants had IQ gains of one standard deviation or more, with three experiencing increases close to or above two standard deviations (i.e., 28 or above). Figure 3 illustrates the baseline and post-intervention IQ scores graphically. It can be seen from this graph that increases in FSIQ are consistent and large across participants.

***Insert Figure 3 Around Here***

The significance of these IQ rises in an educational or clinical context can be further underlined by considering change in percentile rank, which reflects the performance of a child on an IQ test relative to the entire population of their peers. Before training, only three participants had full scale IQs above the 60th percentile for the population. After training, however, only three participants did not have full scale IQs above the 80th percentile. This means that most participants were moved from the mid-percentile ranges to the high-percentile ranges.

RAI score at baseline did not correlate with the extent of IQ rises from pre to post-intervention ($r = 0.002$, $p = $ ns), so it might be tentatively concluded that there was no effect of baseline relational skill levels on the effectiveness of intervention. Similarly, IQ scores at baseline were not related to IQ rises observed for each participant ($r = -0.28$, $p = $ ns). Analysis was also undertaken to determine the relationship between the number of days spent training and increases in IQ scores. The number of days required to complete training did not correlate with IQ increases, $r (15) = .002$, $p = ns$. In summary, while exposure to relational skills training lead to significant increases in IQ, the increase would not appear to have been constrained by the length of time required to complete training, baseline relational ability or IQ scores, at least using the current design. It should be noted, however, that due to the low $n$ employed here, the lack of correlation between these two measures may well be related to the decreased power of the relevant statistical analyses. Interestingly, the RAI scores also failed to correlate with FSIQ scores as theoretically predicted, either at baseline, $r (15) = .128$, $p = ns$, or at post-intervention, $r (15) = .151$, $p = ns$ (see General Discussion).

Because the current study was somewhat exploratory, it was not predicted that each and every subtest of the WISC would show increases from pre to post

intervention. However, in order to identify the main factors underlying the increase in scaled IQ, an analysis of pre to post-intervention changes in the unscaled subtest scores was conducted. Ten separate repeated measures t-tests were run. Bonferroni correction was used to adjust the alpha level to .005 (two-tailed). All of the ten subtests showed significant increases from pre to post intervention ($p < .001$). Table 3 shows the individual participant subtest data and the group means for each subtest, pre and post intervention.

***Insert Table 3 Around Here***

## 4. Experiment 2

The findings of Experiment 1 suggest that significant IQ gains can be achieved using a Relational Framing Skills intervention. However, that experiment was not without its limitations. First, the IQ assessor was not blind to the experimental design or the experimental hypotheses. Second, an increase in IQ does not necessarily imply that an increase in scholastic ability can be expected. An assessment of the effect of increases in relational framing ability on scholastic ability is of immediate theoretical and practical significance.

Experiment 2 addressed these concerns by running a significantly larger trial in which pre and post-intervention measures were administered and scored by a third and fourth party, respectively, neither of whom was aware of the experimental design or was directly involved in the study. The measure employed as the baseline and post-intervention assessment was a widely used scholastic ability test (the Differential Aptitude Test; DAT) that also provides scores in specific cognitive domains relevant to school performance. The verbal and numerical ability domains of the DAT are the domains of most relevance theoretically to the current relational framing skill intervention which focuses only on Same, Opposite, More and Less relational skills

21

proficiency (see Cassidy et al., 2010 for rationale). Training in these skill domains is not expected to improve mechanical or spatial skills as measured on the DAT. In addition, the general index called "educational aptitude" (sometimes referred to as "scholastic ability"), which is of key interest here, is calculated in the DAT on the basis of scores in the numerical ability and verbal reasoning subtests.

## 5. Method

### 5.1 Participants

Fifty-one participants, spanning two full classes of students at a secondary school in Co. Kildare, Ireland, were chosen by the school principal for inclusion in this trial. Eighteen students had to be excluded from the analysis due to a failure to complete baseline or follow-up DATs assessments, either due to absences from school, or due to simply not completing the group-based test as instructed under teacher supervision (see Discussion for full details). Highly erratic or missing scores on any subtest of the DATs were taken to be indicative of inattention to the task as a whole. Data for an entire participant was considered untenable if a DAT subtest score of less than 10 was recorded for any one part of the DAT battery at baseline or at follow up, or if they scored zero on more than one DATs subtest across the entire baseline and follow up assessment batteries. Such low or frequently missing scores were taken to be indicative of inattention to the task and poor engagement with the study, as they fall more than two standard deviations below the mean score, and all students were in the normal intellectual range based on school classifications.

Students who completed training and at least one of the numerical or verbal reasoning subtest assessments at both baseline and follow-up, were retained in the study. Three further participants failed to complete the supervised training by the time the study was complete, due to long or several absences from school. The

remaining 30 students were aged 15-17 years (M = 16.36 years), with an equal number being male and female (n = 15). Written consent for participation was provided by legal guardians in all cases and assent for participation was sought from each individual child. None of the children had any reported learning difficulties of clinical significance.

A post-hoc enquiry revealed that of the 18 students deemed to have untenably varying (up as well as down) DATs index scores, five were known to have general learning difficulties, while two had suspected general learning difficulties. According to school authorities, two of the 18 students had socio-economic disadvantage that may interfere with school performance. Of the three students who failed to complete the training due to long school absences, none had any diagnosed learning difficulties but one had a suspected general intellectual disability. Of the 30 students whose performances were included in the analysis, none had any diagnosed learning disabilities, but one had a suspected learning difficulty. Two had socio-economic disadvantage that may interfere with school performance.

**5.2 Materials**

**5.2.1 The Differential Aptitude Test (5th ed)**. The Differential Aptitude Tests (Bennett, Seashore, & Westman, 1990) were developed in the USA, but the fifth edition has been standardized for use in Irish schools with a sample of over 10,000 school children. The DAT battery is a set of group-administered tests of cognitive ability across eight domains; Verbal Reasoning, Numerical Ability, Clerical Speed and Accuracy, Abstract Reasoning, Mechanical Reasoning, Space Relations, Spelling, and Language Usage. A composite Educational Aptitude (EA) score (sometimes referred to as the *scholastic ability* score) can also be calculated from the subtest data. For theoretical reasons, only the verbal (VR), numerical (NR) and

composite EA scores were of interest in the current study. The DAT has good test-retest reliability for an aptitude test and has a high level of internal validity (see Hersen, 2004). The scales have reliability coefficients of at least 0.8 and most are above 0.9 (see Kline, 1999).

**5.2.2 Relational Abilities Index (RAI).** The Relational Abilities Index assessment was employed in the same manner as described in Experiment 1.

**5.3 Procedure**

The general format of this study had received prior approval from the Maynooth University Ethics Committee. Before commencing the study, ethical approval was provided by the relevant school principal. Detailed information was provided to each participant's parents and written consent was provided from these for each child. Children were tested on the DAT approximately one week before commencing the training, which took on average 43.7 days to complete (Range 27-62 days; SD=8.63). The follow up DAT was administered approximately one week following the completion of the training for each child.

**5.3.1 DAT assessment.** Baseline and follow-up DAT administrations took place in a large quiet room in the school setting, with all students present. The paper and pencil version of the test was administered by a school teacher qualified to administer it for vocational purposes, but who were unaware of the nature of the study and not involved in the study in any way. A different teacher, again uninvolved in the study, administered the post-intervention DATs assessment. Each of the DATs required two two-hour (approx.) sessions to complete. Tests were scored by a blind third party (a commercial psychometric assessment service).

### 5.3.2 Relational Abilities Index Assessment and Training Intervention.

These both followed the same procedure as described in Experiment 1[2].

### 6. Results and Discussion

The performances of the omitted and retained participants (see Participants section) on the VR and NR subtests of the DATs were compared to ascertain whether or not there were pre-existing differences in ability across these two groups. Independent samples t-tests found that VR scores for the retained group (M = 61.13, SD = 24.46) were significantly higher than for the omitted participants (M = 44.76, SD = 25.85), t (49) = 2.312, p < .05.  The NR scores for the retained group (M = 54.6, SD = 22.71) were also found to be significantly higher than for the omitted participants (M = 39.67, SD=22.69), t (49) = 2.298, p < .05.  This suggests that the apparent differences in ability to complete the DATs at baseline and follow-up in a group context (across four two-hour sessions in total) were indicative of pre-existing aptitude differences.

In addition to the necessary removal of 21 participants due to incomplete DATs tests (18 participants) or relational skills training (3 participants), a small number of additional data points are missing from various of the following analyses due to absences of children from school for a single assessment period (of the two baseline and two follow-up assessment periods). Specifically, where only one single data point is missing for a participant they were included in the study, so long as no other subtest score was under 10.

