TITLE: **Can SMART Training Really Increase Intelligence? A Replication Study.**

Accepted for Publication in the ***Journal of Behavioural Education***

Colbert, D., Tyndall, I., Roche, B., & Cassidy, S.

Dylan Colbert, Department of Psychology, Maynooth University, Ireland

Ian Tyndall, Department of Psychology, University of Chichester, UK

Bryan Roche, Department of Psychology, Maynooth University, Ireland

Sarah Cassidy, Department of Psychology, Maynooth University, Ireland

Abstract

A burgeoning research stream supports the efficacy of a novel behavior-analytic intervention, known as SMART training, in raising general intelligence by training a set of crucial cognitive skills, referred to as relational skills. A sample of Irish secondary school students (*n* = 26) was divided into two IQ matched groups, with the experimental group receiving 12 weeks of SMART training delivered in bi-weekly 45-minute sessions. WASI IQ assessments were administered at baseline and follow-up to all participants by blind testers. For each of the three WASI IQ indices and the four IQ subtests, significant follow-up rises were found for the experimental group only. Analyses of variance indicated a significant effect of training on Verbal IQ, Matrix Reasoning and Vocabulary scores. Results lend further support for the efficacy of the SMART training program in enhancing intellectual skills.

Keywords: Strengthening mental abilities with relational training, Relational Frame Theory, derived relational responding, Intelligence, Educational Intervention.

Training interventions designed to improve intellectual function have been developed within many different perspectives and across numerous disciplines, including behavioral psychology (Cohen, Amerine-Dickens, & Smith, 2006; Eikeseth, Smith, Jahr, & Eldevik, 2002; Lovaas, 1987; Remington et al., 2007), cognitive psychology (Au et al., 2015; Buschkuehl et al., 2008; Buschkuehl, & Jaeggi, 2010; Jaeggi, Buschkuehl, Jonides, & Perrig, 2008; Jaeggi, Buschkuehl, Jonides, & Shah, 2011) medical and health sciences (Aberg et al., 2009; Coe, Pivarnik, Womack, Reeves, & Malina, 2006; Cotman, Berchtold, & Christie, 2007; Davis et al., 2007; Tuckman & Hinkel, 1986; Shephard et al., 1984) and psychopharmacology (Elliot et al., 1997; Kimberg, D'Esposito, & Farah, 1997; Rae, Digney, McEwan, & Bates, 2003). More recently, promising behavior-analytic training programs have been developed that are informed by Relational Frame Theory (RFT; Hayes, Barnes-Holmes, & Roche, 2001; see also Dymond & Roche, 2013). RFT aims to provide a bottom-up explanation of high-level cognitive processes, such as intelligence, in terms of basic learning principles. RFT interventions for intellectual skills development typically focus on a key behavioral repertoire known as derived relational responding or relational framing, a skill which has been shown to correlate significantly with a number of aptitudes that are proposed to comprise intelligence (Barnes-Holmes, Barnes-Holmes, Stewart, & Boles, 2010; Cassidy, Roche, & O'Hora, 2010; Colbert, Dobutowitsch, Roche, & Brophy, 2017; Dixon, Whiting, Rowsey, & Belisly, 2014; Gore, Barnes-Holmes, & Murphy, 2010; Moran, Stewart, McElwee, & Ming, 2010; O'Hora, Pelaez, & Barnes-Holmes, 2005, O'Hora et al., 2008; O'Toole, Barnes-Holmes, Murphy, O'Connor, & Barnes-Holmes, 2009; Stewart, Tarbox, Roche, & O'Hora, 2013; see also Andrews & Halford, 1998; Cattell, 1971; Gentner & Loewenstein, 2002; Halford, Wilson, & Phillips, 2010; for parallel research from fields outside behavior analysis).

Derived relational responding refers to the process of discriminating and deriving relationships between stimuli, in accordance with particular relational frames. These relational frames come in a variety of different forms, such as coordination (A is same as B), comparison (A is bigger than B), opposition (A is opposite to B), distinction (A is different to B), hierarchy (A is a type of B), analogy (A is to B as X is to Y), deixis ("I am here and you are there") and temporality (A comes before B). For example, if a child is taught that a dog is larger than a cat and that a cat is larger than a mouse, the child is able to derive the relations between the animals in this network that have not been explicitly trained (i.e., a dog is larger than a mouse and a mouse is smaller than a dog). Derived relational responding has been implicated as a possible explanation for the "language explosion" witnessed in toddlers (Stewart & Roche, 2013), because the ability to derive untrained relationships of coordination and opposition between words can facilitate a rapid expansion of vocabulary. For instance, imagine that a child has learned to use the word "cookie" effectively. They are then taught in a school context to utter the word "biscuit" when presented with a picture of a cookie. The child may now spontaneously ask for a biscuit in a home context, despite this verbal behavior never having been reinforced in the past with the delivery of a sweet treat. This exemplifies the manner in which relational responding can facilitate vocabulary expansion, but, in a way that can be fully traced back to direct learning and the derived relational responding repertoire. Importantly, the skill of derived relational responding must itself be taught, but we will return to that issue in a later section.

RFT proposes that it is our sophistication in discriminating and deriving both trained and untrained relationships between stimuli in accordance with relational frames that may underlie much of human cognition (Hayes, 1991). As outlined by Cassidy et al. (2010), many of the test items that constitute traditional IQ measures can be understood as tests of relational responding. For instance, IQ subtests from the Wechsler Adult Intelligence Scales

(WAIS; Weschler, 1955, 1981, 1997, 2008) and Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949, 1974, 1991, 2003) such as the Vocabulary and Similarities subtest essentially assess relational skills by probing for word-word (e.g., “What does ‘purpose’ mean?”), word-object (e.g., “What is a cart?”) and object-object relations (e.g., “How are a plane and a bus alike?”). The importance of relational responding for IQ test performance extends to non-verbal tasks such as the Wechsler Picture Concept subtest, which requires the child to select a number of pictures from among a wider array based upon a common characteristic or similarity (e.g., choosing a squirrel and a bird when presented with a pictorial array of a crayon, a squirrel, an umbrella, and a bird). The relational frame of comparison is clearly implicated in the Wechsler Arithmetic subtest (e.g., “Jamal has twice as money as Seth. Jamal has 17 pounds. How much money does Seth have?”), as well as any other test of numerical ability (see Marr, 2015). Temporal relations, a subset of comparison relations, are assessed in the Wechsler Information and Picture Arrangement subtests. In the Information subtest, for example, an individual may be asked a simple time-based question, such as “What day comes after Friday?”. The Picture Arrangement subtest asks the participant to place a number of images, each depicting a different moment of a short comic-strip style story, in the correct order. By doing this, the participant is required to construct a story that is logically coherent in real time (e.g., a pizza base cannot land on the chef’s head without him throwing it into the air beforehand). The frame of hierarchy is also assessed by the Wechsler Similarity subtest, in questions such as; “How are a grape and a strawberry alike”, insofar as it probes for member-category relations. In order to answer this question correctly, the participant is required to find the most specific category that contains both members (i.e., to gain full marks, participants must respond with “they are both fruits”, rather than a more general answer, such as “they are both food”). Such a functional overlap in relational responding and intellectual performance would thereby implicate the importance of the former to the latter (see Cassidy et al., 2010, for a full conceptual unpacking of the relevance of relational responding to IQ and psychometrics).

