The Psychological Record, 2005, 55, 155-175

DERIVED RELATIONAL RESPONDING AND PERFORMANCE ON VERBAL SUBTESTS OF THE WAIS-III

DENIS O'HORA
University of Ulster

MARTHA PELAEZ
Florida International University

DERMOT BARNES-HOLMES
National University of Ireland, Maynooth

Twenty-six monolingual and 46 bilingual college students were assigned to 2 groups on the basis of their performance on a complex relational task, an empirical model of instructional control (O'Hora, Barnes-Holmes, Roche, & Smeets, 2004). The subjects were then exposed to the vocabulary, arithmetic, and digit-symbol encoding subtests of the WAIS-III. Subjects (N = 31) who successfully completed the relational task performed significantly better on the vocabulary and arithmetic subtests than those subjects (N = 44) who failed to do so. No significant differences in relational task performances of these 2 groups were obtained on the digit-symbol encoding subtest. In post-hoc statistical analyses, a low but significant correlation was obtained between the vocabulary and arithmetic scores and the percentage of correct responses emitted in 1 particular training phase of the relational task. Monolingual and bilingual subjects' performances were not significantly different in either the relational task or the WAIS subtests. These findings support the position that derived relational performances may provide a behavioral approach to human language abilities.

Derived relational responding, and stimulus equivalence in particular, has been the focus of hundreds of laboratory studies (see Hayes, Barnes-Holmes, & Roche, 2001; Sidman, 1994, for reviews). One reason for this interest was the suggestion by Sidman and Tailby (1982) that such relational responding could provide a behavioral model of semantic or symbolic relations. In particular, according to Sidman (1994):

This research was partially supported by the Florida International University, College of Education, Dean's Grant, Spring, 2002, titled: "Instructional Control and Relational Framing Across Cultures as a Function of Language, Mathematical, and Spatial Skills."

We thank Janet Lubian and Luzmary Amesty for their assistance in data collection and Dr. Joe Spradlin and an anonymous reviewer for their helpful comments on an earlier version of this article. Please mail all correspondence related to this manuscript to Dr. Martha Pelaez, Department of Educational Psychological Studies, College of Education 242-B, Florida International University, Miami, FL 33199. (E-mail: Marthapn@aol.com).

> the equivalence paradigm [one type of derived relational responding] demonstrates . . . one way that words can come to "mean" what they "stand for". The phenomena are real; we see them all about us and we can reproduce them in the laboratory. I think the phenomena are important in their own right, and I think they are a special property of language in the sense that they help make language as powerful as it is (p. 563).

For many years, interventions designed to establish derived relational responding have been used to establish a wide range of novel verbal performances in a wide range of subjects. Recently, however, there has also been considerable empirical analysis of the relationship between derived relational responding and language.

One early observation (Sidman, 1971) was that both derived stimulus relations and the relations between a word in a language and its referent are bidirectional. In a stimulus equivalence experiment, for example, when trained that A is the 'same as' B, adult subjects will typically choose B as the 'same as' A, which is a *derived* or untrained relation. Similarly, if we are told that the English word 'apple' *stands for* an actual apple, then we know that the actual apple *is called* 'apple.' In this way, the derived relations and linguistic relations function in similar ways. Hayes and Bisset (1998) provided evidence of this functional similarity when they demonstrated that priming, a classic property of semantically related stimuli, occurred between members of equivalence classes. Specifically, using a lexical decision task, they found that subjects responded more quickly and more accurately to nonsense words that were related either directly or through derived relations than to words that were unrelated.

Other researchers have examined the physiological correlates of both derived relational responding and complex language performance. In particular, Dickins et al. (2001) employed fMRI technology in order to compare brain activation of 12 subjects during tests for equivalence relations with activation during a test of verbal fluency. Both derived relational responding and the verbal fluency task activated the dorsolateral prefrontal cortex (DLPFC) and posterior parietal cortex bilaterally, but only the verbal fluency task activated Broca's area. Dickins et al. concluded, therefore, that brain activation during derived relational responding resembled activation during semantic processing but not the simple subvocal articulation of stimulus names. Ongoing work (e.g., McIlvane, DiFiore, & Wilkinson, 2003; Staunton, Barnes-Holmes, Whelan, & Barnes-Holmes, 2003) on event-related potentials (ERPs) observed during derived relational responding promises to further elucidate these similarities.

A third strand of research has established that performance on derived relational responding tasks varies with language performance. A number of studies have attempted to establish stimulus equivalence, a type of derived relational performance, in subjects with varying language abilities. For example, Devany, Hayes, and Nelson (1986) and Barnes, McCullagh, and Keenan (1990) found that equivalence responding was absent in language disabled children and Peláez, Gewirtz, Sanchez, and

Mahabir (2000) found similar effects with prelinguistic infants. In a longitudinal study, Lipkens, Hayes, and Hayes (1993) tracked the emergence of a simple repertoire of derived relational responding in a single child and their findings suggested that such responding showed a developmental trend similar to language. Finally, in language-able human subjects, equivalence and other complex derived relational performances emerge readily, but nonhuman subjects require extensive training and testing to demonstrate such performances (e.g., Schusterman & Kastak, 1994). Indeed, due to certain characteristics of the training and testing procedures employed in these nonhuman studies, there remains some debate as to whether these performances satisfy the necessary requirements for derived relational responding.

Derived Relational Responding and the WAIS-III

The current study contributes to the foregoing strand of research by comparing performance on derived relational tasks to performance on an accepted psychometric measure of verbal performance, the Wechsler Adult Intelligence Scale – Third Edition (WAIS-III; Wechsler, 1997). The WAIS-III was employed in the current study in order to provide an accepted measure of verbal performance. This instrument is the latest revision of the Wechsler-Bellevue Intelligence Scale developed by David Wechsler in 1939 and it is currently the most widely used test of adult intelligence. The basic format of the WAIS-III is very similar to its predecessor, the Wechsler Adult Intelligence Scale-Revised, and the previous Wechsler Adult Intelligence Scale. However, the WAIS-III was re-formatted to be similar to the Wechsler Intelligence Scale for Children – Third Edition (WISC-III; Wechsler, 1991). Specifically, similar to the previous scales, the WAIS-III includes verbal, performance, and full-scale IQ scores, but it also includes four broad first-order factors: verbal comprehension, perceptual organization, working memory, and processing speed.

The WAIS-III also provides scaled scores on each of its 14 subtests (M = 10, SD = 3). For the purposes of the current study, we employed three subtests of the WAIS-III; the vocabulary subtest, the arithmetic subtest, and the digit-symbol coding subtest. These subtests contribute to the following factors: The vocabulary subtest contributes to both the verbal comprehension index and verbal IQ factors, the arithmetic subtest contributes to the working memory index and verbal IQ factors, and the digit-symbol coding subtest contributes to the processing speed index and performance IQ factors. The current study investigated whether performance on complex relational tasks would predict performance on the subtests that contribute to the verbal comprehension index and the verbal IQ factor (vocabulary, arithmetic) but not on those subtests that contribute to the performance IQ factor (digit-symbol coding).