A Kolmogorov-Smirnov test indicated that the two DAT subtest scores of interest and the composite Educational Aptitude test scores did not deviate significantly from a normal distribution (p > .05 in all cases). Thus planned

---

[2] By the time this second experiment was conducted the online training site at raiseyouriq.com had become a commercial site.

comparison t-tests were employed to test the three main hypotheses that Verbal and Numerical Reasoning, as well as the composite Educational Aptitude score, would rise as a result of the relational frame training intervention. The RAI score distribution did deviate from normality during the post-intervention assessment ($D =$ .242, $p =$ .000), and thus a non-parametric planned comparison was employed to assess the increase in RAI scores from pre to post-intervention. Bonferroni correction was used to adjust the usual .05 alpha level (to .0125; two-tailed).

The number of days between baseline and follow-up relational ability assessment (i.e., the number of days required to complete training) ranged from 27-62 days (Mean=43.7, SD=8.63), and this is indicated for each participant in Table 3.

***Insert Figure 4 Around Here***

**6.1 RAI Scores.**  A mean relational abilities index score of 44.7 out of 55, or 81.3% correct, (SD = 5.7) was recorded at baseline (see Figure 4). This rose to 51.3 out of 55, or 93.3% correct (SD 3.84) post-intervention. This increase was significant, $z$ (29) = 4.451, $p <$ .001, indicating that scores rose from pre to post-intervention as expected. The Cohen's $d$ effect size for this increase in relational ability was large at 1.52 (Cohen, 1988). This outcome functions as a manipulation check insofar as it confirms that the key skills targeted in the training were themselves successfully improved by the intervention.

**6.2 Verbal Reasoning.**  Table 4 shows that scores on the Verbal Reasoning subtest of the DAT rose from a mean of 63.2 (SD=22.8) to a mean of 69.5 (SD=23.1) from baseline to post-intervention. This increase proved to be significant, $t$ (26) = 2.979, $p =$ .006. A Cohen's $d$ statistic of 0.4 (small) was calculated for this increase in verbal reasoning ability. Figure 5 shows the increases in verbal reasoning ability for each participant.

**6.3 Numerical Reasoning**. Scores on the Numerical Reasoning subtest of the DAT rose from a mean score of 54.5 (SD=21.2) to a mean score of 67.6 (SD=20.0) from baseline to post-intervention (see Table 3). This mean increase of 14.87 points (SD= 14.61) proved to be significant, $t$ (29) = 5.572, $p < .001$. A Cohen's $d$ statistic of 0.68 (medium) was calculated for this increase in numerical reasoning ability. Figure 6 shows the increases in numerical reasoning ability for each participant.

**6.4 Educational Aptitude.** The Educational Aptitude (or Scholastic Ability) score is a composite score calculated for each participant using their Verbal and Numerical Reasoning scores. Scores on this composite scale rose from a mean of 58.9 (SD=20.9) to 68.5 (SD=18.0) from baseline to post-intervention (see Table 3). This mean increase of 17.1 points (SD=17.0) proved to be significant, $t$ (26) = 5.207, $p < .001$. A Cohen's $d$ effect size statistic of 0.73 (medium) was calculated for this increase in overall Educational Aptitude. Figure 7 graphically illustrates this increase in educational aptitude for each participant.

A Pearson's product-moment correlational analysis was conducted to calculate the co-variance between baseline RAI scores and increases in the three DAT metrics of interest. Baseline RAI scores did not correlate with increases in Verbal Reasoning $r$(27) = .233, $p$ =.241, Numerical Reasoning, $r$(30) = .177, $p$ =.35, or Educational Aptitude scores, $r$(27) = .262, $p$ =.186. In summary, the increases in DAT scores were not associated with an individual's pre-existing level of relational skills.

A second and similar correlational analysis was undertaken to determine the relationship between the number of days spent training and increases in DAT scores

(see Table 4). This analysis found that the number of days required to complete training did not correlate with Verbal Reasoning, $r(27) = .177$, $p = .377$, Numerical Reasoning, $r(30) = .332$, $p = .073$, or Educational Aptitude scores, $r(27) = .209$, $p = .296$. In effect, while exposure to relational skills training led to significant increases in all three DAT measures, the magnitude of these increases was not related to the length of time it took participants to complete the training.

As expected, the RAI scores at baseline correlated with Verbal Reasoning, $r(29) = .669$, $p = .000$, Numerical Reasoning, $r(30) = .426$, $p = .019$, and overall Educational Attainment, $r(27) = .664$, $p = .000$, tentatively supporting the RFT position that relational ability may underlie educational ability. Interestingly, however, post-intervention RAI scores did not correlate with post-intervention Verbal Reasoning, $r(29) = -.092$, $p = .642$, Numerical Reasoning, $r(30) = .041$, $p = .829$, or Educational Attainment, $r(27) = .024$, $p = .904$ scores. This may not be surprising, however, because the intervention would not necessarily be expected to produce scholastic domain skill gains precisely in proportion to relational skill gains. The precise nature of this functional relationship is at present not well understood.

In summary, Experiment 2 found that exposure to a relational framing skills intervention significantly increased scores on a widely used standard measure of general educational ability, and specific measures of verbal and numerical ability.

### 7. General Discussion

Experiment 1 showed that general intelligence, as assessed by the WISC IV$^{UK}$ increased significantly following a relational framing skills intervention. Thus, these findings replicate the pilot findings of Cassidy et al. (2011) using a somewhat larger sample. The increases in IQ recorded for all participants in Experiment 1 are not only large numerically, but are of educational and clinical significance. Only one

IQ increase recorded here fell short (by a single point) of a full standard deviation increase.

While no hypotheses were formed regarding changes to specific subtests of the WISC, scores on all subtests increased from pre to post intervention. However, we have in fact learned little from this outcome about the mechanisms underlying the increase in full scale IQ. According to Relational Frame Theory, it might be expected that increases in verbal ability (i.e., Vocabulary subtest), tests assessing semantic equivalences (e.g., Similarities subtest), or tests assessing abstract relational skills, even including spatial relations (e.g., Block design) might be among the largest noted. While this may be more or less the case in the current data set, large increases were also observed for several other subtests. Future studies should carefully examine these relative increases using larger sample sizes and detailed regression methods in an effort to understand the relative roles of the various subtest performance improvements in the overall IQ increases.

Even the smallest of the increases in IQ observed here (i.e., 14 points) is unlikely to be solely accounted for by the effect of re-testing given that an increase of only four points is to be expected on average from test to re-test on the WISC IV for children aged 12-16 years (Kaufman, Flanagan, Alfonso & Mascolo, 2006). Kaufman et al. employed a very brief one-month test-retest period, considerably shorter than the one employed in the current study. The original standardization data for the WISC-IV (Wechsler, 2003) suggest that for children of any age, gains of 5.6 points might be attained on average across slightly more comparable test-retest periods of 13-63 days (N=243). Another study using a longer test-retest period of 11 months (N=43) found FSIQ gains to average only 1.63 points across the test-retest period (Ryan, Glass, & Bartels, 2010). Yet another study, using a three-year test-retest

interval (N=131), recorded a decrease in FSIQ of 0.55 points (Lander, 2010). Most

recently, Watkins & Smith (2013), using a sample of 344 students aged 6.1-14.3

years found no significant change in FSIQ across an average 2.84 year test-retest

interval. In fact FSIQ dropped by 0.12 of a point from test to retest. However, the

authors of that study emphasized that large within-participant variations form test to

retesting are possible for a smaller number of individuals. Specifically, 25% of the

participants had a FSIQ score at retest that differed by more than 10 points from the

test period. Around 6% had changes of 15 points or more.  Importantly, however,

these variations included more decreases than increases. While intra-participant

variation such as this might suggest less test-retest stability than usually reported for

the WISC (especially for specific subtests and indices), it would not lead us to expect

large gains for the typical student or any gains at all at the group level. In the authors'

own words; "…there was considerable stability of the FSIQ score across the 2.84

year test–retest interval for this heterogeneous group of students. This suggests that

FSIQ scores from the WISC–IV can be confidently used for longitudinal screening

decisions and group research (Salvia & Ysseldyke, 2004)" (Watkins & Smith, 2013,

p. 480).  Thus, the use of the FSIQ score from the WISC-IV is acceptable for use as a

pre to post-intervention assessment of IQ gain at the group level, even if it is less

reliable for making clinical judgments about individual gains, for which higher test-

retest coefficients in the region of 0.9 or higher might be required (see Bracken, 1987;

Nunnally & Bernstein, 1994; Salvia & Ysseldyke, 2004).