An extensive body of research has reported high levels of correlation between performance on relational responding tasks and various metrics of general intelligence (Barnes-Holmes et al., 2010; Cassidy et al., 2010; Colbert et al., 2017; Dixon et al., 2014; Gore et al., 2010; Moran et al., 2010; O'Hora et al., 2005, 2008; O'Toole et al., 2009; Stewart et al., 2013). In a comprehensive correlational analysis of performance on a relational responding task and general intelligence metrics, Colbert et al. (2017) analysed performance on the Wechsler Adult Intelligence Scale-III (Wechsler, 1997), a battery of cognitive measures and the Relational Abilities Index (RAI), a 55-item assessment of coordination, opposition and comparison relations. In the first of their studies, RAI test performance correlated significantly with measures of verbal ability (National Adult Reading Test, NART; Nelson, 1992), visuospatial function (the Trail Making Test, TMT; Lezak, 1995) and memory (Rey Auditory Visual Learning Test, RAVLT; Rey, 1958; English version: Taylor, 1959). In their second study, medium-to-strong correlations were reported between RAI scores, and the three WAIS-III indices (Full Scale IQ, Verbal IQ, & Performance IQ), the four IQ subindices (Verbal Comprehension, Working Memory, Perceptual Organisation and Processing Speed) and 10 of 13 IQ subtests. Such correlational studies support the RFT-based claims that intelligence and relational skills may be functionally related or even synonymous, and this has laid the conceptual foundation for the idea that if relational skills can be enhanced, IQ scores should rise as a result.

Because RFT approaches derived relational responding in terms of its function, rather than topography, relational responding is approached as a generalized operant behavior that can be established and shaped (Hayes, 1994). This functional approach is what makes RFT so amenable to developing progressive interventions to increase skills more typically thought of as invariant. Novel interventions based on the RFT approach, such as SMART training (Strengthening Mental Abilities through Relational Training; Cassidy, Roche, & Hayes, 2011) and PEAK training (Promoting the Emergence of Advanced Knowledge; McKeel, Dixon, Daar, Rowsey, & Szekely, 2015; Rowsey, Belisle, & Dixon, 2014), focus on training and increasing the fluency of derived relational responding repertoires. Several studies have now shown that when relational skills repertoires are enhanced, large gains in intelligence quotients, and scores on other tests of general cognitive functioning, are observed (Amd & Roche, in press; Cassidy et al., 2011, Cassidy, Roche, Colbert, Stewart, & Grey, 2016; Hayes & Stewart, 2016; Thirus, Starbrink, & Jansson, 2016). As intellectual performance is amongst the best predictors of a wide range of socially-desirable outcomes, such as academic achievement (Deary, Strand, Smith, & Fernandes, 2007; Mackintosh, 1998; Watson & Monroe, 2009), job performance (Hunter, 1983a; Hunter & Schmidt, 1996; Ree & Earles, 1993), income (Neisser et al., 1996; Lynn & Vanhanen, 2005; Zagorsky, 2007) and even happiness (Ali et al., 2012), interventions which show potential to increase IQ scores harbor genuine social implications.

A key component of an RFT approach to relational skills training is its utilization of Multiple Exemplar Training (MET). MET involves the presentation of a large number of syllogistic tasks (e.g., A is same as B, B is opposite to C, is A the same as C?) each of which requires the participant to derive relations between arbitrary stimuli under specific forms of contextual control (i.e., relational cues such as the words same, opposite, more etc.). Bringing the derived relational responding repertoire under fine contextual control is an important part of making the relational skill repertoire effective in the real world (Barnes-Holmes, Barnes-Holmes, Roche, & Smeets, 2001; Barnes-Holmes, Barnes-Holmes, & Smeets, 2004; Berens & Hayes, 2007; Gomez, Lopez, Martin, Barnes-Holmes, & Barnes-Holmes, 2007; Luciano, Becerra, & Valverde, 2007). That is, a highly developed relational skills repertoire involves is arbitrarily applicable but it is not arbitrarily applied. It is applied only to stimuli where appropriate. The appropriateness of a given derived relational response is controlled by contextual cues, such as words like “Same”, “Opposite”, or other features of the task at hand. Thus, extensive training in the correct derivation of relations in the presence of a wide range of relational cues is required to maximize the fluency and effectiveness of the relational skill set.

During MET training phases, feedback is provided following every response. Because the topography of the stimuli presented across these tasks varies, the emphasis is placed on the relational cue and its arbitrarily applicable nature, rather than on the relationships specified between a given set of stimuli. A participant can, therefore, learn how to respond correctly to a given relational network regardless of the topography of the stimuli that are included within it (i.e., If A is more than B, then B is less than A, regardless of the form that A and B take). This facilitates the generalization of relational responding to a potentially infinite number of stimuli.

In a pilot study of the SMART system, Cassidy et al. (2011) isolated and trained coordination/opposition and comparison (more/less) relations in two studies using small samples of children. The first study administered a training program to eight normally developing children aged between 8 and 12 years old. The training system consisted of five stages: (1) stimulus equivalence training and testing; (2) MET for stimulus equivalence; (3) MET to establish the relational frame of coordination; (4) MET to establish the relational frame of opposition; and (5) MET to establish the relational frames of comparison (more than/less than). This program was implemented in its entirety to experimental participants during ten 90 minute sessions over a period of 6 weeks, while control participants completed only the first stage. In order to complete each level of training, participants were required to produce100% correct responding across a block of 16 trials. Analyses of changes in scores on the WISC (Wechsler, 2003) found average Full Scale IQ rises of 27 points for experimental participants, while control participants displayed a mean decrease of 2 points. Experimental participants also displayed significantly larger rises in Verbal (17 points) and Performance IQ (32 points) in comparison to control participants (0 and -4 respectively).

In the second of their studies, Cassidy et al. (2011) administered an abbreviated version of this training by omitting the first two training phases used previously. Eight children, aged between11 to 12, completed this training in bi-weekly sessions of 6 to 14 weeks (administered over 9 calendar months) in between WISC-IV IQ assessments. As part of this study, an additional metric was composed, called the Relational Abilities Index (RAI), which provided an estimate of an individual's level of relational responding proficiency in accordance with the same coordination, opposition and comparison tasks that comprised the training intervention. Following training, mean RAI scores increased from 58.5% correct responding to 92.4% correct responding. Mean Full Scale IQ scores increased from 83 to 96, alongside significant rises for three of the four IQ subscales; Verbal Comprehension (10 points), Perceptual Reasoning (12 points) and Processing Speed (26 points). The results of this study provided preliminary support for an MET training program as an efficacious means of increasing intellectual performance via relational skills training.

Several studies have subsequently extended and varied upon Cassidy et al. SMART paradigm in examining the effect of relational skills training on cognitive functioning (e.g., Cassidy et al., 2016; Hayes & Stewart, 2016; Thirus et al., 2016). The current 55-level program is an automated multiple exemplar training tool that is designed to increase the fluency and complexity of coordination, opposition and comparison responding across multiple online training sessions over many months. Each of these training levels exposes participants to between 1 and 4 relational premises using nonsense syllables as relata (e.g., BEF is more than CUG), followed by a relational question based on the relational network composed of these premises (e.g., is CUG more than BEF?). Participants respond to these questions by clicking either the “Yes” or “No” response option on screen, with a time limit of 30 seconds imposed for all trials. Each level isolates and trains specific relations within a network (e.g. first relatum – second relatum relations, second relatum – third relatum relations, first relatum – third relatum relations and so on). Training begins at the most basic level of complexity (simple reflexivity tasks; e.g. “ZIG is same as LER. Is LER same as ZIG?”), before gradually increasing in complexity. Task complexity varies by controlling; 1) the number of relational premises (1-4); 2) the order of relational premises (in sequential or random order); 3) the directionality of the relational question (i.e., whether or not the relational question probes for first term-last term relations, or last term-first term relations as specified in the premises); 4) the number of relation types presented in each trial (e.g., only same relations are specified, or a combination of same/opposite); and 5) the presence/absence of the relational cue used in the question in the relational premise(s) (e.g., ZIG is more than LER. Is ZIG less than LER?).

