

Utah State University

DigitalCommons@USU

---

All Graduate Theses and Dissertations

Graduate Studies

---

5-2009

## Combining Information to Answer Questions about Names and Categories

Ginger L. Kelso  
*Utah State University*

Follow this and additional works at: <https://digitalcommons.usu.edu/etd>

 Part of the [Education Commons](#)

---

### Recommended Citation

Kelso, Ginger L., "Combining Information to Answer Questions about Names and Categories" (2009). *All Graduate Theses and Dissertations*. 406.  
<https://digitalcommons.usu.edu/etd/406>

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact [digitalcommons@usu.edu](mailto:digitalcommons@usu.edu).



COMBINING INFORMATION TO ANSWER QUESTIONS ABOUT  
NAMES AND CATEGORIES

by

Ginger L. Kelso

A dissertation submitted in partial fulfillment  
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Disability Disciplines

Approved:

---

Dr. Timothy A. Slocum  
Major Professor

---

Dr. Sandra Gillam  
Committee Member

---

Dr. Charles Salzberg  
Committee Member

---

Dr. Lori Roggman  
Committee Member

---

Dr. Sarah Rule  
Committee Member

---

Dr. Cynthia Rowland  
Committee Member

---

Dr. Byron Burnham  
Dean of Graduate Studies

UTAH STATE UNIVERSITY  
Logan, Utah

2009

Copyright © Ginger L. Kelso 2009

All Rights Reserved

## ABSTRACT

Combining Information to Answer Questions about  
Names and Categories

by

Ginger L. Kelso, Doctor of Philosophy

Utah State University, 2009

Major Professor: Dr. Timothy A. Slocum  
Department: Special Education and Rehabilitation

Children's language and world knowledge grows explosively in the preschool years. One critical contributor to this growth is their developing ability to infer relations beyond those that have been directly taught or modeled. Categorization is one type of skill commonly taught in preschool in which inference is an important aspect. This study explored the development of specific types of inferences within a categorization relation: those among naming items and categories, selecting items based on their names and categories, and answering questions that relate names and categories. Children learned names and categories for a set of unfamiliar cartoon characters through one of two training protocols: (a) *Listener training* involved selecting a picture upon hearing an item name or category; (b) *Expressive training* involved saying an item name or category upon seeing a picture. Following training, we tested whether children derived several kinds of untrained responses. Those children who received Expressive training (saying names) completed tests of listener responses (selecting pictures); similarly, those children who

received Listener training (selecting pictures) completed tests of expressive responses (saying names). Next, children answered oral questions in the absence of pictures. Results show that children receiving Expressive and Listener training produce naming and question answering responses at levels above chance. However, many children failed to answer all questions correctly. The Expressive group produced naming and question answering responses at significantly higher levels than the Listener group. This suggests that Listener training is a weaker form of instruction when the goal of instruction is the production of untrained responses. However, these results are tentative because unequal proportions of children completed each type of training. Finally, we examined the relationship between naming and question answering. Few children answered questions at a higher level than they produced names. This study shows that children learn to infer responses from both Listener and Expressive trainings. This study also suggests that naming and question answering responses are related responses. The current study highlights the need for later research on teaching inference skills such as naming and question answering to those who do not develop them in the absence of specific instruction.

## CONTENTS

	Page
ABSTRACT.....	iii
LIST OF TABLES.....	vii
LIST OF FIGURES.....	ix
CHAPTER	
I. INTRODUCTION.....	1
Skinner’s Analysis of Verbal Behavior.....	2
Relational Frame Theory .....	19
Reinterpreting the Literature Based on RFT.....	24
Existing Research on Mutual Entailment.....	27
Conclusions.....	31
II. METHODS.....	34
Participants and Setting.....	34
Materials.....	35
Procedures.....	36
Dependent Variables.....	41
Social Validity.....	45
III. RESULTS.....	46
Treatment Fidelity and Reliability of Measures.....	46
Descriptive Analysis.....	47
Research Question 1.....	56
Research Question 2.....	61
Research Question 3.....	62
Research Question 4.....	68
Research Question 5.....	75
Social Validity.....	76
IV. DISCUSSION.....	85
Overview of Study.....	85
Findings.....	86
Alternative Explanations.....	97

Implications for Future Research.....	107
REFERENCES.....	113
APPENDICES.....	117
Appendix A: Example of Test Stimuli Page.....	118
Appendix B: Example of Stamp Cards.....	120
Appendix C: Data Sheets.....	122
Appendix D: Treatment Fidelity Checklists.....	127
Appendix E: Statistical Procedures.....	129
Appendix F: Delayed Mastery Subanalysis.....	132
Appendix G: Calculation of Chance.....	143
Appendix H: Contingency Table Analysis .....	146
CURRICULUM VITAE.....	151

## LIST OF TABLES

Table		Page
1	Antecedents and Consequences of Verbal Behavior.....	5
2	Comparison of Study Characteristics.....	16
3	Aspects of Relational Responding.....	21
4	Comparison of Study Characteristics using RFT.....	26
5	Examples of Pictures, Names, and Categories.....	37
6	Completion Rates of PES and ELL Subgroups.....	49
7	Age of Participants Assigned to and Completing Training.....	49
8	Mean (SD) of Dependent Variables by Training Group.....	50
9	Mean (SD) of Dependent Variables for PES and ELL Subgroups.....	54
10	Wilcoxon Signed Ranks Test Comparisons to Chance for Expressive Training.....	59
11	Wilcoxon Signed Ranks Test Comparisons to Chance for Listener Training.....	64
12	Kolmogorov-Smirnov Test Comparing Listener and Expressive Training.....	69
13	Correlations Between Working Memory and other Dependent Variables.....	77
14	Correlations Between Matching and other Dependent Variables.....	83
15	Comparisons of Head Start Benchmarks to Current Study Outcomes.....	84
F1	Kolmogorov-Smirnov Test Comparing Expressive and Listener Groups.....	136
F2	Correlations Between Working Memory and Dependent Variables.....	141



