

The Psychological Record, 2009, 59, 119-132

## THREE CHRONOMETRIC INDICES OF RELATIONAL RESPONDING AS PREDICTORS OF PERFORMANCE ON A BRIEF INTELLIGENCE TEST: THE IMPORTANCE OF RELATIONAL FLEXIBILITY

Catriona O'Toole and Dermot Barnes-Holmes

National University of Ireland, Maynooth

*Participants completed a before/after and a similar/different relational task, using the Implicit Relational Assessment Procedure (IRAP), and subsequently took the Kaufman Brief Intelligence Test (K-BIT). For each relational task, response latencies were measured first on consistent trials, where participants responded in accordance with preestablished verbal relations, and then on inconsistent trials, where they responded against these relations. A difference-score was calculated by subtracting consistent from inconsistent response latencies. The inconsistent trials and the difference-score provided measures of relational flexibility. Results showed that faster responding on the IRAP and smaller difference-scores predicted higher IQ. These findings suggest that relational flexibility is an important component of intelligence and might therefore be targeted in educational settings.*

For humans, the ability to identify relations between objects and events is an integral part of everyday life. They routinely recognize similarities between various stimuli, and they have the ability to make comparisons, learn from analogies, understand a sequence of events, and plan for the future. Each of these abilities relies on relational thought, which Gentner and Loewenstein (2002b) referred to as the *sine qua non* of human cognition. Consistent with this view is the observation that a child's cognitive development is typically observed in the increasing relational complexity of the language and concepts this child employs as he or she grows older (Andrews & Halford, 1998, 2002; Gentner & Ratterman, 1991; Hayes, Barnes-Holmes, & Roche, 2001).

Although often considered as predominantly the subject matter of cognitive psychology, relational abilities can also be understood from the perspective of behavior analysis. Modern behavioral research into complex relational performances was greatly enhanced by the seminal work of Murray Sidman and colleagues, who developed a methodology for examining what is now known as stimulus equivalence (see Sidman, 1994, for a review). In a typical equivalence experiment, explicit training is given in a number of interrelated stimulus relations, which then produces a number of predicted but untrained relations.

For example, if they are trained in A-B and A-C relations, most verbally able humans will spontaneously show B-A, C-B, A-C, and C-A relations. When these untaught relational responses emerge, the three stimuli A, B, and C are said to participate in an equivalence class or a relation of similarity.

Although the equivalence relation is the one most widely researched, it is of course just one of the relations that seem to occur in human language and cognition. Among the other relations are those of comparison and difference, as well as hierarchical, spatial, and temporal relations. One of the primary reasons why behavioral researchers have shown such interest in these relational abilities is the apparent relevance to human language and cognition (see Hayes, Barnes-Holmes, & Roche, 2001, for a book-length review).

For some time now, the suggested relationship between language and stimulus relations has attracted considerable attention. More recently, however, some researchers have suggested that derived relations may be relevant to our understanding of human intelligence, or at least those forms of human intelligence that are verbally based (see Hayes, Barnes-Holmes, & Roche, 2001). However, to our knowledge, only one published study has tested this idea (O'Hora, Peláez, & Barnes-Holmes, 2005). In that study, performances on derived relations tasks were compared with performances on a number of subtests of the *Wechsler Adult Intelligence Scale-Third Edition* (WAIS-III; Wechsler, 1997). Specifically, participants were presented with a complex relational task in which they had to learn to relate novel stimuli in accordance with similar/different and temporal relations. The results showed that participants who passed a subsequent test for the predicted novel patterns of relational responding also produced significantly higher scores on the verbal and arithmetic subtests (but not on the digit-symbol coding subtest) than the participants who failed the relational test. Furthermore, the data also showed a significant correlation between the number of accurate responses on the relational training phase and the number of correct responses on both the vocabulary and arithmetic subtests of the WAIS (see O'Hora et al., 2008, for a replication and extension of this work).

The pattern of results obtained was predicted based on arguments made by Y. Barnes-Holmes and colleagues (2001). Specifically, they suggested that:

Persons with highly elaborated vocabulary will tend to have highly elaborated relational repertoires. Nevertheless, it is the relational skills that are key, not merely verbal content in a formal sense. A task such as learning to spell is far less relationally rich than learning word meanings, and thus it is no surprise that spelling correlates less well with overall levels of intellectual behavior than vocabulary, even though both tasks involve verbal material. (p. 160)

The authors predicted that higher levels of proficiency on relational responding tasks would be a better predictor of performance on relationally rich subtests (such as vocabulary and arithmetic) than on subtests that are less relational (such as the digit-symbol coding subtest, which measures processing speed), and this is exactly what O'Hora et al. (2005) found.

