Training Relational Language to Improve Reading Comprehension

Kendra Brooks Newsome,
Fit Learning, Nevada, USA

Kimberly Nix Berens,
Fit Learning, New York, USA

Patrick M. Ghezzi,
University of Nevada, Reno

Tiffany Aninao, and William D. Newsome
Fit Learning, Nevada, USA

The purpose of this study was to assess the impact of strengthening participants' relational abilities on measures of comprehension. We evaluated the effects of two phases of multiple exemplar training on reading comprehension: hierarchical relational training, and hierarchical framing under the contextual cues of "same" and "different". Five participants with poor reading comprehension, but strong reading abilities, were trained in four components of hierarchical relational training. This phase was designed to increase the frequency and flexibility with which participants related properties to stimuli, and to assess if training in hierarchical relations improved participant's ability to derive hierarchical relations between stimuli. The next phase of training required participants to discriminate relevant relations under contextual cues of "same" and "different". The effects of training were examined on measures of reading comprehension and written expression. All but one participant showed improvements on one or more academic measures indicating a functional relationship between the language skills acquired and comprehension and written expression. These data have important implications for guiding research endeavors in behavior analysis and for informing practices in education. Limitations and directions for future research are discussed.

Key words: Reading comprehension, Relational Frame Theory, education, language, precision teaching

Reading comprehension is of paramount importance to academic success and is pivotal in the development of other academic skills (National Reading Panel, 2000). Reading comprehension means a student can behave effectively with respect to what is read, and this is the primary means through which new information is acquired. Deficits in reading comprehension, therefore, can impact a student's ability to access an appropriate education (Catts & Kamhi, 2005).

Historically, it was thought that as students acquired reading abilities, comprehension of reading material would emerge without explicit instruction. This theory led to a passive approach to comprehension. Durkin (1978; 1979) conducted a detailed analysis

Correspondence concerning this article should be addressed to: Kendra Brooks Newsome 3953 S. McCarran Blvd Reno, NV 89502. E-mail at drkendra@fitlearners.com

of comprehension practices of 40 intermediate grade teachers. Results indicated that less than 1 of nearly 300 instructional hours contained explicit instruction in reading comprehension. Students spent the majority of time answering questions about a passage, a means of assessing comprehension, not training it (cf. Pearson & Gallagher, 1983). Most time during comprehension activities was allocated to answering comprehension questions; only 2.4% of time was spent instructing comprehension strategies. More recently, it has become clear that reading with understanding is the result of a convalescence of variables.

Several factors have been linked to reading comprehension abilities (e.g., Catts, Hogan, Adlof, 2005; NRP, 2000; Snow, Burns, & Griffin, 1999). These include, (1) reading fluency and phonemic awareness (e.g., decoding), (2) vocabulary, (3) the ability to use strategies, and (4) general language abilities. Thus, research and efforts to remediate comprehension deficits have fallen primarily into three categories: 1) methods for improving phonemic awareness and reading fluency, 2) vocabulary instruction and pre-teaching of concepts, and 3) instruction on comprehension strategies.

Poor reading fluency and poor phonemic awareness is often implicated in poor reading comprehension (National Reading Panel, 2000). Following improvements in these aspects of the reading process, comprehension frequently improves. These findings have informed the development of best practices for remediating deficits in decoding and improving phonemic awareness (e.g., Joseph, 2008; NRP, 2000; Pikulski & Chard, 2005; Rasinski & Hoffman, 2003).

Despite the breakthroughs in reading instruction, however, about two thirds of fourth graders still struggle with comprehension of grade-level material (National Center for Education Statistics, 2004). This statistic indicates that there are aspects of reading comprehension that are independent of decoding and reading fluency (Williams, 2005). Gough & Tunmer (1986) posit that there are two aspects to the reading process: decoding and language comprehension. Over the past decade, much attention has been given to the study of decoding with little focus on the latter.

In 2000, the National Reading Panel (NRP) emphasized two additional factors of reading comprehension: vocabulary and the application of strategies to improve comprehension. Efforts for improving comprehension have since been expanded to include explicit teaching of vocabulary, and direct teaching of comprehension strategies. Pearson & Gallagher (1983) in their review of comprehension strategies noted that vocabulary training often included pre-teaching of new words, while strategy training included teaching students how to construct webs and hierarchies of information to form links between known information and new concepts. Other studies focused on teaching students to identify text structure (Williams, 2005). Here, students are taught how to identify themes and common patterns like compare/contrast. The results from text structure studies indicate that this can be an effective means of improving reading comprehension for some students (Gersten, Fuchs, Williams, & Baker, 2001; Williams, 2005). It has been noted, however, that strategies can be difficult to teach in a way that produces lasting impact, and preparing teachers to implement strategy training can be challenging (National Reading Panel, 2000; Williams, 2002; Williams, 2005).

The shift in our understanding of comprehension is a pivotal one for education. Illuminating factors related to comprehension has facilitated the development of more effective technologies. Early reading instruction emphasizes decoding and reading fluency; pre-reading activities to teach new vocabulary are incorporated, and children are taught how to apply strategies while reading. Nonetheless, there are students who fail to improve in comprehension despite these efforts (e.g., Pearson & Gallagher, 1983; Nicholson & Thompson, 1999; Williams, 2005). Williams (2005) in a program evaluation focused on teaching text-structure strategies, noted that there were some students who failed to benefit from the training. Williams states, " regardless of which criteria we used, we found that the students who had not performed as well [with strategy training] as the others had lower listening and lower reading comprehension scores," (pg. 14).

The students who fail to benefit from strategy training indicate a need for further consideration of the factors underlying comprehension. The success of strategy training depends on other important components. Williams (2005) states that specific structures in text are not limited to text only; rather, it is put forth that these structures reflect universal cognitive processes commonly exhibited in the thinking of young children. She asserts, "by the time children enter school, they tell stories, compare and contrast objects, order events in a temporal sequence, and attribute causality," (pg. 7) (cf., Carey, 1990). Children who fail to understand what they read despite strong reading abilities and strategy training do not readily demonstrate these cognitive processes, or the relative strength of these processes is insufficient for reading comprehension. Approximately 10% of primary-school children struggle with reading comprehension despite strong abilities to read and decode text (Nation & Snowling, 1997). This phenomenon has been labeled *specific reading comprehension difficulties* and relates to the second component of the reading process: language comprehension.

To engage meaningfully in the reading process a reader predicts, summarizes, paraphrases, infers, integrates text, and reacts personally to what is read (Leslie & Caldwell, 2006). These behaviors entail sophisticated language, and these language abilities are tantamount to the cognitive processes described by Williams (2005). Children who have strong word recognition and reading skills, but struggle with comprehension, present with weak verbal repertoires insufficient for predicting, inferring, integrating and summarizing. Bishop & Adams (1990) assessed language and literacy skills in 83 children and found that those with language impairments between the ages of 4 and 6 years old were more likely to have later comprehension deficits. Similarly, Hart and Risley (1995) evaluated 42 American families over two years and found that future language abilities were substantially influenced by the frequency of language-based experiences children were exposed to in the first three years of life. They found that particular types of interactions were so critical for future success that children without that early experience showed academic deficits years later.

Few studies to date have focused on ameliorating language deficits as a means of improving reading comprehension. In one such study, improvements in comprehension resulted when students were taught to relate what they already knew to events in the story, predict aspects of the story, and then relate prior knowledge and predictions (Hansen, 1981). Similarly, Clarke et al., (2010) assessed three different methods to improve reading comprehension: text-comprehension training (TC), oral-language (OL) training, and the two methods combined (COM). They found that oral language training aimed at improving participants' vocabulary and figurative language skills resulted in superior outcomes relative to the other efforts. While all intervention groups showed improvements relative to no treatment, the OL group showed the greatest comprehension gains at follow up. The OL and COM groups also showed significant growth in expressive vocabulary. The authors suggest that the improvements in expressive vocabulary were a mediator of the improved comprehension abilities for these groups. Taken together, these findings suggest that language is a viable unit of analysis in the search for effective technologies for improving comprehension.

From the view of contextual behavior science, language is behavior that is social in nature and evolves through interactions in and with the environment. On this view, children with language delays lack a sufficient language-based history or have failed to benefit from their histories, and the means for reducing the risks associated with language delays center on the remediation of core language processes. Thus, a comprehensive theory of language can offer the precision, scope and depth required to tackle the analysis from which effective technologies can emerge. Such an approach can have profound impacts on resulting comprehension research. In light of this, Relational Frame Theory (RFT) stands out for the behavioral researcher interested in language for comprehension.

Relational Frame Theory (Hayes & Barnes, 1997; Hayes, Barnes- Holmes et al., 2001; Hayes & Berens, 2004; Hayes, Fox, et al., 2001; Dymond & Roche, 2013) is a theory of language and cognition rooted in behavior science and provides a framework for understanding language for comprehension. Much evidence supports that arbitrarily applicable derived relational responding is operant behavior and that relational responding is the core process underlying language and cognition (e.g., Dymond & Barnes, 1995; Healy, Barnes-Holmes, & Smeets, 2000; Lipkens, Hayes & Hayes, 1991; Steele & Hayes, 1991; Wulfert & Hayes, 1988). The ability to derive and arbitrarily apply relational frames is learned through a history of exposure to multiple exemplars (see Hayes et al., 2001). This notion is important because it means that when particular ways of relating are established, the ability to relate events in such a way are not dependent on the events or their physical properties; rather, relational responses are occasioned by the context and can come to bear on any event (see Hayes et al., 2001; Törneke, 2010, and Dymond & Roche, 2013).

RFT asserts that relating is a generalized operant, and that there are some basic relational operants (e.g., similarity, distinction, hierarchy, comparison). In other words, it defines some specific languaging skills that can be trained in isolation and whose relationship to/impact on comprehension can be systematically measured. Other frameworks (see lit review above) either lack the identification of specific response classes to train or are too lacking in scope to accomplish this analysis. For example, "relating to prior knowledge" and "predicting" are skills identified by Clark, et al., but this prescriptive approach does not match the scope offered by "relating as a generalized operant" nor the precision of specific, fundamental relational operants identified by RFT.