Given the foregoing, it is at least tenable that a change in the underlying

intelligence trait may best explain the IQ gains reported in Experiment 1, although

further research using control groups is required to fully explore this interpretation.

That possibility notwithstanding, the fact that the psychometric tester at baseline and

follow-up was not blind to experimental hypotheses cannot be ruled out as a potential source of IQ score variation across the baseline to follow-up periods.

The significant increase in RAI scores from baseline to post-intervention in both experiments at least theoretically corroborates the position that the increase in IQ (Experiment 1) and Educational Aptitude (Experiment 2) were mediated by an increase in relational skills, although the extent of improvement in one did not correlate with the extent of improvement in the other. In addition, the RAI scores did not correlate with FSIQ scores either at baseline or at post-intervention. These two outcomes run counter to theoretical predictions and may challenge the idea that relational skills underlie IQ. On balance, the RAI is not a standardized test, and so scores cannot be expected to discriminate individuals reliably, and as a consequence scores may also fail to correlate with conceptually related standardized measures. It is also likely that the low sample size employed in Experiment 1 resulted in insufficient power in the correlational analyses. A larger sample may reveal more interesting patterns of relationship between baseline RAI scores and IQ, as well as between RAI scores and IQ increases resulting from relational skills training.

Another issue worth considering is that the RAI itself has a "hard ceiling" in its score range, insofar as a maximum score of 55 is possible, with scores not reflective of speed of responding. A more complex scoring system that integrated accuracy with speed to produce a fluency score, may produce a more discriminating index of relational skills ability, as well as broaden the score range.

The foregoing issues also apply to Experiment 2, although in that case correlations were observed between baseline RAI scores and all three baseline DATs measures. This supports the RFT position that relational ability is related functionally to educational aptitude. However, baseline RAI and DATs measures

failed to predict *increases* in DATs scores, where this may be expected. It is difficult to ascertain at this stage, whether or not this is due to the psychometric properties of the RAI, the sample size, or the fact that a relational skills intervention is unhindered in its impact by baseline abilities.

One related matter pertains to the potentially diminishing returns on skills training effort likely as baseline IQ and RAI scores increase. If indeed a functional relationship were to be more reliably established between baseline IQ or RAI scores and post-intervention increases, it may well indicate that score increases become smaller as baseline scores rise. This interpretation is supported anecdotally by the current research insofar as a negative correlation between baseline IQ and IQ increases (-.28) was found in Experiment 1. In other words, while not significant, increasing baseline IQ was associated with smaller IQ gains as a result of the intervention. Pursuing this issue is an important matter for future studies.

As was the case for the WISC-IV, the pre to post-intervention gains in the verbal and numerical reasoning subtests of the DAT, as well as for the Educational Aptitude composite score in Experiment 2, are unlikely to be accounted for entirely by test-retest effects alone. All tests on the DAT battery have test-retest coefficients of at least 0.8 or above, and most are above 0.9, making it very stable even compared to the WISC-IV. Once again, the variations in scores observed here were not random, as one expects for test-retest variation, but were almost entirely in an upward direction, averaging around +10 points for Educational Aptitude. Specifically, only three of the 27 students for whom an Educational Aptitude score was calculated showed a decrease from pre to post-intervention, and these were among the smallest changes observed, averaging only 1 point. Nevertheless, a control group is required to more fully explore test-retest effects under current intervention conditions.

32

The improvements in DAT scores reported in Experiment 2 are interesting precisely because this battery is widely used to identify scholastic aptitudes and correlates well with school grades. It should be noted, however, that the DAT battery is not without its limitations with regard to its validity. Researchers have questioned its real power to differentiate students in terms of pure aptitudes rather than in terms of underlying intellectual ability. Some researchers have suggested that its scores are loaded too heavily by general intelligence (g) and fluid intelligence ($g_f$) factors (see Kline, 1999). Importantly, however, this is not critical to the current study insofar as the DAT was intended only as an outcome measure of the intervention that would speak directly to educators.

One important issue requiring consideration is the attrition rate observed in Experiment 2. Twenty-one of 51 students were omitted from the study due almost entirely to failure to complete the DATs battery satisfactorily. The DATs sessions were two hours each and two sessions were required for each of the baseline and follow-up test periods (i.e., eight hours of DATs assessment in total). This was a considerable demand for a cohort of children who were required to complete the test in a large group setting (i.e., all 51 students). The drop out rates (i.e., refusal to complete sections, missing assessment sessions entirely) observed may not be so surprising when this level of demand, and the logistics of securing attendance by all participants at four two-hour DATs assessment sessions are considered. Nevertheless, this attrition has undeniably compromised the findings of Experiment 2. More specifically, while only three indices from the DATs were of interest theoretically (VR, NR and EA), the full DATs battery was in fact administered to each student. It was observed that many students had simply failed to complete sections of the test, or were missing from school for the assessments. As suggested

earlier, it was noted that when one score on any subtest of the DATs, at either baseline or follow-up, was untenably low (below 10) the variations in test scores across indices and from baseline to follow-up were also often untenable and highly variable across subtests and sessions. An untenably low score on any section of the assessment was deemed to bring the entire performance of that individual into question. While deemed necessary for stabilizing data, removing outliers from the study undoubtedly raises concerns about "cherry picking", and for this reason we would like to describe the excluded data in more detail.

Participant data sets were eliminated if the participant was absent from the assessment room for more than one full subtest at either baseline or follow up, and if any score on any subtest was lower than 10 (whether or not that subtest was of theoretical interest to the current study). For example, one student received a score of 7 on the baseline VR subtest, and at follow up scored only 4. These scores could not possibly reflect the child's true score, because all children were functioning normally in a mainstream school at the time of the study. The same child showed a large increase in NR from 21 to 43 across the intervention. Another child scored 9 on NR at baseline and 27 at follow up. A further three participants were eliminated simply because they did not complete the training and so in effect did not even qualify as participants. Importantly, 14 of the 21 eliminated participants had higher VR scores and NR scores at follow-up compared to baseline and so the elimination procedure did not consist in large part of the elimination of participants whose NR and/or VR score changes were not in line with experimental hypotheses.

The current data analysis revealed that the students who were eliminated from the study were in fact weaker in VR and NR indices than those who were retained. This may not be surprising given that poor school attendance and inability to attend to

the DATs battery across four two-hour sessions may well be expected to be related to educational aptitude. To this extent the retained sample may not be representative of a typical sample of mainstream school children and this may compromise the generalizability of our findings. Experiment 1, in contrast, involved a 100% adherence to the training protocol and the assessment procedure. In that sample, there was a wide range of baseline IQ scores, and yet all children benefitted from the training in terms of full-scale IQ increases, albeit not necessarily to the same degree.

One other factor that we should consider in interventions of this kind is the role of motivation or 'buy in". The nature of the extended intervention described here is such that it requires significant teacher and guardian support. As such, it may be that involvement in such an intervention increases the attention paid to a child's intellectual and educational development more generally. In addition, success at each of the 55 stages of training, and feedback delivered during individual training trials, may serve to reinforce the child's efforts and commitment and, in some way not yet understood, lead indirectly to gains in aptitude or IQ scores. For example, the training may lead to the coming-to-light of intellectual weaknesses that then receive extra attention and consideration in school or at home.

A small literature base is emerging which appears to suggest that incentives offered to families for engagement in academic activity (e.g., money) lead to improvements in cognitive skills in their children. In the 1998 PROGRESSA project in Mexico, for example, families were incentivized for both their children's attendance at school and for scholastic attainment (Skoufias 2005). Schultz (2000) reported that the program had an accumulated effect of 0.66 years additional schooling for typical lower SES students, which at face value represents a 10% increase in schooling attainment. Opportunity NYC was a further study modeled on

the PROGRESSA framework, but it failed to find any effects on scholastic outcomes for incentivizing parents to take on more responsibilities towards their children (Riccio et al., 2013) or for direct incentives to students to improve school grades (Fryer, 2011). However, Fryer and Holden (2012) found that when the target outcomes of such incentive programs are aligned well between parents, teachers and students, outcomes are very positive, at least in the context of maths attainment (effects on reading achievement were negative). In effect, the possibility that student and parent commitment to the current training may have at least some effect on outcomes can only be ruled out by a more carefully controlled study involving a well-matched control group that receives a placebo intervention.