The purpose of the current study was to again test the relationship between relational responding and intelligence. However, the present research differed in two major respects. First, a brief but complete IQ test was administered to participants to determine to what extent relational

responding correlated with full-scale IQ.<sup>1</sup> Second, a different methodology was used to assess relational responding, namely, the Implicit Relational Assessment Procedure (IRAP). (See Barnes-Holmes et al., 2006; McKenna, Barnes-Holmes, Barnes-Holmes, & Stewart, 2007.) The IRAP requires participants to respond accurately and rapidly to a series of relational tasks. In some blocks of tasks, participants are asked to respond in a way that is consistent with previously learned relations (is a Shoe similar to a Sandal? = True); and in other blocks, responding in an inconsistent pattern is required (is a Shoe similar to a Sandal? = False).

The use of the IRAP as a methodology was deemed important because it provides a number of measures of relational responding. In the study by O'Hara et al. (2005), only response accuracy was used as a measure of performance on the relational tasks. In contrast, the IRAP provides a measure of both the accuracy *and* the speed at which participants respond to the tasks. Measuring response speed seemed particularly relevant because it may provide an indication of the fluency with which individuals can respond relationally. It also taps into other variables, such as processing speed, attention, and working memory, each of them also widely considered important components of intelligence (e.g., Sternberg, 2000).

Another advantage of using the IRAP as a relational task is that it requires two patterns of responding, one consistent and the other inconsistent with previously established relations. The general assumption is that participants will respond more rapidly on consistent than on inconsistent trials, and this prediction has been supported across numerous studies (e.g., Barnes-Holmes et al., 2006). Critically, however, performance on the inconsistent trials is unlikely to be a well-practiced or firmly established skill (individuals rarely practice incorrect responding for protracted periods of time). Consequently, response speed on these trials may provide a possibly useful measure of relational or cognitive flexibility. In other words, the faster an individual can produce responses that contradict previously well-established verbal relations (by the wider community), the more flexible the behavior. Assessing relational flexibility may be particularly advantageous because flexibility is widely regarded as an important component of human cognitive abilities (e.g., Cattell, 1971; Kyllonen, Lohman, & Woltz, 1984; Premack, 2004).

On balance, the response latencies data obtained on inconsistent trials may provide a measure of not just relational flexibility but also the other variables mentioned above (e.g., processing speed). To obtain a relatively "pure" measure of relational flexibility, the effects of these extraneous variables need to be controlled. Insofar as these variables contribute to both types of trials, the difference between the two, known as the difference-score, may therefore provide an uncontaminated measure of relational flexibility (i.e., the smaller the difference-score between consistent and inconsistent trials, the greater the relational flexibility). IRAP requires participants to switch back and forth repeatedly between blocks of consistent and inconsistent trials, so the difference-score would reflect the relational flexibility produced across the numerous shifts in the IRAP contingencies.

---

1 It is important to note that in using IQ to assess intelligence, we do not suggest that an IQ score captures a real or underlying entity. Rather, we are working on the premise that standardized IQ tests are the best predictors that we currently have of the range of diverse and adaptive behaviors collectively and conveniently termed "intelligence" (Martinez, 2000). Thus, although the focus of our interest is upon these diverse behaviors, IQ scores can provide a reasonable approximation of these behaviors in group-based research.

Participants were presented with two IRAP tasks: One assessed similar/different relations; the other, before/after relations. Participants were subsequently exposed to the *Kaufman Brief Intelligence Test* (K-BIT; Kaufman & Kaufman, 1990). Insofar as relational responding and flexibility are defining features of human intelligence, participants with higher scores on an intelligence test would be predicted to respond more quickly on the relational tasks than those with lower scores, particularly on inconsistent relative to consistent trials. Furthermore, smaller difference-scores, indicating greater relational flexibility, would predict higher IQ scores.

## Method

### *Participants*

Participants were 62 (15 male, 47 female) undergraduate students from various faculties of the National University of Ireland, Maynooth. None of the participants had previous exposure to the IRAP protocols that were used, nor were they familiar with the IQ test that was subsequently administered. The data from 1 female participant who reported being dyslexic on a subject profile form (see below) were not included in the final analyses. In addition, to control for any confounding effects that might occur because of a speed/accuracy trade-off, only the data from participants who produced at least 80% correct responding on the IRAPs were included in the final analyses. Fifty-five participants met this criterion on one or both of the IRAPs. They ranged in age from 18-55 years (mode = 18;  $M = 23$ ).

### *Materials and Stimuli*

Materials included the IRAP software (available from the authors upon request), the *Kaufman Brief Intelligence Test* (K-BIT), and a subject profile form. The IRAP was presented on a Dell Personal Computer with a Pentium 4 Processor and a standard keyboard and monitor. The software was used to present the instructions and the stimuli and to record responses. Two separate IRAP tasks were presented, differing only in the stimuli employed. Table 1 shows the details of the stimuli used in each IRAP.

