Piloting a Brief Relational Operant Training Program: Analyses of Response Latencies and

Intelligence Test Performance

Author Note

Abstract

Previous research suggests that training relational operant responding using the SMART (Strengthening Mental Abilities with Relational Training) program over several months can result in improved performance on cognitive intelligence tests. This study aimed to investigate whether engaging in a 3-week relational training program would improve (i) scores and (ii) reaction times on a standardised intelligence test, and (iii) to pilot a new multiple exemplar training procedure targeting complex analogical operant responding (SMARTA; SMART for Analogy). We administered the Kaufman Brief Intelligence Test (KBIT-2) to eight adults across four time points. Control: Time 1–4: No intervention. Experimental: Time 1–2: No intervention. Time 2–3: SMART relational operant training. Time 3–4: SMARTA analogical relational operant training. Experimental participants demonstrated greater improvements in terms of both (i) response latencies and (ii) response fluencies on the Verbal Knowledge subscale of the KBIT-2.

Keywords: SMART training, analogy, derived relational responding, response latencies, intelligence, relational frame theory

A Brief Relational Operant Training Program: Analyses of Response Latencies and Intelligence

Test Performance

The challenge of both predicting and influencing the capacity for language and cognition is

particularly important within the broader field of psychology (Stewart, 2016). Indeed, Hayes,

Barnes-Holmes, and Roche (2001) suggested that language and cognition underpin what is

referred to, more generally, as *intelligence*. Traditionally, intelligence testing has been primarily

the preserve of cognitive and educational psychology. Indeed, several popular tests (e.g.,

Wechsler Intelligence Scales; Stanford-Binet Scales) are comprised of several different subtests

supposedly reflective of distinct cognitive domains such as working memory, perceptual

reasoning, and verbal comprehension. It seems, however, that there is no general consensus

within the broad field of cognitive psychology on what the specific components (i.e., internal

structures) of intelligence are, with many conflicting theories abound (e.g., Cattell, 1963;

Gardner, 1983; Guildford, 1967; Sternberg, 1985; Thurstone, 1938). Nonetheless, cognitive

intelligence tests remain popular as many of them (e.g., Academic Aptitude Test, Mental Survey

Test, Raven's Progressive Matrices, Stanford-Binet, Terman Group Intelligence Test, and

Wechsler-Bellevue Test) have shown some predictive validity in numerous contexts, including

for predicting salary, education level, and occupation (see Strenze, 2007, for a full review). Thus,

interventions designed to enhance performance on such intelligence tests would be useful and

desirable tools to develop.

Psychologists have, thus far, generally struggled to significantly increase performance

scores on these standardised tests (see Flynn, 2016 for a book-length discussion; see also Jäeggi,

Buschkuehl, Jonides, & Perrig, 2008; Jäeggi, Buschkuehl, Jonides & Shah, 2011; Jäeggi, Studer-

Luethi, Buschkuehl, Su, Jonides & Perrig, 2010). From a behaviour-analytic point of view,

inferences of cognitive psychology regarding intelligence as a quantifiable mental construct based on factor analyses of answers on a set of arbitrary questions are not an adequate explanation of what human intelligence entails (see Schlinger, 2003). For example, Schlinger (1992) suggested that attempts to simulate human intelligence should focus on behaviour-environment relations and not on hypothetical internal structures such as working memory. In this vein, applied behaviour analysis has been broadly successful in terms of assessing and training patterns of adaptive behaviour (e.g., operant responding) that might be considered hallmarks of what we mean by "intelligence", as a lay term. For instance, Carpentier, Smeets, and Barnes-Holmes (2003) successfully trained 5-year-old children who had not reached the developmental cusp of being able to respond analogically (i.e., to relate stimulus relations) in that pattern of behaviour. Similarly, Nuzzulo-Gomez and Greer (2004) used multiple exemplar training to train children how to mand (a "requesting" operant).

According to one track of behaviour-analytic research known as Relational Frame Theory (RFT; Hayes et al., 2001), human language and cognition are synonymous, and both are characterised by the ability to respond symbolically. According to RFT, the key skill underlying the complexity and generativity in language and cognition is arbitrarily applicable relational responding (AARR). AARR, writ large, is the behaviour of responding to one event in terms of another based on their symbolic properties. For example, "left" is only understood relative to "right" (spatial responding; May, Stewart, Baez, Freegard, & Dymond, 2017), "more" relative to "less" (comparative responding; Dymond & Barnes, 1995), "before" relative to "after" (temporal responding; O'Hora, Pelaez, Barnes-Holmes, Rae, Robinson, & Chaudhary, 2008), "part" relative to "whole" (hierarchical responding; Slattery & Stewart, 2014), or "same" relative to "different" (equivalence; Sidman, 1971). According to RFT, AARR is the behaviour of

responding relationally in accordance with arbitrary (non-physical) properties of behavioural events. This pattern of behaviour develops as a function of responding in accordance with formal (physical) properties of these events (non-arbitrarily applicable relational responding; NAARR) across multiple exemplars. For example, it is possible to condition either a human or a non-human for choosing a picture of a ball in the presence of another picture of a ball. This is simple matching behaviour based on physical similarity of stimuli. After multiple of exemplars of such non-arbitrary relational responding wherein a relational pattern of behaviour has been consistently reinforced by the verbal community, humans have the unique capacity to abstract the pattern of behaviour and apply it to symbols, such as words, as a generalised operant (see Barnes-Holmes & Barnes-Holmes, 2000). Following on from the above example, if several instances of matching behaviour (e.g., matching shapes to shapes, matching colours to colours) are reinforced at the same time as the utterance of the word "same", then matching behavior will come under stimulus control of the word "same" such that it becomes a generalised operant. Subsequently, I might be able to use the word "same" to indicate that formally unrelated stimuli (e.g., a ball and the word "pelota" – Spanish word for ball) should be treated as having the same functions, which is not necessarily apparent based on their formal properties.