## 7.1 Conclusion

Relational Frame Theory is only just embarking on the development of interventions to assist in the acquisition of scholastic skills, such as speaking (including vocabulary), reading, numerical reasoning and problem solving. The current intervention was a beta-phase intervention that might be described as a one-size-fits-all approach to relational skills training. That is, it focused only on the most pervasive relational skills used in intellectual tasks and at a level suitable only for children with a relatively well-established baseline relational skills repertoire (i.e., around mental age 8 years or above). However, more is known about additional relational skills involved in mathematics (e.g., hierarchical relations), higher-level language phenomena such as metaphor and humor (e.g., the ability to relate relations to other relations), logic (e.g., if-then and temporal relations), and creative insight and perspective skills (i.e., deictic relations such as here-there and now-then). None of these foundational relational skills were trained to fluency in the current intervention

but they are of great importance for cognitive and scholastic performance (see Hayes et al. 2001; Stewart, Tarbox, Roche, & O'Hora, 2013).

In addition, more fundamental lower level relational skills can be taught to younger children, such as basic symmetry or mutual entailment relations (e.g., if A=B then B=A, If A is more than B then B is less than A) which are foundational to initial vocabulary acquisition and the most rudimentary forms of generative language. Future research will focus on incorporating these extra modules of relational skills training into the intervention battery and assessing the effects of such elaborated interventions.

## References

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. *The Psychological Record 60*, 527-542.

Barnes, D., McCullagh, P.D., & Keenan, M. (1990). Equivalence class formation in non-hearing-impaired children and hearing-impaired children. *The Analysis of Verbal Behavior, 8*, 19-30.

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,* 765–777. doi:10.1037/ a0013494

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

Borella, E., Carretti, B., Riboldi, F., & De Beni, R. (2010). Working memory training in older adults: Evidence of transfer and maintenance effects. *Psychology and Aging, 25,* 767–778. doi:10.1037/a0020683

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, 49, 229-240.

Bracken, B. A. (1987). Limitations of preschool instruments and standards for minimal levels of technical adequacy. *Journal of Psychoeducational Assessment, 5,* 313–326. doi:10.1177/073428298700500402

Brinch, C. and T. Galloway (2012), 'Schooling in adolescence raises IQ scores', Proceedings of the National Academy of Science 109, 425–430.

Campbell, F. A., Pungello, E. P., Miller-Johnson, S., Burchinal, M., & Ramey, C. T. (2001). The development of cognitive and academic abilities: Growth curves

from an early childhood educational experi- ment. *Developmental Psychology, 37,* 231–242. doi:10.1037/0012- 1649.37.2.231

Campbell, F. A., Ramey, C. T., Pungello, E., Sparling, J., & Miller- Johnson, S. (2002). Early childhood education: Young adult outcomes from the Abecedarian Project. *Applied Developmental Science, 6,* 42– 57. doi:10.1207/S1532480XADS0601_05

Cassidy, S., Roche, B., & Hayes, S. C. (2011). A relational frame training intervention to raise Intelligence Quotients: A pilot study. *The Psychological Record, 61*, 173-198.

Cassidy, S., Roche, B., & O'Hora, D. (2010). Relational Frame Theory and human intelligence. *European Journal of Behavior Analysis, 11,* 37-51.

Ceci, S. J. (1991). How much does schooling influence general intelli- gence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27,* 703–722. doi:10.1037/0012-1649.27 .5.703

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences (2nd Ed.).* Lawrence Erlbaum Associates.

Devany, J. M., Hayes, S.C., & Nelson, R.O. (1986). Equivalence class formation in language-able and language-disabled children. *Journal of Experimental Analysis of Behavior, 46*, 243-257.

Dixon, M.R., Whiting, S.W., Rowsey, K., Belisle, J. (2014) Assessing the relationship between intelligence and the PEAK Relational Training System. *Research in Autism Spectrum Disorders, 8*, 1208–1213.

Dymond, S. & Roche, B. (2013). *Advances in Relational Frame Theory: Research and Application*. CA: New Harbinger Publications, Inc.

Feldt, L. S. & Brennan, R. L. (1993). Reliability. In R. L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 105–146). Phoenix, AZ: Oryx Press.

Fryer, R. G. (2011). Financial Incentives and Student Achievement: Evidence from Randomized Trials. *Quarterly Journal of Economics*, *126,* 1755-1798.

Fryer, R. G, & Holden, R. (2012). Multitasking, Incentives, and Learning: A Cautionary Tale. *NBER WP No. 17752.*

Galizio, M ., Stewart, K. L., & Pilgrim, C. Clustering in artificial categories: An equivalence analysis. *Psychonomic Bulletin & Review 8,* 609-614.

Garber, H. L. (1988). The Milwaukee Project: Preventing mental retardation in children at risk. Washington, DC: American Association on Mental Retardation.

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*, 1-17.

Groth-Marnat, G. (2009). *Handbook of psychological assessment (5th Ed.)*. New York: Wiley.

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

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Plenum Press.

Herrnstein, R. J., Nickerson, R. S., Sanchez, M., & Swets, J. A. (1986). Teaching thinking skills. *American Psychologist, 41,* 1279–1289. doi: 10.1037/0003-066X.41.11.1279

Hersen, M. (2004). *Comprehensive handbook of psychological assessment, industrial and organizational assessment*. NJ: John Wiley & Sons.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid

intelligence with training on working memory. *Pro- ceedings of the National

Academy of Sciences, USA, 105,* 6829 – 6833. doi:10.1073/pnas.0801268105

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,* 625–635.

doi:10.1016/j.intell.2010.09.001

Jäeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term

benefits of cognitive training. *Proceedings of the National Academy of Science, 108*
10081-10086.

Kaufman, A.S., Flanagan, D.P., Alfonso, V.C., & Mascolo, J.T. (2006). Test review:

Wechsler Intelligence Scale for Children, Fourth Edition (WISC-IV). *Journal

of Psychoeducational Assessment, 24*, 278-295.

Jencks, C., Smith, M., Acland, H., Bane, M. J., Cohen, D., Gintis, H., Heyns, B., &

Michelson, S. (1972). *Inequality: A reassessment of the effect of family and

schooling in America.* New York, NY: Harper & Row.

Kline, P. (1999). Handbook of Psychological Testing. New York: Routledge.

Klingberg, T., Forssberg, H., & Westerberg, H. (2002). Training of working memory

in children with ADHD. *Journal of Clinical and Experimental

Neuropsychology, 24,* 781–791. doi:10.1076/jcen.24 .6.781.8395

Lander, J. (2010). Long-term stability of scores on the Wechsler Intelligence Scale

for Children—Fourth Edition in children with learning disabilities.

*Dissertation Abstracts International: Section A. Humanities and Social

Sciences.*

41

Lipkens, R., Hayes, S.C., & Hayes, L. (1993). Longitudinal study of the development

    of derived relations in an infant. *Journal of Experimental Analysis of*

    *Behavior, 56*, 201-239.

Lovaas, O. (1987). Behavioral treatment and normal educational and intellectual

    functioning in young autistic children. *Journal of Consulting Clinical*

    *Psychology, 55*, 3-9.


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,* 582–

    590. doi:10.1111/j.1467-7687.2010.01005x

Mackey, A. P., Miller-Singley, A. T., & Bunge, S. A. (2013). Intensive reasoning

    training alters patterns of brain connectivity at rest**.** *The Journal of*

    *Neuroscience, 33,* 4796 – 4803**.**

Mackey, A. P., Whittaker, K. J., & Bunge, S. A. (2012). Experience-dependent

    plasticity in white matter microstructure: reasoning training alters structural

    connectivity. Frontiers in Neuroanatomy, 6, 32.

Margolis, E., & Laurence, S. (1999). *Concepts: Core readings*. Cambridge, MA:

    MIT Press.

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

    80.

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*,

    1149-1153.

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2014). Relational ability and

language performance in children with autism spectrum disorders & typically developing children: A further test of the TARPA protocol. *The Psychological Record, 64*, 233-251.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A., Payne, R., et al. (2006). Transformation of mathematical and stimulus functions. *Journal of Applied Behavior Analysis, 39*, 299–321.

Ninness, C., Dixon, M., Barnes-Holmes, D., Rehfeldt, R.A., Rumph, R., McCuller, G., Holland, J., Smith, R., Ninness, S. K., & McGinty, J. (2009). Constructing and deriving reciprocal trigonometric relations: A functional analytic approach. *Journal of Applied Behavior Analysis, 42(2)*, 191-208.

Ninness, C., Rumph, R., McCuller, G., Harrison, C.,Ford, A. M., & Ninness, S. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis, 38*, 1–22.