The K-BIT is a brief, individually-administered measure of verbal and nonverbal intelligence suitable for use with individuals aged 4-90 years. It consists of two subtests, a vocabulary subtest (consisting of Part A, Expressive vocabulary, and Part B, Definitions) and a matrices subtest. The vocabulary subtest measures verbal, school-related skills by assessing a person's word knowledge and verbal concept formation. The matrices subtest measures nonverbal skills and the ability to solve new problems by assessing a person's ability to perceive relationships and to complete analogies. All matrices items involve pictures or abstract designs rather than words. Additional materials included the K-BIT Individual Test Record Form and a stopwatch to ensure that a maximum of 30 s was allowed for each item on the definitions task. Kaufman and Kaufman (1990) report a correlational coefficient of .75 between K-BIT Composite IQ scores and full-scale IQ scores on the *Wechsler Adult Intelligence Scale-Revised* (WAIS-R; Wechsler, 1981). Furthermore, the vocabulary subtest of the K-BIT correlates .60 with verbal IQ on the WAIS-R, and the matrices subtest correlates .52 with the performance IQ on the

Table 1  
*Sample and Target Stimuli Used in the Before/After and the Similar/Different IRAPs*

Before/After IRAP				Similar/Different IRAP			
Sample 1		Sample 2		Sample 1		Sample 2	
Before		After		Similar		Different	
Targets congruent with Sample 1		Targets congruent with Sample 2		Targets congruent with Sample 1		Targets congruent with Sample 2	
Spring	Summer	Summer	Spring	Oven	Grill	Oven	Tree
Child	Adult	Adult	Child	Book	Journal	Book	Dog
Engagement	Marriage	Marriage	Engagement	Wall	Fence	Wall	Parrot
Crawl	Walk	Walk	Crawl	Chair	Seat	Chair	Lion
Problem	Solution	Solution	Problem	Table	Desk	Table	Cat
Effort	Reward	Reward	Effort	House	Cottage	House	Star
Treatment	Cure	Cure	Treatment	Boat	Ship	Boat	Shelf
Introduction	Conclusion	Conclusion	Introduction	Shoe	Sandal	Shoe	Apple

WAIS-R. These coefficients, it has been argued, provide strong support for the construct validity of the K-BIT (Kaufman & Kaufman, 1990). Excellent split-half reliability coefficients are also reported for the K-BIT IQ composite, as well as for the subtests. The reliability of the K-BIT IQ composite, for instance, averages .93 across the entire age span.

### *Procedure*

The experiment was split into two sessions. In Session 1, participants completed the two IRAP protocols. In Session 2 they were exposed to the K-BIT. These sessions were conducted on separate days but within one working week of each other. Both sessions were conducted in individual cubicles in the experimental psychology laboratory at NUI Maynooth.

*Session 1: IRAPs.* Upon arrival at the laboratory, participants completed a subject-profile form, which contained a series of questions on demographic information (i.e., age, gender, ethnicity, and education level/occupation). They were also asked to state how often they use computers (whether daily, weekly, monthly, annually, or never), and whether or not they had any learning difficulties. After filling out this form, participants were directed toward an open display page on the computer screen containing the instructions necessary for completing the similar/different or the before/after IRAP tasks (the full instructions are available from the authors upon request). The display page contained some general instructions and a consent statement. It also contained specific instructions and illustrations designed to explain how participants should respond to the tasks. Using these illustrations, the experimenter verbally explained the nature of the task. For example, participants beginning with the before/after IRAP were referred to the first illustration on the before/after instruction page. This illustration contained the word *before* at the top of the display, the words *spring* and *summer* in the center, and the words *true* and *false* at the bottom. Participants were instructed to read *spring before summer* and then to respond appropriately by pressing the designated *true* or *false* key. Participants were also informed that