There is ample evidence that relational responding is relevant to intellectual ability and educational attainment (e.g., Smith, Smith, Taylor & Hobby, 2005; Cassidy, Roche, & Hayes, 2011; Cassidy, Roche & O'Hora, 2010; Leader & Barnes-Holmes, 2001). Studies have shown that improving targeted relational operants can lead to improvements in IQ scores and other intelligence measures (e.g., Cassidy, Roche, & Hayes, 2011; Cassidy, Roche, & O'Hora, 2010). Moreover, Ninness and colleagues have demonstrated the success of a RFT framework to teach complex mathematical skills in algebra and trigonometry (Ninness et al., 2005; Ninness et al., 2005; Ninness et al., 2006; Ninness et al., 2009). RFT has also informed teaching practices for children with autism, a population defined by deficits in language. Persicke, Tarbox, Ranick & St Clair (2012) taught children with autism to solve metaphors using multiple exemplar training. Participants were able to generalize the skill to solve novel metaphors. Taken together, these findings suggest that RFT might prove fruitful for informing practices in education towards ameliorating comprehension deficits.

There are various relational operants and relational operants can interact. While not an exhaustive list, some proposed relational frames include coordination, opposition, distinction, comparison, hierarchical relations, spatial relations, conditionality and causality, and deictic relations (see Hayes et al., 2001 for detailed discussion).

Some relational operants are more rudimentary than others (for a more detailed discussion, see Luciano et al., 2008; Dymond & Roche, 2013). Three relational frames are described here and informed the curriculum package employed: frames of coordination, frames of distinction and hierarchical frames.

A fundamental way of relating is in terms of sameness, or similarity. This is termed frames of coordination in RFT. Understanding text requires that a reader respond to words in a symbolic way. Thus, frames of coordination give us linguistic reference and coordinate framing is the path by which vocabulary expands and organizes. The word dog shares no physical features with an actual dog, yet through a particular history, the printed word dog and a real dog will occasion many of the same responses given an appropriate context.

Distinction is a relation that involves "responding to one event in terms of its differences from another, typically along some specified dimension" (Hayes, Barnes-Holmes, et al., 2001, p. 36). As with frames of coordination, frames of distinction can be based on physical properties of the stimuli or may be based on arbitrary properties. With respect to comprehension, the ability to relate things in terms of distinction is critical. Coordinated and distinct relations between words and objects lie at the foundation of reading comprehension and are an important foundation to the ability to integrate, compare and contrast, make predictions, understand metaphor and engage in inferences.

Hierarchical framing is how we organize and classify the world around us. Hierarchical frames, it has been suggested, are "the ideal playground for building relational flexibility" (Luciano et al., 2008). Luciano et al., (2008) state, "...it seems clear that [hierarchical frames] may be a very useful frame in terms of more fully elaborating the flexibility and contextual sensitivity of an individual's relational repertoire. Furthermore, becoming skilled in hierarchical frames should benefit other frames" (p.23). Stimulus events can be members of many hierarchical frames, and include "part-whole" or "attribute of" relations. For example, "dogs are animals" or "dogs have fur" are both examples of hierarchical frames (Hayes et al. 2001).

Hierarchical relations have not been the subject of much research; nonetheless, the metaphor training conducted by Persicke et al., (2012) entailed an element of hierarchical framing. Children were taught to state properties of the stimuli in the metaphor, therefore relating properties of each stimulus in the metaphor (cf. Dymond & Roche, 2013). This is an instance of hierarchical framing and illuminates how critical the hierarchical foundation is for cognition. If the children could not relate properties to each stimulus in the metaphor, we might expect the training to be less successful.

The purpose of this study was to assess the impact of strengthening participants' relational abilities on measures of comprehension. We evaluated the effects of two phases of training on reading comprehension: hierarchical relational training, and hierarchical framing under the contextual cues of "same" and "different". First, we increased the frequency and variability with which participants engaged in hierarchical relations along some specified dimension (described below). Participants were taught to label properties of everyday objects. These properties included the category of the object, the parts, sensory elements and their functions. We were concerned with whether this training would improve participant's ability to derive coordinated or distinct relations of hierarchy between stimuli.

Secondly, participants were taught to compare two items along some hierarchical dimension under cues for coordination (i.e., how the items are the same) or distinction (i.e., how the items differ from one another). The intervention was evaluated in terms of its effect on untrained relational tasks and conventional measures of reading comprehension and written expression.

Method

Design and Analysis Overview

This study employed an A/B/C design with a multiple probe element and a constant series control. The A phase constituted baseline. Phase B consisted of training on four specified dimensions of hierarchy. During both the A and B phases, weekly probes were administered to test for derived relations of hierarchy under cues of "same" and "different". The multiple-probe element was instated to assess the effects of frequency and flexibility building in hierarchical relations on participants' ability to derive hierarchical frames under specific contextual cues.

Following the B phase, isolated coordination and distinction sets were trained for all participants, constituting the C phase of the study. One participant served as a constant-series control. The purpose of the constant-series-control was to evaluate the extent to which effects observed were due to maturation, experiential events such as interventions at school, or other language-based experiences to which participants may be exposed. Once the design logic was satisfied for this element, this individual participated in the same training as the experimental participants. Pre and post comprehension and written expression measures were obtained to evaluate the effects of the independent variables on comprehension.

Materials and Sessions

Various materials were used to determine participant eligibility and were administered as pre and post assessments. AIMSweb™ passages were used to assess and determine fluent reading levels. AIMSweb™ is a standardized universal screening tool often used for progress monitoring in school districts in the United States. Reading assessments are conducted at specified intervals (e.g., weekly, monthly, quarterly) using grade-specific passages and yield a word per minute score (WPM). The WPM scores can be used to determine a student's standing relative to the national norms in the United States. This tool was selected because administration is easy and brief, and AIMSweb™ assessments have been shown to be sensitive to improvement (http://www.aimsweb.com/measures-2/reading-cbm/). AIMSweb™ story starters were used in pre/post expressive writing measures. Story starters are writing prompts that consist of a partial sentence and indicate the writing start point for a student (e.g., A rocket ship landed on the moon and...).

The Qualitative Reading Inventory 4 (QRI 4), an evaluative tool for assessing comprehension of reading material, was administered to measure reading comprehension. The QRI 4 is a tool that can facilitate the documentation of learners' growth when administered prior to beginning and at the conclusion of an intervention (Leslie & Caldwell, 2006).

Noun-picture cards from the categories, 1) animals, 2) toys, 3) food, 4) vehicles, 5) furniture, and 6) household items were used during both phases of hierarchical relation training. Digital timers were used to conduct timings and to track participants' cumulative time in training. Sessions were conducted four days per week and took place in the participant's homes or in a private room at a learning center in Reno, NV. Session duration ranged from 15-45 minutes.

Participants

Reading and comprehension assessments were administered to assess eligibility. Children who read accurately and fluently, but had poor reading comprehension were included in the study. The reading criteria required participants to read with 90% or better accuracy and 80 words per minute (wpm) on a 1st grade or higher passage. Thus, the assessment level matched the child's functional reading level, not necessarily grade level. If these criteria were met, the corresponding level of the Qualitative Reading Inventory 4 was administered. Comprehension was measured in two ways:

retelling and answering comprehension questions. Participants who had retell scores at 50% accuracy or below were read the comprehension questions from the passage. Participants who answered comprehension questions with 69% accuracy or below met the inclusion criteria.

Five children between ages of 9 and 12 were selected for participation. Irene (age 9), Thomas (age 9) and Tina (age 12) were of typical development, but all struggled with comprehension according to parent and teacher reports. Thomas had an Individual Education Plan (IEP) at school and received pullout services for comprehension on a daily basis. Irene had a history of seizures that were reported to have affected her language skills. Irene's teacher reported that Irene struggled with reading comprehension, particularly inference, and had difficulty with expressive writing tasks. Tina had a fetal addiction to amphetamines and methamphetamines. Tina's mother and teacher reported that Tina struggled with comprehension tasks, content courses, and expressive writing. Mary (age 9) had a diagnosis of autism and received services in an early intervention program. Mary required significant supports in the classroom due to the discrepancy between her reading level and comprehension abilities. She also had an IEP. Ivan (age 12) had a diagnosis of Down's syndrome. Ivan also had an IEP, received speech services, and had a tutor to help with homework. Despite these supports, Ivan struggled with comprehension tasks, content courses and expressive writing.

Independent Variables

Hierarchical relational training. Hierarchical relational training (HRT) constituted the B phase of the study. The HRT included four "part-whole" and "attribute of" elements, and was tested in a series of pilot studies to evaluate its treatment utility for comprehension. In the first component, *Categories*, participants were trained to label an item's general category (e.g., toy, vehicle, food). The second component trained participants to label *Concrete Features,* or parts of objects that could be readily seen (e.g., a dog has eyes; a flower has petals). *Expanded Features*, the third component, taught participants to identify places an item might be found (e.g., you can find a sandwich in a lunchbox), times when an item might be used (e.g., you eat a sandwich when you are hungry), sensory functions (e.g., a sandwich can be soft, crunchy, and delicious) and other items that might be found with or near the target item (e.g., a sandwich might be found with a pickle or chips). The final component taught participants to describe *Common* and *Creative Functions* of stimuli. *Common Functions* were defined as an object's primary use (e.g., a pencil is used to write). *Creative Functions* were defined as other ways an object can be used (e.g., a pencil can be used to scratch an itch, or squish a spider).

Hierarchical framing under cues of "same" and "different". Training in hierarchical frames under the cues "same" and "different (HCSD) constituted the C phase of the study. The C phase was also informed through a series of additional pilot implementations demonstrating a relationship between these relational abilities and improvements in reading and listening comprehension. There were 9 relational tasks (5 sets presented under cues of "same", and 4 sets presented under cues of "different") that varied along several dimensions: experimental history with stimuli, stimulus modality, and relational complexity (See Table 1 for a summary of all HCSD tasks). For similarity, Set 1 pairs consisted of two stimuli from the same overarching category (e.g., two toys, two animals, two foods). Stimuli from Set 1 pairs were also presented in the context of category identification, concrete features, expanded features and functions. Stimuli from sets other than Set 1 for both similarity and distinction were not used during the B phase of training. Set 2 stimuli consisted of pairs from the same overarching category. For Set 3, stimulus pairs were drawn from different overarching categories (e.g., one toy and one food; one animal and one toy).