There are three core properties of AARR. The first is mutual entailment: given the relation "A is to B" (hereinafter A:B), I can derive B:A. For relations such as "same", "different", and "opposite", the A:B relation will be the same as the derived relation, B:A (e.g., if A is the same as B, then B is the same as A; if A is opposite to B, then B is opposite to A). For other relations such as "more/less", the generalised operant follows an asymmetrical pattern of behaviour (e.g., if A is more than B, B is not also more than A). The second property of AARR is combinatorial entailment. Here, given an A:B relation and a B:C relation, it becomes possible

not just to derive the mutually entailed relations in both cases (i.e., B:A and C:B) but also the combinatorial A:C and C:A relations. For instance, if one dollar (A) is less than one Euro (B), and one Euro (B) is less than one British pound (C), then I can derive that one pound (C) is more than one dollar (A). The third property of AARR is transformation of stimulus functions, where stimulus functions change in accordance with their relations with other stimuli. In this case, I might derive the C:A relation that one pound is more than one dollar, but I will also behave in accordance with that relation by treating the pound as more valuable by choosing it given the choice between having a pound or a dollar.

However, AARR is also a necessarily contextually controlled set of behaviours. For example, if I am in the USA with British pounds and have no currency exchange available, it will not be true to say that the pound is more valuable in that context. The function and context of behaviour dictate the derived relation. This uniqueness of each organism's history of reinforcement is one reason given for the pragmatic nature of human behaviour (see Barnes-Holmes, 2000). Nonetheless, it appears that multiple exemplar training of generalised operants allows the organism to adapt and function effectively across multiple environments / contexts.

One program of research (outlined below) uses an automated online relational operant training program known as "Strengthening Mental Abilities with Relational Training" (SMART). This program trains AARR via multiple exemplar training in a gamified format. Users can earn progress points and various medals based on their progress through the program, which is comprised of 70 stages of incrementing complexity, training both symmetrical (same/opposite) and asymmetrical (more/less) AARR. Each stage is made up of 16 training trials, followed by a block of 16 test trials. To pass a stage, a user must respond correctly on 16 consecutive trials. During training, reinforcement is provided after each trial (by way of a

progress-bar that resets upon answering incorrectly, in addition to correct/incorrect specific

sound effects). Reinforcement is not provided during testing block trials. Relational skills are

assessed both before and after the training in an assessment comprised of a single trial from each

of the first 55 stages of the training program (no feedback is provided). Relational skill is then

determined using an accuracy criterion (number of answers correct out of 55).

The utility of SMART has been tested in various ways. For example, Cassidy, Roche, and

Hayes (2011) assessed four children (aged 10-12) using the Wechsler Intelligence Scale for

Children (WISC III-UK) at baseline, following stimulus equivalence training, and again after

taking part in the SMART program. Four other children (aged 8-12) were assessed at each of

three time-points but they did not receive training as control participants. Experimental

participants showed significant rises in IQ following both stimulus equivalence training and after

AARR training using SMART. Control participants' IQ did not improve. Cassidy et al. (2011)

then conducted a second experiment with children with educational difficulties, but with no

stimulus equivalence training this time. There was a significant increase in IQ after SMART

training such that participants with educational difficulties were indistinguishable from their

typically developing peers. These initial results have since been corroborated. For example, in

Cassidy, Roche, Colbert, Stewart, and Grey (2016), fifteen 11-12 year olds were given SMART

training and the mean IQ (as measured using the WISC IV-UK) rose from 97 to 120. In the

second experiment of Cassidy et al. (2016), they administered the same intervention to 15-17

year old children ($n = 30$). This time, however, they were assessed using the Differential

Aptitude Test, fifth edition (DAT 5). Significant improvements were observed on both the verbal

and numerical subscales of the DAT 5 in almost all cases.

Hayes and Stewart (2016) compared the efficacy of SMART training and an active control group intervention which was comprised of computer programming training (also considered to involve abstract relational reasoning) for improving performance in a range of tasks related to memory, literacy, and numeracy in children aged 10-11 years. Significant post-training improvements were observed on the digit span letter/number sequencing task, spelling, reading, and numerical operations. The improvements on these tests were significantly greater in the SMART training group compared to the coding group. A number of other relational skills training interventions have also been of benefit for improving various other abilities (see Parra & Ruiz, 2016; Ruiz & Luciano, 2011; Thirus, Starbrink, & Jansson; 2016; Vizcaíno-Torres et al., 2015). Taken together, these studies show some evidence for the utility of relational skills training and for RFT's assertion that AARR is the key skill underlying human intelligence (see also O'Hora, Pelaez, & Barnes-Holmes, 2005; O'Hora et al., 2008; O'Toole, Murphy, Barnes-Holmes, O'Connor, & Barnes-Holmes, 2009; but see Cassidy, Roche, & O'Hora, 2010 for a critical discussion of IQ test items in terms of relational skills). Nonetheless, there are some issues which merit addressing.

Thorndike's (1920, 1926) Anarchic Theory of Intelligence is perhaps the earliest of the sophisticated behavioural accounts of the main constituents of intelligent behaviour in humans. In this account, he outlined three main types of intelligence: (i) abstract intelligence, the ability to deal with symbols; (ii) concrete intelligence, the ability to deal with stimuli on the basis of physical properties; and (iii) social intelligence, the ability to deal with other people. In addition, Thorndike (1926) proposed four attributes of each intelligence: (i) "level", or the difficulty of task which can be undertaken; (ii) "range", or the variety of exemplars of a task of a given

difficulty which can be dealt with; (iii) "area", the overall variety of task to which one can respond effectively; (iv) "speed", the response latencies on given task types.

Thorndike's three main types of intelligence can be understood in RFT terms. Abstract intelligence and AARR are synonymous, as are concrete intelligence and NAARR (see Cassidy et al., 2010). According to RFT, social intelligence may involve more complex examples of AARR, such as self/other relational responding (Barnes-Holmes, McHugh, & Barnes-Holmes, 2004) or the relating of relational networks (McLoughlin & Stewart, 2017). Given that NAARR precedes AARR (see Hayes et al., 2001), one might argue that it is focusing on training complex AARR that is most important in terms of appropriately assessing the major facets of intelligence in accordance with Thorndike's (1920, 1926) guidelines. SMART focuses on training AARR. However, RFT and SMART research to date has not measured all of the attributes of AARR that Thorndike (1926) suggested.