H1	Contingency Table Comparing Mastery of Mutual Entailment and Intraverbals.....	148
H2	Contingency Table Comparing Mastery of Mutual Entailment and Intraverbals (PES).....	150

## LIST OF FIGURES

Figure	Page
1	Mutual entailment between picture and item name..... 23
2	Combinatorial entailment among picture, item name, and category label..... 23
3	Sequences of training and testing.....39
4	Training and testing for item names.....42
5	Training and testing for category labels.....42
6	Dot plot distributions for mutual entailment.....51
7	Dot plot distributions for intraverbals.....52
8	Dot plot distributions for delayed mastery.....52
9	Comparison between ELL and PES means (Listener).....53
10	Comparison between ELL and PES means (Expressive).....54
11	Comparisons to chance for all participants (Expressive).....57
12	Comparisons to chance for PES (Expressive)..... 58
13	Comparisons to chance for ELL (Expressive)..... 58
14	Comparisons to chance for all participants (Listener)..... 62
15	Comparisons to chance for PES (Listener)..... 63
16	Comparisons to chance for ELL (Listener)..... 63
17	Dot plot distributions of mutual entailment and intraverbals for all participants (Listener and Expressive)..... 66
18	Dot plot distributions of mutual entailment and intraverbals for PES (Listener and Expressive)..... 67

19	Dot plot distributions of mutual entailment and intraverbals for ELL (Listener and Expressive).....	67
20	Scatterplot of mutual entailment and intraverbals for all participants (Listener).....	71
21	Scatterplot of mutual entailment and intraverbals for all participants (Expressive).....	72
22	Scatterplot of mutual entailment and intraverbals for PES (Listener).....	72
23	Scatterplot of mutual entailment and intraverbals for PES (Expressive).....	73
24	Scatterplot of mutual entailment and intraverbals for ELL (Listener).....	74
25	Scatterplot of mutual entailment and intraverbals for ELL (Expressive).....	75
26	Scatterplot of mutual entailment and working memory (Listener).....	78
27	Scatterplot of intraverbals and working memory (Listener).....	78
28	Dot plot distributions for matching (Listener and Expressive).....	79
29	Scatterplot of matching and mutual entailment (Listener).....	80
30	Scatterplot of matching and intraverbals (Listener).....	81
31	Scatterplot of matching and mutual entailment (Expressive).....	81
32	Scatterplot of matching and intraverbals (Expressive).....	82
A1	Example stimuli page used in training and testing procedures.....	119
B1	Example of stamp card used in motivation procedures.....	121
D1	Assessment and training steps fidelity checklist.....	128
D2	Working memory fidelity checklist.....	128
F1	Dot plot distributions of mutual entailment and intraverbals for all participants (Listener and Expressive). ....	134
F2	Dot plot distributions of mutual entailment and intraverbals for PES (Listener and Expressive). ....	134

F3	Dot plot distributions of mutual entailment and intraverbals for ELL (Listener and Expressive).....	135
F4	Scatterplot of mutual entailment and intraverbals for all participants (Listener).....	137
F5	Scatterplot of mutual entailment and intraverbals for all participants (Expressive). ....	138
F6	Scatterplot of mutual entailment and intraverbals for PES (Listener).....	139
F7	Scatterplot of mutual entailment and intraverbals for PES (Expressive).....	139
F8	Scatterplot of mutual entailment and intraverbals for ELL (Listener).....	140
F9	Scatterplot of mutual entailment and intraverbals for ELL (Expressive).....	141
F10	Scatterplot of mutual entailment and working memory (Listener).....	142
F11	Scatterplot of mutual entailment and working memory (Listener).....	142
G1	Possible mutual entailment responses for chance calculation.....	144
G2	Possible intraverbal responses for chance calculation.....	144

## CHAPTER I

### INTRODUCTION

Language is the behavior that most clearly separates humans from other animal species. Language allows us to engage in many complex interactions and communications that would not be possible otherwise. On a general level, language allows us to interact more effectively in a society. Language allows us to learn through instructions (Hayes, Zettle, & Rosenfarb, 1989). We alter our behavior based on our understanding of words without having to directly experience the events described. We do not have to learn everything through the effects of consequences or through imitation. Instead, we can follow instructions to correctly perform a vast variety of behaviors including performing novel behaviors performed for the first time and improving previously learned skills. Instructions can establish discriminative stimuli by specifying the relationship between antecedent conditions and consequences of behaviors (e.g., Skinner, 1966). On a broader scale, knowledge and culture can be passed down through generations using language. Our behavior is in part a result of the experience of others impacting us through their language. The value of language is evident because language provides a means to generate or infer knowledge not learned through direct teaching. Theories of language development help to explain the role and importance of language. Numerous scholars from a wide variety of psychological traditions have attempted theoretical analyses of language to map out what language is, how it is learned, and how it affects our behavior. In this study, we will focus primarily on behavioral theories of language.

## Skinner's Analysis of Verbal Behavior

B.F. Skinner provided a functional analysis of verbal behavior (Skinner, 1957). His fundamental observations were that language (verbal behavior) is operant behavior, people engage in these behaviors because of a variety of consequences and in a variety of contexts, and that instances of verbal behavior can be classified according to their function and context. Here, we will focus on just three types of verbal behavior defined by Skinner – mands, tacts, and intraverbals. The mand is an instance of verbal behavior that is a result of a motivating operation (such as a state of deprivation or a need). In non-technical terms, mands are frequently called requests, commands, or pleas. The mand is reinforced by access to the thing requested. For example, a very simple mand for a candy bar may take the form of “Give me a candy bar” or “That candy bar sure looks good.” If these behaviors are typically reinforced by access to the candy bar, they are mands, even if they are not formally phrased as requests (as in the second example).