Ninness, C., Rumph, R., McCuller, G., Vasquez, E., Harrison, C., Ford, A. M., et al. (2005). A relational frame and artificial neural network approach to computer-interactive mathematics. *The Psychological Record, 51*, 561–570.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count.* New York, NY: Norton.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F.,  & Turkheimer, E.  (2012).  Intelligence: New findings and theoretical developments. *American Psychologist, 67,* 130-159.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). New York, NY: McGraw-Hill.

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. *The Psychological Record 59*, 53-74.

O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal sub-tests of the WAIS-III. *The Psychological Record, 55*, 155-175.

O'Hora, D., Pelaez, M., Barnes-Holmes, D., Rae, G., Robinson, K., & Chaudary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record, 58*, 569-584.

Reed, P., Osborne, L., & Corness, M. (2005). The effectiveness of early intervention programmes for autistic spectrum disorders. *A Report for the South East Regional Special Educational Needs Partnership*. Research Partners: Bexley, Brighton & Hove, East Sussex, Kent, Midway, Surrey, West Sussex.

Riccio, J., Dechausay, N., Miller, C., Nunez, S., Verma, N., & Yang, E. (2013). Conditional cash transfers in New York City: The continuing story of the Opportunity NYC-Family Rewards Demonstration. *MDRC*.

Roche, B., Cassidy, S., & Stewart, I. (2013). Nurturing genius: Realizing a foundational aim of Psychology. In Kashdan, T & Ciarrochi, J. (Eds.), *Cultivating well-being: Treatment innovations in Positive Psychology, Acceptance and Commitment Therapy, and beyond* (pp. 267-302). Oakland, CA: New Harbinger.

Ryan J. J., Glass L.A., & Bartels J.M. (2010). Stability of the WISC-IV in a sample of elementary and middle school children. *Applied Neuropsychology, 17*, 68-72. doi: 10.1080/09084280903297933.

Sallows, G. O., & Graupner, T. D. (2005). *Replicating Lovaas' treatment and findings: Preliminary results.* PEACH. Putting research into Practice Conference, London.

Salvia, J., & Ysseldyke, J. E. (2004). *Assessment in special and inclusive education* (9th Ed.). Boston, MA: Houghton Mifflin Company.

Sanz de Acedo Lizarraga, M. L., Ugarte, M. D., Iriarte, M. D., & Sanz de Acedo Baquedano, M. T. (2003). Immediate and long-term effects of a cognitive intervention of intelligence, self-regulation, and academic achievement. *European Journal of Psychology of Education, 18,* 59– 74. doi:10.1007/BF03173604

Schlinger, H.D. (1993). Separating discriminative and function-altering effects of verbal stimuli. *The Behavior Analyst, 16*, 9-23.

Schultz, T. P. (2000). Impact of PROGRESA on school attendance rates in the sampled population. *February. Report submitted to PROGRESA. International Food Policy Research Institute, Washington, DC.*

Schweinhart, L. J., Montie, J., Xiang, Z., Barnett, W. S., Belfield, C. R., & Nores, M. (2005). *Lifetime effects: The High/Scope Perry Preschool Study through age 40* (Monographs of the HighScope Educational Research Foundation, No. 14). Ypsilanti, MI: High/Scope Press.

Schweinhart, L. J., & Weikart, D. P. (1980). *Young children grow up: The effects of the Perry Preschool Program on youths through age 15* (Monographs of the HighScope Educational Research Foundation, No. 7). Ypsilanti, MI: High/Scope Press.

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

Skoufias, E. (2005). *PROGRESA and its impacts on the welfare of rural households in Mexico* (Vol. 139). International Food Policy Research Institute.

Smith, T., Eikeseth, S., Klevstrand, M., & Lovaas, O. (1997). Intensive behavioral treatment for preschoolers with severe mental retardation and pervasive developmental disorder. *American Journal on Mental Retardation, 102,* 238-249.

Smith, B. L., Smith, T. D., Taylor, L., & Hobby, M. (2005). Relationship between intelligence and vocabulary. *Perception and Motor Skills, 100*, 101-108.

Stephenson, C. L., & Halpern, D. F. (2012). *Improved fluid intelligence limited to increasing working memory capacity using intensive n-back tasks with a visuospatial component.* Manuscript submitted for publication.

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

Sundberg, M. & Michael, J. (2001). The benefits of Skinner's analysis of verbal behavior for children with autism. *Behavior Modification, 5*, 698-724.

Vitale, A., Barnes-Holmes, Y., Barnes-Holmes, D., & Campbell, C. (2008). Facilitating responding in accordance with the relational frame of comparison II: Methodological analyses. *The Psychological Record, 58*, 365-390.

Watkins, W. M., & Smith, L. G. (2013). Long-term stability of the Wechsler intelligence scale for children (4th Ed.). *Psychological Assessment, 25*, 477–483.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—Fourth Edition (WISC–IV) technical and interpretive manual.* San Antonio, TX: Psychological Corporation.

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

Wechsler, D., Kaplan, E., Fein, D., Morris, R., Kramer, J. H., Maerlender, A. et al. (2004). *The Wechsler Intelligence Scale for Children-Fourth Edition Integrated*. San Antonio, TX: The Psychological Corporation.

Figure 1. Nine sample relational training tasks taken from a single block of training (stage 29). Note that all tasks have in common that they involve three premises, presented in a sequential order, but involving more than one relational concept. The questions also vary in terms of which relation is specified. Both one and two-node derived relations probes were presented randomly across trials within this particular single level of training complexity. In the RAI, a random selection of one of these tasks was chosen as a representative testing task for a given level of relational complexity. In this case, a task of the kind above constituted question number 29 in the RAI.
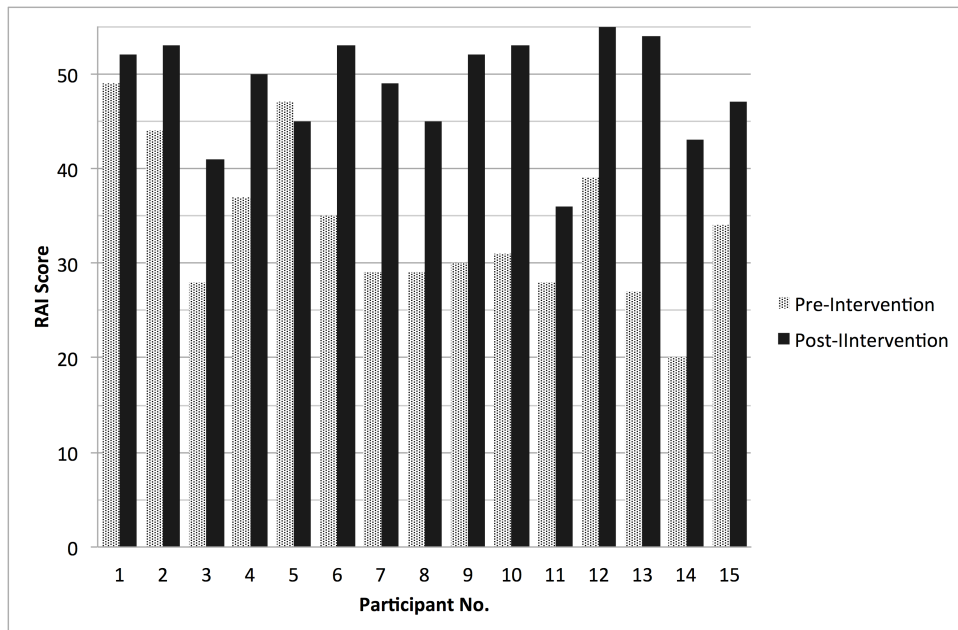
*Figure 2*. Graphical representation of the baseline and follow-up relational ability index scores for each participant in Experiment 1.
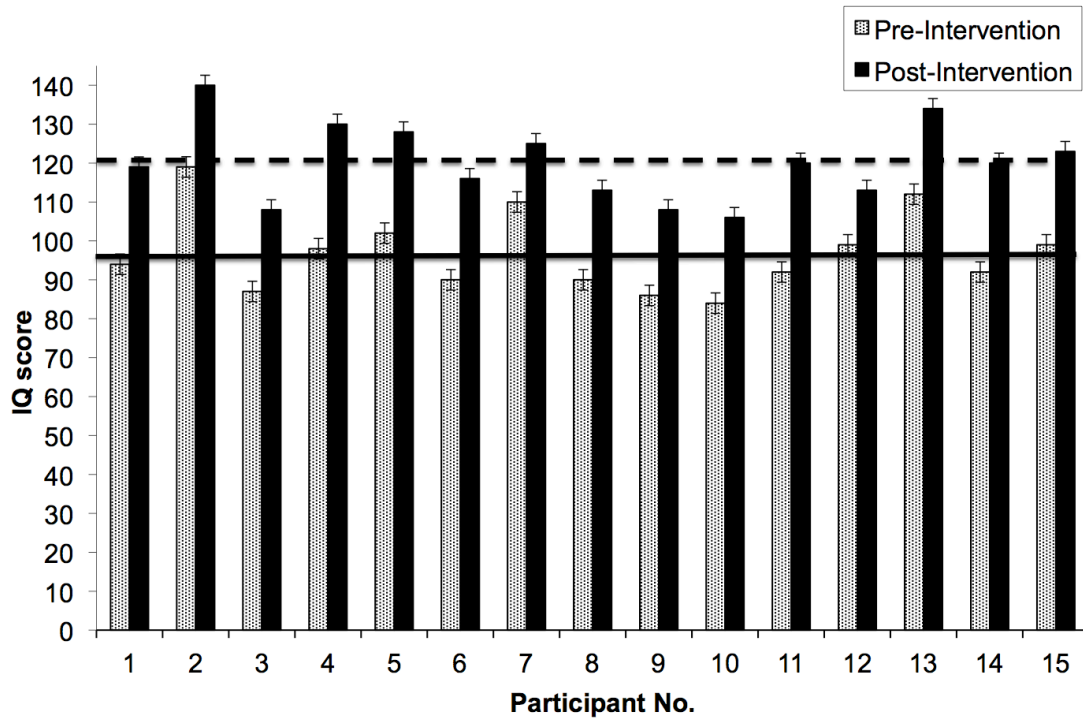
*Figure 3*. Graphical representation of the baseline and follow-up Full Scale IQ scores