Using the online SMART tool as an intervention method, Cassidy et al. (2016) found a mean WISC-IV Full Scale IQ increase of 23 points in a sample of 15 school children aged 11-12 years, with the smallest increase for any participant being 14 points (i.e., approx. one standard deviation). Mean RAI scores also increased almost 15 points from 33.8 to 48.5 out of 55. In the second experiment, the effect of SMART on a widely-used group administered measure of educational aptitude, the Differential Aptitude Test (DAT; Bennet, Seashore, & Wesman, 1990) was assessed. SMART training was found to result in significant increases on three of the key DAT subscales (Verbal Reasoning, Numerical Ability and the Educational Aptitude composite score), as well as the RAI in a sample of thirty 15-17-year-olds.

Hayes and Stewart (2016) compared the efficacy of SMART training in improving performance across an extensive battery of intellectual and scholastic assessments using a sample of twenty-eight 10 and 11-year-old children. SMART was administered to experimental participants in 29 bi-weekly 1-hour sessions, while a control group matched for baseline ability completed the same amount of training using a computer coding training program called *Scratch™*. The testing battery consisted of the RAI, four Wechsler Abbreviated Scale of Intelligence (WASi, Wechsler, 1999) subtests, two Wechsler Intelligence Scale for Children (WISC) subtests, three Wechsler Individual Achievement Test (WIAT, Wechsler, 2005) scales, the Drumcondra Primary Reading Test – Revised (Educational Research Centre, 2007) and the Drumcondra Primary Mathematics Test – Revised (Educational Research Centre, 2006). Significant improvements were found in the SMART group alone for all three WIAT scales, WASi Block Design, WISC Digit Span & Letter/Number Sequencing and for DPMT-R. Despite failures to reach statistical significance for the other measures, performance on all measures improved to a greater extent for the SMART group when compared to the Scratch group.

Thirus et al. (2016) investigated the effect of SMART training on mathematical and logical reasoning in a controlled study using a sample of 21 high school students aged 16-18 years old. Following 8 to 10 weeks of relational training, experimental participants showed significantly greater increases in intellectual performance, as measured by Ravens Standard Progressive Matrices (Raven, 1981). However, SMART training did not appear to significantly increase performance on the non-standardized measures of mathematical performance administered. The authors propose that the SMART relational tasks may not be sufficiently complex to exert a positive impact on the high-level mathematical performance required of high school students. In addition, the authors point to a high attrition rate as a potential explanation for this lack of effect, insofar as only half of the experimental group completed the entire training program. When training progress was accounted for there was a significant relationship between training completion and gains in non-standardized tests of mathematical skills.

Most recently, Amd and Roche (in press) reported on a study that involved allowing a sample of vulnerable and underprivileged children in Bangladesh to access SMART training several times per week for several months. Time spent training and progress during training varied considerably across children due to social, political, and personal factors. Nevertheless, significant gains in IQ, as assessed by standard Raven's Progressive Matrices, were reported for those children who naturally fell into a high engagement cohort, and progress with the training program was significantly correlated with IQ gain, which in turn was not accounted for by baseline IQ.

The Current Study.

While the studies outlined here have produced promising results, reports of the kinds outlined above deserve a special kind of critical attention. That is, spurious claims that various training methods or practices can increase intelligence (e.g., The "Mozart" effect) have plagued psychology for decades and the popularity of such methods usually outlives emerging evidence that no such effects can be substantiated. With regard to SMART, studies from a small number of separate laboratories have been published, but each of these studies suffers from various methodological limitations. Specifically, all, barring Cassidy et al. (2016; Experiment 2), involved non-blinded and non-independent testers pre and post intervention. Indeed, in a comprehensive meta-analysis of the effect of blindedness on post-intervention treatment effects in randomly controlled trails, Schulz, Chalmers, Hayes and Altman (1995), report that such effects are exaggerated by approximately 17% due to non-blind tester bias. There is also an absence of control groups in both of the Cassidy et al.

(2011, 2016) studies and the Amd and Roche (in press) study, which has been identified as a key criticism of many intervention studies attempting to increase intelligence (Melby-Lervag et al., 2016; Redick, 2015; Redick et al., 2013; Shipstead, Redick & Engle, 2012). As such, a further and improved replication of the reported 'SMART' effect is required, with random participant assignment and blinded testers pre and post intervention.

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

Method

Participants

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

Settings & Materials

All assessments took place in a small room (3m x 3m approx.) within one of the school's two main buildings.

Wechsler Abbreviated Scale of Intelligence

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

Relational Training Protocol

The relational training intervention replicated the online SMART program implemented previously by Cassidy et al. (2016). This trained and tested relational responding proficiency in coordination/opposition and comparison relations), across 55 levels of progressive difficulty using an automated game-based MET method. The SMART program presents participants with relational networks, as specified by sets of relational premises which contain nonsense syllables as relata. It then asks participants to respond to a relational question based on this network. Answers are provided by clicking on the Yes or No buttons onscreen, with a time limit of 30 seconds for all trials. The use of MET training and nonsense syllables as trial stimuli facilitates the generalization of coordination/opposition and comparison relational responding to novel stimulus sets using during intermittent testing phases.

Each of the 55 training and testing levels consists of a distinct training stage and testing stage, each of which consists of relational tasks which isolate a particular type of relational responding (see Appendix 1 for more detail). To pass each training stage, correct responses are required on 16 consecutive tasks, each of which varies slightly in form within training level parameters, and always in terms of the arbitrary stimuli employed. Corrective feedback is provided on all tasks during all training stages. Following successful completion of the training stage, the following testing stage requires participants to respond correctly to all trials in a finite 16-trial block presenting tasks of the same type employed in the training that preceded it. Stimuli are novel on all trials and feedback is not provided. If the participant fails to produce 16 correct responses in the first run of 16 test trials, they are returned to the previous straining stage where training to criterion is again administered using yet more novel stimuli as relata. This cycle is related *ad infinitum*, using novel stimuli on every trial, until the participant passes both a training and successive testing phase. In this way, the user can pass all 55 stages of increasingly complex training and testing. For a full description of the procedural details of the SMART program, see Cassidy et al. (2016).

General Experimental Procedure

All participants were administered WASi IQ assessments at baseline. To ensure that the experimenters were blind to group membership, participants were then divided into two IQ matched groups by school authorities. The experimental group was then administered the SMART program in bi-weekly, 45-minute sessions within school hours over a 12-week period. During these sessions, the control group continued with their regular classroom activities. Following this training period, all participants were then retested using the WASi. Once the study was completed, access to the SMART program was offered to the control group for ethical reasons (i.e., not to deny treatment). This procedure conformed to the procedures approved by the Maynooth University Research Ethics committee for studies into SMART training involving children, as well as those laid down by the Psychological Society of Ireland.