Table 1. *Stimulus Set Characteristics for Similarity and Distinction Sets*

Probe Set	Experimental History	Stimulus Characteristics	Stimulus Modality	Number of Stimuli in Each Trial
Set 1 – Similarity	Presented in B phase	Same Category	Visual	2
Set 2 – Similarity	Not presented during B phase	Same Category	Visual	2
Set 3 – Similarity	Not presented during B phase	Different Categories	Visual	2
Set 4 – Similarity	Not presented during B phase	Different Categories	Visual	3
Set 5 – Similarity	Not presented during B phase	Activities from Same or Different Categories	Auditory	2
Set 1 – Distinction	Presented in B phase	Different Categories	Visual	2
Set 2 – Distinction	Not presented during B phase	Different Categories	Visual	2
Set 3 – Distinction	Not presented during B phase	Same Categories	Visual	2
Set 4 – Distinction	Not presented during B phase	Activities from Same or Different Categories	Auditory	2

For Set 4, three stimuli were presented during each trial, each drawn from different overarching categories (e.g., one toy, one animal, one furniture). During Set 5 probes, participants were asked to describe how activities are similar to other activities. Distinction sets were modeled after similarity sets with the exception of the set containing 3 stimuli. (See Appendix A-D for a list of all HCSD stimuli).

Of the 9 tasks that served as probes throughout HRT, either 1 or 3 sets were selected for training, serving as the second independent variable and the C phase of the study. Thomas and Tina received training on one set, Coordination Set 1. Irene, Mary and Ivan received training on 3 of the 9 sets, Coordination Sets 1 and 3, and Distinction Set 1. The remaining tasks were probed weekly to maintain the multi-probe element of the design.

Dependent Variables

Performance on untrained HCSD sets. The multiple-probe element served to evaluate relating as a generalized operant. Weekly probes were administered to evaluate whether hierarchical relational training improved participants' accuracy and frequency of derived hierarchical relations under cues of "same" and "different". Improvements on untrained sets might predict that participants could generalize hierarchical framing along specific dimensions in the context of reading activities. As mentioned previously, all participants received training on some sets of HCSD tasks.

Weekly probes continued on sets that did not receive training.

QRI 4. The Qualitative Reading Inventory 4 was administered before and subsequent to training to detect improvements in reading comprehension. Different passages and questions were administered for pre and post. Comprehension was assessed following an oral reading task in two ways: retelling and answering comprehension questions. Retelling was measured by asking participants to verbally recreate the story as if s/he were telling the story to someone who had not heard it before. A retelling score was generated by calculating the frequency of nouns, verbs, adjectives and adverbs emitted that were related to the read passage in a 30-sec timing. Following retelling, participants were read a sequence of implicit and explicit comprehension questions. Participants provided a vocal response to the question and were not allowed to refer to the passage once s/he completed reading. A percent correct score was obtained for the comprehension questions.

Expressive writing. A pre/post writing measure was administered to evaluate whether the HRT and HCSD had an effect on participant's written output. A story starter was provided as a writing prompt, and participants were asked to continue writing the story for a period of 3 min. Written products were evaluated with respect to the frequency of words written per minute.

Procedures

Precision Teaching. Our study employed a Precision Teaching paradigm for instruction. Precision Teaching is an instructional paradigm that employs measures of frequency and rate of change, or celeration, as its basic datum (Lindsley, 1992; Maloney, Blearley & Preece, 1998; Kubina & Yurich, 2012). Frequencies obtained are plotted on Standard Celeration charts, and trends yield measures of celeration. Celeration is a measure of learning that reflects changes in the rate of behavior over time. It is expressed numerically as a multiplication (i.e., acceleration X) or division (i.e., deceleration /) in frequency per unit of time (Binder, 1996).

Frequency aims. Frequency ranges, or aims, for the number of responses per minute were established for all skills and were identified in two manners: functional aim analysis for skills in hierarchical relation training, and normative sampling for the probe sets. Functional aims were identified through an analysis of clinical and pilot data. Namely, data collected from several clients at a learning center were evaluated to identify frequency ranges on the hierarchical relation targets that correlated with increases in frequencies on coordination and distinction probes. These ranges (see Table 2) were used in the current study.

Frequency aims for each of the probe sets were obtained through a normative sampling procedure. Samples were collected from 10 individuals employed at a learning center, who had extensive experience administering the probes, and research assistants who helped develop the curriculum. Multiple timings were run on each set with each individual to determine frequency ranges. Frequency aims were set at the low end of the frequency range (see Table 2 for all aims).

Baseline and skill introduction. In baseline, the A phase, the investigator read scripted instructions about the task and the specific feature for training to participants during each session. During the A phase, three, 15-sec timings were conducted on category identification, the first skill to enter into training. No programmed feedback was given on performance during baseline. Reinforcement was not delivered for improvements in performance, but was provided for appropriate behavior pertinent to the instructional situation. Namely, praise and points were provided for appropriate sitting (e.g., forward orientation, feet on ground, buttocks in chair, hands at sides or resting on table top, and chair legs on the floor), positive statements (e.g., "I will try my best," "This is fun," "I want to try again"). Praise and points were delivered on an average of every three appropriate behaviors during the A phase.

Table 2. *Frequency Ranges for Training Targets and Probes*

TARGET	Frequency Range per Min
Category Identification	60–65
All Discrete Trial Targets	16–20
Concrete Features – FO	20–30
Expanded Features – FO	20–30
Common and Creative Functions – FO	18–25
Similarity Set 1	25–30
Similarity Set 2	25–30
Similarity Set 3	20–25
Similarity Set 4	15–20
Similarity Set 5	20 –25
Distinction Set 1	20–25
Distinction Set 2	20–25
Distinction Set 3	20–25
Distinction Set 4	20–25

Baseline sessions for categories continued until a stability criterion was met (X1.2 celeration or less), and until a minimum of two probe points were collected for the other skills. Phase A was also terminated based on other aspects of the data. For example, an increasing trend in errors may have warranted moving categories into the training phase. All remaining skills were probed once weekly until they either entered the training phase, or the study concluded. The data obtained during weekly probes served as the baseline data for these skills when they entered into training.

HRT targets were introduced sequentially when the frequency goal was met for the preceding skill. Training on each component continued until the frequency aim was reached and errors did not exceed 2/min for the skill in training for two consecutive sessions.

HRT. Hierarchical relation training, the B phase, consisted of the four components described above in the *Independent Variables* section. The *Categories* component was the first trained. While *Categories* was in training, all other HRT components were probed weekly, along with the similarity and distinction sets. Fifteen noun-picture cards were randomly selected for category identification for each session. The 15 cards were composed of three cards from each group of animals, toys, food, vehicles and furniture. Participants were asked to vocally identify the category of the item on the card during three, 15-sec timings. The investigator held and flipped the cards. The number of categories correctly labeled was counted after the 15-s timing to record the frequency. The cards were shuffled for randomization before the next timing. Following *Categories* training, participants were trained to discriminate

Concrete Features, Expanded Features, and *Common and Creative Functions.* These components were introduced sequentially in the order listed above and the training procedures were identical across these three components (See Appendix E for a sample of training data). Training included, 1) concept instruction to introduce the type of description/feature to the participant 2) scripted discrimination training using multiple exemplars and 3) free-operant timings for the participant to apply the descriptive concepts to stimuli. Three noun-picture cards were randomly selected for each session.

First, participants received instruction on the feature identified for training (see Appendix F for a sample of instructions). Participants were presented with a definition of the target feature and examples and non-examples of the feature. Following concept instruction, scripted questions with both examples and non-examples of target features were read aloud to the participant in two, 1-min timings (see Appendix G for a sample of scripted questions). Participants responded with a yes or no response, and were required to answer in complete sentences (e.g., yes, you can find it with a baseball; no, you cannot eat it for breakfast). The final element had participants practice applying the descriptive feature to stimuli in free-operant timings. Two, 1-min free-operant timings were run in each session. One of the three randomly selected stimuli was placed on the table. Participants were asked to respond by providing target descriptive features in training. Each of the three cards selected for that session were presented for 20 s of the 1-min timing.

Hierarchical framing under cues of "same" and "different". The C phase entailed training hierarchical relating under contextual cues of same or different (HCSD) on particular sets of stimuli. Four pairs of stimuli from each set were randomly selected from a pre-determined group for each training session. The training sequence for the C phase was identical to the B phase sequence. Prior to training, a sample stimulus pair was presented with brief instructions and a model for how to perform the task, (e.g., I want you to tell me how these are the same, or I want you to tell me how these are different). A sample instructional script is included in Appendix H. Four stimulus pairs were randomly selected for each training session. Scripted examples and non-examples were presented inside of discrete-trial timings (see Appendix G) and two free-operant timings followed. In free-operant timings, one of the stimulus pairs was presented and participants were asked to label either similarities or differences between the two stimuli. Each of the four pairs of stimuli was presented for 15 s of the 1-min timing.

The effects obtained on remaining probe sets determined the number of sets trained for each participant. Thus, if following training on one HCSD set, a visual effect (e.g., ≥ X1.4 celeration, deceleration in errors) was obtained on all remaining probe sets for a participant, training on HCSD sets ceased, post measures were administered and the study was concluded. In contrast, if an effect was not observed on remaining probe sets as a function of training on other sets, a new set was introduced when the mastery criteria were met for the training set. Samples of individual data sets obtained on similarity and distinction probes are provided in Appendix I and Appendix J.

Goal setting and reinforcement. Throughout training, goal setting and reinforcement was provided for increases in the frequency and accuracy of responding, and continued until the participant met the frequency aim for each skill (See Table 2 for frequency aims). Frequency goals were set for each session by adding one to the median response for the last three sessions. Repeated responses in the same timing were counted as errors. Error correction immediately followed timings and included a representation of the stimulus and modeling of a correct response;

the participant was then required to engage in the correct response. Goals for decreases in errors were established by subtracting one from the total number of errors made during the previous timing. A goal was stated before each timing. Following a timing, immediate feedback was provided with respect to the frequency of correct and incorrect responses. Points were provided for goal attainment and were exchangeable for a variety of back up items identified as preferred in an interview with each participant.

Data analysis. Data were analyzed using two versions of the Standard Celeration Chart (SCC): the daily per minute and the weekly per minute chart (Pennypacker, Guiterrez & Lindsley, 1972). During phases A, B and C, data were analyzed using the daily per minute SCC. Weekly SCCs were employed for the HCSD probes. All data were analyzed with respect to trend, level, variability and celeration. Moreover, retention checks to assess a participant's ability to perform the task after a period of no practice were conducted for all HRT skills. Retention checks were scheduled after the passing criteria were met. Retention checks were administered one week following training and subsequent checks occurred at various intervals after the first week.