The WAIS-III was standardized on 2450 adult subjects, selected according to 1995 U.S census data, and stratified according to age, gender, race/ethnicity, geographic region, and education level. The

WAIS-III thus constitutes a highly reliable measure of ability or intelligence as traditionally defined. The WAIS-III is also widely respected by clinical psychologists and psychiatrists who employ it in order to investigate deficits in particular cognitive abilities. Nevertheless, the WAIS-III instrument was designed to measure intelligence and thus, one might wonder why behavioral researchers would employ such an instrument. However, previous empirical studies within the behavioral literature have employed psychometric tests to measure rigidity (Wulfert, Greenway, Farkas, Hayes, & Dougher, 1994), and personality behaviors (Harzem, 1984). Also, in the applied arena, it is commonplace to use psychometric instruments to approximate a client's behavioral history.

In the current study, we did not employ the subtests of the WAIS-III to measure subjects' internal and unchangeable abilities. Rather, we employed the verbal subtests as measures of subjects' performance on accepted measures of verbal responding. Not only does this contribute to the burgeoning literature on derived stimulus relations, but it may also provide the first steps to increasing the relevance of behavioral approaches to language within mainstream psychology.

Multiple Stimulus Relations

Many studies have compared language performance to responding solely in accordance with equivalence relations. However, one approach to stimulus equivalence, Relational Frame Theory (RFT; Hayes et al., 2001) suggests that equivalence is just one of many possible derived relations. Previous work has indeed demonstrated that human subjects can be trained to respond to a variety of derived stimulus relations, including: Same and Opposite (Dymond & Barnes, 1996; Roche & Barnes, 1996a, 1997; Roche, Barnes-Holmes, Smeets, Barnes-Holmes, & McGeady, 2000; Steele & Hayes, 1991), Different (Roche & Barnes, 1996a; Steele & Hayes, 1991), and More than and Less than (Dymond & Barnes, 1995; O'Hora, Roche, Barnes-Holmes & Smeets, 2002). In addition, empirical models of many complex language phenomena have been based on multiple derived stimulus relations. These include metaphor and analogy (Stewart & Barnes-Holmes, 2001; Stewart, Barnes-Holmes, Roche, & Smeets, 2001), instructional control (O'Hora, Barnes-Holmes, Roche, & Smeets, 2004), "Theory of Mind" (McHugh, Barnes-Holmes, & Barnes-Holmes, 2004), and sexual behavior (Roche & Barnes, 1995; 1996a; 1996b). Finally, in reference to vocabulary in particular, Barnes-Holmes et al. (2001) suggested that:

> Persons with a 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 performance will correlate less with overall levels of intellectual behavior than will vocabulary even though both tasks involve verbal material. (p. 160)

Thus, Relational Frame Theory predicts that higher levels of proficiency on relational responding tests should be a better predictor of performance on a vocabulary subtest than on other, less relationally rich subtests.

A computer-based empirical model of instructional control developed by O'Hora, Barnes-Holmes, Roche, & Smeets (2004) was employed in the current research. According to this methodology, subjects are first trained to respond in accordance with Same, Different, Before, and After relations in the presence of particular arbitrary stimuli across four training and testing phases. Following successful completion of these four phases, subjects were then exposed to a test for instructional control, in which these derived relational responses are combined to control a sequence of four responses. A final phase tests the generalization of this derived relational performance in the presence of stimuli from 24 novel stimulus sets.

In the current study, subjects were first asked to complete a monolingual/bilingual assessment questionnaire. Subjects were then exposed to the training and testing procedures involved in the empirical model of instructional control. Subjects who passed or failed this relational task were subsequently exposed to three subtests of the WAIS-III abilities scale: the vocabulary subtest, the arithmetic subtest, and the digit-symbol coding subtest. The performances on these three subtests of bilingual and monolingual subjects and of subjects who passed or failed the derived relational task were then compared.

Method

Subjects

Seventy-five undergraduate students (55 female and 20 male) recruited from the Florida International University in Miami participated in the current study. Subjects ranged in age between 18 and 54 years old (M = 25.8 yr) and received course credit for their participation. All but 4 of the subjects were from courses other than psychology and none of the subjects was familiar with either the WAIS-III scale or with the study of derived relational responding.

Setting and Apparatus

The study was conducted in a three-room experimental suite at Florida International University. On arrival at the laboratory, subjects were seated in the waiting room (14 x 6 ft) and filled out an informed consent form. They responded to a series of questions on demographic information (i.e., age, gender, ethnicity, place of birth, major) and to whether they were bilingual or monolingual (see Appendix 1).

Upon completion of these two forms, each subject was individually brought to an adjacent control-observational room (18 x 8 ft). The windowless control-observational room contained a large table, two file cabinets, three Mac computers, and two chairs. In this room each subject completed the monolingual-bilingual assessment and the relational responding task, which lasted approximately 1 hour and 30 minutes. The six-phase relational task was presented on an Apple iMac® computer with a 14" display. The letters Z, C, B, and M were covered by different colored squares of masking tape (green, red, blue, and yellow, respectively). Presentation of stimuli and recording of responses were controlled by the experiment-generating software application PsyScope (Cohen, MacWhinney, Flatt, & Provost, 1993).

Subsequently, each subject entered an adjacent experimental room for the WAIS assessment. Subjects were assessed on three subtests from The Wechsler Adult Intelligence Scale-III (Wechsler, 1997); the vocabulary subtest, the arithmetic subtest, and the digit-symbol coding subtest. This experimental room contained a small table placed against the wall and two chairs. The subject and the examinee could be observed from the control room through a one-way mirror for intermittent check-ups on the application of the tests.

General Procedure

Subjects were exposed to the following three tasks. First, the experimenter conducted an assessment of language comprehension. On the basis of this assessment, subjects were rated as bilingual or monolingual. Second, subjects were exposed to a complex relational task, a computer-based empirical model of instructional control. Subjects who successfully completed the relational task were assigned to the relational consistent responding (RCR) group; subjects who failed any phase were assigned to the relational inconsistent responding (RIR) group. Third, subjects were exposed to three subtests of the WAIS III: the vocabulary subtest, the arithmetic subtest, and the digit-symbol coding subtest. After completing these, subjects were debriefed and they left the laboratory. Each of the three tasks will now be described in turn.

Monolingual-Bilingual Assessment

In order to test second language ability, subjects were assessed using the Spanish Language Comprehension Assessment, developed by one of the main experimenters who is fully bilingual. The questions were validated for consistency with a number of pilot subjects. All 75 subjects were examined by two fully bilingual school psychologist graduate students working at the Learning Laboratory at Florida International University. The assessment consists of 5 questions in Spanish selected randomly from a total pull of 10 questions (see Appendix 1).