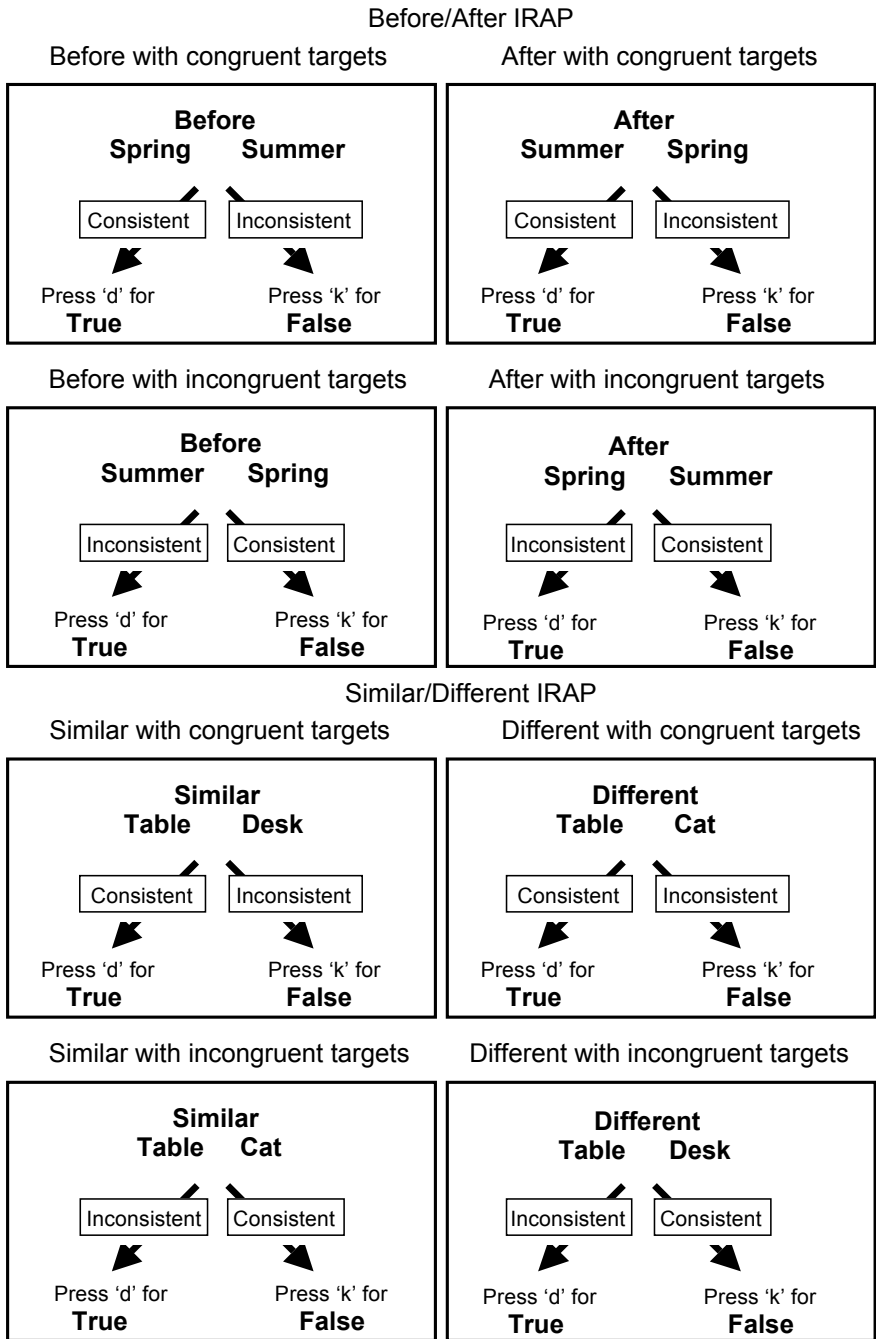
sometimes they would be required to respond to the tasks in a way that agrees with what they believe, and at other times they would be required to respond in a way that disagrees with what they believe.

Consistent IRAP blocks were defined as those requiring responses that were deemed to be generally consistent with common verbal practices (e.g., choosing *true* when presented with the target *spring before summer*). Inconsistent IRAP blocks required responses that were inconsistent with common verbal practices (e.g., choosing *true* when presented with the target *summer before spring*). Participants were informed that the program would alternate between the two types of blocks and that the first two blocks of trials were for practice, but thereafter they should respond as quickly and as accurately as possible. Finally, they were instructed that they should keep their left and right index fingers resting on the *d* and *k* keys, respectively, because these keys corresponded to the true and false response options (see below). The experimenter then left the room, and the participants were free to scroll through the instruction page at their own pace. They were required to press a button on the computer screen when they were sure they fully understood the task. After they pressed this button, another display page was presented, stating that when a error was made, a red X would appear below the target stimulus; to clear the red X and continue with the trials, they would need to provide the correct response by pressing the appropriate *d* or *k* key. The next part of the message invited the participant to press the space-bar to proceed with the tasks.

Each IRAP consisted of two blocks of 32 practice trials and six blocks of 32 test trials. Each participant was randomly assigned to one of two conditions, consistent-relations-first (CF) or inconsistent-relations-first (IF). Participants assigned to the CF condition began both IRAPs with a consistent trial block and thereafter alternated between inconsistent and consistent blocks. The experimental sequence for the CF group was then consistent practice, inconsistent practice, consistent test 1, inconsistent test 1, consistent test 2, inconsistent test 2, consistent test 3, and inconsistent test 3. Participants assigned to the IF group started both IRAPs with an inconsistent block and subsequently alternated between consistent and inconsistent blocks. The order in which the two IRAPs (similar/different and before/after) were presented was counterbalanced across participants.