The level/difficulty of task is well accounted for in the SMART program as the critical test of relational ability involves a test of AARR of increasing complexity. SMART also ensures that users can deal with a range of exemplars of each task type because SMART involves deriving relations between nonsense syllables, none of which appear more than once for each user. The overall variety of tasks to which one can respond effectively is not particularly high in that there are only modules for same/opposite and more/less, when RFT also proposes several other kinds of AARR which might be trained (see above). One might argue that training in symmetrical (same/opposite) and asymmetrical (more/less) AARR would be enough to increase intellect, since many other forms of relational responding follow the same behavioural patterns (e.g., member/class, before/after, above/below all involve asymmetrical responding). The utility

of measuring and training multiple asymmetrical patterns of AARR is ultimately an empirical

question.

The final attribute of intelligence that Thorndike (1926) proposed as being important to

measure is response latency. To date, SMART studies and most IQ tests do not take response

latency into account when determining relational skill or intelligence respectively (although time

limits are often imposed). However, Ramey et al. (2015) argue that increasing one's response

speed is a crucial outcome of behavioural skills training. To develop a better understanding of

how AARR relates to intelligence, it will be important to measure the effect of relational operant

training on both speed and accuracy of responding to trials (i.e., IQ test items). In the current

study, we also calculated how many correct responses participants emitted per unit of time (i.e.,

accuracy divided by response latency). This is commonly referred to as a *fluency* outcome

measure and many behavioural researchers consider changes in fluency to be an important

measure of the effectiveness of precision teaching (Binder, 1996; Binder, Haughton, & Bateman,

2002; Murphy & Barnes-Holmes, 2017). Indeed, the Digit Symbol Substitution Test of response

fluency is one of the sub-tests of the Wechsler Adult Intelligence Scales (see e.g., Crowe et al.,

1999).

One final critique of SMART training is that it only trains the ability to perform AARR

behaviours in a linear format, which does not fully represent the complexity of AARR in the real

world (particularly when it comes to complex ideas such as "social intelligence"). For example,

relating derived stimulus relations is a distinct skill (commonly referred to as analogical

responding) which develops later than the ability to derive relations linearly, but this ability can

be trained using multiple exemplar training (see Carpentier et al., 2003). One recent study

(McLoughlin & Stewart, 2017) demonstrated that it is possible to derive stimulus relations and

stimulus functions across complex relational networks beyond the definitional scope of analogical responding. For example, analogical responding is said to occur when two arbitrary stimulus relations are treated as being the same or not (a binary decision: same or not same), but it is also possible to specify that relations can be different (anything not the same) or opposite (specifically antithetical to) one another. To that end, the current experiment involved training participants using both (i) the SMART program and (ii) a new pilot intervention targeting complex relational networking skills (SMARTA: SMART for analogical responding), and measured performance, including both accuracy and response latencies, on a brief mainstream test of intellectual ability.

Method

Participants

Eight volunteers participated in this experiment (see Table 1 for demographics). Four participants were randomly assigned to take part in the Experimental Condition (relational training), while four were assigned to a Control Condition (no training). No participants had previously diagnosed specific learning difficulties.

*** Table 1 ***

Materials

We measured the intelligence construct using the Kaufman Brief Intelligence Test, Second Edition (KBIT-2) at four time points. The KBIT-2 is a brief test of intelligence for people aged 4-90 years. It consisted of three subtests: (i) Verbal Knowledge, (ii) Matrices, and (iii) Riddles. A Verbal IQ composite is calculated based on the Verbal Knowledge and Riddles subscales, while a Non-Verbal IQ is calculated based on Matrices only. The Verbal Knowledge subtest of the KBIT-2 involves choosing between multiple answers following a discriminative

stimulus (SD). The Matrices subtest involved choosing between multiple possible answers in

order to complete a picture. The Riddles subtest involved asking open-ended questions (e.g.,

name something that bleats; again, not an actual item). Kaufman and Kaufman (2004) reported a

correlation coefficient of .89 between the overall KBIT-2 IQ composite and the Wechsler Adult

Intelligence Scale, Third Edition (WAIS III; Wechsler, 1997) full-scale IQ composite. The

KBIT-2 has a mean test-retest reliability of .83 for the Non-Verbal sub-scale, .91 for the Verbal

scale, and .90 for the overall composite score (Kaufman & Kaufman, 2004). We chose the

KBIT-2 because it can be administered in only 20 minutes and our assessment time for this study

was limited due to the amount of training already required.

We administered an automated version of the KBIT-2. The software recorded accuracy

and reaction time and then we computed a fluency score by dividing the percentage of correct

answers by the response latency.

Design and Procedure

Control participants did not receive an experimental intervention across this three-week

study. Experimental Condition participants did not complete any training during Week 1; they

were requested to complete the SMART relational operant training online on four days of their

choosing, for a minimum of 30 minutes per session in Week 2; then finally, they completed the

SMARTA relational operant training on four days negotiated with the experimenter, during a 30

minute session supervised by the experimenter, in Week 3. The independent variable for this

experiment was *relational operant training* with three levels: (i) no training, (ii) simple relations

(SMART), and (iii) complex relations (SMARTA) and the dependent variables were response

accuracy (number of answers correct), response speed (time taken to completion), and response

fluency (accuracy / speed) in each sub-section of the cognitive ability test. Table 2 summarises this procedure.

*** Table 2 ***

For each sitting of the KBIT-2 assessment, participants completed the assessment at their own pace. We administered this test at the start of Week 1, at the end of Week 1, at the end of Week 2, and at the end of Week 3 (hereinafter Time 1, Time 2, Time 3, and Time 4, respectively).

The experimenter programmed the SMARTA software for the training in Week 3 in MS Visual Basic 6.