A second type of verbal behavior defined by Skinner is the tact. Tacting is frequently called naming, identifying, or describing. The tact is under the control of a nonverbal stimulus in the environment (e.g., an object) and is reinforced by a generalized social form of reinforcement such as attention. For example, a child may see a cat and say “cat.” This is reinforced by the parent giving attention, saying “great,” and so on. In this example, the response “cat” is under the control of the object cat and is reinforced with praise, not access to the cat. It is important to note that the mand and the tact may have the same form, but still have different functions. For example, if the child says “milk” to obtain milk, the word is functioning as a mand. If the child is not thirsty but sees some

milk and says “milk” for general attention and praise, the word “milk” is serving a different function, that of a tact.

The third type of verbal behavior that will be addressed here is the intraverbal. The intraverbal is a verbal response to a verbal stimulus (i.e., it is a response to someone else’s or one’s own statement or question) reinforced through generalized reinforcement. For example, given the stimulus “Who is your teacher?” the response “Mrs. Jones” is an intraverbal. This response is reinforced by a generalized form of reinforcement. Many categorization responses are also intraverbals. For example, the child says “cat” when told to “name an animal.” Or the child says “animal” when asked “What is a cat?” Each response is socially reinforced and is not controlled by the presence of the object (as in a tact) or a need for the object (as in a mand). Many of the common things we say throughout the day are intraverbal responses. Saying the word “plants” when asked “What do herbivores eat?” is an example of an intraverbal response. “Seraphim” is an intraverbal response to “Name a celestial being.” On a more complex level, logical syllogisms require intraverbal responses. For example, “If the bell rings, then there is someone at the door. The bell did not ring. Is there someone at the door?” You can produce the correct response to this question through a series of intraverbal responses based on the rules of logic. These responses are intraverbal because each response is based on a verbal antecedent and are typically followed by a social form of reinforcement. As these examples suggest, the class of intraverbals is large and varied.

Skinner continued his analysis beyond classifying instances of verbal behavior based on their contexts and functions. He also analyzed the ways in which each type of verbal behavior is likely learned. Skinner proposed that different types of verbal

behaviors initially develop independent of each other despite the fact that they might have the same topographical form (i.e., use the same word) – this has been referred to as functional independence. For example, a child may learn to name milk when seeing a glass of milk on the table, but not be able to request it when thirsty. In fact, it is a common observation among language teachers and researchers that language taught out of context may not be used in its real world context. The possibility of various types of verbal behavior developing independently is plausible if we consider the different conditions under which each type of response is produced. Table 1 contains a list of types of verbal behavior (mand, tact, and intraverbal) along with the conditions under which each type occurs. Scanning down the *antecedent* column shows the important differences in the contexts of each of these forms of verbal behavior. The mand for milk occurs when the speaker *wants* milk, the tact occurs when the speaker *sees* milk, and the intraverbal occurs when the speaker hears a question (or other verbal stimulus) that calls for “milk” as a response. These antecedents for the three types of responses have no physical characteristics in common. Based on these differences in antecedents (and consequences), Skinner argued that the different types of responses cannot be assumed to automatically interconnect. If a person can use a word in multiple ways, each usage of the word must be accounted for separately.

A substantial amount of research exists on functional independence of naming and requesting. (To aid the reader, from this point on, mands will be referred to as requests and tacts will be referred to as names.) For example, Lamarre and Holland (1985) examined whether young children could transfer responses from requests to



Table 1

*Antecedents and Consequences of Verbal Behavior*

Type of verbal Behavior	Antecedent	Behavior	Consequence
Mand (requesting)	”Desire” for a thing. (e.g., thirsty for milk)	Saying “milk”	Access to milk
Tact (naming)	The presence of a <i>thing</i> (e.g., glass of milk)	Saying “milk”	Social reinforcement (adult smiles)
Intraverbal	Verbal stimulus (e.g., “What is cheese made from?”)	Saying “milk”	Social reinforcement (adult smiles)

names, and vice versa. Nine typically developing children between the ages of three and five were taught to either name or request the placement of an object. Five children were trained to name the placement. They were shown two objects placed side by side and were asked “Where is the (object)?” The children were trained to respond either “on the left” or “on the right.” Following training, each child was tested to determine if s/he could also request the placement of the object by answering the question “Where do you want me to put the (object)?” The child was expected to answer either “on the left” or “on the right.” Following training on object placement names, these children did not produce requests for object placement. Four additional children were directly trained to request object placement. They were shown two objects, and were asked “Where do you want me

to put the (object)?" They were taught to say either "on the left" or "on the right." When the child said one of these phrases the objects were placed in the requested position. Following training, these four children did not produce the corresponding naming response for placement of the object. Overall, Lamarre and Holland found that nine typically developing preschool-age children did not show transfer between naming and requesting behaviors. Given these outcomes, the Lamarre and Holland study supports the argument that different types of verbal behavior develop independently in young children. However, spatial positions or placements are abstract concepts. It is not known whether the children would have performed differently if they were asked to name or request concrete objects instead of object placements. Additionally, it is not clear whether the children were sufficiently motivated to request placement of an object. The children may have performed differently given sufficient motivation.