Celeration collections were employed to quickly analyze performance of all participants on HCSD sets. Celeration collections involve superimposing celeration lines on top of one another and permit a quick visual analysis of trend and level in each participant's performance for each HCSD set, as well as an analysis across participants. The solid trend lines are the celerations obtained across correct response frequencies while the dashed trends are celerations across error frequencies. The horizontal line across the top of the chart indicates the low end of the frequency aim for HCSD sets.

Agreement and procedural integrity data were collected for 30% of sessions. Inter-observer agreement was calculated by determining the number of agreements and disagreements for each block of timings on all targets and probes. The percentage of agreement for the sessions scored was 96.5%. For procedural integrity, a percentage score was calculated for the number of correct procedural implementations out of the total possible implementations for each phase of the study. Procedural integrity scores ranged from 89% - 100% with an overall score yielding 99% accuracy.

Results

All participants met the passing criteria for each component in HRT. Training data obtained during the B and C phases for each participant has been omitted, however, a sample of Irene's training data for *expanded features* is included in Appendix E so the reader can glean the training process. Additional training data can be obtained by contacting the first author.

All participants received training on at least one HCSD set. Thomas and Tina received training on Similarity Set 1, while Irene, Ivan and Mary received training on Similarity Sets 1 and 3, and Distinction Set 1. The remaining HCSD sets remained as weekly probes throughout the duration of the study. Celeration collections obtained from HCSD tasks are included here, as well as the individual pre and post comprehension measures.

Celeration Collections

Celeration collections were generated by extracting celeration lines from the individual HCSD charts and yield a picture of the differences in changes of behavior over time on these tasks. A sample of individual data obtained on all HCSD tasks is provided in Appendix Appendix J and I. For Irene, Tina and Thomas, base levels of responding on all HCSD sets were higher than Mary and Ivan. For Irene, Tina and Thomas, performance is differentiated with respect to correct and errors, and these participants achieved higher frequencies overall.

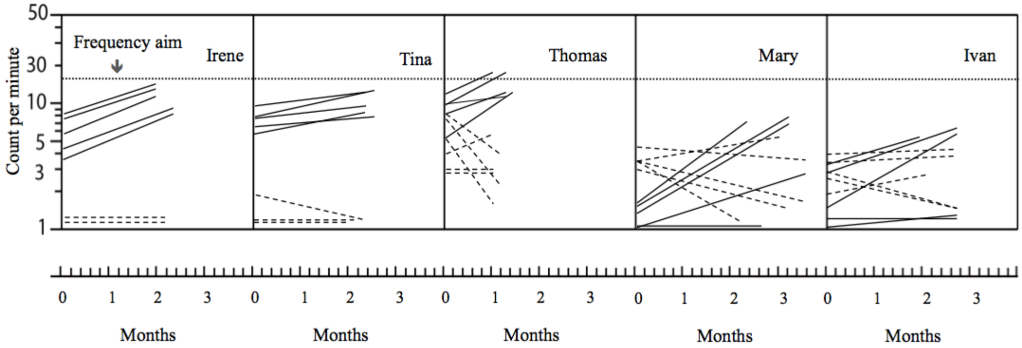


Figure 1. Celeration collections across similarity sets for all participants. Solid celeration lines represent the slope obtained across correct response frequencies. Dashed celeration lines represent the slope obtained across incorrect response frequencies. The horizontal dashed line indicates the frequency aim for similarity probes.

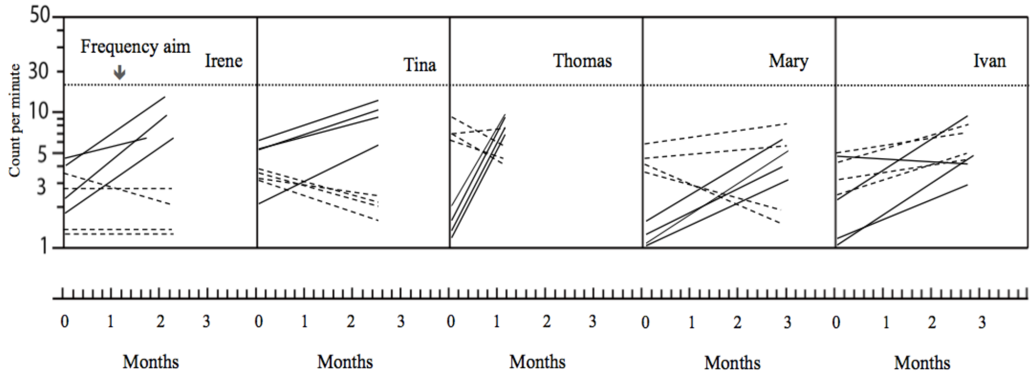


Figure 2. Celeration collections across distinction sets for all participants. Solid celeration lines represent the slope obtained across correct response frequencies. Dashed celeration lines represent the slope obtained across incorrect response frequencies. The horizontal line indicates the frequency aim for distinction probes.

Although Thomas' errors were occurring at high levels following his time as the constant-series control, they rapidly decreased when training was introduced, resulting in very steep celerations across correct and incorrect response frequencies. Also readily apparent in the collections is the brevity of Thomas' training following the pre-training phase. Thomas consistently obtained higher levels of correct responses on all probes than the other participants. Moreover, he obtained these levels in a fraction of the time – 7 weeks as compared to 12 and 14 weeks for Irene and Tina, and 15 and 18 weeks for Ivan and Mary. For Mary and Ivan, correct and error responses remained largely undifferentiated on several sets. Nonetheless, there are steep performance trends, demonstrating their improved ability to engage with these tasks.

Time spent in training was also measured with respect to the total amount of

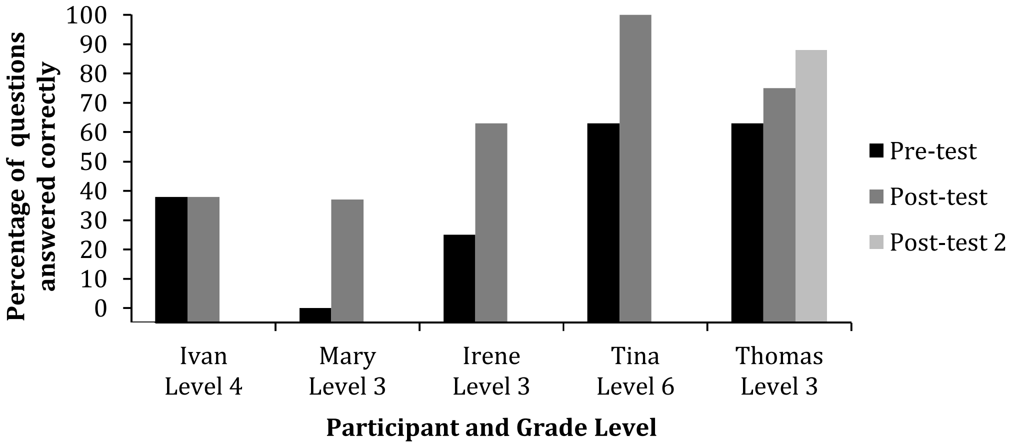


Figure 3. Pre and posttest results for all participants for the percent correct of comprehension on the QRI 4

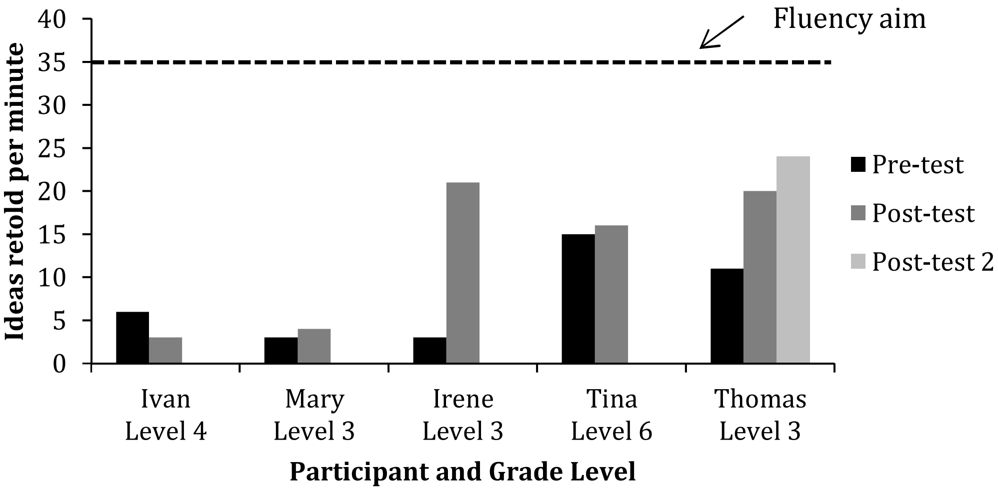


Figure 4. Pre an posttest results for all participants for the number of ideas retold per minute on the story retell component of the QRI 4

time spent in instruction (e.g., goal-setting, feedback and error correction). This measure also shows that results were produced with Thomas with fewer exemplars than were required with other participants. The total time spent in training for each participant was 8:35 (eight hours and 35 minutes) for Irene, 8:29 for Tina, 5:42 for Thomas, 16:09 for Mary and 13:18 for Ivan.

Pre-Post Measures

Three pre and post academic measures were obtained for each participant: percent correct on comprehension questions from

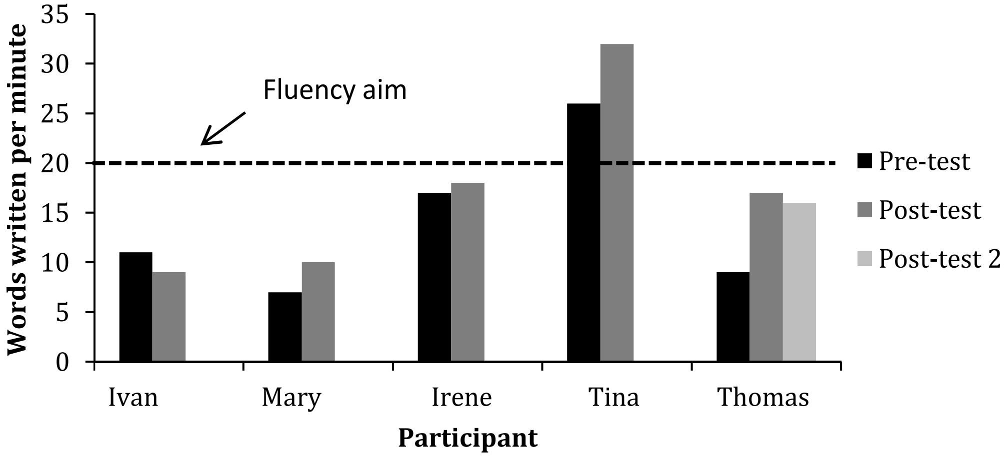


Figure 5. Pre and post testresult for all participants for words written per minute during the expressive writing measure.

the QRI 4, story retell, and written expression. The results for all participants are presented in Figures 3-5 and are described below.