Results

In total, just over half the experimental group (*n*=7) completed all training levels, with the mean number of completed levels being 41.6 out of 55. Mean Full Scale IQ scores were in the average range at baseline for both the Experimental (*M* = 99.2, *SD* = 16.25) and the Control groups (M = 98.9, SD = 8.4). Mean Verbal IQ scores were also within the average range for the Experimental (M = 100.6, SD = 17.84) and the Control group (*M* = 98.7, *SD* = 8.9). This was also the case for Performance IQ scores (Experimental: *M* = 97.8, SD = 13.3, Control: *M* = 98.1, *SD* = 9.17). Full descriptive statistics for IQ scores at baseline and follow-up are displayed in Table 1.

(Insert Table 1 here)

Full Scale IQ

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

(Insert Figure 1 here)

On average, Full Scale IQ scores recorded for the SMART intervention group increased by more than one full standard deviation (i.e., 18.8 points > 15 points) post intervention, which demonstrates a mean percentile rank increase of over 31% from approximately the 47^{th} percentile (*M* = 99.14) at baseline to the 88^{th} percentile (*M* = 117.9) at follow-up. This increase moved the average group IQ classification band from *average* to *high average*. There was no significant correlation between Full Scale, Verbal or Performance IQ score at baseline and the *change* IQ score which suggests that pre-test IQ score did not predict or account for the change in IQ score. For the experimental group, number of training levels completed did not correlate with subsequent IQ change. However, when taking the sample as a whole, there was a very strong significant correlation between these two metrics (r = . 84, p < .001). Mean scores for Full Scale, Verbal and Performance IQ are displayed in Figure 2.

(Insert Figure 2 here)

Verbal IQ and Performance IQ

A two-way (condition, time) mixed ANOVA was conducted for the Verbal IQ (VIQ) composite scores at Time 1 and Time 2. There was a within-subjects effect of time, $F(1, 24) = 41.17$, $p < .001$, $\eta_p^2 = .632$, and interaction of time*condition, $F(1, 24) = 44.22$, $p < .001$, $\eta_p^2 = .648$. The between-subjects effect of condition reached statistical significance, $F(1, 24) = 5.89$, $p = .023$, $\eta_p^2 = .197$. Follow-up paired samples *t*-tests found a significant increase in VIQ for the Experimental Condition from pre- (*M* = 100.58, *SD* = 17.83) to post-intervention (*M* = 120.58 , *SD* = 16.54), $t(11) = -7.78$, $p < .001$, 95% CI [-25.66, -14.34], but no significant difference between Time 1 (*M* = 98.71, SD = 8.89) and Time 2 (*M* = 98.4, *SD* = 8) VIQ for the Control Condition, $t(13) = 4.22$, $p = .845$, 95% CI [-3.5; 4.22].

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

IQ subtests

To gain a deeper understanding of the specific increases that mediate the observed effects of relational training on full scale IQ indices, a series of mixed between-within ANOVAs were conducted to assess changes in performance for each of the IQ subtests. For the Vocabulary subtest, there was a significant within-subjects effect of time, $F(1, 24) = 22.1$, $p < .001$, $\eta_p^2 = .479$, and an interaction of time*condition $F(1, 24) = 29.8$, $p < .001$, $\eta_p^2 = .554$. The between-subjects effect of condition was also significant, $F(1, 24) = 5.47$, $p = .028$, $\eta_p^2 = .185$.Paired samples *t*-tests found a significant increase in Vocabulary subtest scores from Time 1 ($M = 50.5$, $SD = 11$) to Time 2 ($M = 62$, $SD = 9.64$) for the Experimental group, $t(11) = -6.19$, $p < .001$, 95% CI [-15.59; -7.41], but the Control group failed to display significant rises from Time 1 ($M = 49.5$, $SD = 6.55$) and Time 2 ($M = 48.64$, $SD = 5.62$), $t(13) = .627$, $p = .54$, 95% CI [-2.09; 3.81].

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

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

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

In summary, results indicate a significant effect of SMART training in increasing scores for Full Scale IQ (*M* = 18.4 points), Verbal IQ (*M* = 20) and Performance IQ (*M* = 13.5), while scores for the Control group remained virtually unchanged for each of these measures. While the between subjects effect for Full Scale IQ was not statistically significant (*p* = .06), between-group effects were significant for Verbal IQ, Vocabulary and Matrix Reasoning scores. However, paired samples t-test results show significant increases for the three IQ indices and each of the four IQ subtests for the Experimental group, while the Control group did not display significant rises for any of these metrics. These results combine to add further support to assertions that relational skills training may be an efficacious mean of improving intellectual performance.

Discussion

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

Previous analyses have reported increases in verbal intelligence following relational skills training interventions, a finding which is replicated in the current analysis (Cassidy et al., 2011; 2016; Hayes & Stewart, 2016). The Verbal IQ rises reported in the current study (*M*=19.6 points) are similar to the 18 point increase reported in Cassidy et al. (2011) which used a small sample of 8- to 12-year old children. Hayes and Stewart (2016) also report significant rises in a number of verbal indices, such as WIAT Spelling, WIAT Reading, WISC Letter Number Sequencing and WISC Digit Span following SMART training, indicating that improvements in IQ scores following SMART training may further extend to increments in scholastic aptitude.

The efficacy of the SMART program in improving aspects of verbal intelligence is predicted by an extensive theoretical and empirical literature base proposing the importance of relational skill to language development and proficiency (Colbert et al., 2017; Sidman, 1994; Gore et al., 2010; Hayes et al., 2001; O'Connor, Rafferty, Barnes-Holmes, & Barnes-Holmes, 2009; O'Hora et al., 2008; Stewart & Roche, 2013). From an RFT perspective, word-word and object-word relations underpin language development (Stewart & Roche, 2013) and serve as the basis for linguistic reference. As such, relational responding proficiency would appear to facilitate the verbal intellectual performance as assessed by Verbal IQ subtests. As outlined by Cassidy et al., (2010) many of the Verbal subtests can be understood as tests of relational responding to a greater or lesser degree, and therefore predict gains for this IQ index. Indeed, numerous studies have reported significant correlations between measures of relational responding and Verbal IQ items (Colbert et al., 2017; Gore et al., 2010; O'Hora et al., 2008). In a correlational analysis of relational responding and scores on the WAIS-III, Colbert et al. (2017) reported moderate-to-strong statistically significant correlations between RAI scores and Verbal IQ, both Verbal IQ subindices (Working Memory & Verbal Comprehension) and 6 of 7 Verbal IQ subtests, indicating a wide ranging relationship between relational responding and virtually all aspects of verbal intelligence as assessed by the WAIS-III.

Of particular interest in the current discussion is the finding that while there are a number of intervention studies reporting success in improving intellectual function, very few have been able to produce such improvements as large and widespread as the SMART training program. There have been numerous training programs that have been proposed to improve intellectual function by targeting performance in very specific cognitive domains, such as working memory (e.g. Klingberg et al., 2002b; Klingberg et al., 2005), attention (Rueda et al., 2004), mental planning and strategy (Basak, Boot, Voss & Kramer, 2008) and general problem solving and creativity (Tranter & Koutstaal, 2008). However, none of these studies have demonstrated reliable rises using a full scale IQ assessment.