Petursdottir, Carr, and Michael (2005) suggest that the functional independence demonstrated in the Lamarre and Holland (1985) study may have been due to methodological features of the study described. To test their conclusions, Petursdottir et al. conducted a replication study. Typically developing 2 to 4-year-old children were trained to complete two puzzles. Correct placement of each piece of the puzzle was established as a reinforcer by providing the children with stickers upon puzzle completion. Stickers were assumed to be reinforcing because the children chose the specific stickers to be earned prior to each session. Delivery of stickers was always paired with praise. After the children could reliably complete the puzzle, they were taught to name each piece of one puzzle. Nonsense words were used as names. The experimenter held up each piece and asked "What is this?" Correct responses were reinforced with

praise and stickers. Incorrect responses were followed by a model of the correct name. Once the children had mastered names, they were instructed to complete the puzzle. However, the puzzle pieces were kept out of the child's sight. The child was expected to request each puzzle piece. If no request occurred, the experimenter prompted a request by asking "What do you need?" Following training in naming responses, two of the five children reliably requested the puzzle pieces. Of the three children who did not reliably request puzzle pieces, two produced requests intermittently while one child failed to produce any requests. For the second puzzle, requesting of pieces was directly trained. The procedure was identical to the requesting test for the first puzzle except different names were used for pieces of the second puzzle. Following mastery of requesting, the experimenter held up each puzzle piece and asked, "What is this?" Following training for requests, all children reliably produced untrained naming responses. This study shows that children even younger than those in the Lamarre and Holland study are able to transfer responses from requesting to naming with some children also transferring names to requests when (1) concrete objects were named and requested, and (2) the children were motivated to request the objects.

Hall and Sundberg (1987) examined the independence of tact and mand repertoires using concrete objects and a task similar to that used by Petursdottir et al. (2005). Two teenagers with mental retardation who could name numerous objects but showed very low rates of requesting behavior participated in this study. They were tested to ensure that they could name a set of items, and were then taught to perform a chain of behaviors that required those items. For example, one student was taught the steps for making instant soup: tear open the package of instant soup, pour it into the bowl, pour hot

water on it, stir the soup, and eat it with a spoon. In the test, the students were told to start the chain of behaviors, but one necessary item was missing. In the instant soup example, the bowl was missing. The students would need to request this item to finish the chain of behaviors. Although each of the students could reliably name each of the missing items, they did not request these items when they came to the point in the task that the item was necessary. This strongly supports the functional independence of naming and requesting repertoires for this population. Following this test, the students were trained to request the needed items until all students reliably requested the necessary item for each chain of behaviors. After the students had mastered this training, they were told to complete the same chain of behaviors again, but a different item was missing. The students were able to request these new missing items having only been trained to name them. These studies (Hall & Sundberg, 1987; Lamarre & Holland, 1985) show that, for individuals with rudimentary language skills, naming and requesting behaviors can develop independently of each other. The Petursdottir et al. study as well as the training phase of the Hall and Sundberg study demonstrate that, with training and practice, naming and requesting can become interdependent – that is, children can be taught the relationship between naming and requesting.

Similar to the outcomes of the Petursdottir et al. (2005) and Hall and Sundberg (1987), others (e.g., Hanley, Iwata, & Lindberg, 1999; Kelley, Shillingsburg, Castro, Addison, & LaRue, 2007) have found that under certain circumstances, training in one type of verbal response produces others without additional training. In fact, Skinner (1957) described several of the circumstances that make transfer from one type of verbal behavior to another more likely. These circumstances include:

1. The request may have been learned in the presence of the target object. For example, a child may see a glass of milk sitting on the table. The child says “milk” because she is thirsty, and an adult hands the milk to the child. This is clearly a request (mand) since the child was thirsty. But, it may help the child learn to name (tact) the milk since the milk was present prior to responding (i.e., the conditions for a tact were present).
2. The consequence for requesting a target object (obtaining object) is similar to the antecedent conditions for naming the object (presence of object). For example, if the child does not see any milk but requests it by saying “milk,” the child obtains the milk. This is also clearly a request (mand), but the consequence (obtaining milk) is similar to the conditions for naming (tact) – presence of the milk. This may help the child learn to name the milk.
3. A listener may accidentally reinforce a naming response as if it were a request, and vice versa. It is possible that a child, who is not thirsty, may see milk on the table and say “milk” to receive praise. This is clearly a naming (tact) response. But, the adult may misunderstand the child and hand the milk to the child as if the child had requested it (mand). This may help the child to learn to request the milk in the future when she is thirsty.
4. The child may learn an intervening behavior that enables the child to transfer a response from a name to a request. For example, the child may already reliably name milk by saying “milk” as well as request other objects by saying “Please give me...” This child, when thirsty in the presence of milk, may request a behavior of the listener (“Please give me...”) and then name the milk (“milk”).

The entire spoken phrase “Please give me milk” is then reinforced as a request (milk is obtained). The response “milk” would likely be produced in the future as a request due to this history of reinforcement.

While a substantial amount of research has focused on the independence of naming and requesting, little research has explored the relationship between naming and intraverbal behaviors. According to Skinner, early in language acquisition these types of verbal behaviors should also develop independently since they are under the control of different stimuli. As described in Table 1, the tact (naming) is controlled by a nonverbal stimulus (e.g., an object) while the intraverbal is under the control of a verbal stimulus (someone else’s or one’s own verbal behavior). The situation in which a child says “milk” upon seeing a glass of milk is very different from the situation in which the child answers the question “What is cheese made from?” It is plausible that we might learn to name an item without being able to answer a question about the item.

One area of language in which the relationship between naming and intraverbals has been addressed is categorization. Earlier we mentioned intraverbal categorization responses such as saying “cat” when told to “name an animal” and saying “animal” when asked “What is a cat?” If the development of naming and intraverbals is completely independent, then each naming response and intraverbal response must be taught separately. But, if naming and intraverbal repertoires are interdependent, learners could, under some circumstances, make intraverbal responses without direct training. It is important to determine what circumstances may produce interdependence.