Training Interventions. There were two training interventions in this study. Both involve multiple exemplar training of the ability to derive relations between arbitrary nonsense syllables based on English language contextual cues.

SMART. This training system typically focuses on training the ability to derive relations of coordination (i.e., equivalence), opposition (see Figure 1, upper panel), more than, and less than (see Figure 1, lower panel). This training is typically carried out over several months and involves progressing users through up to 70 stages of incrementing complexity. In that training format, users unlock new stages of increased complexity upon mastery of the previous stage. Training in that format limits the user to progressing through a maximum of five new stages per day. However, we were interested in testing the utility of SMART as a brief intervention, so participants were instructed to use the *Brain Agility* mini-game included as an extra feature of the training software. In this, trials are presented at random from the full array of 70 trial types which vary in complexity. Forty-one percent (i.e., 41%) of the randomised trials were SAME/OPPOSITE trials, while the rest were MORE/LESS trials. The network nodal distances

were random, ranging from a nodal distance of one (an A-B network) to four (an A-E network).

During *training* blocks, the training software provided corrective feedback provided after each

response. In order to "pass" a training block, participants were required to answer 16 trials in a

row correctly. During a test phase, the same array of trial types were presented, but without

corrective feedback. No two relational networks were seen twice across either training or testing

phases; each nonsense syllable stimulus was only ever seen in one trial.

<center>*** Figure 1 ***</center>

By default on each trial, users initially had a maximum of 30 seconds in which to respond

to the trial. If the time limit lapsed, the answer was recorded as incorrect and a new trial was

presented. The 30-second time limit could be adjusted in order to either turn the time limit off or

to reduce it in order to provide more of a challenge. These data were not recorded in this study.

However, participants were instructed to use the features of the Brain Agility mini-game to

challenge themselves: Once they could comfortably answer all trials correctly at a given level of

complexity within the 30-second time limit, they should reduce the time allowed in an effort to

maintain their response accuracy under increasingly stringent time constraints.

SMARTA. Trials in this training program involved the same basic format. The main

difference was that were required to derive two relations per trial and subsequently identify the

relation that holds between those stimulus relations. For example, in Figure 2 (upper panel), it is

possible to derive the A:B relation (ENE is more than FOP) and the C:D relation (ANJ is more

than ENE). The critical question in this trial is whether the A:B relation is opposite the C:D

relation. Specifically, in this case, "is a 'more than' relation opposite to a 'more than' relation?"

is the question to be answered correctly by selecting "NO". The trials always involved deriving

two relations in a format identical to the one used in the SMART program before subsequently

relating those relations. In other words, the relating of stimulus relations itself accounted for a maximum of one third of each SMARTA trial, and less if the relata were derived over a larger nodal distance.

*** Figure 2 ***

In the SMARTA program, participants first completed an initial block of 48 unique trial types with no feedback provided. These trials were presented in quasi-random order; no trial type was presented twice. After this initial assessment, the data output identified which trial types were answered incorrectly. Participants could then train (i.e., with feedback) and test (no feedback) specific problematic trial types before returning to the 48 trial assessment to practice a broader array of trial types (see Figure 2, lower panel). We employed a training and testing criterion of answering 16 trials correctly in a row in the SMARTA training stages, as was the case in the SMART intervention before it. After each block of 48 trials, the experimenter checked the data output in order to identify trial types for practice.

Results

Mixed design analyses of variance assessed the effectiveness of SMART and SMARTA respectively. These figures should be interpreted with caution due to the low sample sizes in this study. Table 3 and Figure 3 allow readers to interpret these results at face value in a similar manner to a single subject design report.

Fluencies

We conducted a series of 2*4 mixed design ANOVAs to examine fluencies (see above for a description) on (i) the Verbal Knowledge KBIT-2 subscale, (ii) the Matrices KBIT-2 Subscale, (iii) the Riddles KBIT-2 subscale, and (iv) the Full KBIT-2 assessment across the

between-subjects factor: Condition (Experimental, Control) and the within-subjects factor: Time

(Time 1, Time 2, Time 3, Time 4).

Verbal Knowledge. Results showed a significant main effect of Time on Fluency on the

"Verbal Knowledge" sub-test of the KBIT-2 ($F(1.23, 7.38) = .02$, $MSE = .006$, $p = .002$, $\eta_p^2 =$

.777). There was no significant main effect of Condition. However, there was an interaction

effect of Time and Condition on Verbal Knowledge Fluency ($F(1.23, 7.38) = 10.10$, $MSE = .01$,

$p = .012$, $\eta_p^2 = .63$). We used post-hoc pairwise comparisons using a Bonferroni correction to

examine simple comparisons in more detail.

Experimental group. From Time 1 to Time 2 there was no difference in KBIT-2 Verbal

Knowledge Fluency. From Time 2 ($M = .30$, $SD = .03$) to Time 3 ($M = .49$, $SD = .08$) there was a

significant increase in KBIT-2 Verbal Knowledge Fluency (95% CI [-.08 to .04], $p = .001$). From

Time 3 to Time 4 ($M = .73$, $SD = .21$) there was a marginally significant increase in KBIT-2

Verbal Knowledge Fluency (95% CI [-.48 to .01], $p = .055$). From Time 2 to Time 4 there was a

significant increase in KBIT-2 Verbal Knowledge Fluency (95% CI [-.70 to -.15], $p = .006$).

Control group. From Time 1 ($M = .41$, $SD = .15$) to Time 2 ($M = .48$, $SD = .16$) there

was a significant increase in KBIT-2 Verbal Knowledge Fluency (95% CI [-.12 to -.01], $p = .03$).

Across all other levels of the Time condition, there were no significant differences in KBIT-2

Verbal Knowledge Fluencies.

Matrices. Results showed no significant main effect of Time or Condition on Fluency on

the "Matrices" sub-test of the KBIT-2.

Riddles. Results showed a significant main effect of Time on Fluency on the "Riddles"

sub-test of the KBIT-2 ($F(3, 18) = 16.82$, $MSE < .001$, $p < .001$, $\eta_p^2 = .74$). There was no main

effect of Condition and no main interaction effect of Time and Condition on Riddles Fluency.