From those five questions, three were read aloud by the examiner to the subject, and the remaining two were presented in written form. All answers were spoken aloud and required a minimum of three sentences in Spanish in order to assess *expressive* language. Subjects varied in the manner in which they replied to questions given by the bilingual evaluator. In fact, some monolingual subjects indicated a complete lack of Spanish understanding while others, with some receptive comprehension, translated the meaning of the questions to English. Their level of comprehension indicated the presence of *receptive* knowledge of the

Spanish language. Each answer was scored according to a 1-5 (5-25) Likert-type scale ranging from "not at all" to "excellent" understanding of Spanish. The lowest score that could therefore be obtained by a subject was 5 and the highest was 25. Subjects were assigned to the following groups on the basis of their scores: 5-7 = M+ (fully monolingual); 8-10 = M- (Monolingual); 11-15 = B- (bilingual); 16-25 = B+ (fully bilingual). The questionnaire yielded two scores; one for expressive language and a second for receptive language. The lower of the two scores, usually expressive, was taken as the score on which the subject was categorized. For instance, to be considered fully bilingual, both expressive and receptive scores needed to be at least 16 points.

Relational Task

The relational task employed for the current study was the empirical model of instructional control developed by O'Hora et al. (2004). The experimental sequence and mastery criteria for each phase are presented in Table 1. Subjects were first trained to respond according to Before and

Table 1

Sequence of Training and Testing Phases and Mastery Criteria in Relational Task

Training or Testing Phase	Mastery Criterion
Pretraining for Before and After relational responding	14/16 on last block
Testing for Before and After relational responding	30/32
Pretraining for Same and Different relational responding	14/16 on last block
Testing for Same and Different relational responding	15/16
Test for instructional control	20/24 on last 24 probes
Test for generalization	20/24

After relations using an REP procedure (see Figure 1). When subjects satisfied the mastery criterion on a test for Before and After relational responding, they were exposed to pretraining for Same and Different relational responding. Responding in accordance with Same and Different was achieved by exposing subjects to a modified match-to-sample procedure (see Figure 2). When subjects satisfied the mastery criterion on a test for Same and Different relational responding they were exposed to the test for instructional control, in which subjects were presented with a number of networks of Same and Before or After relations (see Figure 3). Finally, subjects were exposed to a test for generalization of instructional control in the presence of 24 novel stimulus sets.

All trials were presented on the computer monitor. Feedback followed responses on all training trials, which was followed in turn by an intertrial interval (i.e., the screen remained blank for 2.5 s). Following a correct response, the screen cleared, and the word "Correct" appeared accompanied by a high-pitched tone from the computer. Following an incorrect response, the word "Wrong" appeared accompanied by a low-pitched tone.

Pretraining for Before and After Relational Responding

Minimal instructions (i.e., orientation and response requirements only) were presented to subjects before beginning pretraining and testing for Before and After relational responding. During each pretraining trial and test probe, two complex stimuli were first presented at the bottom left and right corners of the screen. Reading from bottom to top, the complex stimuli consisted of two arbitrary shapes (e.g., a square and a circle) with an arbitrary contextual cue (e.g., ()()) presented between them (e.g., 'circle ()() square' /'square ()() circle'). Both complex stimuli were presented in a sequence such that the first arbitrary shapes presented in both complex stimuli were presented simultaneously at opposite sides of the screen followed by the simultaneous presentation of the contextual cue followed by simultaneous presentation of the second arbitrary shapes (see Figure 1). This was to encourage subjects to 'read' up the screen. After 2.5 s, one of the two arbitrary shapes (e.g., a circle) was presented for 1 s above both complex stimuli, and then disappeared. Following an interstimulus interval of 0.5 s, the second arbitrary shape (e.g., a square) was then presented for 1 s.

In order to establish contextual control of responding in accordance with Before and After relations, choosing one of the two complex stimuli presented at the bottom of the screen was reinforced based upon the order in which the two latter shapes were presented at the top of the screen. For example, in a typical trial to establish the function of 'BEFORE' in the arbitrary contextual cue '()()', the complex stimuli 'circle ()() square' and 'square ()() circle' were first presented in a sequence up the screen as described above. Then, at the top of the screen, a circle flashed up followed by a square. In such a trial, choosing 'circle ()() square' was then reinforced. In this way, responding in accordance with a 'BEFORE' relation was reinforced in the presence of the arbitrary stimulus '()()'. On further trials, given a choice between 'circle :::: square' and 'square :::: circle', and the same nonarbitrary sequence, circle followed by square, choosing 'square :::: circle' was reinforced. In this way, the second arbitrary contextual cue, '::::', acquired the function of AFTER.

The mastery criterion for pretraining was set at 14/16 on the last block of trials. Following pretraining, subjects were exposed to two blocks of test probes that utilized two novel stimulus pairs and the mastery criterion was 30/32 probes correct. If subjects satisfied the mastery criterion on the test for Before and After relational responding, they were exposed to pretraining and testing for Same and Different relations. Subjects who did not satisfy the mastery criterion on this test were reexposed to pretraining for Before and After relations. If subjects continued to fail to demonstrate

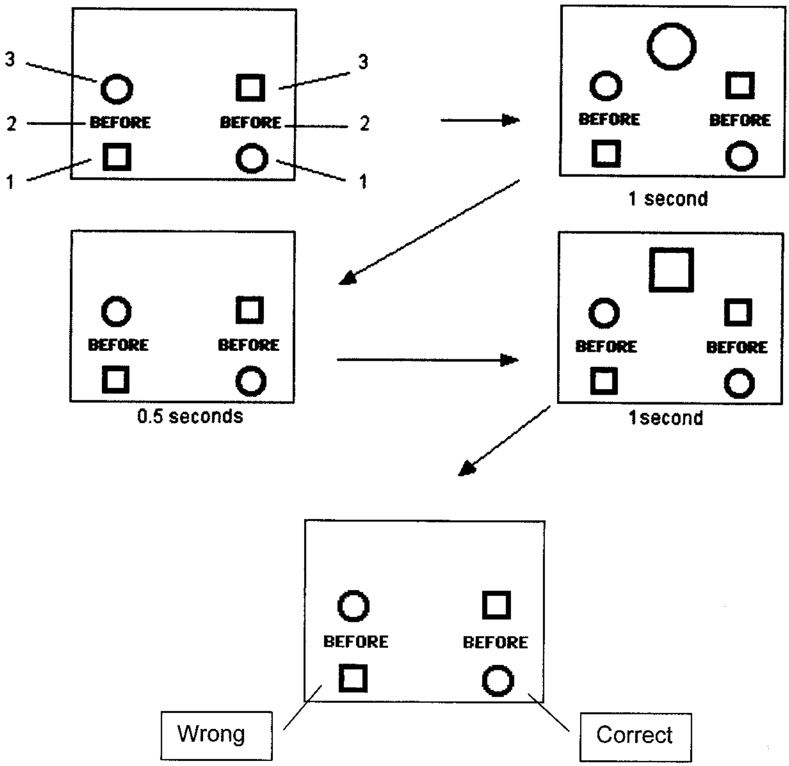

Figure 1. Flow chart outlining a representative task from pretraining for Before and After relational responding. Note that the elements on the initial screen appeared in the order indicated (i.e., the bottom elements were presented first, followed 0.5 s later by the middle elements and then 0.5 s later by the top elements. The words Before and After were not presented to subjects, but are used here to indicate the functions established in arbitrary contextual cues. Test probes were identical in form but novel arbitrary shapes were employed and no reinforcement was provided.