Several researchers have assessed children’s ability to produce naming and intraverbal responses. Luciano (1986) investigated functional independence of naming

and intraverbal responses for three teenage students with mental retardation. Stimulus items consisted of eight to twelve objects or pictures for each of several categories (e.g., food, drinks, and clothes). All students, prior to beginning the study, were able to say each item name when shown the picture or object (e.g., bread, shirt). However, the researchers did not confirm whether the students could also produce the category label (e.g., food, clothes) when shown each picture or object. Category labels were neither tested nor trained. The participants were tested on their ability to produce intraverbal responses. The intraverbal behavior was defined as producing the names of items (e.g. bread, cheese, and banana) after a cue such as “Tell me the names of foods.” During baseline, most students produced very low levels of correct intraverbal responses despite being able to name each item. Following baseline, intraverbal responses were trained directly. If the student failed to produce intraverbals during training, a picture prompt was used. Results showed that following training in intraverbal responses, students increased the number of correct directly trained intraverbals produced. This study shows that the ability to name the items alone was insufficient to result in high levels of correct intraverbal responses. However, since it is unknown whether the participants could produce category labels when presented with pictures or objects, students may have lacked logically necessary information to make the intraverbal responses. Thus, this study was not a strong test of independence of naming (tacting) and intraverbals.

Partington and Bailey (1993) replicated and extended Luciano’s (1986) work. In Experiment 1, they directly replicated Luciano’s methods and obtained similar results. In Experiment 2, however, Partington and Bailey repeated the methods of Experiment 1, and added instruction in category labels. Four typically developing 4-year-old children

who did not participate in Experiment 1 learned item names and category labels for five pictures in each of four categories. For each item, the children were first trained to say the item name (e.g., “apple”). After the correct name was produced for each item, the experimenter provided a prompt for the child to say the category label (e.g., “and it’s a fruit”). This prompt was then faded so that the child was producing both the item name and category label as a single response upon seeing the picture (e.g., “apple and it’s a fruit”). Intraverbal responses were tested before and after this training. Training in both item names and category labels resulted in a noticeable increase in production of untrained intraverbal responses for two of the four participants in some categories. Compared to baseline, these students produced from one to three additional intraverbals out of five possible responses. This result should be interpreted with caution, however, because the item name and category label were taught as a single response. The juxtaposition of names and category labels within a single response could have facilitated the production of intraverbals. The intraverbals were then directly trained and all children improved in trained intraverbal responses. Intraverbal responses were not directly trained in one category (furniture) in order to measure generalization to untrained intraverbals. Compared to responses emitted prior to the introduction of training on other categories, these children produced from one to three additional intraverbals out of five possible responses. This experiment shows that when preschool-age children are taught both category labels and item names, they may produce some untrained intraverbal responses based on those names and categories. Furthermore, the study suggested that direct training of intraverbals in several categories can improve untaught intraverbals in other categories. That is, intraverbal skills generalized across categories. These findings



demonstrate that it is crucial to ensure that participants have mastery of all relevant labels (in this case item names and categories) before testing for related intraverbals.

While the two previously described studies provided information on the sufficiency of training item names and category labels in producing untrained intraverbals, a third study (Petursdottir, Carr, Lechago, & Almason, 2008) added another variable to the analysis. This study brought the modality of responding (expressive or listener) into question. Training of items can involve expressive responses (i.e., the child says the name) or listener responses (i.e., the child points to a named object). Children between the ages of three and four were taught to produce item names and category labels for arbitrary and unfamiliar stimuli - outline shapes of several African countries and characters from non-Roman alphabets. Arbitrary and unfamiliar stimuli were used to prevent a child's history with a particular item from either enhancing or interfering with learning. The outlines of countries were divided into four sets of two with category labels of North and South or East and West. The non-Roman alphabet characters were also divided into four sets of two with category labels of Greek and Cyrillic or Kata and Hira. Children were taught to name items as in previously described studies (e.g., "What is this?"), which is an expressive response. Unlike previously described studies, a listener response was also taught for item names. This took the form of selecting the picture from an array of three after hearing the item name (e.g., "Which one is Sudan?"). The child was required to consistently produce both the naming and selecting responses to master this part of the training. At this point, category labels had not been taught. In this study, the intraverbal response is different than those described in previous studies. The intraverbal was defined as saying the category label when presented with a name (e.g.,

“Sudan is ...”). Following training in item names, no intraverbal responses were produced. Next, the children were trained in category labels. In previously described studies the category label was trained as an expressive response of saying the category label when the picture is presented. However, in this study, the category label was only trained as listener responding. When the appropriate cue was provided (“Which one is North?”), the child would select a picture from an array of three. Following this training, intraverbals were again tested. Two of the three children showed no increases in intraverbals, and the remaining child showed only very small increases in intraverbal responses. This study suggests that when the item name is taught by both saying the name and selecting the picture, and the category label is taught only by selecting the picture, this is insufficient to produce substantial untrained intraverbal responding in preschool children. Compared to Experiment 2 of Partington and Bailey (1993), in which both item names and category labels were trained expressively, this study showed a smaller increase in intraverbal responses. This may mean that teaching the category label as an expressive skill (i.e., saying the label when shown a picture) is necessary to produce substantial increases in intraverbal responding in this population.