Full Fluency. Fluency changes for the full KBIT-2 are illustrated in Figure 3. Results showed a significant main effect of Time on Fluency on the full KBIT-2 ($F(1.13, 6.81) = 16.03$, $MSE = .01$, $p = .002$, $\eta_p^2 = .73$). There was no significant main effect of Condition. There was also a significant overall interaction effect of Time and Condition on Full Fluency ($F(1.13, 6.81) = 7.30$, $MSE = .002$, $p = .029$, $\eta_p^2 = .55$). We used post-hoc pairwise comparisons using a Bonferroni correction in order to examine simple comparisons in more detail.

Experimental group. From Time 1 ($M = .17$, $SD = .02$) to Time 2 ($M = .19$, $SD = .02$) there was no significant increase in KBIT-2 Fluency. From Time 2 to Time 3 ($M = .28$, $SD = .06$) there was a significant increase in KBIT-2 Riddles Fluency (95% CI[-.15 to -.03], $p = .009$). From Time 3 to Time 4 ($M = .41$, $SD = .13$) there was a significant increase in KBIT-2 Fluency (95% CI[-.26 to <-.001], $p = .05$). From Time 2 to Time 4 there was a significant increase in KBIT-2 Fluency (95% CI[-.39 to -.04], $p = .017$).

Control group. From Time 1 ($M = .24$, $SD = .09$) to Time 2 ($M = .27$, $SD = .10$) there was a significant increase in KBIT-2 Fluency (95% CI[-.06 to -.01], $p = .019$). Across all other levels of the Time condition, there were no significant differences in KBIT-2 Fluencies.

*** Figure 3 ***

KBIT-2 Intelligence (accuracy) and Latency Scores Summary

We conducted a series of 2*4 mixed design ANOVAs in an experiment where (i) the non-verbal intelligence score, (ii) the verbal intelligence score, and (iii) the full KBIT-2 intelligence score were recorded and compared across a between-subjects factor: Condition (Experimental, Control) and a within-subjects factor Time (Time 1, Time 2, Time 3, Time 4). There were no statistically significant changes in IQ scores (i.e., the accuracy criterion). We conducted a similar analysis of response latencies across the four time points and found that the

changes in fluency were due to reductions in response latencies (see Table 3) and not due to increases in accuracy.

*** Table 3 ***

Furthermore, a qualitative view of changes in reaction speed amongst individual participants shows that one control participant (CON2 in Table 4) became significantly faster from Time 1 to Time 4 overall. In this case, it is possible that this is the result of testing error at Time 1 because they did not demonstrate notable improvements after Time 2.

Table 4

Discussion

This study sought to implement a similar design to Cassidy and colleagues (2011), but with shorter training intervals. Here we will first discuss fluency changes on the KBIT-2 and its subscales, and then attempt to determine which of the constituents of fluency (accuracy or latency) were changed due to training. Answering a particular question or set of questions quickly is only desirable if they are answered correctly. To that end, we considered fluency scores to be the main focus of the present analysis.

On the Verbal Knowledge subtest of the KBIT-2, there was no change in fluency in the Experimental group from Time 1 to Time 2 (no intervention), as expected. From Time 2 to Time 3 (after SMART) and from Time 3 to Time 4 (after SMARTA) there was a significant increase in fluency. Unexpectedly, there was an increase in the fluency of Control participants from Time 1 to Time 2 (no intervention) which requires explanation. From the current data, one might suggest that this could be a practice effect which might have presented itself just as easily from Time 1 to Time 2 in either the Experimental or Control group on any particular subtest, particularly where the outcome variable takes response latency into account. As a matter of

logistical necessity, participants were re-tested at weekly intervals, making a practice effect all the more likely. To overcome this in future, studies should ideally conduct their repeated measures temporally further apart, or perhaps use measures of behavioural performance that use arbitrary stimuli (e.g., working memory tasks, as mentioned above) instead of recall-based tests wherein the experimenter has little control over both previous learning opportunities and the salience of the items to be remembered.

There was no change in fluency on the matrices subtest of the KBIT-2 in either the Experimental or Control group. This is unsurprising because there was also change in neither the Matrices accuracy nor the Matrices response latency, the two constituents of fluency. It is possible that the relational skills trained (same/opposite, more/less, analogy) are less relevant to the Matrices tasks than to spatial relational responding (see May et al., 2017).

There was a significant main effect of Time on Riddles fluency, and no main effect of Condition, but no interaction effect of Time and Condition. This suggests that although fluency improved over time, there were improvements in both the Experimental and Control groups. As previously mentioned, the Riddles subtest is a test with multiple assumptions regarding participants' learning histories and with vague and competing sources of stimulus control over answering quickly and correctly that are not controlled or accounted for. A practice effect may have occurred, but relational skills training was not demonstrably effective for improving performance on the Riddles subtest; simple re-exposure at weekly intervals was enough to improve fluency. Given the multiple determinants of Riddles performance, there are also multiple extraneous sources of error. In future, it may be desirable to study the effects of relational training on Riddles performance using larger samples and better control conditions in order to better control for error.

Overall fluency changes (see Figure 3) across all the subscales of the KBIT-2 bore out more or less as expected. In the Experimental group, fluency increased from Time 2 to Time 3 (SMART) and from Time 3 to Time 4 (SMARTA), but not from Time 1 to Time 2 (no intervention). The opposite was true in the Control group: Fluency increased significantly from Time 1 to Time 2, but not from Time 2 to Time 3, nor from Time 3 to Time 4. The latter suggests that any practice effect might have been based on initial familiarity with the test format and therefore the practice effect peaked at Time 2 for the Control group. The overall changes in fluency scores suggest that even though brief interventions may not be able to significantly improve scores on intelligence tests (accuracy), response latency may be an outcome variable that is particularly malleable such that improvements can be observed in the short term. We suggest fluency as a particularly useful variable to measure because of the importance of not conceding accuracy in pursuit of responding more quickly to either a question on a test or a real-world problem.