Some of the most noteworthy research on intellectual enhancement in recent times has focused on improving levels of working meory using what is called the *dual n-back* procedure (e.g. Buschkuehl, & Jaeggi, 2010; Jaeggi et al., 2008; 2011; Buschkuehl et al., 2008). This research reports more modest gains, typically of just a few standardized points, on a specific domain of intellectual performance (fluid intelligence) as assessed by matrix reasoning tasks. In addition, doubts have been raised over the generalizability of such findings to other intellectual domains, (Ackerman, Beier & Boyle, 2005; Colom, Abad, Quiroga, Shih & Flores-Mendoza, 2008; Kane, Hambrick & Conway, 2005; Moody, 2009) as such studies have shown insufficient evidence of far transfer (Colom et al., 2013; Lampit, Hallock, & Valenzuela, 2014; Melby-Lervag et al., 2016; Rapport, Orban, Kofler, & Friedman, 2013). Furthermore, a number attempts at replicating these results have not been successful (Chooi & Thompson, 2012; Lawlor-Savage & Goghari, 2016; Owen et al., 2010; Redick, 2015). Further doubts over the validity of Jaeggi and colleagues' findings have been raised due to methodological issues and procedural inconsistencies. For example, Jaeggi et al. (2008) did not administer the same assessment of working memory to each of their groups, with one group completing Ravens Advanced Progressive Matrices (RAPM; Raven, 1990) while the others were administered the Bochumer Matrices Test (BOMAT; Hossiep, Hasella, & Turck, 1999). In addition, the administration time allotted for the BOMAT was drastically reduced from 45 minutes to 10 minutes, without adequate rationale for doing so. As the score increases were found only for the three groups tested using the BOMAT, an inappropriately administered, less established measure of working memory, the reliability of such increases must be doubted. In an analysis of working memory training programs, Redick (2015) proposed that some post-intervention between-group differences in working memory were dependent on decreases in scores for the control group, rather than increases in scores for the experimental group.

In light of such limitations, the current experiment is added to the growing collection of studies (e.g., Amd & Roche, in press; Cassidy et al., 2011; 2016; Hayes & Stewart, 2016), supporting the efficacy of SMART training in improving intellectual performance as measured by gold-standard full-scale IQ assessments. While this result suggests that working memory training may not be the only potential means of increasing intellectual function, it also indicates the relational skills training may be a more effective and more consistent method of doing so. However, crucially, there is a burgeoning evidence base proposing that these increments in ability are not restricted to relational ability or IQ test performance alone, but have positive implications for practical applications such as academic ability (Cassidy et al., 2016; Thirius & Starbrink, 2016, Hayes & Stewart, 2016). This may be in part explained by the emphasis placed by the SMART program on improving *derived* relational responding, as, by definition, an increased sophistication in this skill facilitates *generalised* application to completely novel stimuli. The finding that a training program appears to boost scores on tasks which differ broadly in topography to the skills being trained (e.g. WASi Block Design and Matrix Reasoning) is somewhat of a rarity in interventions designed to increase intelligence.

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

Another limitation of the current study design may be the lack of a manipulation check of the variable being manipulated (i.e., relational skills). Studies generally take some form of relational skills assessment, such as the recently developed Relational Abilities Index (RAI; Colbert et al., 2017) at baseline and at follow-up in order to see if skills improvements have been made on a direct measure of the very skill being trained. However, because the training was administered by school authorities and not the researchers, this was not feasible. Having access to such measures would allow for more complex statistical analyses of the relationship between IQ gains and relational skills improvements and should be a feature of all future studies.

Not all participants completed the SMART training due to time limitations imposed on the training program and imposed for practical reasons by the relevant school. It can only be assumed at this stage that IQ gains may have been more impressive had every student completed the 55 stages of training typically required in other studies. The average number of levels completed for those who did not complete the training was approximately 26, which comprises almost all of the Same/Opposite task training levels and just under half of the complete training program (which also involves More/Less training). However, a post-hoc analysis found that the collective Full Scale IQ gain found for participants who failed to complete the training program was 18.2 points, an average increase only marginally lower than that found for those who completed all 55 levels (M = 19.1 points). As such, the number of training levels completed was not a predictor of subsequent IQ gains (as confirmed by correlational analyses), which is inconsistent with the "dosage effects" reported in the Amd & Roche (in press) study. From an RFT point of view, it may well be that as a primary and more utilized form of relational responding in daily life and education, increased proficiency at Same and Opposite derived relational responding on its own is sufficient to enhance a wider range of intellectual skills important during the teenage years (i.e., syllogistic reasoning, vocabulary expansion). Furthermore, while there is a clear effect of completing at least some training as opposed to no training at all (as evinced by between-group effects as well as correlational analyses), this lack of a dosage effect may indicate that while there is a relationship between SMART training and IQ gain, this relationship may not be linear. The number of students who failed to complete training in this study, however, is too small, to conduct in-depth analyses of the relationship between training progress and IQ gain. As such, further investigations should aim to isolate and investigate the effect of different levels of training on subsequent IQ gain.

The current SMART program is relatively limited in the breadth of relational skills training it provides. As discussed by Colbert et al. (2017), the inclusion of training procedures for a wider range of relational frames (such as temporal, categorical and analogical frames) may further extend the efficacy of the training program. Correlational studies have indicated that these frames are linked to various aspects of intelligence (Gore et al., 2010; O'Hora et al., 2008; McHugh et al., 2004). Therefore, improving proficiency in responding to such frames may expand the reach of SMART in terms of the intellectual domains it can benefit. Future studies should aim to further investigate the potential impact of training the wider range of relational frames on intellectual performance, as well as elucidating the relationship between proficiency in other forms of derived relational responding and IQ.

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

References

Aberg, M. A., Pedersen, N. L., Toren, K., Svartengren, M., Backstrand, B., Johnsson, T. et al. (2009). Cardiovascular fitness is associated with cognition in young adulthood. Proceedings of the National Academy of Sciences USA, *106*, 20906–20911. http://dx.doi.org/10.1073/pnas.0905307106

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

Ali, A., Ambler, G., Strydom, A., Rai, D., Cooper, C., & McManus, S. et al. (2012). The relationship between happiness and intelligent quotient: the contribution of socio-economic and clinical factors. *Psychological Medicine*, *43*(06), 1303-1312. http://dx.doi.org/10.1017/S0033291712002139

Amd, M. & Roche, B. (in press) Assessing the Effects of a Relational Training Intervention on Fluid Intelligence Among a Sample of Socially Disadvantaged Children in Bangladesh. *The Psychological Record*.

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

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

Barnes-Holmes, Y., Barnes-Holmes, D., & Smeets, P.M. (2004). Establishing relational responding in accordance with opposite as generalized operant behavior in young children. *International Journal of Psychology and Psychological Therapy, 4*, 559-586.

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

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

Barnett, S., & Ceci, S. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612-637. http://dx.doi.org/10.1037/0033-2909.128.4.612

Basak C., Boot W. R., Voss M., & Kramer A. F. (2008). Can training in a real-time strategy videogame attenuate cognitive decline in older adults? *Psychology and Aging, 23,* 765–777.

Bennett, G., Seashore, H,. & Wesman, A. (1990). *Manual for the Differential Aptitude Tests*. San Antonio: Psychological Corporation.

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

Buschkuehl, M., & Jaeggi, S. M. (2010). Improving intelligence: a literature review. *Swiss Medical Weekly, 140*(19–20), 266–272.