For each trial, a sample stimulus and a pair of target stimuli were presented on each IRAP; two response options, *true* and *false*, were also presented. The sample stimulus was presented at the top of the computer screen, the target words were presented in the center, and the *true* and *false* response options were presented in the bottom left- and right-hand corners. The response options switched sides unpredictably across trials. The various combinations of sample stimuli with target stimuli served to generate four different trial-types. Specifically, one trial-type was generated by presenting Sample 1 with congruent targets (e.g., *before* with *spring summer*); another, by presenting Sample 2 with congruent targets (e.g., *after* with *summer spring*); a third, by presenting Sample 1 with incongruent targets (e.g., *before* with *summer spring*); and a fourth, by presenting Sample 2 with incongruent targets (e.g., *after* with *spring summer*). For each block of 32 trials, each target stimulus was presented in a quasi-random order such that each target was presented twice, once in the presence of each sample stimulus (i.e., the four trial-types were presented eight times).

Figure 1 illustrates examples of the four different trial-types on both the before/after (top section) and the similar/different (bottom section) IRAPs as



*Figure 1.* Examples of four trial-types on the before/after IRAP (upper section) and the similar/different IRAP (lower section). Sample stimuli, target words, and response options were presented simultaneously on the screen. Note that the superimposed arrows and text boxes, used here to illustrate consistent and inconsistent responses, did not appear on the screen during the IRAP.

they would appear on the computer screen. All the stimuli (i.e., the relational term, target stimuli, and the response options) appeared simultaneously and remained on the screen until the participant emitted a response (i.e., pressed either the *d* or *k* key). If the participant emitted a correct response (i.e., consistent during consistent blocks and inconsistent during inconsistent blocks), the screen cleared and a new trial was presented after a 400-ms delay. Following an incorrect response (i.e., consistent during inconsistent and inconsistent during consistent blocks), a red X appeared on the screen immediately below the target word, and the participant was required to make the correct response to clear the screen and to proceed to the 400-ms inter-trial interval. Although participants were required to self-correct, the IRAP program recorded these trials as incorrect. Following each 32-trial block, the computer presented feedback for that block on the percentage of correct responses and the median response latency. It also informed participants that during the next block of trials the previously correct and incorrect answers would be reversed. When participants had completed the two practice blocks, an additional message appeared on the screen informing them they would now be completing a test block, and they were instructed to respond quickly and accurately. Following the final trial of the final block, the screen cleared and a message appeared asking participants to report to the experimenter. The experimenter then returned to the room and loaded the second IRAP program. The procedure for the second IRAP was similar to that of the first (i.e., participants proceeded through instructions, practice blocks, and test blocks in the manner described above). Following completion of the final trial in the final block, the participant was again instructed to report to the experimenter. This marked the end of the first session.

*Session 2: IQ test.* The K-BIT was administered and scored in accordance with the standardized procedures set out in the K-BIT manual (Kaufman & Kaufman, 1990). For all participants, testing started with the expressive vocabulary subtest and proceeded to definitions and finally to matrices. In the definitions subtest, participants had a maximum of 30s to respond to each item. No time limit was set for either the expressive vocabulary or matrices subtests. All participants started each subtest at the designated starting point for adult age groups. For each subtest, testing continued until all items were completed or until the discontinue criterion was met (i.e., until a participant failed every item in a unit). During administration, the individual's test record sheet was shielded from view, and no feedback was given until the participant had completed the entire test.

## Results

### *Preliminary Data Screening*

Fifty-five participants achieved 80% accuracy on the test blocks for one or both IRAPs. Further, 11 participants met the accuracy criterion on only one IRAP: 4 on before/after and 7 on similar/different. The data from an IRAP exposure that produced a below-criterion performance were removed from subsequent analyses. Since response accuracy was used to screen out entire data sets falling below 80% correct, the assumption was that accuracy would not discriminate between high and average IQ participants. Preliminary



statistical analyses confirmed this prediction (all  $ps > .55$ ). Furthermore, none of the three latency measures, described subsequently, correlated with the accuracy scores (all  $ps > .1$ ).

Response latencies on each IRAP trial are recorded from the point of target onset to the first correct response the participant emitted. To control for extreme outliers, any latencies greater than 10,000 ms were removed from the data. Two mean response latencies were then calculated for each participant, one for the consistent trials and the other for the inconsistent trials. Calculating these two mean response latencies involved collapsing the data across the three consistent and three inconsistent test blocks (this practice is normally adopted in IRAP research, because previous and ongoing studies have found that block sequence does not interfere with the critical IRAP effect).

Two overall group mean latencies were then calculated for consistent and inconsistent trials for each IRAP. The data were then checked to determine that the four mean latencies for each participant did not exceed 2 standard deviations above the corresponding group mean. If a participant produced one or more mean latencies that exceeded 2 standard deviations, all data for that participant were excluded from further analyses. The data for 3 participants were removed on this basis. Overall, therefore, 45 and 48 participants contributed data toward the before/after and similar/different IRAP measures, respectively. Since the results of the Shapiro-Wilk test used to investigate the response latency and IQ data, showed no violation of the assumption of normality, the data were analyzed using parametric statistics.