for each participant in Experiment 1. The solid and hashed horizontal lines indicate

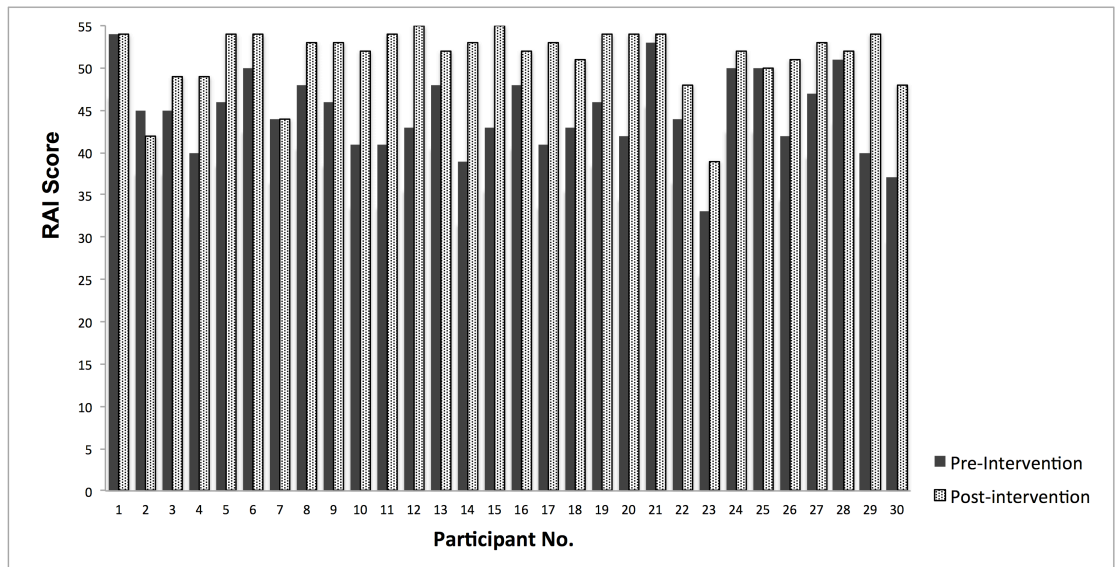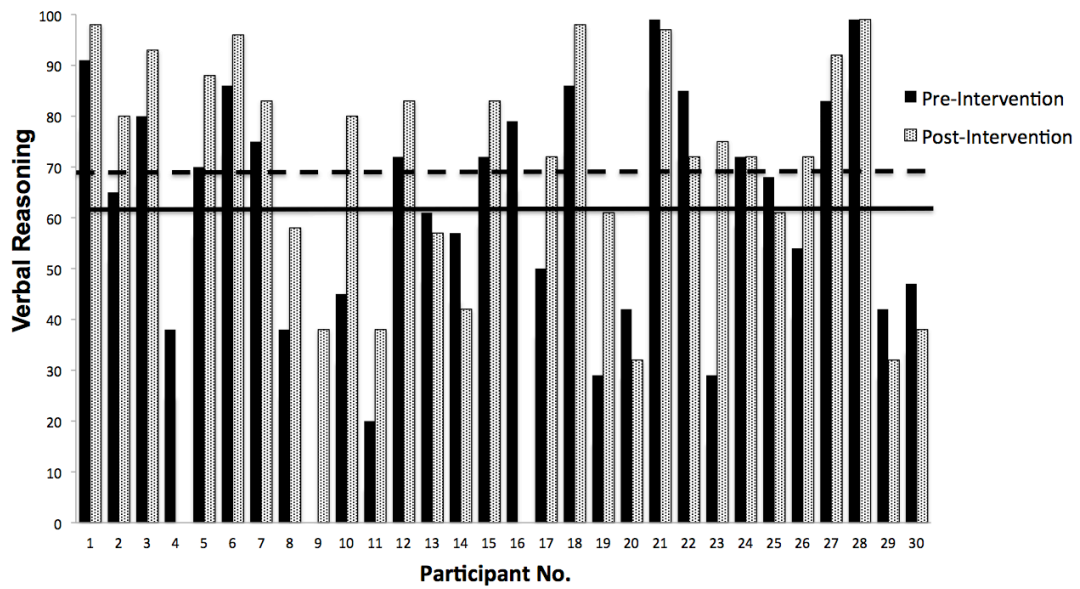the mean Full Scale IQ scores pre and post-intervention, respectively.

*Figure 4.* Graphical representation of the baseline and follow-up relational ability

index scores for each participant in Experiment 2.

*Figure 5.* Graphical representation of the baseline and follow-up Verbal Reasoning scores for each participant in Experiment 2. The solid and hashed horizontal lines indicate the VR scores pre and post-intervention, respectively.
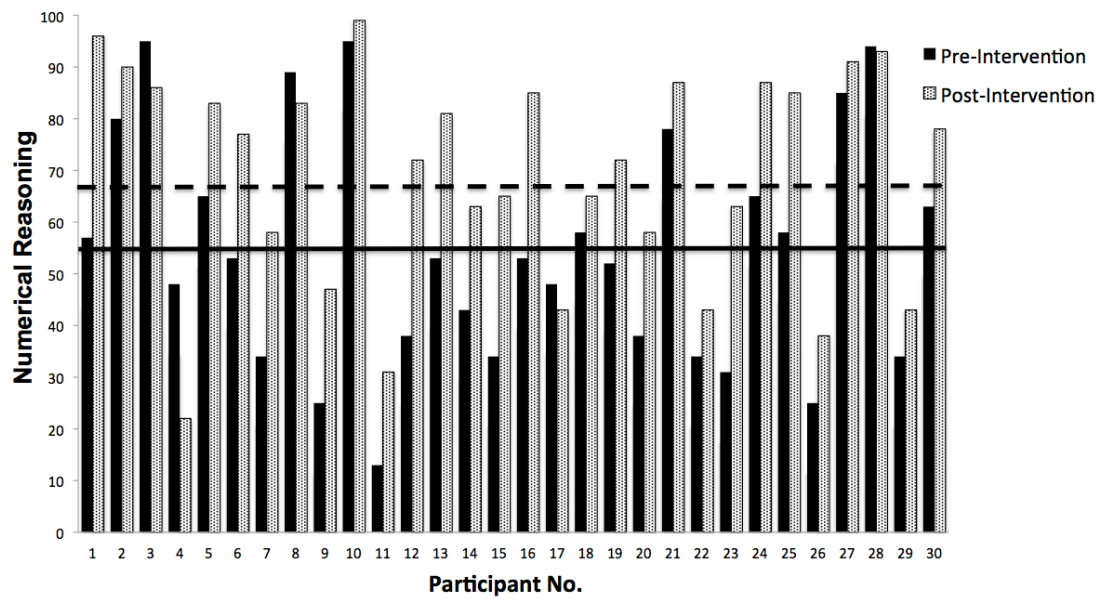
*Figure 6.* Graphical representation of the baseline and follow-up Numerical Reasoning scores for each participant in Experiment 2. The solid and hashed horizontal lines indicate the NR scores pre and post-intervention, respectively.
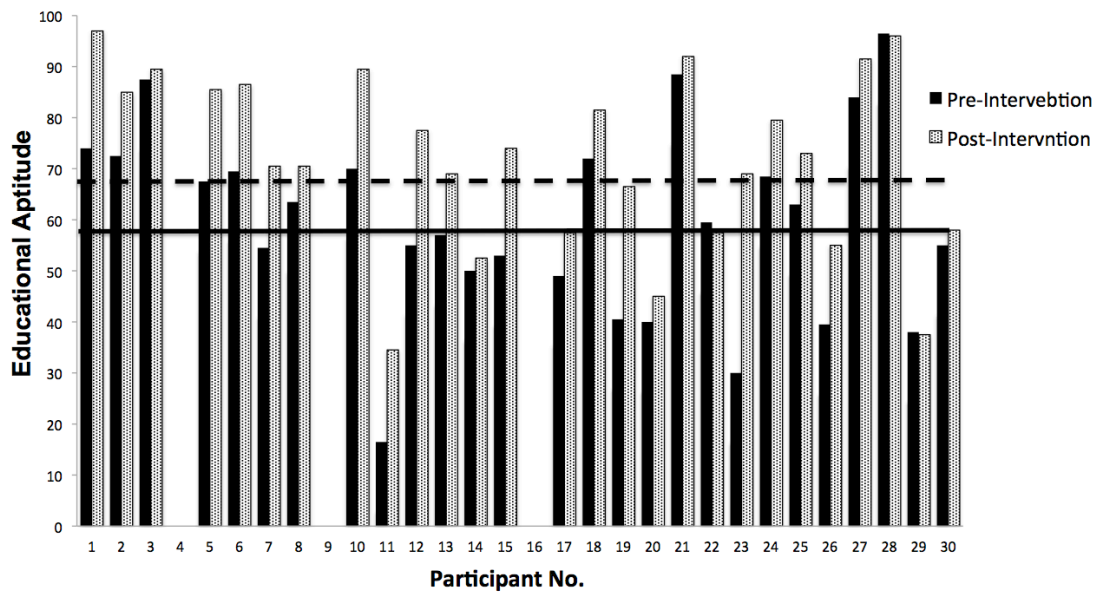
*Figure 7.* Graphical representation of the baseline and follow-up Educational