Neither relational skills training intervention, SMART nor SMARTA, significantly altered intelligence scores as measured using the KBIT-2 in the present study. In previous studies, a similar kind of intervention (albeit with SMART only) was implemented over a considerably longer time period (e.g., 15 hours of training sessions in Cassidy et al., 2011; or criterion-based training of about 3 months in Cassidy et al., 2016) and large changes in intelligence as measured by other intelligence and aptitude tests were observed. Thus, the current findings require some considered examination.

The first possible explanation is that the training provided in the current study was too brief for the strengthened operant behavioral repertoires to permeate the participants' environments in this specific context. In this case, the quality of a person's interaction with its

environment might be deemed irrelevant if the learning opportunities were not presented between the strengthening of the operant and the subsequent retest. For example, a "working memory" test might therefore constitute a measure of intelligent behavior whereby the experimenter could control for learning opportunities.

It is also possible that the training was too brief to strengthen operant responding itself. While this is at least partially unsupported by our data suggesting that the training group improved their reaction speeds (discussed below), it must be considered carefully. It is still possible that the *magnitude* of change in operant responding was considerably less than in comparable studies carried out over a longer period, to the extent that "accuracy" of responding (what the KBIT-2 measures) was a later cusp we did not reach. There is reason to suppose that the benefits of multiple exemplar training in operant responding might be analogous to the benefits of physical exercise in that sense. Just as the efficiency of one's muscles and respiratory system improve with physical exercise, one's nervous system may become more efficient given the correct type of 'exercise' over a sufficient period. On the other hand, there are several studies showing that relatively few training exemplars in patterns of relational responding are sometimes enough for such patterns to successfully generalise to new stimuli or contexts, as commonly demonstrated via probes for generalisation (e.g., deriving novel analogies in Carpentier et al., 2003; or deriving novel mands in Murphy, Barnes-Holmes, & Barnes-Holmes, 2005). Therefore, it would seem likely that relational operant behaviors were indeed strengthened but did not have time to permeate the environment such that the KBIT-2 accuracy scores would improve. The response latency data corroborate this assertion. On the other hand, there were reductions in response latencies associated with AARR training of a magnitude such that fluencies increased even though accuracy scores did not (see Table 3).

In the current study, there was no change in fluency nor its constituents on the Matrices subtest of the KBIT-2, and the change observed in the Riddles subtest of the KBIT-2 was not significantly different across groups. As mentioned above, performance on these tests may rely on particular learning histories that are uncontrolled for in the KBIT-2, and therefore that performance may tell us more (and yet, a limited amount) about an individual's previous circumstances than of their environmental adaptability. The Verbal Knowledge subtest is the only subtest that arguably measures one discrete process: matching a name or a function to a picture. Nonetheless, the relevance of being able to answer each of the individual items varies across individuals in accordance with their salience, and as such, they are arbitrarily included and performance on each task is not functional across all environments. In this sense, even though a change in response latency was observed on this subtest, it is doubtful that performance on it (especially with accuracy as the DV) would predict success across domains for all individuals, such as in education or at work.

Perhaps the biggest limitation of this study is its design. In order to compare the two training procedures, independent measures would have been desirable. In the present study, it was not possible to recruit enough participants for an independent group design who were also willing to make the required 4-hour minimum training commitment. Nonetheless, the predicted change in trajectory of KBIT-2 fluency does appear at Time 3 (see Figure 3), indicating that the inclusion of relating stimulus relations training may indeed be a useful addition to the overall SMART research program.

For the SMART training period, we did not supervise the training, so we relied on participants' self-reports of completing the agreed amount of training. However, the SMART program does record "points earned" for completing training. Participants in this study created

their own accounts and associated passwords, so we did not have access to their "points earned"

data which are only available through logging into each individual account. In future, perhaps

this would be a more accurate and higher resolution way of quantifying the amount of training

undertaken than independent group analyses. This would allow researchers to use natural

variation in "points earned" as an indicator of "amount of training completed". In larger samples,

this continuous independent variable may allow for a more accurate interpretation of the effects

of relational skills training using more complex regression models.

In several instances, we have reported quite large observed effect sizes. While these may

corroborate existing literature on the efficacy of relational operant fluency training, readers

should exercise caution in their interpretation due to the small sample size in this study.

Finally, the participants in the Experimental group were somewhat older. While this

would normally affect results on an intelligence test because results are standardised for each age

category, the *accuracy* outcome variable analysed in this study was based on raw scores and thus

individuals are comparable across groups. On the other hand, the age differential may explain the

baseline difference in mean *response latencies* across conditions (see Figure 3 and Table 3).

In light of the weaknesses of the KBIT-2 and of intelligence tests generally, there were

several limitations of the current study which merit discussion when planning future research in

this area. Below, we provide two suggestions for future research and detail where they might be

of benefit. Perhaps discrete tests of relational skills fluency wherein the complexity of the task is

well-controlled would be a better predictor of real-world outcomes than intelligence tests such as

the KBIT-2. Such a test would also be more suitable for repeated measures experiments because

they use arbitrary stimuli and rely solely on the behavioural performance of the individual rather

than assuming the functional utility of the individual items to be equal across individuals. In

particular, if such a test could be standardised, it would provide a tool for testing RFT's assertion that AARR underlies cognition writ-large by correlating relational skills with various important real-world outcomes. It would also allow researchers to overcome some of the difficulties encountered during the present study. For example, the training takes too long for it to be practical for the researcher in terms of recruitment (e.g., before running out of funding, or in terms of avoiding high levels of attrition), but also for potential beneficiaries who cannot commit the time (e.g., in schools where they must also cover a prescribed curriculum). A test for an individual's current level of relational fluency would also allow the researcher to target particularly deficient (relative to the norm) skills for training, rather than giving the same training to all, irrespective of their current level.