### *Initial IRAP Analyses*

The mean response latencies on the consistent blocks were shorter on the similar/different task ( $M = 2,230$ ) relative to the before/after task ( $M = 2,722$ ), and the same pattern was observed for the inconsistent blocks (similar/different,  $M = 2,871$  versus before/after,  $M = 3,139$ ). Paired-samples  $t$ -tests indicated that these differences were significant: for consistent ( $t[40] = 9.99$ ,  $p < .0001$ ,  $\eta^2 = .71$ ) and inconsistent blocks ( $t[40] = 3.21$ ,  $p = .003$ ,  $\eta^2 = .20$ ). These results indicate that before/after relational responding was more difficult (because it required more time) than similar/different responding.

The overall mean response latencies for the consistent blocks on both IRAPs were shorter than on their respective inconsistent blocks (before/after,  $M = 2,722$ ,  $SE = 70$  versus  $M = 3,139$ ,  $SE = 89$ ; and similar/different,  $M = 2,230$ ,  $SE = 53$  versus  $M = 2,871$ ,  $SE = 75$ ). Paired-samples  $t$  tests showed that both differences were significant: before/after ( $t[44] = -9.53$ ,  $p < .0001$ ,  $\eta^2 = .67$ ) and similar/different ( $t[47] = -14.22$ ,  $p < .0001$ ,  $\eta^2 = .81$ ). As predicted, therefore, the current study produced a typical IRAP effect, consistent with that observed in previously reported IRAP studies.

A difference-score was calculated for each participant by subtracting response latencies on consistent trials from latencies on inconsistent trials for each of the IRAPs. The mean difference-score on the similar/different IRAP was greater than that for the before/after IRAP (before/after = 417, and similar different = 641). A paired-samples  $t$  test comparing the two sets of difference-scores proved to be significant ( $t[40] = -4.64$ ,  $p < .0001$ ,  $\eta^2 = .51$ ).

### *IQ Data*

The IQ scores obtained in the current sample ranged from 95-130. The mean IQ score was above average ( $N = 52$ ,  $M = 110$ ), which is unsurprising since the participants were university students. Furthermore, the IQ scores in the current sample ( $SD = 8$ ) showed less variation than is typically observed in the general population ( $SD = 15$ ; Kaufman & Kaufman, 1990). The IQ score generated by the K-BIT is a composite score, which corresponds to the sum of the standard scores on the verbal and matrices subtests. The mean standard score on the vocabulary subtest was 108 ( $SD = 8$ ), and on the matrices subtest it was 109 ( $SD = 9$ ).

### *Correlations Between IRAP and IQ*

As noted previously, both the difference-score and response latencies on inconsistent trials may provide a measure of relational flexibility, and such flexibility could be considered an important component of IQ. If this view is correct, negative correlations would be expected between the difference-score and performances on the IQ test. More evidence for correlations between response latencies and IQ scores would also be expected for the inconsistent, relative to the consistent, trials. A series of correlational analyses were conducted to test these predictions. Table 2 presents these results.

Table 2  
*Correlations Between the IRAP Measures and the Intelligence Test Measures*

IRAP Measure	IQ	Verbal	Matrices
<b>Response Latency</b>			
Before/after Consistent	-0.27	-0.09	-0.25
Similar/different Consistent	-0.32*	-0.24	-0.24
Before/after Inconsistent	-0.38**	-0.23	-0.32*
Similar/different Inconsistent	-0.35*	-0.43**	-0.16
<b>Difference-score</b>			
Before/after	-0.32*	-0.31*	-0.24
Similar/different	-0.21	-0.44**	0.02

\* $p < .05$ ; \*\* $p < .01$

Only one significant correlation was observed between IQ and consistent trials (i.e., the correlation between full-scale IQ and response latency on the similar/different IRAP). In contrast, four significant correlations were observed between the IQ measures and the inconsistent response latencies. Specifically, significant correlations were observed between full-scale IQ and inconsistent trials on both the before/after and similar/different tasks. Furthermore, on similar/different inconsistent trials, a significant correlation was observed with standard scores on the verbal subtest; but the correlation with standard scores on matrices was nonsignificant. Interestingly, the opposite pattern was observed for the before/after inconsistent trials (i.e., a significant correlation with matrices but not with verbal).

Significant correlations were obtained between the before/after difference-scores and both full-scale IQ and the verbal subtest, but the correlation with matrices was nonsignificant. The difference-scores for the similar/different

IRAP also correlated significantly with verbal but not with IQ or matrices. So, in contrast with the correlations from the absolute response-latency data, neither the similar/different nor the before/after difference-scores correlated significantly with the matrices subtest; they were loaded instead onto the verbal subtest.

## Discussion

The current study showed that individuals who produced higher IQ scores on a standard measure of intelligence tended to respond faster on the relational tasks presented on the IRAP. Importantly, the inconsistent IRAP trials produced a larger number of significant correlations than the consistent tasks; furthermore, only the inconsistent trials produced significant correlations across both similar/different and before/after IRAPs. These results indicate that individuals who performed better on the intelligence test not only were faster at responding relationally but also demonstrated a greater degree of relational flexibility.