Buschkuehl, M., Jaeggi, S., Hutchison, S., Perrig-Chiello, P., Däpp, C., & Müller, M. et al. (2008). Impact of working memory training on memory performance in old-old adults. *Psychology And Aging*, *23*(4), 743-753. http://dx.doi.org/10.1037/a0014342

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

Cassidy, S., Roche, B., Colbert, D., Stewart, I., & Grey, I. (2016). A Relational Frame Skills Training Intervention to Increase General Intelligence and Scholastic Aptitude. *Learning and Individual Differences*, *47*, 222–235. http://dx.doi.org/10.1016/j.lindif.2016.03.001

Cassidy, S., Roche, B. & Hayes, S. C. (2011). A relational frame training intervention to raise Intelligence Quotients: A pilot study. *The Psychological Record, 61*, 173-198. http://dx.doi.org/10.1007/BF03395755

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

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

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

Coe, D. P., Pivarnik, J. M., Womack, C. J., Reeves, M. J., & Malina, R. M. (2006). Effect of physical education and activity levels on academic achievement in children. *Medicine and Science in Sports and Exercise, 38*, 1515–1519. http://dx.doi.org/10.1249/01.mss.0000227537.13175.1b

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

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

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

Colom, R., Roman, F. J., Abad, F. J., Shih, P. C., Privado, J., Froufe, M., Escorial, S., Martinez, K., Burgaleta, M., Quiroga, M. A., Karama, S., Haier, R. J., Thompson, P. M., & Jaeggi, S. (2013b). Adaptive nback training does not improve fluid intelligence at the construct level: gains on individual tests suggest that training may enhance visuospatial processing. *Intelligence, 41*, 712–727. http://dx.doi.org/10.1016/j.tins.2007.06.011

Cotman, C.W., Berchtold, N.C., & Christie, L.A. (2007). Exercise builds brain health: key roles of growth factor cascades and inflammation. *Trends in Neuroscience. 30*, 464 – 472. http://dx.doi.org/10.1016/j.tins.2007.06.011

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

Cuijpers, P., Smit, F., Bohlmeijer, E. T., Hollon, S. D., & Andersson, G. (2010). Is the efficacy of cognitive behaviour therapy and other psychological treatments for adult depression overestimated? A meta-analytic study of publication bias. *The British Journal of Psychiatry, 196*, 173−178.

Davis, C., Tomporowski P., Boyle C., Waller J., Miller P., Naglieri J., & Gregoski, M. (2007) Effects of aerobic exercise on overweight children's cognitive functioning: A randomized controlled trial. *Research Quarterly for Exercise and Sport, 78*(5), 510-519.

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

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

Dymond, S., & Roche, B. (2013). *Advances in relational frame theory*. Oakland, CA: Context Press.

Educational Research Centre (2006). *Drumcondra primary mathematics test – Revised. Levels 3–6 administration manual and technical manual*. Dublin, OH: Educational Research Centre, St Patrick's College.

Educational Research Centre (2007) *Drumcondra primary reading test – Revised. Levels 3–6 administration manual and technical manual*. Dublin, OH: Educational Research Centre, St Patrick's College.

Eikeseth, S., Smith, T., Jahr, E., & Eldevik, S. (2002). Intensive Behavioral Treatment at School for 4- to 7-Year-Old Children with Autism: A 1-Year Comparison Controlled Study. *Behavior Modification*, *26*(1), 49-68. http://dx.doi.org/10.1177/0145445502026001004

Elliott, R., Sahakian, B.J., Matthews, K., Bannerjea, A., Rimmer, J., & Robbins, T.W. (1997) Effects of methylphenidate on spatial working memory and planning in healthy young adults. *Psychopharmacology, 131*, 196–206. http://dx.doi.org/10.1007/s002130050284

Gentner, D., & Loewenstein, J. (2002). Relational language and relational thought. In J. Byrnes, & E. Amsel (Eds.), *Language, Literacy, and Cognitive Development* (pp. 87-120). Mahwah, NJ: LEA.

Gomez, S., Lopez, F., Martin, C.B., Barnes-Holmes, Y., & Barnes-Holmes, D. (2007) Exemplar training and a derived transformation of functions in accordance with symmetry and equivalence. *The Psychological Record, 57*, 273-293. http://dx.doi.org/10.1007/BF03395577

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

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

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

Hayes, S. (1994). *Behavior analysis of language and cognition*. Reno, NV: Context Press.

Hayes, S., Barnes-Holmes, D., & Roche, B. (2001). *Relational frame theory*. New York: Kluwer Academic/Plenum Publishers. http://dx.doi.org/10.1007/b108413

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

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

Holmes, J., Gathercole, S.E., & Dunning, D.L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*, F9–F15. http://dx.doi.org/10.1111/j.1467-7687.2009.00848.x

Hossiep, R., Hasella, M., & Turck, D. (1999). *BOMAT - advanced*. Göttingen [u.a.]: Hogrefe, Verl. für Psychologie.

Howard, J. S., Sparkman, C. R., Cohen, H. G., Green, G., & Stanislaw, H. (2005). A comparison of intensive behavior analytic and eclectic treatments for young children with autism. *Research in Developmental Disabilities*, *26*, 359–383. http://dx.doi.org/10.1016/j.ridd.2004.09.005

Hunter, J. E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement theory* (pp. 257-266). Hillsdale, NJ: Erlbaum

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447–472. http://dx.doi.org/10.1037/1076-8971.2.3-4.447

Jacobson, J. W., Mulick, J. A., & Green, G. (1998). Cost-benefit estimates for early intensive behavioral intervention for young children with autism: General model and single state case. *Behavioral Interventions*, *13*, 201–226. http://dx.doi.org/10.1002/(SICI)1099-078X(199811)13:4%3C201::AID-BIN17%3E3.0.CO;2-R

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

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

Kane, M., Hambrick, D., & Conway, A. (2005). Working Memory Capacity and Fluid Intelligence Are Strongly Related Constructs: Comment on Ackerman, Beier, and Boyle (2005). *Psychological Bulletin*, *131*(1), 66-71.

Kimberg, D., D'Esposito, M., & Farah, M. (1997). Effects of bromocriptine on human subjects depend on working memory capacity. *Neuroreport*, *8*(16), 3581-3585. http://dx.doi.org/10.1097/00001756-199711100-00032

Klingberg, T., Fernell, E., Olesen, P., Johnson, M., Gustafsson, P., Dahlstrıˆm, K., et al. (2005). Computerized training of working memory in children with ADHD—a randomized, controlled trial. *Journal of the American Academy of Child and Adolescent Psychiatry*, *44*(2), 177–186.

Klingberg T., Forssberg H., & Westerberg, H. (2002b). Training of working memory in children with ADHD. *Journal of Clinical Experimental Neuropsychology*, *24*, 781–791. http://dx.doi.org/10.1076/jcen.24.6.781.8395

Lampit, A., Hallock, H., & Valenzuela, M.J. (2014). Computerized Cognitive Training in Cognitively Healthy Older Adults: A Systematic Review and Meta-Analysis of Effect Modifiers. *PLoS Medicine*, *11*(11), e1001756. http://dx.doi.org/10.1371/journal.pmed.1001756

Lawlor-Savage, L., & Goghari, V. (2016). Dual N-Back Working Memory Training in Healthy Adults: A Randomized Comparison to Processing Speed Training. *PLOS ONE*, *11*(4), e0151817.