QRI 4. All participants except Ivan showed an improvement over pretest scores in the percentage of correct questions answered (Figure 3). Irene's score improved from 25% to 63%, Tina's from 63% to 100%, and Mary's from 0% to 37%. Thomas, who served as the constant-series control, had three sets of measures; measures were obtained prior to the study, at the completion of the constant-series control phase, and subsequent to HRT and training on HCSD tasks. Thomas' scores improved following both posttests. At pretest, he scored 63% accuracy on comprehension questions. His post scores were 75% and 88% on posttests 1 and 2, respectively.

IOA data were obtained for 37% of the comprehension questions scored during the QRI 4 pre and posttests. Agreement was calculated by dividing the total number of agreements by the total number of disagreements and multiplying that number by 100. A score of 91% agreement was obtained. Story retell results are presented in Figure 4. A rate measure was obtained by by counting the number of nouns, verbs and modifiers emitted (adjectives, adverbs). Improvements were observed for Irene and Thomas, but not for the other three participants. Irene's rate increased from 3 responses per min to 21 per min at posttest. Thomas' rate of responses increased from 11 per min at pretest to 20 and 24 at posttest 1 and posttest 2.

Written expression. Pre and posttest results for the expressive writing measure are presented in Figure 5. Irene, Mary, Tina and Thomas all showed increases in the number of words written per min. For Irene and Mary, this was a marginal increase: 17 to 18 wpm for Irene, and 7 to 10 wpm for Mary. Tina and Thomas, however, showed a more notable increase. Tina's total wpm increased from 26 to 32, and Thomas' from 9 at pretest to 16-17 at posttests 1 and 2.

Discussion

The ability to behave effectively to what is read is an ability linked to language and influenced by other variables. As factors related to comprehension have been illuminated, effective technologies have emerged; nonetheless, these technologies fail to improve comprehension for many students. Few studies to-date have focused on ameliorating language deficits as a means of improving reading comprehension. This study provides some evidence, however, that such an approach can be both effective and efficient. Improvements in comprehension were obtained for all but one participant. The results obtained herein, therefore, have important implications for educational practices and research.

The results obtained provide support for the notion that increasing the frequency and flexibility of hierarchical relations is important for more complex language skills to emerge, such as discriminating particular hierarchical relations under relevant contexts. The responding occasioned during HCSD, which required participants to discriminate relevant dimensions of hierarchical relations under contextual cues of "same" and "different", shares features with the responding of a reader who connects aspects of text, compares and contrasts, makes predictions, and who bring one's personal history to bear. Thus, increasing the frequency and variability of hierarchical relations appears to have a generative effect on related academic repertoires like comprehension and written expression.

The findings of this study suggest that Relational Frame Theory is a pragmatic theoretical framework for designing comprehension interventions. Comprehension is complex language that includes predicting, summarizing, paraphrasing, inferring, integrating text, and responding personally to what we read (Leslie & Caldwell, 2006). Children with weak verbal repertoires insufficient for comprehension will benefit from efficient and effective interventions. The training employed rapidly proliferated language-based experience, resulted in derived higher-order relational frames and improved performance on important academic measures for all but one participant.

There is some support for the notion that fluency in relational behavior is important (e.g., Berens & Hayes, 2009; Cassidy et al., 2011). Preliminary findings here suggest that developing fluency in relational responding may lead to differential outcomes in academic performance. For instance, baseline rates of responding for Thomas, Tina and Irene demonstrate that they could respond with some level of accuracy and at moderate frequencies on HCSD probes. In some instances (e.g., Thomas), base frequencies increased throughout baseline. Nonetheless, the extent to which these repertoires were improving during baseline was of a much smaller magnitude than the improvements following training. Though these three participants presented with accurate, previously established hierarchical framing repertoires, improvements on comprehension measures at posttest correlated with increased rates of deriving. This point has relevancy to a discussion of the sufficiency of accuracy alone with respect to relational operants.

Moreover, Thomas' pre/post measures provide some precursory support for the notion that, while improvements in the accuracy of derived relational responding are good, derived relating at optimal frequencies may facilitate better outcomes for comprehension Thomas' pre/post measures provide some precursory support for the notion that, while improvements in the *accuracy* of derived relational responding are good, derived relating at optimal *frequencies* may facilitate better outcomes for comprehension. Improvements were obtained on Thomas' first posttest, though he received no explicit training and no programmed feedback. These improvements occurred following an 11-week baseline phase while he served as the constant-series control. Further improvements were obtained on posttest 2. Though improvements occurred on both posttests, the ratio of growth to time is worth noting. Gains of a greater magnitude were made in a shorter period of time (6 weeks as compared to 11 weeks) once Thomas was included as an experimental participant. It is possible that differential growth rates on comprehension measures were due to the extent to which Thomas achieved optimal response frequencies with HRT and HCSD training.

Table 3. *Transcription Sample for Mary on Similarity Probes*

Stimuli and Contextual Cue: "Same as"	**Participant Responses**
Responses to HCSD sets during baseline: Stimuli: *Socks, Shoes*	They are the same socks They are the same shoes They are not different They are the same
Responses to HCSD sets following HRT: Stimuli: *Socks, Shoes*	Both are warm Both are clothes You wear them You take them off at home You wear them in the afternoon Both have blue Find them in your room
Responses to HCSD sets during baseline: Stimuli: *Tiger, Pig*	A tiger has black and white stripes A tiger lives in the jungle A pig lives in the zoo
Responses to HCSD sets following HRT: Stimuli: *Tiger, Pig*	Both are animals Both have fur Both have tails They both have a nose They have eyes They have ears They have legs They live on a farm

Thus, fluency may play an important role in evaluating the strength of the relational repertoire. Certainly much more research is needed to explore the relationship between relational fluency and intellectual abilities.

The results obtained further support the premise that relational responding functions as a generalized operant. With multiple exemplar training, both the fluency and novelty of responding increased on probe measures as a result of training a language operant that generalized across stimulus classes. This can be seen in the qualitative measures obtained for each participant. Sample transcriptions are provided in Tables 3 and 4. The transcription reveals responses obtained on HCSD probe sets during baseline, and following training. The frequency-building procedures across multiple exemplars was a successful means of expanding the contextual field-stimuli that previously occasioned only a small set of relational responses later occasioned a multitude of responses. As participants were exposed to the four components of the HRT, novel relations were derived during HCSD.

Despite the profitability of RFT for informing education practices, few translations of RFT have been adopted in education. While there is a great deal more to learn, the results of this study suggest that relational

Table 4. *Transcription samle for Tina on Distinction Probes.*

Stimuli and Contextual Cue: "Different than"	**Participant Responses**
Responses to HCSD sets during baseline: Stimuli: *Rice, Watermelon*	Rice you eat with a spoon Watermelon you don't eat with a spoon
Responses to HCSD sets following HRT: Stimuli: *Rice, Watermelon*	Rice is smaller than watermelon Watermelon has seeds Watermelon is juicy Rice is seen in a bowl Watermelon has a part you don't eat
Responses to HCSD sets during baseline: Stimuli: *Skirt, Broccoli*	Broccoli you eat A skirt you don't eat A skirt has buttons Broccoli doesn't have buttons Broccoli is on a plate A skirt you wear
Responses to HCSD sets following HRT: Stimuli: *Skirt, Broccoli*	One is clothing One is food One is hanging on pins One could be eaten for breakfast One could be eaten for lunch One could be eaten for dinner One could be worn to school One is pink One is green One has buttons One is seen on a plate

framing comes to bear in comprehension and written expression. Thus, the results obtained have important implications for education. This study provides one of the first models of a curriculum package that can be easily implemented in applied settings to improve language-based academic deficits. The procedures are easy to implement, do not require special equipment and could be easily incorporated into an academic curriculum. Moreover, the training was efficient. Time spent in training ranged from 5 hours and 42 minutes to 16 hours and 9 minutes, supporting the efficacy of the training package for expanding language-based experience in a timely manner. As such, this methodology demonstrates how interventions based on RFT might be translated into academic environments and treatment facilities.

While the findings presented here are promising, there are limitations and future directions that warrant discussion. First, only five participants were included in the investigation, and a language package was employed. The study included a training package so as to approximate as closely as possible the conditions where these effects were initially observed in non-research clinical practice. A research program starts by replicating the effect of interest in a lab setting – picking the event apart only comes after it can be produced reliably. The small *n* time-series represents an early step in a to-be-completed loop of activities as depicted in Hayes & Barlow's scientist practitioner model (1999). The next steps in the model (tinkered program evaluation, RCTs, etc.) remain to be completed.

The most robust effects on outcome measures were obtained with Tina, Thomas and Irene. Mary and Ivan, however, failed to achieve equally robust gains on the various outcome measures. Mary's comprehension improved from 0% to 37% at posttest, and though this score is far from ample, it does suggest that training did impact Mary's reading comprehension abilities. Training at a more rudimentary level may have been necessary for Mary and Ivan given the extent of their language impairments. It is possible, for example, that skill building in other relations, such as non-arbitrary relations (i.e., relations based on physical features of stimuli), may have improved Mary and Ivan's performance on relational tasks when stimuli were relatable in only subtle, arbitrary ways (i.e., little or no shared physical features). Cassidy et al., (2011) employed a similar strategy when participants failed to pass training on arbitrary relations. It is also possible that an insufficient number of exemplars were trained with Mary and Ivan. More extensive multiple exemplar training may have resulted in a more robust effects. Future research should evaluate both of these possibilities.