Aptitude scores for each participant in Experiment 2. The solid and hashed horizontal

lines indicate the EA scores pre and post-intervention, respectively.

Table 1.  The sequence of relational skill probes and relational skill training and testing stages employed in Experiments 1 and 2.

| Stage # | No. of Premises | Relations in Premises | Premises Direction | Relations in Questions | Question Direction | No. of Nodes |
|---|---|---|---|---|---|---|
| 1 | 1 | Same | Forwards | Mixed | Mixed | 0 |
| 2 | 1 | Opposite | Forwards | Mixed | Mixed | 0 |
| 3 | 2 | Both Same | Forwards | Same | Forwards | 0, 1 |
| 4 | 2 | Both Same | Forwards | Same | Backwards | 0, 1 |
| 5 | 2 | Both Same | Forwards | Same | Mixed | 1 |
| 6 | 2 | Both Opposite | Forwards | Same | Forwards | 0, 1 |
| 7 | 2 | Both Opposite | Forwards | Same | Backwards | 0, 1 |
| 8 | 2 | Both Opposite | Forwards | Same | Mixed | 1 |
| 9 | 2 | Both Same | Forwards | Mixed | Mixed | 1 |
| 10 | 2 | Both Same | Mixed | Same | Mixed | 1 |
| 11 | 2 | Both Same | Mixed | Opposite | Mixed | 1 |
| 12 | 2 | Both Same | Mixed | Mixed | Mixed | 1 |
| 13 | 2 | Both Opposite | Forwards | Mixed | Mixed | 1 |
| 14 | 2 | Both Opposite | Forwards | Mixed | Mixed | 0, 1 |
| 15 | 2 | Both Opposite | Mixed | Same | Mixed | 0, 1 |
| 16 | 2 | Both Opposite | Mixed | Opposite | Mixed | 0, 1 |
| 17 | 2 | Both Opposite | Mixed | Mixed | Mixed | 0, 1 |
| 18 | 2 | Mixed | Forwards | Same | Mixed | 1 |
| 19 | 2 | Mixed | Forwards | Opposite | Mixed | 1 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 20 | 2 | Mixed | Forwards | Mixed | Mixed | 0, 1 |
| 21 | 2 | Mixed | Mixed | Mixed | Mixed | 0, 1 |
| 22 | 3 | Same | Forwards | Mixed | Mixed | 1 |
| 23 | 3 | Same | Forwards | Mixed | Mixed | 1, 2 |
| 24 | 3 | Opposite | Forwards | Mixed | Mixed | 1 |
| 25 | 3 | Opposite | Forwards | Mixed | Mixed | 1, 2 |
| 26 | 3 | Mixed Type 1 (S/S/O) | Forwards | Mixed | Mixed | 1 |
| 27 | 3 | Mixed Type 2 (S/O/O) | Forwards | Mixed | Mixed | 1, 2 |
| 28 | 3 | Mixed Type 3 (O/S/S) | Forwards | Mixed | Mixed | 1, 2 |
| 29 | 3 | Mixed Type 4 (O/O/S) | Forwards | Mixed | Mixed | 1, 2 |
| 30 | 1 | More | Forwards | Mixed | Mixed | 0 |
| 31 | 1 | Less | Forwards | Mixed | Mixed | 0 |
| 32 | 2 | Both More | Forwards | More | Forwards | 0, 1 |
| 33 | 2 | Both More | Forwards | More | Mixed | 0, 1 |
| 34 | 2 | Both More | Forwards | Less | Mixed | 0, 1 |
| 35 | 2 | Both More | Forwards | Mixed | Mixed | 1 |
| 36 | 2 | Both Less | Forwards | Less | Forwards | 1 |
| 37 | 2 | Both Less | Forwards | Less | Mixed | 0, 1 |
| 38 | 2 | Both Less | Forwards | More | Mixed | 0, 1 |
| 39 | 2 | Both Less | Forwards | Mixed | Mixed | 1 |
| 40 | 2 | Both More | Mixed | More | Mixed | 0, 1 |
| 41 | 2 | Both More | Mixed | Less | Mixed | 0, 1 |
| 42 | 2 | Both More | Mixed | Mixed | Mixed | 1 |
| 43 | 2 | Both Less | Mixed | Less | Mixed | 0, 1 |
| 44 | 2 | Both Less | Mixed | More | Mixed | 0, 1 |

| 45 | 2 | Both Less | Mixed | Mixed | Mixed | 1 |
| 46 | 3 | All More | Forwards | Mixed | Mixed | 0, 1 |
| 47 | 3 | All More | Forwards | Mixed | Mixed | 1, 2 |
| 48 | 3 | All Less | Forwards | Mixed | Mixed | 0, 1 |
| 49 | 3 | All Less | Forwards | Mixed | Mixed | 1, 2 |
| 50 | 3 | All More | Mixed | Mixed | Mixed | 1 |
| 51 | 3 | All More | Mixed | Mixed | Mixed | 2 |
| 52 | 3 | All More | Mixed | Mixed | Mixed | 0, 1, 2 |
| 53 | 3 | All Less | Mixed | Mixed | Mixed | 1 |
| 54 | 3 | All Less | Mixed | Mixed | Mixed | 2 |
| 55 | 3 | All Less | Mixed | Mixed | Mixed | 0, 1, 2 |

Each stage incrementally increased in complexity along one or more of the dimensions listed in the rows from left to right. Each stage definition allowed for multiple trial sub-types that varied slightly form each other within the stated parameters for that stage. Individual trials were selected randomly on the fly within each stage. Multiple premises presented in a forwards direction were presented linearly (e.g., A is the same as B, B is the same as C). Premises presented in a mixed direction were presented in a random order within and across premises (e.g., C is the same as B, A is the same as B). Questions asked in a forwards direction always followed the sequence of linear syllogistic reasoning (e.g., Is A the same as C?). Questions asked in mixed directions did not follow a linear syllogistic pattern (e.g., Is C the same as A?). The nodal distances between relata in the questions is indicated in the far right column by comma-separated numerals (e.g., 0, 1 signifies that both zero and one node derived relations were probed for). The signifiers O and S in parentheses refer to relations of Same and Opposite, respectively. The RAI assessment consisted of one sample probe trial from each stage, selected at random (i.e., 55 trials in total).