These three studies give us only preliminary and tentative evidence about whether intraverbals will emerge untrained following training in item names and category labels for preschool-age children. A summary of studies can be found in Table 2. If naming and intraverbal repertoires develop independently we would expect to see that training in item names and category labels is insufficient to result in untrained intraverbal responses. But, if these repertoires are inherently interdependent or have become interdependent by the time children are preschool age, we would expect to see this type of training result in

increases in untrained intraverbal responses. These studies suggest that the degree of independence of intraverbal and naming responses may depend on the type of name and category label training provided. Expressive training (saying the item names and labels) may produce some untrained intraverbals while listener training (selecting the items when presented with item names or category labels) may not result in untrained intraverbal responses.

Miguel, Petursdottir, and Carr (2005) studied the effects of Listener training, Expressive training, and intraverbal training on intraverbal responses in typically developing children between the ages of three and five. Ten pictures of objects from each of two categories were trained for each child. Categories included kitchen items, musical instruments, and tools. In the Expressive training, subjects learned to say the item name (e.g. “hammer”) when shown a picture. They also learned to say the category label (e.g. “tool”) when shown the picture. The Listener training required the children to point to the picture of an item when presented with the item name or the category label. Three children received the Expressive training and then the Listener training. The other three children received Listener training first followed by Expressive training. All six children participated in intraverbal training last. Intraverbal responses were tested before and after each method of training. In this study, children made intraverbal responses by saying the names of items when given a cue such as “What are some tools?” When Expressive training was implemented first, it resulted in small increases in intraverbal responses. One participant produced only one intraverbal response following Expressive training while two participants produced as many as four responses. When Listener training was implemented first, it also resulted in small increases in intraverbal responses. One

Table 2

*Comparison of Study Characteristics*

Study	Participants	Stimuli	Skills trained (training type)	Transfer to intraverbals	Generalization across categories
Luciano (1986)	Teenagers with mental retardation	Familiar	Names (Expressive)	No	Not tested
Partington & Bailey (1993) Exp. 1	Typically developing preschoolers	Familiar	Names (Expressive)	No	No
Partington & Bailey (1993) Exp. 2	Typically developing preschoolers	Familiar	Names (Expressive) Category (Expressive)	1 – 3 out of 5 responses	Yes
Petursdottir et al. (2008)	Typically developing preschoolers	Unfamiliar	Names (Both) Category (Listener)	No	Not tested
Miguel et al. (2005)	Typically developing preschoolers	Familiar	Names (Both) Category (Both)	0 – 7 out of 10 responses	Yes

participant produced no untrained intraverbal responses following Listener training while two participants produced between five and seven responses. Whichever form of training was implemented second resulted in no improvement in intraverbal responses over that achieved in the first training. This finding seems unusual but may be explained by methodological features that will be described later. The final intervention in which the intraverbal was directly trained resulted in substantial increases in the trained intraverbal responding. This study shows that Expressive (naming) and Listener (selecting) trainings both resulted in small increases in intraverbal responses. The authors conclude that these findings support functional independence of naming and intraverbal repertoires. However, several aspects of this study may limit confident conclusions.

The pattern of responding obtained on intraverbal tests complicates the results of the Miguel et al. (2005) study. First, following training in either expressive or listener responses, some of the participants showed an initial increase in untrained intraverbal responses followed by a quick decline in correct intraverbals. This was also the case when children received direct training in intraverbal responses. A possible explanation for this pattern of responding may be found in the reinforcement schedule used during training and intraverbal testing. During all three forms of training, every correct response was followed by praise. But in intraverbal testing that followed training, no praise was provided for correct responding. It is possible that the pattern of responding during intraverbal testing (initial increases followed by decline in correct responding) was due to this abrupt change in the amount of praise provided in training and in testing. The absence of reinforcement (praise) during testing may have resulted in extinction of the

intraverbal response. It is possible that levels of intraverbal responding would have continued to increase if such responses were reinforced.

Another aspect of the Miguel et al. (2005) study that adds complication to drawing confident conclusions is the presence of generalization to untrained categories. For four out of the five children who received direct training in intraverbal responses, a pattern of responding emerged. For each child, intraverbal testing occurred on two categories at the same time, but intraverbal training began at different times for each category. When the training began on the first category, increases in intraverbal responses were shown in both the trained category as well as the untrained category. This could be explained in two ways. First, this could signal that the child was already capable of producing intraverbal responses prior to direct intraverbal training on the second category but was not sufficiently motivated until praise was encountered for intraverbal responding on the first category. Or the child may not have understood the task until correct responding was praised on the first category. On the other hand, this may suggest that the child learned the underlying relationships between the naming and intraverbal responses through training on the first category, and thus could infer the correct intraverbal response in the second category without direct training.

In combination, the four studies discussed show that in some situations, training in item names and category labels is sufficient to produce untrained intraverbal responses for some preschool-age children. Evidence from Partington and Bailey (1993) and Miguel et al. (2005) may help us begin to understand the circumstances necessary to produce untrained intraverbal responses. Refer to Table 2 for a comparison of these studies. In both of these studies some untrained intraverbals do emerge following training

in item names and category labels. In Partington and Bailey the naming responses were taught expressively (saying item names and category labels). In Miguel et al. the naming responses were taught both expressively and as listener responses (selecting pictures when presented with item names and category labels). When the results from these two studies are combined, it looks as if the expressive form of responding in both item names and category labels supports untrained intraverbal responses. This finding is supported by the Petursdottir et al. study in which category labels were trained only as listener responding and untrained intraverbals were not reliably produced. The Partington and Bailey and the Miguel et al. studies also showed evidence that untrained intraverbal responses increased in later categories following direct intraverbal training in earlier categories. This suggests that the children may have learned the underlying relationship between naming responses and intraverbal responses. A second behaviorally based theory of language may shed some light on how this underlying relationship was learned.