the required performance after exposure to 12 blocks of 16 training trials, they were assigned to the RIR (relational inconsistent responding) group.

Pretraining and Testing for Same and Different Relations

Pretraining and testing for Same and Different relations employed a match-to-sample type procedure (see Figure 2). Similar to the previous phase, only minimal instructions were presented to subjects. First, two comparison stimuli (e.g., a long line and a short line) appeared in the bottom left and right corners of the computer screen, then a sample stimulus (e.g., a long line) appeared in the middle of the screen, and finally a contextual cue (e.g., %%% or ///) appeared in the center top third of the screen. The contextual cue, sample, and comparisons remained until a response was recorded. In the presence of one contextual cue (%%%), choosing the comparison stimulus (long line) that was the same as the sample stimulus (long line) was reinforced. In the presence of another contextual cue (///), choosing the comparison (short line) that was different from the sample stimulus (long line) was reinforced. In this way, the arbitrary contextual cues came to control responding according to Same (%%%) or Different (///) relations between sample and comparison stimuli.

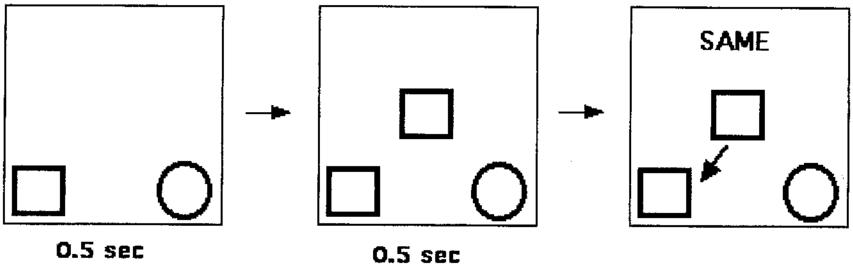

Figure 2. A diagrammatic representation of a task from the nonarbitrary relational pretraining for Same and Different relations. Unlike the tasks described in Chapter 3, stimuli appeared on the screen sequentially from bottom to top (i.e., comparison stimuli first, then sample, then contextual cue). Subjects were reinforced for choosing the same comparison stimulus as the sample stimulus in the presence of the SAME contextual cue and for choosing the comparison that was different from the sample in the presence of the DIFFERENT contextual cue.

Initially, subjects were exposed to two blocks of eight trials, and subsequently, subjects were exposed to single blocks of eight training trials. The mastery criterion for pretraining for Same and Different relations was seven correct responses out of eight on the last block of pretraining trials presented. When subjects achieved the mastery criterion, they were exposed to a testing session that consisted of two blocks of eight trials, during which no reinforcement was provided and the samples and comparisons were novel stimuli. If subjects satisfied the mastery criterion on the test for Same and Different relational responding (15/16 correct), they were exposed to the test for instructional control. If subjects failed to achieve this level of responding, they were reexposed to pretraining for Same and Different relations. If subjects continued to fail to demonstrate the required performance after exposure to 12 blocks of eight trials, they were assigned to the RIR group.

Test for Instructional Control

Given the above experimental history of reinforcement, four contextual cues had been established for responding in accordance with Same, Different, Before, and After relations. In the next phase of the experiment, subjects were presented with 36 'instruction' probes in the form of networks of Same, Different, Before, and After relations without reinforcement (see Figure 3). The following minimal instructions were presented to subjects before they were exposed to the test for instructional control:

> In a moment a series of images will appear at the bottom of this screen. A second series of images will then appear above those images. You must press the colored keys on the keyboard in a particular sequence based on the images on the computer screen. When you are finished pressing the colored keys, you must press the RETURN key to proceed.
>
> Hit any key when you are ready to begin.

At the start of each probe, C stimuli were presented vertically such that all four C stimuli (C1, C2, C3, or C4) were presented in a random order. In between each pair of C stimuli either Before or After contextual cues were presented (e.g., reading upwards: C1 Before C2 Before C3 Before C4).

Following an interval of 1 s, A, B, and C stimuli and Same or Different contextual cues were presented at the top of the screen. Each C stimulus

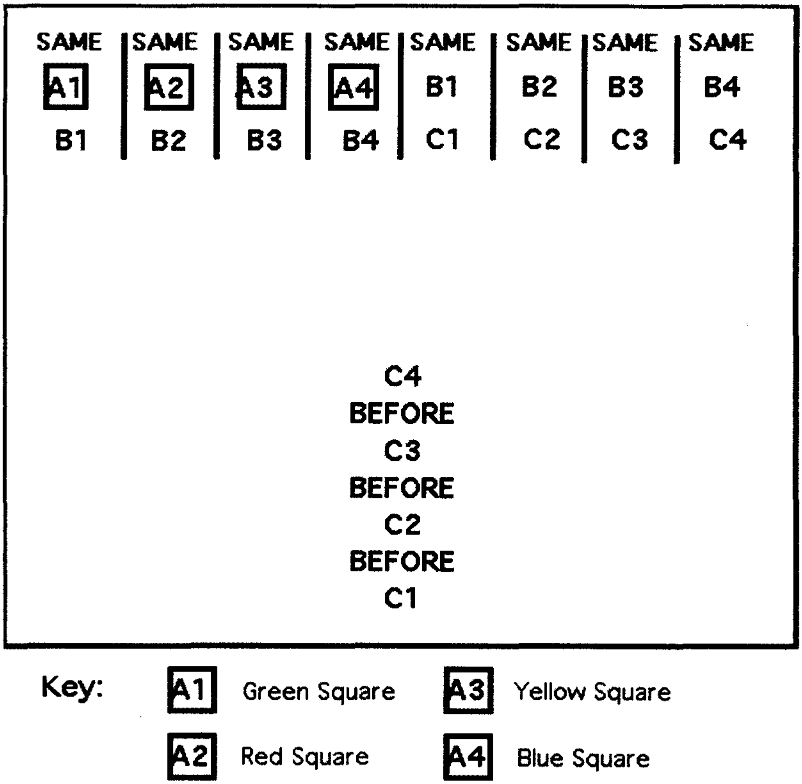

Figure 3. Diagrammatic representation of a test probe from the test for instructional control. Four computer keys were designated response keys and were colored green, red, yellow, and blue. Sequences of responses were predicted based on the C stimuli that were the SAME as the A stimuli according to the top of the screen (e.g., A1, green, is the same as B1, which is the same as C1) and the order of C stimuli and the presence of BEFORE or AFTER cues (e.g., C1 is before C2, which is before C3 and so on). The correct response to the above probe was Green – Red – Yellow – Blue. No reinforcement was provided during the test.

was presented beneath a B stimulus and above these, either a Same or Different contextual cue was presented (e.g., reading upwards: C1 B1 Same). On the left-hand side of the screen, each B stimulus was presented beneath an A stimulus (a colored square) and above these a Same contextual cue was presented (e.g., reading upwards: B1 A1 Same).