Thomas' inclusion as a constant-series did not serve as a "pure" control. He was exposed to the same procedures as the other participants during the control phase. No feedback or goal setting was employed; nonetheless, his performance improved. Frequencies increased as a function of practice in the absence of goal setting and reinforcement (e.g., Berens & Hayes, 2009; Brooks & Boyce, 2005; Halligan, Berens & Ghezzi, 2010; Kaelin, Berens & Ghezzi, 2008). The improvement on Thomas' pre/post measures could be due to maturation or other events outside of the experimental context. However, there were no outside remediation efforts during the time of his participation in the study. Thomas' grades in reading comprehension and expressive writing after participation showed improvement over his three previous grading periods. His teacher noted that Thomas' ability to express himself had improved. Though conclusive statements regarding the nature of Thomas' improvement on pre/post measures are not permitted, the results obtained herein are promising and warrant further investigation.

We were concerned here with only a small set of relations, and we only evaluated the effects of training on comprehension. Comprehension of reading material likely requires relational frames beyond what was trained here. Currently a much broader array of relational operants are targeted in clinical practice at our center, including many hypothesized about in Hayes, et al., 2001. A systematic approach to understanding the effects and interactions resulting from fluency building of these various relational operants is needed. There is still much work to be done in evaluating the impact of fluency building in other relations and the subsequent effects on academic skills.

Conclusion

In sum, contextual behavior science is uniquely situated to contribute to the understanding of the core language processes that enable comprehension. Language is tantamount to comprehension. This preliminary investigation illuminates the utility of RFT as a viable theoretical model for understanding the complexities of language, and therefore, the components of comprehension. This study demonstrates that core language processes – arbitrarily applicable derived relational responding- can be directly trained to improve comprehension. Therefore, RFT provides a useful theoretical foundation from which effective technologies can emerge. We are hopeful that our study will inspire more research guided by RFT, and will stimulate educational research in the behavior analytic community.

References

Baer, D.M., Petersen, R.F., & Sherman, J.A. (1967). The development of imitation by reinforcing behavioral similarity to a model. *Journal of the Experimental Analysis of Behavior, 10*, 405-416.

Baer, D. M., & Sherman, J. A. (1964). Reinforcement control of generalized imitation in young children. *Journal of Experimental Child Psychology, 1*, 37-49.

Berens, N. M., & Hayes, S. C. (2009). *Using a vocabulary-building program to increase the rate of arbitrarily applicable derived relational responding* (Unpublished doctoral dissertation). University of Nevada, Reno.

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

Bishop, D. V., & Adams, C. (1990). A prospective study of the relationship between specific language impairment, phonological disorders and reading retardation. *Journal of Child Psychology and Psychiatry, 31*, 1027-1050.

Brooks, K. L., & Boyce, T. E. (2005). *An evaluation of progressive vs. fixed intervals to build academic frequencies among elementary school students* (Unpublished master's thesis). University of Nevada, Reno.

Carey, S. (1990). Cognitive development. In D. N. Osherson & E. E. Smith (Eds.), *Thinking: An invitation to cognitive science* (Vol. 3, pp. 147-172). Cambridge, MA: MIT Press.

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

Cassidy, S., Roche, B., O'Hora, D. (2010). Relational frame theory and human intelligence. *European Journal of Behavior Analysis, 11*,1-15.

Catts, H. W., Hogan, T. P., & Adlof, S. M. (2005). Developmental changes in reading and reading disabilities. In H. Catts & A. Kamhi (Eds.), *The connections between language and reading disabilities* (pp. 23-36). Mahwah, NJ: Erlbaum

Catts, H. W., & Kamhi, A. G. (Eds.) (2005). *The connections between language and reading disabilities*. Mahwah, NJ: Erlbaum

Clarke, P. J., Snowling, M. J., Truelove, E., & Hulme, C. (2010). Ameliorating children's reading-comprehension difficulties: A randomized controlled trial. *Psychological Science, 21,* 1106-1116.

Durkin, D. (1978-79). What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly,* 14, 481-533.

Dymond, S., & Barnes, D. (1995). A transformation of self-discrimination response functions in accordance with the arbitrarily applicable relations of sameness, more than, and less than. *Journal of the Experimental Analysis of Behavior, 54,* 163-184.

Dymond, S., & Roche, B. (Eds.). (2013). *Advances in relational frame theory: Research and application*. Oakland CA: New Harbinge.

Gersten, R., Fuchs, L. S., Williams, J. P., & Baker, S. (2001). Teaching reading comprehension strategies to students with learning disabilities: A review of research. *Review of Educational Research, 71*, 279-320.

Gough, P. B, & Tunmer, W. E. (1986). Decoding, reading, and reading disability. *Remedial and Special Education, 7*, 6-10.

Halligan, M., Berens, K. N., Ghezzi, P. M. (2010). *Evaluating the role of frequency building on the establishment of an emergent decoding repertoire.* Unpublished master's thesis, University of Nevada, Reno.

Hansen, J. (1981). The effects of inference training and practice on young children's reading comprehension. *Reading Research Quarterly, 16,* 391-417.

Hart, B. & Risley, T. (1995). *Meaningful differences in the everyday experiences of American children.* Baltimore, MD: Brookes

Hayes, S. C., Barlow, D. H., & Nelson-Gray, R. O. (1999). *The scientist-practitioner: Research and accountability in the age of managed care.* Boston MA: Allyon and Bacon.

Hayes, S. C., & Barnes, D. (1997). Analyzing derived stimulus relations requires more than the concept of stimulus class. *Journal of the Experimental Analysis of Behavior, 68*, 235-244.

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

Hayes, S. C., & Berens, N. M. (2004). Why relational frame theory alters the relationship between basic and applied behavioral psychology. *International Journal of Psychology & Psychological Therapy, 4*, 341-353.

Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., Healy, O. (2001). Derived relational responding as learned behavior. In S. C. Hayes, D. Barnes-Holmes &, B. Roche (Eds.), *Relational frame theory: A post-Skinnerian account of human language and cognition* (pp. 21-49). New York, NY: Kluwer Academic/Plenum.

Joseph, L. M. (2008). Best practices on interventions for students with reading problems. *Best Practices in School Psychology*, 4, 1163-1180.

Kaelin, K., Berens, K. N., Ghezzi, P. M. (2012). An investigation of rapid automatic naming as a generalized operant. *Journal of Precision Teaching and Celeration, 28*, 3-16.

Kubina, R.M., & Yurich, K.L. (2012). *The precision teaching book.* Lemont, PA: Greatness Achieved

Leader, G., & Barnes-Holmes, D. (2001). Matching to sample and respondent-type teaching as methods for producing equivalence relations: Isolating the critical variable. *The Psychological Record, 51*, 429-444.

Leslie, L., & Caldwell, J. (2006). *Qualitative reading inventory.* Boston MA: Pearson Education

Lindsley, O. R. (1992). Precision teaching: Discoveries and effects. *Journal of Applied Behavior Analysis, 25*, 51-57. doi:10.1901/jaba.1992.25-51

Lipkens, G., Hayes, S. C., & Hayes, L. J. (1991). Longitudinal study of derived stimulus relations in an infant. *Journal of Experimental Child Psychology*, 56, 201-239.

Luciano, C., Berens, N. M., Valdivia, S., Rodriguez, M., Mañas, I. I., Ruiz, F. J. (2008). Acquiring the earliest relational operants: Coordination, difference, opposition, comparison, and hierarchy. In R. A. Rehfeldt (Ed.), *Derived relational responding applications for learners with autism and other developmental disabilities: A progressive guide to change (pp.149-170).* Oakland, CA: New Harbinger.

Maloney, M., Blearley, L., & Preece, J. (1998). *Teach Your Children Well.* Cambridge, MA: Cambridge Center for Behavioral Studies

Nation, K., & Snowling, M. (1997). Assessing reading difficulties: The validity and utility of current measures of reading skill. *British Journal of Educational Psychology, 67*, 359-370.

National Center for Education Statistics. (2004). *The condition of education* 2004. Washington, DC: U.S. Government Printing Office.

National Reading Panel. (2000). *Teaching children to read: An evidence-based of the scientific research literature on reading and* *its implications for reading instruction.* Washington, DC: National Reading Panel.

Neilsen, A., Rennie, B., & Conell, A. (1982). Allocation of instructional time to reading comprehension and study skills in intermediate grade social studies classrooms. In J. Niles & L. Harries (Eds.), *New inquiries in reading research and instruction*. Rochester, N.Y.

Healy, O., Barnes-Holmes, D., Smeets, P. M. (2000). Derived relational responding as generalized operant behavior. *Journal of the Experimental Analysis of Behavior, 74*, 207-227.

Joseph, L. M. (2008). Best practices on interventions for students with reading problems. *Best Practices in School Psychology, 4,* 1163-1180

Nicholson, T., & Thompson, G. B. (1999). Reading comprehension processes. In G. B. Thompson & T. Nicholson (Eds.), *Learning to read: Beyond phonics and whole language* (pp. 127-149). New York, NY: Teachers College Press International Reading Association.

Ninness, C., Barnes-Holmes, D., Rumph, R., McCuller, G., Ford, A., Payne, R., Elliott, M. P. (2006). Transformation of mathematical and stimulus functions. *Journal of Applied Behavior Analysis, 39,* 299-321. doi: 10.1901/jaba.2006.139-05

Ninness, C., Dixon, M., Barnes-Holmes, D., Rehfeldt, R. A., Rumph, R., McCuller, G., Holland, J., Smith, R., Ninness, S. K., & McGinty, J. (2009). Constructing and deriving reciprocal trigonometric relations: A functional analytic approach. *Journal of Applied Behavior Analysis, 42*, 191-208. doi: 10.1901/jaba.2009.42-191

Ninness, C., Rumph, R., McCuller, G., Harrison, C., Ford, A. M., & Ninness, S. (2005). A functional analytic approach to computer-interactive mathematics. *Journal of Applied Behavior Analysis, 38,* 1-22. doi: 10.1901/jaba.2005.2-04

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

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

Pearson, P. D., & Gallagher, M. (1983). The instruction of reading comprehension. *Contemporary Educational Psychology, 8,* 317-344.

Pennypacker, H. S., Guiterrez, A., & Lindsley, O.R. (1972). *Handbook of the Standard Behavior Chart: Deluxe Edition.* Gainesville, FL.: H. S. Pennypacker.