One other difficulty encountered is that there is currently no data to suggest a point at which participants would stop benefitting from relational skills training. An adequately controlled causal test of the benefits of relational skills training may require the experimenter to train relational skills and provide a follow-up after several years in order to allow for learning opportunities to present themselves evenly to the (newly) more relationally fluent participants over time. For example, in the UK education system, GCSE performance relative to the norm could be compared with A-Level performance relative to the norm but only if a relational skills improvement occurred after GCSE exams but before the A-Level curriculum had begun. To provide training after much of the curriculum had already been taught may increase one's capacity to learn from future learning opportunities, but there is no reason to believe that the training would help students learn material that was presented while they were less relationally skilled.

Cassidy et al. (2011) examined a change in trajectory of IQ scores at Time 3 (when they changed the training intervention from stimulus equivalence training to SMART training) in order to compare their two interventions wherein the first intervention (stimulus equivalence) featured as a component of the second (SMART). In this study, we observed the change in the IQ test fluency trajectory at Time 3 in order to compare the efficacy of SMART with SMARTA.

It is not clear from the present data whether SMARTA, a new intervention targeting higher-order operant repertoires, is as effective as SMART training for increasing intelligence or attainment for the general population. However, the data observed herein suggest that both SMART and SMARTA were effective in terms of reducing response latencies on recall tasks. Perhaps this is unsurprising given that the SMARTA tasks involved the same kind of AARR practice as SMART, but with a small addition that participants should also relate two derived relations. In Figure 3, there is a small deviation in trajectory of fluency increases among experimental participants after SMARTA training, but further research is needed in order to explore this further. Nonetheless, the present study was the first time the SMARTA intervention was used and the results suggest that it may be another useful way to train relational skills at a particular level. Furthermore, it demonstrates that it is possible to train complex forms of relational responding such as analogical responding and relational differentiation, corroborating McLoughlin and Stewart (in press). Importantly, this suggests that there may be no limit to the levels of complexity (i.e., either nodal distance or the type of derived relation) at which one might be taught to derive relations.

A brief relational skills training intervention was not enough to significantly change performance on the KBIT-2 intelligence test using the prescribed outcome variable of accuracy. While training interventions with more intensive dosages (e.g., in Experiment 1 of Cassidy et al.,

2011, each participant received 15 hours of training) have reported accuracy changes after

relational skills training, a large-scale replication of these data (e.g., a RCT) is difficult to

implement. We proposed several alternative ways of answering pertinent questions herein. Using

a measure of relational skills fluency *as* a measure of intelligence may help to address several

limitations of typical intelligence tests. Before doing that, it is important to first provide a robust

test of RFT's prediction that relational skill levels predict real-world outcomes of interest. To

date, there are several small-scale studies demonstrating that SMART training is useful in terms

of improving intelligence and other outcomes of interest. There are no studies testing these

hypotheses on a large scale. A cross-sectional study of relational skills and how they relate to

outcomes of interest with a view to producing a standardised measure of relational skills would

seem a promising next step for this research program that would either build a case for

implementing relational skills training on a larger scale, or potentially provide evidence to refute

RFT's assertion that AARR is at the core of cognition.

References

Barnes-Holmes, D. (2000). Behavioral pragmatism: No place for reality and truth. *The Behavior Analyst*, *23*, 191-202.

Barnes-Holmes, Y., McHugh, L., & Barnes-Holmes, D. (2004). Perspective-taking and Theory of Mind: A relational frame account. *The Behavior Analyst Today*, *5*, 15-25.

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

Binder, C., Haughton, E., & Bateman, B. (2002). Fluency: Achieving true mastery in the learning process. *Professional Papers in Special Education*, 2-20.

Carpentier, F., Smeets, P.M., & Barnes-Holmes, D. (2003). Equivalence-equivalence as a model of analogy: Further analyses. *The Psychological Record*, *53*, 349-372.

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.

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*(1), 37–51. doi: 10.1080/15021149.2010.11434333

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22. http://dx.doi.org/10.1037/h0046743

Crowe, S. F., Benedict, T., Enrico, J., Mancuso, N., Matthews, C., & Wallace, J. (1999). Cognitive determinants of performance on the digit symbol-coding test, and the symbol

search test of the WAIS-III, and the symbol digit modalities test: an analysis in a healthy sample. *Australian Psychologist*, *34*(3), 204-210.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more-than, and less-than. *Journal of the Experimental Analysis of Behavior*, *64*, 163-184.

Flynn, J.R. (2016). *Does Your Family Make You Smarter?: Nature*, *Nurture*, *and Human Autonomy*. New York: Cambridge University Press.

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

Guilford, J.P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

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*(3), 397-411. DOI: 10.1111/bjep.12114.

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

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Perrig, W. J. (2008). Improving fluid intelligence with training on working memory. *Proceedings of the National Academy of Sciences*, *105*,, 6829-6833.

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.

Jäeggi, S. M., Studer-Luethi, B., Buschkuehl, M., Su, Y., Jonides, J., & Perrig, W. J. (2010). The relationship between n-back performance and matrix reasoning: Implications for training and transfer. *Intelligence, 38,* 625-635.

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

May, R., Stewart, I., Baez, L., Freegard, G., & Dymond, S. (2017). An empirical demonstration
    of spatial relational responding. *Journal of the Experimental Analysis of Behavior*,
    107(2), 234-257.

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

Murphy, C., & Barnes-Holmes, D. (2017). Teaching Important Relational Skills for Children
    with Autism Spectrum Disorder and Intellectual Disability Using Freely Available (GO-
    IRAP) Software. *Austin Journal of Autism & Related Disabilities*, *3*(2), 1041.

Murphy, C, Barnes-Holmes, D, & Barnes-Holmes, Y. (2005). Derived manding in children with
    autism: Synthesizing Skinner's verbal behavior with relational frame theory. *Journal of
    Applied Behavior Analysis, 38*(4), 445–462.

Nuzzolo-Gomez, R., & Greer, R. D. (2004). Emergence of untaught mands or tacts of novel
    adjective-object pairs as a function of instructional history. *The Analysis of Verbal
    Behavior*, *20*(1), 63-76.

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

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

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.