Each exposure to the test for instructional control included three types of test probes; Same Sequential probes, Same Nonsequential probes, and Different Nonsequential probes.

Same Sequential probes. Only Same contextual cues were presented at the top of the screen with the pairs of A and B stimuli and pairs of B and C stimuli. At the bottom of the screen, the C stimuli and Before or After contextual cues were presented in a temporal sequence from bottom to top.

Same Nonsequential probes. These probes also included only Same contextual cues, but C stimuli and Before or After contextual cues appeared simultaneously at the bottom of the screen.

Different Nonsequential probes. C stimuli and Before or After contextual cues appeared simultaneously at the bottom of the screen, but Different contextual cues were presented at the top of the screen with the pairs of A and B stimuli and B and C stimuli. Following the initial presentation of four Same Sequential probes, subjects were exposed to a combined block of 24 Nonsequential test probes, which was comprised of 12 Same Nonsequential and 12 Different Nonsequential test probes presented in no particular order.

For test probes that included only Same contextual cues at the top of the screen, a particular four key sequence constituted a correct response. For example, given the stimuli C1 Before C2 Before C3 Before C4 and the stimuli C1 B1 Same/ C2 B2 Same/ C3 B3 Same/ C4 B4 Same, and B1 A1 (green) Same/ B2 A2 (red) Same/ B3 A3 (yellow) Same/ B4 A4 (blue) Same, it was expected that subjects would emit the following four key response: Green – Red – Yellow – Blue, followed by the [ENTER] key. Pressing the [ENTER] key signaled the end of the test probe and all stimuli remained on the screen until the [ENTER] key was pressed.

In the presence of the Different contextual cues, however, no specific response was prescribed by the relational network presented. Consider the following as an example: Given the stimuli C1 Before C2 Before C3 Before C4 (reading upwards), and the stimuli C1 B1 Different/ C2 B2 Different/ C3 B3 Different/ C4 B4 Different, and B1 A1 (green) Same/ B2 A2 (red) Same/ B3 A3 (yellow) Same/ B4 C4 (blue) Same, no specific response may be predicted. Rather, in this case, any response other than Green – Red – Yellow – Blue, followed by the [ENTER] key was considered correct.

The mastery criterion for the test for instructional control was a minimum of 20/24 correct responses on the 24 nonsequential probes. If subjects satisfied this mastery criterion, they were exposed to the final stage, the test for generalization. If subjects failed to demonstrate the required performance, they were assigned to the RIR group.

Test for Generalization

When subjects reached criterion on the test for instructional control (Phase 5), they were then exposed to a variation of this test that included 24 novel stimulus sets. Otherwise, the method of presentation was identical to that in the test for instructional control. The first eight probes consisted of novel sets of nonsense syllables. The second eight stimulus sets consisted of nonsense shapes, and the final eight stimulus sets consisted of clip art pictures of particular themes acquired from the Appleworks software package (e.g., fish, dinosaurs, cars). The mastery criterion for this test was a minimum of 20/24 correct responses and if subjects demonstrated this performance they were assigned to the RCR (relation consistent responding) group. If a subject failed this test or any previous training or testing stage, he or she was assigned to the RIR (relation inconsistent responding) group.

Wechsler Adult Intelligence Scale-Third Edition (WAIS-III)

After each subject had completed the second language assessment and the relational responding tasks, three subtests from the WAIS-III intelligence test were administered individually. The WAIS-III is a clinical instrument for assessing the intellectual ability of adults aged 16 through 89. It contains 14 subtests. For the purpose of the current study, however, only the vocabulary, arithmetic, and digital symbol subtests were administered. In order to score the subtests of the WAIS III, the scores obtained from each test were scaled to the range of equivalent scores according to the subjects' ages, of which the minimum was 1 and the maximum was 19. A thorough description of these tests is beyond the remit of the current paper, but a summary of each is provided here (see Wechsler, 1997, for further details).

Vocabulary subtest. Subjects were asked to define a series of 33 words, which were presented orally and visually by a trained experimenter. Scores were calculated by matching the examinee's responses to the sample responses for that item found in the WAIS-III manual (pp. 70-90). If the examinee's response was too vague or unclear to be readily scored, the examiner asked: "Tell me more about it or explain what you mean." No other questioning was permitted.

Arithmetic subtest. This subtest included a series of 20 arithmetic problems that the examinee solved mentally and answered orally within a set time limit. The examiner, who was seated in front of the subject, timed each item from immediately after the problem was first read to the subject. The subject's answer and time taken to solve each problem was recorded and scored if it was correct and within the completion time limit permitted for the particular item.

Digit-symbol-coding subtest. In this subtest, a series of numbers was paired with corresponding hieroglyphic-like symbols. The experimenter wrote the first symbol corresponding to its number. The subject was then asked to draw the symbols for as many numbers as possible within the 120-s (2-min) time limit. The experimenter recorded the number of correctly drawn symbols.

Results

Descriptive Analysis of the Data

The large number of subjects employed in the current study precludes the presentation of individual data. Rather, salient characteristics of subjects' performances on monolingual/bilingual assessment, the relational task, and the WAIS-III subtests are discussed.

Monolingual/bilingual assessment. The numbers of subjects assigned to each monolingual/bilingual category was as follows: fully monolingual (5-7): 15, monolingual (8-10): 11, bilingual (11-15): 8, fully bilingual (16-25): 41. Due to the small numbers in the bilingual group, subjects in both the fully monolingual and monolingual categories were treated as monolingual (N = 26) and subjects in both the fully bilingual and bilingual categories were treated as bilingual (N = 49) for statistical analyses.