Persicke, A., Tarbox, J., Ranick, J., & St Clair, M. (2012). Establishing metaphorical reasoning in children with autism. *Research in Autism Spectrum Disorders, 6,* 913-920.

Pikulski, J. J., & Chard, D. J. (2005). Fluency: Bridge between decoding and reading comprehension. *Reading Teacher, 58*, 510-519.

Rasinski, T. V., & Hoffman, J. V. (2003). Oral reading in the school literacy curriculum. *Reading Research Quarterly, 38,* 510-522.

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

Snow, C. E., Burns, M. S., & Griffin, P. (1999). Preventing reading difficulties in young children. In C. E. Snow, M. S. Burns, & P. Griffin (Eds.), *Reading research: Anthology: The why? of reading instruction* (pp. 148-155). Novato, CA: Arena Press.

Steele, D.L., & Hayes, S.C. (1991). Stimulus equivalence and arbitrarily applicable relational responding. *Journal of the Experimental Analysis of Behavior, 56*, 519-555.

Törneke, N. (2010). *Learning RFT: An intro-*

duction to relational frame theory and its clinical applications. Oakland, CA: New Harbinger

Williams, J. P. (2002). Reading comprehension strategies and teacher preparation. In A. E. Farstrup & J. Samuels (Eds.), *What research has to say about reading instruction* (pp. 243-260). Newark, DE: International Reading Association, Inc.

Williams, J. P. (2005). Instruction in reading comprehension for primary-grade students: A focus on text structure. *Journal of Special Education, 39,* 6-18.

Wulfert, E., & Hayes, S.C. (1988). The transfer of conditional sequencing through conditional equivalence classes. *Journal of the Experimental Analysis of Behavior, 50,* 125-144.

Appendix A

List of Stimuli for Similarity Sets 1-4

Set 1
Bagel- Crackers
Turtle- Alligator
Socks-Tennis Shoes
Belt-Pants
Dog-Goat
Shorts –Skirt
Broccoli- Banana
Baseball- Swing
Crib-Chair
Bus-Ship
Spoon-Plate
Brush-Hand Soap
Grapefruit- Carrots
Milk- Jell-o
Popsicle- Frosted Animal Crackers

Set 2
Hat- Shirt
Jeep- Tractor
Bucket- Football
Cake- Cookies
Wagon- Tricycle
Van- Ambulance
Duck- Chicken
Drawers-Coffee Table
Hamburger- Spaghetti
Pig- Tiger
Chicken Nuggets- Soda
Phone- Lamp
Toothbrush- Towel
Stove- Knife
Eggs- Fried Chicken

Set 3
Motorcycle- Zebra
Dinosaurs- Gummy Bears
Bed- Boots
Cup-Sink
Snake- Play Dough
Blocks- Gloves
Car- Computer
Sail Boat- Elephant
Baseball Glove- Fries
Couch- Popcorn
Cheetos- Stairs
Peach- Leaf
Lettuce- Lunch Box
Grapes- Rocks

Set 4
Horse- Apple- Soccer Ball
Plane- Bird- Bubbles
Pillow- Teddy Bear- Cat
Lizard- Beach Ball- Bread
Bunk Beds- Rhino- Dump Truck
Bulldozer- Cheese- Puzzle
Bunny- Doll- Jacket
Slide- Hippo- Mail Truck
Backpack- Swim Shorts- Paints
Lawn Chair- Stroller- Cop Car

Appendix B

List of Stimuli for Similarity Set 5

The investigator randomly selects four items to present during probes by drawing a number. One item will be read to the participant every 15 seconds.

1. How is taking care of a bird like taking care of a dog?
2. How is eating ice cream like eating Jell-O?
3. How is eating ketchup like eating mustard?
4. How is an eating pretzel like eating popcorn?
5. How is playing basketball like playing baseball?
6. How is riding a bike like driving in a car?
7. How is playing catch like playing video games?
8. How is playing with a beach ball like playing with a soccer ball?
9. How is going to the park like going to the beach?
10. How is riding a school bus like riding a train?
11. How is building a tree house like playing with legos?
12. How is playing with play dough like putting together a puzzle?
13. How is rollerblading like playing football?
14. How is playing at the park like swimming in the pool?
15. How is riding in a helicopter like riding on a plane?
16. How is eating an orange like eating an apple?

Appendix C

List of Stimuli for Distinction Sets 1-3

Set1
Bus-Dog
Turtle- Chair
Crib-Swing
Banana-Shoes
Cracker- Alligator
Shorts- Goat
Baseball- Socks
Belt- Ship
Pants- Bagel
Skirt- Broccoli
Plate- Popsicle
Brush- Frosted Animal Crackers
Spoon- Carrots
Grapefruit- Hand soap
Milk- Hammer

Set 2
Bat- Strawberries
Orange- Cassette Tape
Desk- Hot Wheels Car
Tomatoes- Bear
Keys- Turkey
Giraffe- Pan
Donuts- Glasses
Umbrella- Sandwich
Sweater- Mouse
Hot Dog- Rollerblades
Legos- Sunflower
Peanut Butter- High Heels
Toilet- Toast
Corn- Camera
Soup- Viewfinder

Set 3
Barbie- Bicycle Helmet
Cereal- Apple Juice
Kangaroo- Cow
Ketchup- Corn dog
Frog- Lion
Candy- Peas
Shovel- Crayons
Tennis Ball- Cassette Player
Tow Truck- Bicycle
Underwear- Swim Suit
Cupcake- tacos
Fruit Roll Up- Butter
Ice cream- Chips
Rice- Watermelon
Balloons- Mr. Potato Head
Pizza- Pretzels

Appendix D

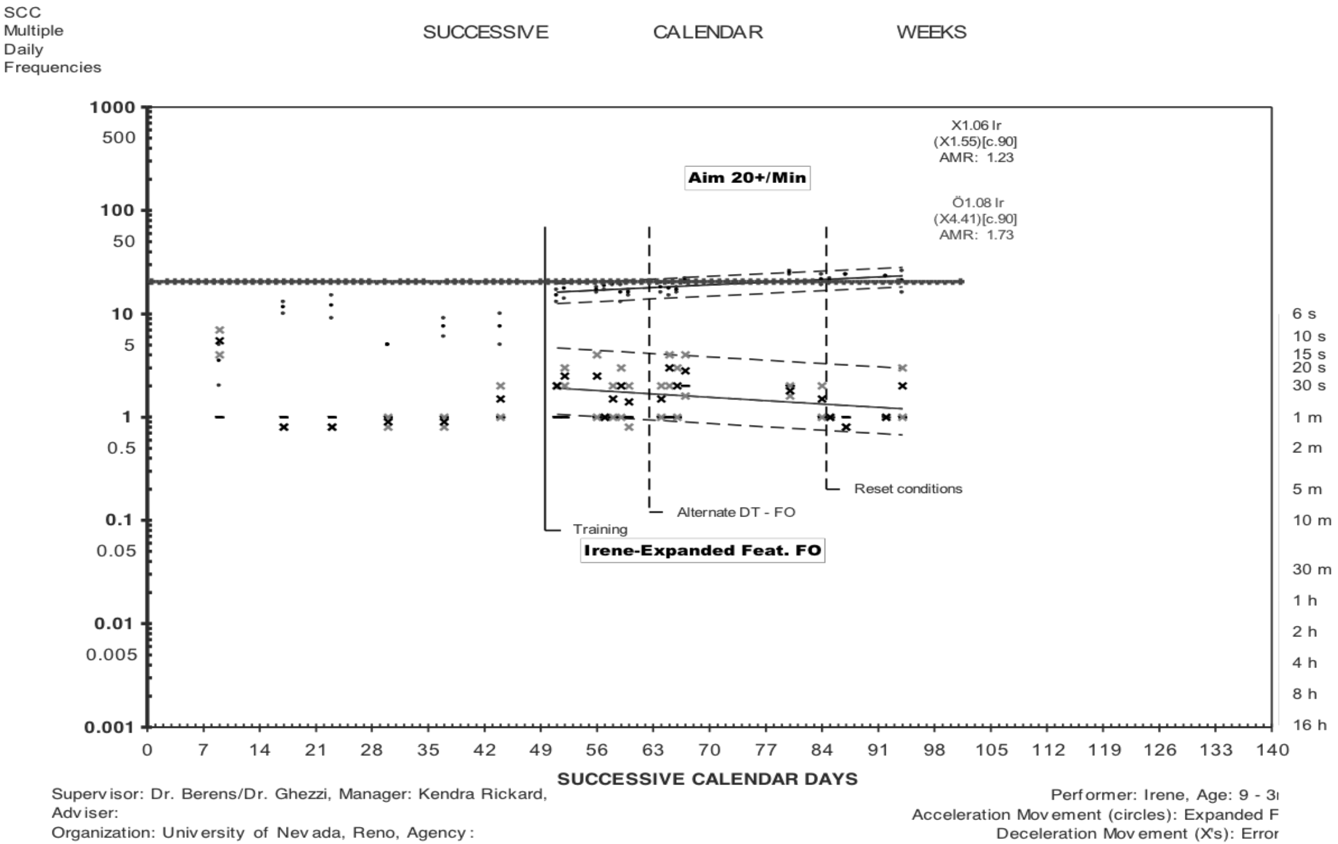
List of Stimuli for Distinction Set 4

The investigator randomly selects four items to present during probes by drawing numbers. One item will be read to the participant every 15 seconds.

1. How is taking care of a bird different than taking care of a dog?
2. How is eating ice cream different than eating Jell-O?
3. How is eating ketchup different than eating mustard?
4. How is an eating pretzel different than eating popcorn?
5. How is playing basketball different than playing baseball?
6. How is riding a bike different than driving in a car?
7. How is playing catch different than playing video games?
8. How is playing with a beach ball different than playing with a soccer ball?
9. How is going to the park different than going to the beach?
10. How is riding a school bus different than riding a train?
11. How is building a tree house different than playing with legos?
12. How is playing with play dough different than putting together a puzzle?
13. How is rollerblading different than playing football?
14. How is playing at the park different than swimming in the pool?
15. How is riding in a helicopter different than riding on a plane?
16. How is eating an orange different than eating an apple?

Appendix E

Expanded Features Training for Irene



Appendix F

Scripted Instruction for Expanded Features

1. During the probe phase, the investigator presents a picture of an apple, a baseball glove and a dog to run the scripted instruction.
2. When expanded features enter into training, the investigator randomly selects the three stimuli to present during training trials. These stimuli are presented during the scripted instruction.