The current study extends the literature on the relationship between relational responding and human cognitive abilities. Most notably, the findings extend the work of O'Hora et al. (2005) in demonstrating that performances on response-time-based similar/different and temporal relational tasks predict overall IQ as well as subtest scores. Furthermore, unlike O'Hora et al. (2005, 2008), the current study employed three different response-time-based measures of relational responding (consistent trials, inconsistent trials, and difference-scores); interestingly, different patterns of significant correlations with IQ were observed across the measures. Regardless of how one interprets these results, the current findings serve to highlight the importance of using multiple measures of relational responding, rather than relying solely on response accuracy.

The IRAP procedure, as described previously, requires participants to respond as quickly as possible; so it might be argued that correlations may emerge only between performance on the IRAP and performance on tests or subtests requiring relatively rapid responses (e.g., the symbol search or the digit symbol-coding subtest on the WAIS-III). Critically, however, the K-BIT has no such component, and yet one or more of the three IRAP measures correlated with the overall IQ and/or subtest measures. (Part B of the verbal subtest does employ a 30-s cutoff for each question; but this subtest was designed to measure word knowledge and verbal concept formation independent of processing speed; Kaufman & Kaufman, 1990.) Furthermore, the current findings may be seen as particularly important given that the K-BIT was designed as a brief IQ test and so included only the core subcomponents thought to be particularly crucial in the assessment of intelligence. In other words, the observed relationship between the IRAP and K-BIT supports the claim that relational processes, *per se*, are important factors in certain critical aspects of human intellectual ability.

Response latencies for the similar/different relation were shorter than those for the before/after relation, indicating that the former relational responding may be at greater strength than the latter responding (at least in the context of the relatively simple and commonly encountered relational stimuli employed here). This result is consistent with the argument that the equivalence relation is the most fundamental class of relational responding, and it is likely established very early in a child's verbal repertoire (Hayes,

1991). Furthermore, the current data are consistent with the results of a previous study, which showed that response latencies were shorter on tasks that probed for derived sameness/difference relations relative to temporal relations (O'Hora, Roche, & Barnes-Holmes, 2002). Also interesting to note is that the difference-score was greater for similar/different relative to the before/after relation, which indicates that reversing the former relation was more difficult than reversing the latter relation. This result would be expected if reversing the similar/different relation does indeed involve responding against the most fundamental and well-established verbal relation. Critically, the current results also revealed a relatively strong correlation between the similar/different difference-score and the verbal subtest. This suggests that the ability to respond rapidly against a highly dominant verbal response (i.e., greater relational flexibility) may be an important feature of verbal intelligence.

The current data revealed that on inconsistent trials the similar/different task loaded onto the verbal subtest, whereas the before/after task loaded onto the matrices subtest. However, the difference-score may provide a relatively uncontaminated measure of relational flexibility. Interestingly, the difference-score was found to load heavily onto the verbal subscale for both the similar/different and the before/after tasks. These findings indicate that *flexibility* in relational responding loads almost exclusively onto the verbal as opposed to the performance domain. On balance, however, the fact that flexibility in relational responding loads onto the verbal and not the matrices subscale may simply be a function of the specific type of verbal and matrices tasks employed in the K-BIT. Perhaps a more extensive intelligence test, such as the WAIS-III, which incorporates a number of verbal and performance subtests, might indicate that relational flexibility loads onto only specific performance measures. In fact, O'Hora et al. (2008) found that accuracy on a temporal relations task did indeed correlate with some performance tasks (e.g., block design) but not with others (e.g., matrices and picture completion).

The finding that the two measures of relational flexibility (i.e., the difference-score and response latencies on inconsistent trials) produced a range of significant correlations with IQ may have important practical implications. If the current findings are applied to educational settings, for instance, a possible conclusion might be that in addition to training correct and fluent relational responding, building *flexible* repertoires may also be crucially important in terms of promoting intelligent behaviors. Indeed, this conclusion is consistent with RFT (e.g., Barnes, Hegarty, & Smeets, 1997) and preliminary findings in applied educational settings. For instance, O'Connor (2004) found that introducing an intervention program designed to promote flexible response repertoires facilitated the acquisition of new skills in children with autism. Moreover, an abundance of evidence suggests that cognitive rigidity interferes with learning, and this interference has been observed across numerous populations (e.g., autism, Turner 1999; Attention Deficit/Hyperactivity Disorder, Lovecky, 2004; schizophrenia, Pishkin & Williams, 1997; and normal adults, Wulfert, Greenway, Frakas, Hayes, & Dougher, 1994). Insofar as rigidity is the antithesis of flexibility, then these studies and the current work lend support to the idea that promoting relational flexibility may be crucial in any learning environment.