Parra, I., & Ruiz, F.J. (2016). The effect on Intelligence Quotient of Training Fluency in

  Relational Frames of Coordination. *International Journal of Psychology & Psychological*

  Therapy, *16*(1), 1-12.

Ramey, D., Lydon, S., Healy, O., McCoy, A., Holloway, J., & Mulhern, T. (2016). A systematic

  review of the effectiveness of precision teaching for individuals with developmental

  disabilities. *Review Journal of Autism and Developmental Disorders, 3*, 179–195.

Ruiz, F.J., & Luciano, C. (2011). Cross-domain analogies as relating derived relations among

  two separate relational networks. *Journal of the Experimental Analysis of Behavior*, *95*,

  369-385. DOI: 10.1901/jeab.2011.95-369

Schlinger, H.D. (1992). Intelligence: Real or Artificial? *The Analysis of Verbal Behavior*, *10*,

  125-133.

Schlinger, H. D. (2003). The myth of intelligence. *The Psychological Record, 53*(1), 15–32.

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

  Research, *14*, 5-13.

Slattery, B., & Stewart, I. (2014). Hierarchical classification as hierarchical framing. *Journal of*

  the Experimental Analysis of Behavior, *101*, 61-75.

Sternberg, R.J. (1985). *Beyond IQ: A Triarchic Theory of Intelligence*. Cambridge: Cambridge

  University Press.

Stewart, I. (2016). The fruits of a functional approach for psychological science. *International*

  Journal of Psychology, *51*(1), 15-27. DOI:10.1002/ijop.12184

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of

    longitudinal research. *Intelligence*, *35*, 401-426.

Thirus, J., Starbrink, M., & Jansson, B. (2016). Relational frame theory, mathematical and

    logical skills: A multiple exemplar training intervention to enhance intellectual

    performance. *International Journal of Psychology and Psychological Therapy, 16*(2),

    141–155.

Thorndike, E.L. (1920). Intelligence and its use. *Harper Magazine*, *140*, 227–235.

Thorndike, E.L., Bregman, E.D., Cobb, M.V., & Woodyard, E.I. (1926). *The measurement of

    intelligence.* New York: Teachers College Press

Thurstone, L.L. (1938). *Primary mental abilities.* Chicago: University of Chicago Press.

Vizcaíno-Torres, R.M., Ruiz, F.J., Luciano, C., López-López, J.C., Barbero-Rubio, A., & Gil, E.

    (2015). The effect of relational training on intelligence quotient: A case study.

    Psichothema, *27*(2), 120-127. DOI: 10.7334/psicothema2014.149

Wechsler, D. (1997). *Wechsler Adult Intelligence Scale, Third Edition*. San Antonio, TX: The

    Psychological Corporation

Table 1

Participant demographics

Participant	Age	Gender	Condition	Occupation	Education
1	26	Male	Experimental	Chemist	B.Sc.
2	39	Female	Experimental	Media	B.A.
3	42	Female	Experimental	Home-maker	Secondary school
4	52	Male	Experimental	Labourer	Secondary school
5	28	Female	Control	Student	B.A.
6	29	Female	Control	Teacher (primary)	B.Ed.
7	24	Female	Control	Administration	Diploma
8	22	Male	Control	Electrician	2 Year Trade

Table 2

Procedure

Participant	Time 1	Time 2	Time 3	Time 4
EXP1	--	Begin SMART	*End SMART* Begin SMARTA	*End SMARTA*
EXP2	--	Begin SMART	*End SMART* Begin SMARTA	*End SMARTA*
EXP3	--	Begin SMART	*End SMART* Begin SMARTA	*End SMARTA*
EXP4	--	Begin SMART	*End SMART* Begin SMARTA	*End SMARTA*

Note. SMART = Strengthening Mental Abilities with Relational Training; SMARTA = SMART for analogical responding; All control participants did no training from Time 1 to Time 4.

Table 3

Mean Response Latencies (in seconds) for Experimental and Control Participants

Participant / Scale	Time 1	Time 2	Time 3	Time 4
EXP / VR	308.35	282.2	183.65	131.9
EXP / MAT	800.13	671.68	454.05	330.4
EXP / RID	728.48	687.78	632.88	533.88
CON / VR	236.03	202.8	195.65	186.68
CON / MAT	581.6	604.18	576.28	628.98
CON / RID	696.65	584.2	547.0	505.95

Note. EXP = Experimental participants; CON = Control participants; VR = Verbal Reasoning sub-scale; MAT = Matrices subscale; RID = Riddles subscale.

Table 4

Full-Scale KBIT-2 raw accuracy, latency, and fluency for individual participants.

Participant	Time 1	Time 2	Time 3	Time 4
	Acc/Lat/Flu	**Acc/Lat/Flu**	**Acc/Lat/Flu**	**Acc/Lat/Flu**
EXP1	128/1700/.16	130/1596/.18	120/1328/.23	131/1141/.29
EXP2	143/1994/.19	145/1580/.21	144/999/.36	145/819/.55
EXP3	125/1774/.19	121/1628/.19	127/1400/.25	129/1184/.30
EXP4	136/1880/.17	135/1762/.18	140/1356/.26	144/840/.49
CON1	128/795/.37	134/749/.41	136/730/.42	138/693/.45
CON2	136/1534/.26	137/1211/.31	134/1154/.32	137/1100/.33
CON3	131/1830/.18	134/1854/.18	138/1722/.19	135/1871/.18
CON4	129/1898/.16	131/1751/.20	132/1669/.22	131/1623/.24

Note. EXP = Experimental participants; CON = Control participants; Acc = raw accuracy across all three subscales, Lat = latency across all three subscales (in seconds), Flu = Fluency.

Figure Captions

Figure 1

Upper panel: A complex SAME/OPPOSITE trial from SMART training. Lower Panel: A simple

MORE/LESS trial from SMART training.

Figure 2

Upper panel: An example of a relating relations trial from SMARTA training. Lower Panel: The

Main Menu of SMARTA training, including the 48-trial assessment and specific trial training

and testing options.

Figure 3

A comparison of mean fluency changes from Time 1 through Time 4 for both the experimental

and control participants.