*Table 2.*

Each participant's baseline and post-intervention full scale IQ score (scaled), along with the increase in IQ recorded for each participant and the total number of days from pre to post-intervention (days in training).

| P. No. | Pre-Intervention | Post-Intervention | Increase in FSIQ | Days in Training* |
|--------|------------------|-------------------|------------------|-------------------|
| 1 | 94 | 119 | +25 | 41 |
| 2 | 119 | 140 | +21 | 98 |
| 3 | 87 | 108 | +21 | 51 |
| 4 | 98 | 130 | +32 | 141 |
| 5 | 102 | 128 | +26 | 38 |
| 6 | 90 | 116 | +26 | 96 |
| 7 | 110 | 125 | +15 | 37 |
| 8 | 90 | 113 | +23 | 43 |
| 9 | 86 | 108 | +22 | 46 |
| 10 | 84 | 106 | +22 | 139 |
| 11 | 92 | 120 | +28 | 43 |
| 12 | 99 | 113 | +14 | 45 |
| 13 | 112 | 134 | +22 | 141 |
| 14 | 92 | 120 | +28 | 176 |
| 15 | 99 | 123 | +24 | 222 |
| **Mean** | **96.93** | **120.2** | **23.27** | **90.47** |
| **SD** | **10.2** | **9.97** | **4.7** | **59.96** |

Note that days in training does not refer to the number of sessions, which were typically run 2-3 times per week.

Table 3. Each participant's subtest scores at baseline and post-intervention across each of the 10 subtests of the WISC III[uk]. The shaded column under each subtest heading represents the post-intervention score.

| P No. | Block Design | | Similarities | | Digit Span | | Picture Concepts | | Coding | | Vocabulary | | Letter-Number Sequencing | | Matrix Reasoning | | Comprehension | | Symbol Search | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 31 | 46 | 17 | 31 | 21 | 24 | 19 | 26 | 39 | 48 | 27 | 41 | 17 | 20 | 20 | 27 | 19 | 29 | 23 | 27 |
| 2 | 42 | 49 | 34 | 41 | 29 | 31 | 23 | 28 | 40 | 53 | 50 | 58 | 20 | 25 | 23 | 26 | 34 | 40 | 26 | 30 |
| 3 | 26 | 44 | 17 | 29 | 13 | 15 | 17 | 21 | 39 | 47 | 30 | 45 | 17 | 19 | 19 | 24 | 22 | 33 | 19 | 23 |
| 4 | 26 | 46 | 24 | 29 | 21 | 27 | 17 | 24 | 56 | 63 | 31 | 40 | 18 | 22 | 17 | 29 | 26 | 37 | 29 | 43 |
| 5 | 39 | 57 | 17 | 30 | 22 | 27 | 19 | 25 | 51 | 68 | 35 | 43 | 19 | 21 | 23 | 28 | 24 | 34 | 27 | 25 |
| 6 | 35 | 45 | 17 | 32 | 23 | 25 | 17 | 23 | 41 | 47 | 39 | 47 | 18 | 21 | 22 | 23 | 23 | 34 | 18 | 25 |
| 7 | 43 | 52 | 22 | 29 | 22 | 24 | 19 | 24 | 44 | 58 | 34 | 43 | 18 | 20 | 24 | 27 | 29 | 34 | 29 | 30 |
| 8 | 26 | 34 | 17 | 27 | 16 | 19 | 15 | 23 | 44 | 60 | 23 | 33 | 17 | 18 | 17 | 25 | 21 | 27 | 23 | 29 |
| 9 | 18 | 30 | 27 | 32 | 17 | 21 | 17 | 23 | 35 | 56 | 37 | 46 | 17 | 19 | 18 | 24 | 27 | 32 | 16 | 19 |
| 10 | 18 | 26 | 19 | 29 | 16 | 20 | 18 | 21 | 49 | 51 | 35 | 43 | 19 | 21 | 16 | 23 | 21 | 32 | 22 | 28 |
| 11 | 22 | 51 | 25 | 34 | 15 | 19 | 21 | 26 | 49 | 52 | 35 | 46 | 15 | 18 | 20 | 24 | 21 | 32 | 26 | 35 |
| 12 | 30 | 52 | 18 | 28 | 21 | 22 | 16 | 20 | 56 | 62 | 26 | 34 | 17 | 18 | 21 | 26 | 25 | 28 | 25 | 29 |
| 13 | 52 | 57 | 29 | 37 | 17 | 20 | 24 | 24 | 48 | 48 | 46 | 54 | 17 | 20 | 23 | 28 | 31 | 38 | 21 | 28 |
| 14 | 22 | 38 | 25 | 34 | 19 | 24 | 16 | 24 | 60 | 67 | 33 | 42 | 17 | 20 | 12 | 24 | 22 | 30 | 19 | 26 |
| 15 | 50 | 60 | 25 | 36 | 15 | 19 | 22 | 25 | 42 | 56 | 37 | 46 | 16 | 19 | 20 | 28 | 24 | 31 | 26 | 28 |
| Mean | 32.0 | 45.8 | 22.2 | 31.8 | 19.1 | 22.5 | 18.7 | 36.5 | 46.2 | 55.7 | 34.5 | 44.1 | 17.5 | 20.1 | 19.7 | 25.7 | 24.6 | 32.7 | 23.3 | 28.3 |

*Table 4.*
Verbal reasoning (VR), numerical reasoning (NR) and educational aptitude (EA) for each of the 30 participants in Experiment 2 at baseline (1) and post intervention (2).

| P. No. | VR1 | VR2 | NR1 | NR2 | EA1 | EA2 | Days in Training* |
|---|---|---|---|---|---|---|---|
| 1 | 91 | 98 | 57 | 96 | 74 | 97 | 27 |
| 2 | 65 | 80 | 80 | 90 | 72.5 | 85 | 36 |
| 3 | 80 | 93 | 95 | 86 | 87.5 | 89.5 | 44 |
| 4 | 38 | | 48 | 22 | - | - | 62 |
| 5 | 70 | 88 | 65 | 83 | 67.5 | 85.5 | 32 |
| 6 | 86 | 96 | 53 | 77 | 69.5 | 86.5 | 39 |
| 7 | 75 | 83 | 34 | 58 | 54.5 | 70.5 | 51 |
| 8 | 38 | 58 | 89 | 83 | 63.5 | 70.5 | 51 |
| 9 | | 38 | 25 | 47 | | | 41 |
| 10 | 45 | 80 | 95 | 99 | 70 | 89.5 | 34 |
| 11 | 20 | 38 | 13 | 31 | 16.5 | 34.5 | 58 |
| 12 | 72 | 83 | 38 | 72 | 55 | 77.5 | 37 |
| 13 | 61 | 57 | 53 | 81 | 57 | 69 | 48 |
| 14 | 57 | 42 | 43 | 63 | 50 | 52.5 | 32 |
| 15 | 72 | 83 | 34 | 65 | 53 | 74 | 41 |
| 16 | 79 | - | 53 | 85 | - | - | 54 |
| 17 | 50 | 72 | 48 | 43 | 49 | 57.5 | 58 |
| 18 | 86 | 98 | 58 | 65 | 72 | 81.5 | 39 |
| 19 | 29 | 61 | 52 | 72 | 40.5 | 66.5 | 46 |
| 20 | 42 | 32 | 38 | 58 | 40 | 45 | 44 |
| 21 | 99 | 97 | 78 | 87 | 88.5 | 92 | 41 |
| 22 | 85 | 72 | 34 | 43 | 59.5 | 57.5 | 55 |
| 23 | 29 | 75 | 31 | 63 | 30 | 69 | 49 |
| 24 | 72 | 72 | 65 | 87 | 68.5 | 79.5 | 41 |
| 25 | 68 | 61 | 58 | 85 | 63 | 73 | 39 |
| 26 | 54 | 72 | 25 | 38 | 39.5 | 55 | 54 |
| 27 | 83 | 92 | 85 | 91 | 84 | 91.5 | 39 |
| 28 | 99 | 99 | 94 | 93 | 96.5 | 96 | 39 |
| 29 | 42 | 32 | 34 | 43 | 38 | 37.5 | 39 |
| 30 | 47 | 38 | 63 | 78 | 55 | 58 | 41 |
| **Mean** | 63.21 | 69.5 | 54.5 | 67.57 | 58.86 | 68.54 | 43.7 |
| **SD** | 22.8 | 23.12 | 21.18 | 19.97 | 20.86 | 17.97 | 8.63 |

Blank cells in all but the far right column represent data missing due to school absence. Educational Aptitude composite scores are calculated based on VR and NR subtest scores, so empty cells in the final two columns result from missing data in any of the other columns. Note that days in training does not refer to the number of sessions, which were typically run 2-3 times per week.