Investigator: I would like you to name the objects in the pictures for me.

Investigator: One way that we can describe things is by their expanded features.

Investigator: What is one way that we can describe things?

Investigator: Expanded features are different than concrete features because they aren't objects or parts of objects that we can see.

Investigator: Can we see expanded features the same way we can see concrete features, yes or no?

Investigator: Are expanded features parts of things like concrete features are, yes or no?

Investigator: Expanded features might be where we can find an object.

Investigator: So an expanded feature might be where we_______________.

Investigator: For example, an expanded feature of an apple might be that we might find it at the store.

Investigator: What might be an expanded feature of an apple?
Investigator: Other expanded features might include how we experience something. For example, an expanded feature of an apple might be that it is crunchy. What might be an expanded feature of an apple?

Investigator: Other expanded features might be things we find with other things. For example, we might have an apple with peanut butter on it. What kind of feature is that?

Investigator: Or we might find a baseball glove with a baseball bat. That can be an expanded feature of a baseball glove.

Investigator: Expanded features might include features that we know because we have had contact with an object. For example, an expanded feature of a dog might be that dogs growl or chase cats. What kind of features are those?

Investigator: Now I want you to name expanded features of the objects in the pictures.

Appendix G

Sample Scripts for Expanded Feature

Bike
Could you see it in the garage?
Can it have training wheels?
Could you see someone sleeping on it?
Can it have a bell on it?
Can a child ride it?
Do you find it at a bike shop?
Could you find it in a fridge?
Can you see it at a sporting goods store?
Can you see it riding on a road?
Could you see it in the mountains?
Are they made by different companies?
Could you ride a bike at a park?
Could you see it in the sky?
Is it found in a backyard?
Can you see adults riding it?
Do you play with it outside?
Is it soft?
Do people usually ride them when the weather is nice?
Can the seat be moved higher or lower?
Can it have pegs on it?
Could you see one with streamers on the handlebars?
Can it have a basket on the front?
Can it tell you the time?
Can you save up money to buy one?
Could you give your friend a ride on it?
Can it take you places?
Can it breathe?
Could you ride it to school?
Could you see it with a bike lock on it?
Could you see it on a bike rack?
Can it do your homework?
Can you ride it to the store?
Can it be a mountain bike?
Can you race on a bike?
Can people do tricks on bikes?
Can you grow flowers on it?
Are there girls' bikes and boys' bikes?
Do people wear spandex when they ride bikes?

Appendix H

Scripted Instruction for Similarity Probes

Similarity Probes: Sets 1-4

Investigator: I would like you to name the objects in the pictures for me.

Investigator: I want you to tell me how these objects are the same or like each other. What will you tell me?

Investigator: Only tell me how they are the same. Don't tell me how they are different. Will you tell me how they are the same, yes or no?

Investigator: Will you tell me how they are different, yes or no?

Investigator: I want you to say, "They are both …, or they are all…". What will you say?

Investigator: For example, if I asked you how a dog was like a goat, you might say," They are both animals, or they both have fur." How might you tell me a dog is the same as a goat?

Similarity Probes: Set 5

Investigator: I am going to ask you to tell me how two activities are the same. What am I going to ask you?

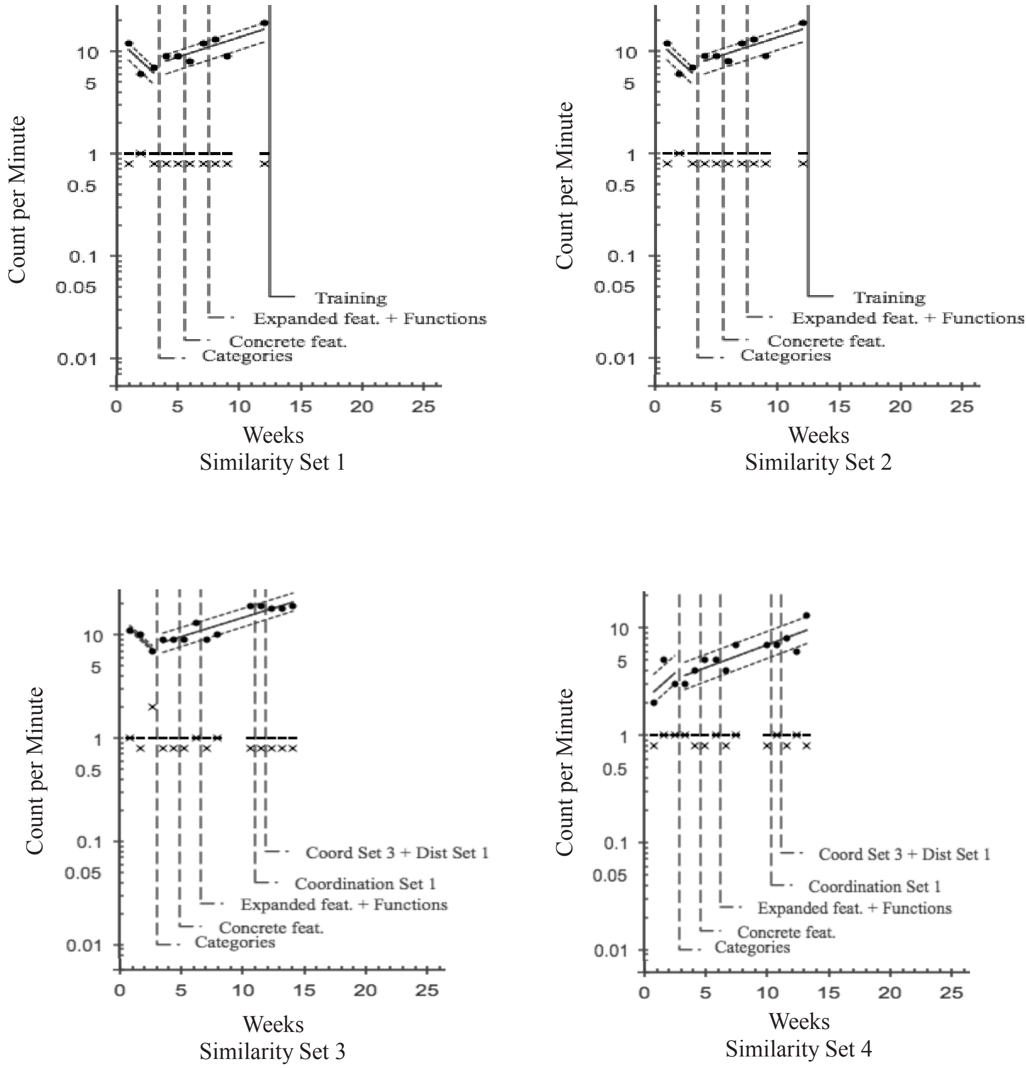
Investigator: I will ask you how one activity is like another activity. I only want you tell me how they are the same. Don't tell me how they are different. Will you tell me how they are the same, yes or no?

Investigator: Will you tell me how they are different, yes or no?

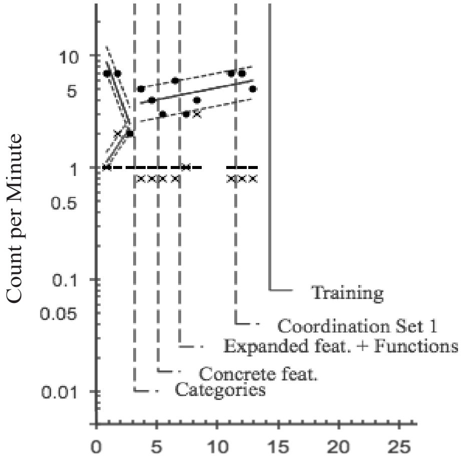
Investigator: I want you to say, "They are both…". What will you say?

Investigator: For example, if I asked you how a dog was like a goat, you might say," They are both animals, or they both have fur." How might you tell me a dog is the same as a goat?

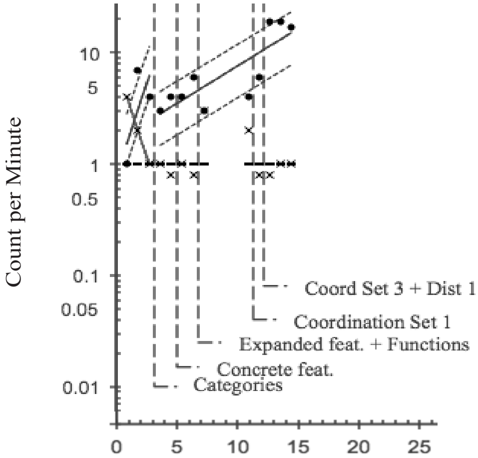
Appendix I
Similarity Probes for Irene



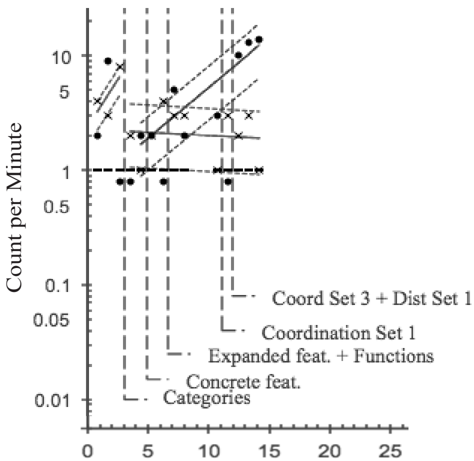
Appendix J
Distinction Probes for Irene



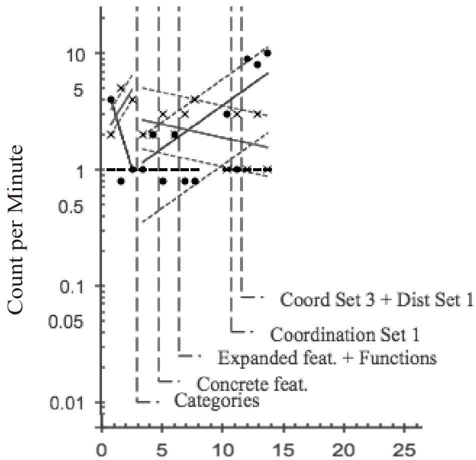
Weeks
Distinction Set 1



Weeks
Distinction Set 2



Weeks
Distinction Set 3



Weeks
Distinction Set 4