Computer-based model of instructional control. Thirty-two subjects failed to achieve the mastery criterion on relational training for Before and After. The remaining 43 subjects achieved the mastery criterion on this phase and also satisfied the mastery criterion for the test for Before and After relational responding, relational training for Same and Different, and the test for Same and Different relational responding at their first exposure. Of these 43 subjects, 31 passed the test for instructional control and also the test for generalization with 24 novel stimulus sets. Therefore, 31 subjects successfully completed the derived relational task (relational-consistent responding, RCR) and 44 failed (relational-inconsistent subjects, RIR).

The Wechsler Adult Intelligence Scale-Third Edition (WAIS-III). Mean scores of each of the four experimental groups were calculated on each of the three WAIS-III subtests. On the vocabulary subtest, the highest mean score was obtained from the monolingual RCR group (M = 13.1, SD = 1.85), and the lowest from the monolingual RIR group (M = 10.6, SD = 1.78). On the arithmetic subtest, the highest mean score was obtained from the monolingual RCR group (M = 12.0, SD = 1.94), and the lowest from the monolingual RIR group (M = 10.2, SD = 2.04). On the digit-symbol encoding subtest, the highest mean score was obtained from the monolingual RCR group (M = 12.4, SD = 2.8), and the lowest from the bilingual RCR group (M = 11.1, SD = 2.62). The mean scores of all groups are presented in graph form in Figure 4.

Statistical Analysis

The first analysis conducted was a multivariate analysis of variance (MANOVA). A 2 (RCR vs. RIR) x 2 (bilingual vs. monolingual) analysis on three dependent measures (i.e., the three WAIS-III subtests: vocabulary, arithmetic, and digit-symbol coding) was conducted. This MANOVA yielded a significant main effect for the RCR (N = 31) vs. RIR (N = 44) groups, $F(3, 69) = 7.31$, $p = .0002$, using Wilk's Lambda (.759), Roy's Greatest Root (.318), Hotelling-Lawley Trace (.318), and Pillai Trace (.241). No significant effect due to bilingual vs. monolingual was found, and no interaction effect between the two factors was obtained.

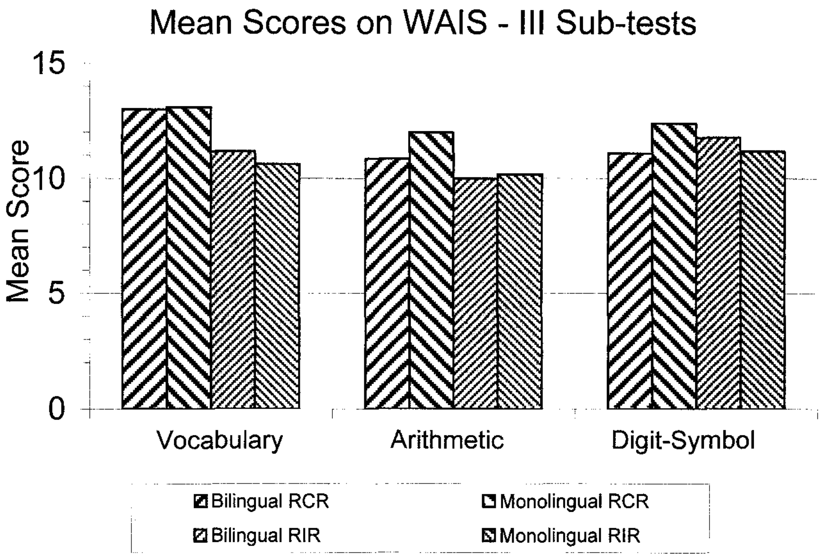

Figure 4. Histogram representing the mean scores on each of the three WAIS-II subtests for each of the four groups of subjects.

Given no difference between monolingual and bilingual subjects, the data of the monolingual and bilingual groups were collapsed (N = 75) for subsequent independent ANOVAs on the three main dependent measures. The first ANOVA yielded a highly significant difference between the performances of the RCR group (N = 31; M = 13.05, SD = 1.81) and the RIR group (N = 44; M = 10.9, SD =1.87) on the vocabulary subtest on the WAIS-III, $F(1, 71) = 21.78$, $p < .0001$. In addition, a significant effect for the arithmetic subtest, $F(1, 71) = 5.90$, $p = .017$, was found between RCR (N = 31; M = 11.42, SD = 2.02) and the RIR group (N = 44; M = 10.09, SD = 2.25). No significant effects ($p > .05$) were obtained between the RCR group (N = 31; M = 11.5, SD = 2.71) and the RIR group (N = 44; M = 11.6, SD = 3.14) on the digit-symbol encoding subtest.

An additional analysis was conducted to measure the relationship between the performance on this first relational training phase and the performance on the vocabulary and arithmetic subtests. To conduct this analysis the percentage of correct responses produced by each subject on the relational training phase was calculated. Two separate Pearson r correlations revealed significant correlations between scores on the vocabulary subtest and the number of correct responses during relational training for Before and After (r = .342, p = .002, N = 74) and between scores on the arithmetic subtest and the number of correct responses on this relational training phase, (r = .231, p = .003, N = 74).

General Discussion

The current study sought to determine if performance on relational tasks would predict performance on verbal or performance subtests of The Wechsler Adult Intelligence Scale Third Edition (WAIS-III). Subjects who completed the relational task performed significantly better on the verbal subtests (i.e., vocabulary and arithmetic) than subjects who failed to do so. There were no significant differences between the performances of those who completed or failed on the digit-symbol encoding subtest and no significant differences were observed in the performances of monolingual and bilingual subjects on any of the three subtests. These findings therefore support those of previous studies that have shown that derived relational responding and performance on complex language tasks are closely related.

Three subtests of the WAIS-III were employed in the current study: vocabulary, arithmetic, and digit-symbol coding. As pointed out previously, the vocabulary subtest contributes to both the verbal comprehension index and verbal IQ scores, whereas the arithmetic subtest contributes to the verbal IQ score alone and the digit-symbol coding subtest does not contribute to either verbal factor. In the current study, passing the relational task was a highly significant predictor of performance on the vocabulary subtest ($p < 0.001$), a significant predictor of performance on the arithmetic subtest ($p < 0.017$), and did not predict performance on the digit-symbol coding subtest. Arguably, therefore, the test most representative of language proficiency was the one best predicted by performance on the complex relational task. This pattern of findings lends further weight to the relevance of derived relational responding in the behavioral understanding of human language. In addition, these data provide support for the suggestion by Barnes-Holmes et al. (2001) that vocabulary and mathematical skills are relationally rich repertoires from an RFT perspective, and that proficiency in these areas should correlate with proficiency in relational responding in laboratory settings.