At least two specific issues raised by the current study might be addressed in future research. First, although they provided a range of

demographic details, the participants were predominately young and female. As a result, the effects of variables such as age and gender could not be examined. Perhaps future studies could explore to what extent such variables of attribute impact the IRAP performance as a predictor of IQ. Second, all the participants were exposed to the K-BIT before they had completed the two IRAPs. Although prior exposure to just two IRAPs would not likely have any significant impact on subsequent IQ test performance, perhaps future studies might also explore the extent to which repeated and extensive IRAP exposures across a range of relational tasks would influence traditional measures of human intelligence.

## References

- 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.
- ANDREWS, G. & HALFORD, G. S. (2002). A cognitive complexity metric applied to cognitive development. *Cognitive Psychology, 45*, 153-219.
- BARNES, D., & HEGARTY, N., & SMEETS, P. M., (1997). Relating equivalence relations to equivalence relations: A relational framing model of complex human functioning. *The Analysis of Verbal Behavior, 14*, 1-27.
- BARNES-HOLMES, D., BARNES-HOLMES, Y., POWER, P., HAYDEN, E., MILNE, R., & STEWART, I. (2006). Do you really know what you believe? Developing the Implicit Relational Assessment Procedure as an indirect measure of implicit beliefs. *The Irish Psychologist, 32*, 169-177.
- BARNES-HOLMES, Y., BARNES-HOLMES, D., ROCHE, B., HEALY, O., LYDDY, F., CULLINAN, V., ET AL. (2001). Psychological development. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. (pp. 157-180). New York: Plenum.
- CATTELL, R. B. (1971). *Abilities: Their structure, growth and action*. Boston: Houghton Mifflin.
- GENTNER, D., & LOEWENSTEIN, J. (2002b). Relational language and relational thought. In J. Byrnes & E. Amsel (Eds.), *Language, literacy, and cognitive development* (pp. 87-120). New Jersey: LEA.
- GENTNER, D., & RATTERMANN, M. J. (1991). Language and the career of similarity. In S. A. Gelman & J. P. Byrnes (Eds.), *Perspectives on thought and language: Interrelations in development* (pp. 225-277). London: Cambridge University Press.
- 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: Context Press.
- HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition*. New York: Plenum.
- KAUFMAN, A. S., & KAUFMAN, N. L. (1990) *The Kaufman Brief Intelligence Test*. Minnesota: American Guidance Service.
- KYLLONEN, P.C., LOHMAN, D. F., & WOLTZ, D. J. (1984). Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology, 76*, 1325-1345.

- LOVECKY, D., V. (2004). *Different minds: Gifted children with AD/HD, Aspergers syndrome, and other learning deficits*. London: Jessica Kingsley Publishers.
- MARTINEZ, M. E. (2000). *Education as the cultivation of intelligence*. New Jersey: Lawrence Erlbaum Associates.
- MCKENNA, I., BARNES-HOLMES, D., BARNES-HOLMES, Y., & STEWART, I. (2007). Testing the fake-ability of the Implicit Relational Assessment Procedure (IRAP): The first study. *International Journal of Psychology and Psychological Therapy*, 7, 253-268.
- O'CONNOR, J. (2004). *Designing procedures to establish repertoires of derived relational responding*. Unpublished doctoral thesis, National University of Ireland, Maynooth, Maynooth, Co. Kildare, Ireland.
- O'HORA, D., PELÁEZ, 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, T., & CHAUDHARY, T. (2008). Temporal relations and intelligence: Correlating relational performance with performance on the WAIS-III. *The Psychological Record*, 58, 569-583.
- O'HORA, D., ROCHE, B., BARNES-HOLMES, D., & SMEETS, P. M. (2002). Response latencies to multiple derived stimulus relations: Testing two predictions of Relational Frame Theory. *The Psychological Record*, 52, 51-75.
- PISKIN, V., & WILLIAMS, W. V. (1977). Cognitive rigidity in information processing of chronic undifferentiated schizophrenics. *Journal of Clinical Psychology*, 33, 625-630.
- PREMACK, D. (2004). Is language the key to human intelligence? *Science*, 303, 318-320.
- SIDMAN, M. (1994). *Stimulus equivalence: A research story*. Boston: Authors Cooperative.
- STERNBERG, R., J. (2000) *Handbook of intelligence*. Cambridge: Cambridge University Press.
- TURNER, M. A. (1999). Generating novel ideas: Fluency performance in high-functioning and learning-disabled individuals with autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, 40, 189-201.
- WESCHLER, D. (1981) *The Weschler Adult Intelligence Scale-Revised*. San Antonio: The Psychological Corporation.
- WESCHLER, D. (1997) *The Weschler Adult Intelligent Scale-Third Edition*. San Antonio: The Psychological Corporation.
- WULFERT, E., GREENWAY, D. E., FARKAS, P., HAYES, S. C., & DOUGHER, M. J. (1994). Correlation between self-reported rigidity and rule-governed insensitivity to operant contingencies. *Journal of Applied Behavior Analysis*, 27, 659-671.