### Relational Frame Theory

It will be useful to reinterpret these studies from the perspective of relational frame theory (RFT); but before offering this analysis, some background on RFT is needed. RFT is a behavior analytic approach to language and cognition (Hayes, Barnes-Holmes, & Roche, 2001). It is based on the premise that language is symbolic, relational, and generative. For our present purposes, the relational feature of language is most relevant. RFT argues that because words relate to things, language is necessarily relational. The relation between words and things is one of identity – the simplest of relations. However, there are many other ways that words, things, actions, and properties

can be related. Some of the more complex relations are conditional (if-then), causal, and hierarchal. Each kind of relation entails different implications for the items that are related. For example, classification is one type of a hierarchical relation. In classification, all of the characteristics of the class are implied for all members of that class. If we know that a barracuda is a fish, and we have learned to respond according to hierarchical relations, then all of our responses to fish come to be related to barracuda. We can infer that it lives in water and breathes with gills even if no one has told us this information directly.

On a more specific level, according to RFT, there are three defining features of relational responding: (a) mutual entailment, (b) combinatorial entailment, and (c) transformation of stimulus functions (Hayes et al., 2001), illustrated in Table 3. Mutual entailment is the ability to respond to relations bidirectionally. For example, if you have learned the faster/slower relation and I told you that Mike runs slower than Adam, you could derive that Adam runs faster than Mike. You are able to respond to Adam as the faster runner even though you were not directly taught this response; you are responding with mutual entailment. Combinatorial entailment is the ability to combine two or more taught relations to derive a third relation. For example, I can tell you that Mike runs slower than Adam and Bobby runs slower than Mike. I did not tell you the relation between Bobby and Adam, but by combining the taught responses, you can derive that Bobby is also slower than Adam. Finally, transformation of stimulus function occurs when a psychological function is transformed through relations. Psychological functions include respondent functions (e.g. aversion, preference, fear, anxiety, etc.) or operant functions (discriminative, reinforcing, etc.). For example, the relations described above



Table 3

*Aspects of Relational Responding*

Relational response	Example	Illustration
Mutual entailment	Taught: Mike runs slower than Adam  Derived: Adam runs faster than Mike	
Combinatorial entailment	Taught: Mike runs slower than Adam and Bobby runs slower than Mike  Derived: Bobby runs slower than Adam	
Transformation of stimulus function	Context: Bet who will lose a race.  Discriminative Stimulus: Bobby  Response: Bet on Bobby  Consequence: You win	

*Note.* Solid lines denote taught relations. Dashed lines denote derived relations.

could transform Bobby into a discriminative stimulus for betting he will lose a race. From an RFT perspective, language competence implies the ability to respond based on mutual entailment, combinatorial entailment, and transformation of stimulus function.

Mutual entailment and combinatorial entailment are the aspects of language most relevant to the relations between naming and intraverbal repertoires. First, mutual entailment may be performed between the picture and the item name or category label. For example, as illustrated in Figure 1, children are taught to respond in one direction such as responding to the word by pointing to the object (solid line). If the child has the skill of mutual entailment for naming relations, she will be able to say the name of the object (dashed line) without additional training. In the same way, if a child was directly taught to say the item name upon seeing the picture, then the mutually entailed response would be to then point to the picture upon hearing the name without direct training. In the studies described above, untrained intraverbal responses are possible only through combinatorial entailment. The children must combine trained responses to item names and category labels to derive the intraverbal response. This is illustrated in Figure 2. If the child (a) can name (tact) a “cat,” (b) can respond to the same object with the category label “animal” (another tact), and (c) has the combinatorial entailment skill for category relations, then s/he should be able to say “cat” when told to “name an animal” and to say “animal” when asked “What is a cat?” even though these responses were not directly trained (dashed lines).

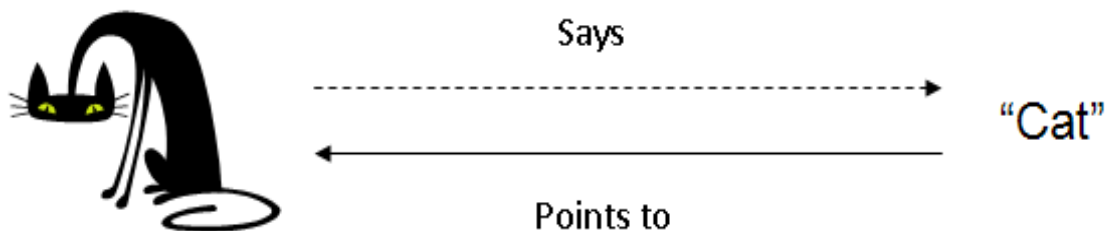


Figure 1. Mutual entailment between picture and item name.

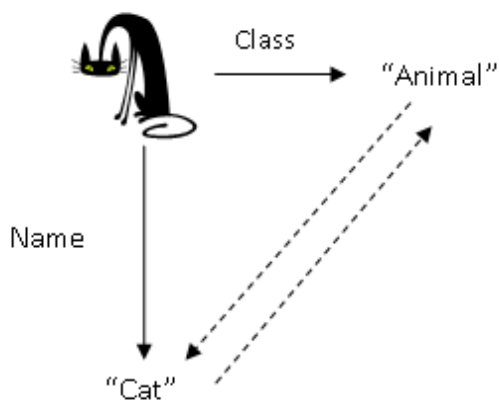


Figure 2. Combinatorial entailment among picture, item name, and category label.

According to RFT, mutual entailment and combinatorial entailment are learned through exposure to multiple exemplars, and this learning is an important part of early language development. Initially, young children would be expected to lack the skills of mutual and combinatorial entailment, but through the process of learning many sets of relations in which these entailment relations hold, they begin to respond according to these patterns. During learning of initial categorical relations, all three relations must be learned independently (item name, category label, and intraverbal). However, after the child has learned multiple categorical relations, new intraverbal responses should emerge following only item name and category label training.