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

Lovaas, O. (1987). Behavioral treatment and normal educational and intellectual functioning in young autistic children. *Journal of Consulting and Clinical Psychology, 55,* 3-9. http://dx.doi.org/10.1037/0022-006X.55.1.3

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

Lynn, R., & Vanhanen, T. (2005). Does IQ Influence the Wealth of Nations? *Psyccritiques, 50*(13). http://dx.doi.org/10.1037/041316

Mackintosh, N. J. (1998). *IQ and human intelligence*. Oxford, UK: Oxford University Press.

Marr, M. (2015). Mathematics as verbal behavior. *Behavioural Processes*, *113*, 75-80. http://dx.doi.org/10.1016/j.beproc.2015.01.005

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

McKeel, A. N., Dixon, M. R., Daar, J. H., Rowsey, K. E., & Szekely, S. (2015). Evaluating the efficacy of the PEAK relational training system using a randomized controlled trial of children with autism. *Journal of Behavioral Education, 24*, 230–241. http://dx.doi.org/10.1007/s10864-015-9219-y

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

Melby-Lervåg, M., Redick, T., & Hulme, C. (2016). Working Memory Training Does Not Improve Performance on Measures of Intelligence or Other Measures of "Far Transfer". *Perspectives On Psychological Science*, *11*(4), 512-534.

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

Moran, L., Stewart, I., McElwee, J., & Ming, S. (2010). Brief Report: The Training and Assessment of Relational Precursors and Abilities (TARPA): A Preliminary Analysis. *Journal of Autism and Developmental Disorders*, *40*(9), 1149-1153. http://dx.doi.org/10.1007/s10803-010-0968-0

Neisser, U., Boodoo, G., Bouchard, T., Boykin, A., Brody, N., & Ceci, S. et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist*, *51*(2), 77-101. http://dx.doi.org/10.1037/0003-066X.51.2.77

Nelson, H. (1982). *National Adult Reading Test (NART): Test manual.* Windsor, U.K.: Nfer-Nelson.

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. http://dx.doi.org/10.1007/BF03395649

O'Hora, D., Pelaez, M., & Banres-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *The Psychological Record*, *55*, 155–175. http://dx.doi.org/10.1007/BF03395504

O'Hora, D., Pelaez, M., Barnes-Holmes, D., Rae, G., Robinson, T., & Chaudhary, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record, 58,* 569–583. http://dx.doi.org/10.1007/BF03395638

Open Science Collaboration (2012) An Open, Large-Scale, Collaborative Effort to Estimate the Reproducibility of Psychological Science. *Perspectives on Psychological Science 7*(6), 657–60. http://dx.doi.org/10.1177/1745691612462588

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

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

Rae, C., Digney, A., McEwan, S., & Bates, T. (2003). Oral creatine monohydrate supplementation improves brain performance: a double-blind, placebo-controlled, cross-over trial. *Proceedings Of The Royal Society B: Biological Sciences*, *270*(1529), 2147-2150. http://dx.doi.org/10.1098/rspb.2003.2492

Rapport, M. D., Orban, S. A., Kofler, M. J., & Friedman, L. M. (2013). Do programs designed to train working memory, other executive functions, and attention benefit children with ADHD? A meta-analytic review of cognitive, academic, and behavioral outcomes. *Clinical Psychology Review, 33*, 1237-1252.

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford: Oxford University Press.

Redick, T. (2015). Working memory training and interpreting interactions in intelligence interventions. *Intelligence, 50*, 14-20. http://dx.doi.org/10.1016/j.intell.2015.01.014

Redick, T. S., Shipstead, Z., Harrison, T. L., Hicks, K. L., Fried, D E., Hambrick, D. Z., Kane, M. J., Engle, R. W. (2013). No evidence of intelligence improvement after working memory training: A randomized, placebo-controlled study. *Journal of Experimental Psychology: General, 142*(2), 359-379.

Ree, M., & Earles, J. (1992). Intelligence Is the Best Predictor of Job Performance. *Current Directions In Psychological Science*, *1*(3), 86-89. http://dx.doi.org/10.1111/1467-8721.ep10768746

Remington, B., Hastings, R., Kovshoff, H., degli Espinosa, F., Jahr, E., & Brown, T. et al. (2007). Early Intensive Behavioral Intervention: Outcomes for Children With Autism and Their Parents After Two Years. *American Journal On Mental Retardation*, *112*(6), 418. http://dx.doi.org/10.1352/0895-8017(2007)112%5B418:EIBIOF%5D2.0.CO;2

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

Rowsey, K., Belisle, J., & Dixon, M. (2014). Principal Component Analysis of the PEAK Relational Training System. *Journal Of Developmental And Physical Disabilities*, *27*(1), 15-23. http://dx.doi.org/10.1007/s10882-014-9398-9

Rueda, M. R., Fan, J., McCandliss, B. D., Halparin, J. D., Gruber, D. B., Lercari, L. P. & Posner, M. I. (2004). Development of attentional networks in childhood.

Neuropsychologia, 42, 1029-1040. http://dx.doi.org/10.1016/j.neuropsychologia.2003.12.012

Scherer, R. W., Langenberg, P., & von Elm, E. (2007). Full publication of results initially presented in abstracts. *Cochrane Database of Systematic Reviews, 2,* Article MR000005. http://dx.doi.org/10.1002/14651858.MR000005.pub3

Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences, *Review of General Psychology, 13*, 90-100. http://dx.doi.org/10.1037/a0015108

Schmidt, F., & Hunter, J. (1996). Measurement error in psychological research: Lessons from 26 research scenarios. *Psychological Methods, 1*(2), 199-223. http://dx.doi.org/10.1037/1082-989X.1.2.199

Schulz, K.F., Chalmers, I., Hayes, R.J., Altman, D.G. (1995) Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *Journal of the American Medical Association, 273*(5), 408-412.

Shephard, R.J., Volle, M., Lavallee, H., LaBarre, R., Jequier, J.C., Rajic, M. (1984). Required physical activity and academic grades: A controlled longitudinal study. In I. Valimaki (Ed). *Children and sport* (pp.58-63). Berlin: Springer. http://dx.doi.org/10.1007/978-3-642-69465-3_8

Shipstead, Z., Redick, T.S., Engle, R. W. (2012) Is working memory training effective? Psychological Bulletin, *138*(4), 628-654. http://dx.doi.org/10.1037/a0027473.

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

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

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

Taatgen, N. (2013). The nature and transfer of cognitive skills. *Psychological Review*, *120*(3), 439-471. http://dx.doi.org/10.1037/a0033138

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

Thirus, J., Starbrink, M. & Jansson, B. (2016). Relational frame theory, mathematical and logical skills: A multiple exemplar training intervention to enhance intellectual performance. *International Journal of Psychology & Psychological Therapy, 16*(2), 141-155.

Thorell, L. B., Lindqvist, S., Bergman Nutley, S.,Bohlin, G. & Klingberg, T. (2009). Training and transfer effects of executive functions in preschool children. *Developmental Science, 12*(1), 106-133. http://dx.doi.org/10.1111/j.1467-7687.2008.00745.x

Tranter, L. J., & Koutstaal, W. (2008). Age and flexible thinking: An experimental demonstration of the beneficial effects of increased cognitively stimulating activity on fluid intelligence in healthy older adults. *Aging, Neuropsychology, and Cognition*, *15*(2), 184-207. http://dx.doi.org/10.1080/13825580701322163

Tuckman, B.W., & Hinkle, J.S. (1986) An experimental study of the physical and psychological effects of aerobic exercise on schoolchildren. *Health Psychology*, *5*(3), 197–207. http://dx.doi.org/10.1037/0278-6133.5.3.197

Watson, A., & Monroe, E. (1990) Academic achievement: A study of relationships of IQ, communication apprehension, and teacher perception. *Communication Reports, 3,* 28–36. http://dx.doi.org/10.1080/08934219009367498

Wechsler, D. (1949). *Wechsler Intelligence Scale for Children*. New York, NY: Psychological Corporation.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale manual*. New York, NY: Psychological Corporation.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children–Revised.* New York: Psychological Corporation.