Interestingly, of the three WAIS-III subtests employed in the current study, the digit-symbol coding subtest most closely resembled the computer-based relational task. The digit-symbol coding subtest required subjects to match abstract symbols to digits and the relational task employed abstract shapes and symbols during all phases. Furthermore, successful performance during the test for instructional control required subjects to match nonsense syllables to colored squares and to order sequences of responses based on arbitrary cues. Nevertheless, performance on the relational task was an extremely poor predictor of performance on the digit-symbol coding subtest. Conversely, successful performance on the relational task was an extremely good predictor of performance on the vocabulary subtest and the face validity of this subtest as a measure of general language ability is quite high. These findings suggest that, although the relational task and the vocabulary subtest were quite distinct formally, they were quite similar functionally (see Carpentier, Smeets, & Barnes-Holmes, 2002, for a relevant discussion). Interestingly, these findings are consistent with those of Dickins et al. (2001), who recorded brain activation in areas involved in semantic processing (the DLPFC) during derived relational responding, but not in areas involved in simple identification or naming (Broca's area).

In the current study, monolingual and bilingual subjects performed equally well in verbal and nonverbal subtests. In previous work, Bialystok (1988, 1999) suggested two components of language processing, analysis and control, and demonstrated that control develops earlier in bilingual children than in comparable monolinguals. Otherwise, differences between monolinguals and bilinguals are not typical (Pena, Bedore, & Rappazzo, 2003) and extrapolations tend to be restricted by the various measures of bilingualism employed. In the current study, we employed a Spanish language assessment questionnaire to evaluate subjects' second language ability. This questionnaire was developed by Marha Peláez who is bilingual, and it was administered by bilingual school psychologists. The questionnaire had a high face validity and categorization of the subjects largely coincided with each of the bilingual researchers' own evaluations.

It should be noted that the WAIS-III instrument was not used in its entirety in the current study. Rather, due to the length of the computer-based relational responding task, three subtests were employed that were each related in varying degrees to the verbal factors of the instrument. Thus, a case might be advanced that the instrument was not intended for this purpose and that the observed effects might not be robust. However, these subtests are also standardized independently and provide scaled scores with a standardized mean and standard deviation. Nevertheless, future studies might well consider employing the full WAIS-III instrument in order to analyze relationships between derived relational performances and both verbal comprehension and verbal IQ factors, and also with each of the full complement of subtests (cf. Taub, 2001). Such work might well isolate specific properties of language performance as traditionally defined that are particularly similar to specific derived relational performances.

A further unresolved issue concerns the order in which subjects were exposed to each stage of the current experiment. These stages were not counterbalanced across subjects and, thus, the results may have been affected by either a practice effect or a fatigue effect. For instance, a practice effect may conceivably have facilitated performance in the vocabulary test in that at least one of the stages previous to it (the monolingual-bilingual assessment) was verbal in nature. However, if these order effects did occur, they would not account for the differences observed across groups and measures in the current study. In fact, if such effects were to occur they would have been more likely to reduce differences between groups and measures because the effect on subjects in all four groups would have been the same and may have dwarfed the group-based effect. On the contrary, the observed group-based effects were remarkably clear.

The relational task employed in the current study was a computer-based empirical model of instructional control. One source of concern centers on the high failure rate of adult human subjects on this relational task. One must consider, however, that at no point were subjects reexposed to the test for instructional control if they failed. Typically, the tasks developed to test for derived relational responding are quite complex. Indeed, there are many studies on stimulus relations that report that subjects fail at some point in the procedures due to the complexity of the tasks. These subjects are sometimes ignored or removed from the study or alternatively, they may be targeted for some form of remedial retraining to facilitate them passing the test. In the current study, we deliberately did not modify training, and chose instead to employ the three WAIS-III subtests to investigate whether successes or failures on our procedures might predict performances on accepted measures of language performance. Future studies, however, might seek to determine if subjects with relatively high verbal scores, but who fail the relational tests, then go on to pass those tests more readily with repeated exposures than subjects with lower verbal scores.

It is clear from the findings of the current study that a behavioral account of language must be informed by the literature on derived relational responding. However, such an account will not benefit only behavioral psychology. From a developmental perspective, for instance, the development of derived relational responding provides an alternative to 'bootstrapping' accounts of children's progress from non-language to language (see Altmann, 2001, for a detailed discussion). From an evolutionary perspective, Dickins and Dickins (2001) have suggested that derived relational responding may be critical to understanding how humans as a species have made the same transition. Finally, such a behavioral account sits well with recent biological research on the plasticity of brain function and the importance of context (e.g., Robertson & Murre, 1999).

In the current study, performance on a complex relational task predicted performance on verbal subtests of The Wechsler Adult Intelligence Scale - Third Edition (WAIS-III). This finding contributes to the growing body of research that suggests that derived relational responding and language are closely related phenomena and also provides support for a specific RFT prediction made by Barnes-Holmes et al. (2001). The broad scope of this research offers an opportunity for thoroughgoing functional analyses of human language and demonstrates that behavioral psychology still has much to offer to the study of complex human behavior.

References

ALTMANN, G. T. M. (2001). The language machine: Psycholinguistics in review. *British Journal of Psychology*, 92, 129-170.

BARNES, D., MCCULLAGH, P. D., & KEENAN, M. (1990). Equivalence class formation in non-hearing impaired children and hearing impaired children. *The Analysis of Verbal Behavior*, 8, 1-11.

BARNES-HOLMES, Y., BARNES-HOLMES, D., ROCHE, B., HEALY, O., LYDDY, F., CULLINAN, V., & HAYES, S. C. (2001). Psychological development. In S. C. Hayes, D. Barnes-Holmes, & B. T. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

BIALYSTOK, E. (1988). Levels of bilingualism and Levels of Linguistic Awareness. *Developmental Psychology,* 24(4), 560-567.

BIALYSTOK, E. (1999). Cognitive complexity and attentional control in the bilingual mind. *Child Development,* 70(3), 636-644.

CARPENTIER, F., SMEETS, P. M., & BARNES-HOLMES, D. (2002). Matching functionally same relations: Implications for equivalence-equivalence as a model for analogical reasoning. *The Psychological Record,* 52, 351-370.

COHEN, J. D., MACWHINNEY, B., FLATT, M., & PROVOST, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavioral Research Methods, Instruments and Computers,* 25(2), 257-271.

DEVANY, J. M., HAYES, S. C., & NELSON, E. O. (1986) Equivalence class formation in language-able and language-disabled children. *Journal of the Experimental Analysis of Behavior,* 46, 243-257.

DICKINS, D. W., SINGH, K. D., ROBERTS, N., BURNS, P., DOWNES, J. J., JIMMIESON, P., & BENTALL, R. P. (2001). An fMRI study of stimulus equivalence. *NeuroReport,* 12 (2), 1-7.

DICKINS, T. E., & DICKINS, D. W. (2001). Symbols, stimulus equivalence and the origins of language. *Behavior and Philosophy,* 29, 221-244.

DYMOND, S., & BARNES, D. (1995). A transformation of self-discrimination functions in accordance with the derived stimulus relations of Sameness, More than, and Less than. *Journal of the Experimental Analysis of Behavior,* 64, 163-184.