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale–Revised manual.* New York, NY: Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children, Third Edition, manual*. San Antonio, TX: Psychological Corporation.

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

Wechsler, D. (1999). *Wechsler Abbreviated Scale of Intelligence manual*. San Antonio, TX: Psychological Corporation.

Wechsler, D. (2003). *Wechsler Intelligence Scale for Children—4th Edition (WISC-IV®)*. San Antonio, TX: Harcourt Assessment.

Wechsler, D. (2005). *Wechsler individual achievement test, 2nd edition (WIAT II)*. London, UK: The Psychological Corporation.

Wechsler, D. (2008). *Wechsler Adult Intelligence Scale, Fourth Edition technical manual*. San Antonio, TX: Psychological Corporation.

Zagorsky, J. (2007). Do you have to be smart to be rich? The impact of IQ on wealth, income and financial distress. *Intelligence*, *35*(5), 489-501. http://dx.doi.org/10.1016/j.intell.2007.02.003

Table 1

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

	Experimental		Control	
Measure	Baseline	Follow-up	Baseline	Follow-up
Full Scale IQ	99.2 (16.3)	117.9 (15.7)	98.9 (8.4)	99.1 (8.2)
Verbal IQ	100.6 (17.8)	120.6 (16.5)	98.7 (8.9)	98.4(8)
Performance IQ	97.8 (13.3)	111.3 (13.5)	98.1 (9.2)	99.9 (8.6)

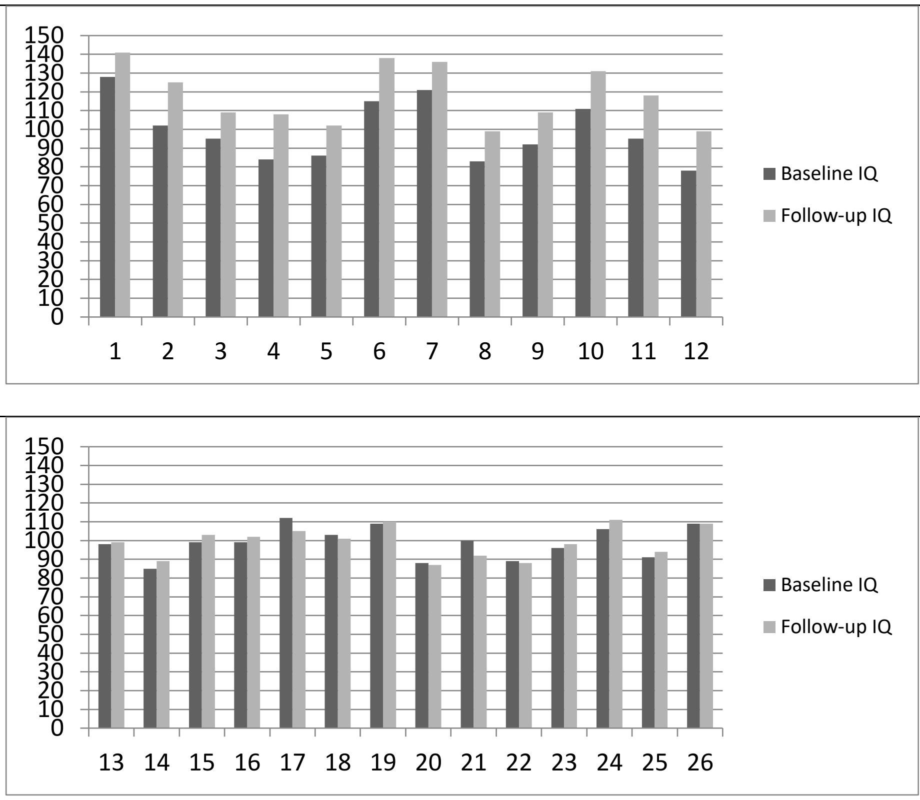

Figure 1. Histograms displaying WASI Full Scale IQ scores for SMART participants (top) and control participants (bottom) at baseline and follow-up.

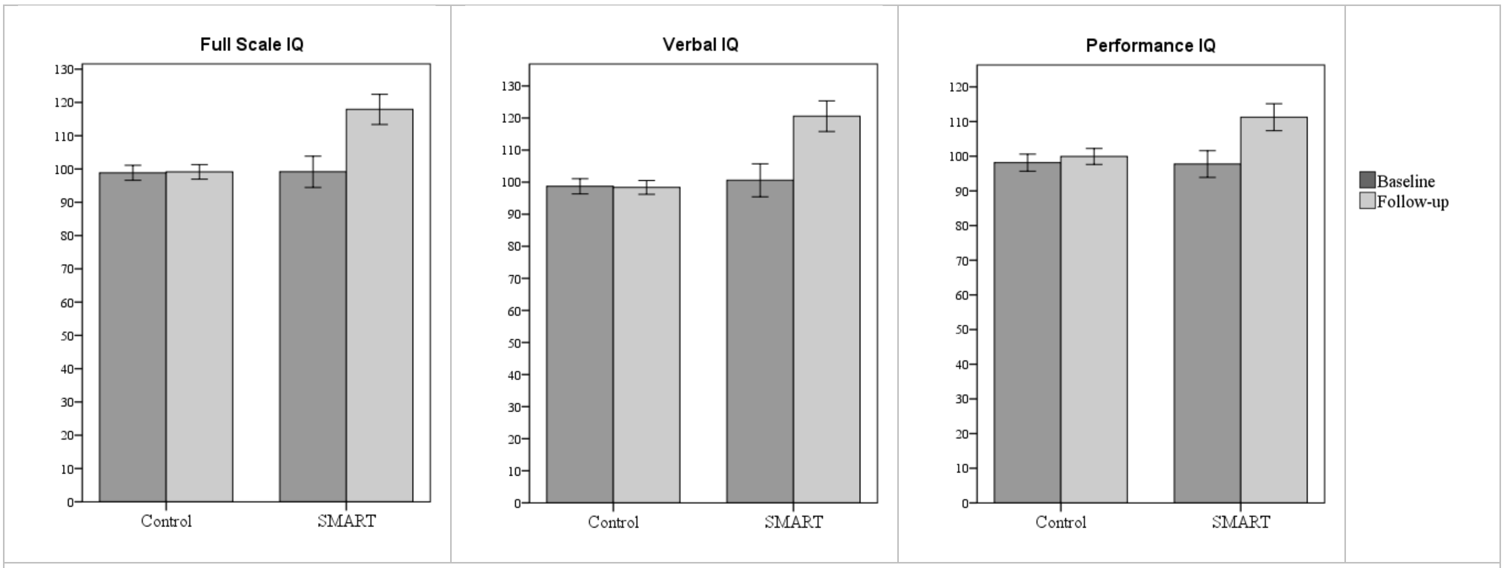

Figure 2. Mean IQ scores at baseline at follow-up for both SMART and control participants. Error bars represent standard error from the mean.