DYMOND, S., & BARNES, D. (1996). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness and opposition. *The Psychological Record,* 46, 271-300.

HARZEM, P. (1984). Experimental analysis of individual differences and personality. *Journal of the Experimental Analysis of Behavior,* 42, 385-395.

HAYES, S. C., BARNES-HOLMES, D., & ROCHE, B. T. (2001). *Relational Frame Theory: A post-Skinnerian account of human language and cognition.* New York: Plenum.

HAYES, S. C., & BISSET, R. T. (1998). Derived stimulus relations produce mediated and episodic priming. *The Psychological Record,* 48, 617-630.

LIPKENS, G., HAYES, S. C., & HAYES, L. J. (1993). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology,* 56, 201-239.

MCHUGH, L., BARNES-HOLMES, Y., & BARNES-HOLMES, D. (2004). Developmental trends in perspective-taking. *The Psychological Record,* 54, 115-144.

MCILVANE, W. J., DIFIORE, A., & WILKINSON, K. (2003, May). *Electrophysiological analysis of stimulus equivalence.* Paper presented at the convention of the Association for Behavior Analysis, San Francisco, CA.

O'HORA, D., BARNES-HOLMES, D., ROCHE, B., & SMEETS, P. (2004). Derived relational networks and control by novel instructions: A possible model of generative verbal responding. *The Psychological Record,* 54, 437-460.

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.

PELÁEZ, M., GEWIRTZ, J. L., SANCHEZ, A., & MAHABIR, N. M. (2000). Exploring stimulus equivalence formation in infants. *Behavior Development Bulletin,* 9, 20–25.

PENA, E., BEDORE, L. M., & RAPPAZZO, C. (2003). Comparison of Spanish, English, and bilingual children's performance across semantic tasks. *Language, Speech and Hearing Services in Schools*, 34 (1), 5-16.
ROBERTSON, I. H., & MURRE, J. M. J. (1999). Rehabilitation of brain damage: Brain plasticity and principles of guided recovery. *Psychological Bulletin*, 125, 544-575.
ROCHE, B., & BARNES, D. (1995). Sexual behaviour as an act in context. *Irish Journal of Psychology*, 16, 409-417.
ROCHE, B., & BARNES, D. (1996a). Arbitrarily applicable relational responding and sexual categorization: A critical test of the derived difference relation. *The Psychological Record*, 46, 489-507.
ROCHE, B., & BARNES, D. (1996b). Sexual fetishism: A modern experimental analogue. *The Clinical Behavior Analyst*, 3, 2-4.
ROCHE, B., & BARNES, D. (1997). A transformation of respondently conditioned stimulus function in accordance with arbitrarily applicable relations. *Journal of the Experimental Analysis of Behavior*, 67, 275-300.
ROCHE, B., BARNES-HOLMES, D., SMEETS, P. M., BARNES-HOLMES, Y., & MCGEADY, S. (2000). Contextual control over the derived transformation of discriminative and sexual arousal functions. *The Psychological Record*, 50, 267-291.
SCHUSTERMAN, R. J., & KASTAK, D. (1993). A California sea lion (*Zalophus californianus*) is capable of forming equivalence relations. *The Psychological Record*, 43, 823-839.
SIDMAN, M. (1971). Reading and auditory-visual equivalences. *Journal of Speech and Hearing Research*, 14, 5-13.
SIDMAN, M. (1994). *Equivalence relations and behavior: A research story.* Boston: Author's Cooperative.
SIDMAN, M., & TAILBY, W. (1982). Conditional discrimination versus matching to sample: An expansion of the testing paradigm. *Journal of the Experimental Analysis of Behavior*, 37, 5-22.
STAUNTON, C., BARNES-HOLMES, D., WHELAN, R., & BARNES-HOLMES, Y. (2003, May). *Priming and Event Related Potentials (ERPs) as Measures of Derived Relational Responding.* Paper presented at the convention of the Association for Behavior Analysis, San Francisco, CA.
STEELE, D., & HAYES, S. C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior*, 56, 519-555.
STEWART, I., & BARNES-HOLMES, D. (2001). Understanding metaphor: A relational frame perspective. *The Behavior Analyst*, 24, 191-199.
STEWART, I., BARNES-HOLMES, D., ROCHE, B., & SMEETS, P. (2001). Generating derived relational networks via the abstraction of common physical properties: A possible model of analogical reasoning. *The Psychological Record*, 51, 381-408.
TAUB, G. E. (2001). A confirmatory analysis of the Wechsler Adult Intelligence Scale-Third Edition: Is the verbal/performance discrepancy justified? *Practical Assessment, Research & Evaluation*, 7(22). Retrieved January 13, 2005 from http://PAREon line.net/getvn.asp?v=7&n=22.
WECHSLER, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition.* San Antonio, TX: The Psychological Corporation.
WECHSLER, D. (1997). *Wechsler Adult Intelligence Scale – Third Edition.* San Antonio, TX: The Psychological Corporation.
WULFERT, E., GREENWAY, D., FARKAS, P., HAYES, S. C., & DOUGHER, M. (1994). Correlation between self-reported rigidity and rule-governed insensitivity to contingencies. *Journal of Applied Behavior Analysis*, 27, 659-671.2.

Appendix 1

Name: ________________ Date: ______________
Subject # ____ Evaluators' Name ______________________________

All subjects were asked the following 4 questions before administering the Spanish Language Comprehension Assessment by an experimenter:

1. Do you consider yourself monolingual or bilingual?
2. Which language(s) other than English can you speak and / or comprehend?
3. Please indicate which you consider to be your second language?
4. How fluent are you in that second language?

Not at all	Somewhat	Well	Really Well	Excellent
1	2	3	4	5

Spanish Language Comprehension Assessment:

Please indicate **(O)** to questions that are read by the subject and **(R)** to questions read to the subjects by the experimenter and use the following scale to measure expressive and receptive language of Spanish.

Not at all	Somewhat	Well	Really Well	Excellent
1	2	3	4	5

	O/R	***Expressive***	***Receptive***
1.Que estas estudiando o que planeas estudiar en el futuro? Porqué?			
2.Que no te gusta de la ciudad de Miami? Porqué?			
3.Tienes algun deporte o actividad favorito? Hablame sobre eso			
4.Cual es tu comida favorita? Porque?			
5.Cual es tu música favorita? Porque te gusta ese tipo de música?			
6.Tienes hermanos y hermanas? Dime sus edades?			
7.Cual fue la ultima película que viste? Hablame un poco de ella			
8.Cual es la razón mas importante en tu vida? Porqué?			
9. Si pudieras tener un trabajo, cual seria?			
10. Te gusta viajar? Porque?			

Total Score

Monolingual (English) _____ Bilingual (Spanish/English) _____