When comparing Skinner's analysis of language and RFT, it is notable that although the two theories approach language from different perspectives, they are not necessarily incompatible. Skinner claimed that types of language like naming and intraverbals develop independently of each other. A child may be able to say a name in response to an object, but not produce that same name in response to a question about the object. RFT does not contradict Skinner's assertion; RFT also assumes that the naming and intraverbal repertoires initially develop independently. However, RFT is much more explicit than Skinner about how the two types of responses become interdependent. RFT claims that the child learns the general relationship between naming and responding to various kinds of intraverbal questions through exposure to multiple exemplars. Therefore, at some point in language development, naming and intraverbal repertoires should become interdependent. Once this interdependence is established, learning to name and give the category label for objects should produce untrained intraverbal responses. Given this background on RFT, we can reinterpret the previously described studies from an RFT perspective.

#### Reinterpreting the Literature Based on RFT

In the Luciano (1986) and Partington and Bailey (1993) studies, intraverbal behavior is first assessed after training in item names, but without systematic training or assessment of category labels. Untaught intraverbals did not emerge under these conditions. This would be expected from an RFT perspective. Combinatorial entailment requires the ability to make all of the logically necessary responses, in this case, naming the object and its category. In Experiment 2, Partington and Bailey taught both item

names and category labels. According to RFT, this should be sufficient to produce untrained intraverbal responses, via combinatorial entailment, in children who have this combinatorial skill. Whether a given preschooler has this skill is an empirical matter that cannot be predicted based on the theory. In fact, untrained intraverbal responses did increase somewhat. Therefore, the findings from these studies are consistent with the RFT perspective of how this type of language skill develops. The incomplete transfer from naming to intraverbal responses could be attributed to the preschoolers' incomplete combinatorial entailment skill with categorical relations or to other factors. Table 4 contains a comparison of study aspects relevant to RFT.

The studies by Petursdottir et al. (2008) and Miguel et al. (2005) both highlight the relationship between the ability to produce mutually entailed responses (produce expressive response after Listener training, and vice versa) and the ability to derive intraverbal responses (combinatorial entailment). In Petursdottir et al., item names are taught as both expressive and listener responses, but only the listener response (selecting the correct picture when the category label is presented) was trained for category labels. A child with strong mutual entailment skills would be able to derive the expressive response for category labels. The three children were tested at the end of the study to determine if they had, in fact, derived the expressive response for category labels. Only one of the three children was able to produce any correct expressive responses (i.e., respond with mutual entailment). Of the three children, he was also the only child to produce any untrained intraverbal responses. From an RFT perspective, this may suggest

Table 4

*Comparison of Study Characteristics using RFT*

Study	Trained two relations?	Mutual entailment?	Transfer to intraverbals (combinatorial entailment)
Luciano (1986)	No	Not tested	No transfer
Partington & Bailey (1993) Exp. 1	No	Not tested	No transfer
Partington & Bailey (1993) Exp. 2	Yes	Not tested	Some transfer
Petursdottir et al. (2008)	Yes	Some	Minimal transfer
Miguel et al. (2005)	Yes	Not tested	Some transfer

that the ability to derive the mutual entailment response for category labels is related to the ability to produce the combinatorial response - the untrained intraverbal.

Miguel et al. (2005) also approached the relationship between mutual and combinatorial entailment by teaching naming either as listener responses (selecting) or as expressive responses (naming). However, in this study, every child received both methods of training sequentially. Some children began with training in both item names and category labels using only listener responses. Some children began with training in only expressive responses. Some increases were evident in untrained intraverbal

responses under each training condition when it was implemented first. Unfortunately, no attempt was made to test whether each child could derive mutually entailed responses. That is, we do not know whether those who first learned listener responses could derive expressive responses prior to receiving training on expressive responses, and vice versa. Therefore, no conclusions about the presence of a relationship between mutual entailment and untrained intraverbals (combinatorial entailment) can be made based on this study.

To date, little research has been conducted to determine whether the ability to perform mutual entailment is related to the ability to perform combinatorial entailment. It is important to determine whether these two repertoires are related. According to Hayes et al. (2001, p. 31) “it seems highly likely that combinatorial entailment usually emerges slightly later in language training than mutual entailment due to its complexity and training history. In principle, however, combinatorial entailment need not be linearly related to mutual entailment.” Research in this area should clarify this relationship and serve as a foundation for developing or enhancing preschool curricula to increase children’s generative use of language. More specifically, research is needed to determine whether mutual entailment is an essential component of learning to produce combinatorially entailed responses such as the untrained intraverbal. An intervention to enhance mutual and combinatorial entailment skills could make instruction more efficient and would allow children to learn more from the environment and daily experiences.

#### Existing Research on Mutual Entailment

While the evidence in the existing literature is inadequate to determine the relationship between mutual entailment and combinatorial entailment, many researchers

have addressed the effects of various training methods on mutual entailment. It has been well established in the literature from various disciplines that names taught as expressive responses (saying names) are likely to transfer to listener responses (selecting the named items) without additional training in many groups of children. Whether names trained as listener responding (i.e., selecting items) transfer to expressive responses without additional training is more variable. For example, Wynn and Smith (2003) taught children with autism between the ages of three and seven names for various objects using either Expressive or Listener training methods. These children showed generalization from Expressive training to listener responses for item names more often than they showed generalization from Listener training to expressive responses. These results are supported by numerous other studies. For example, the relationship between training method and mutual entailment was addressed specifically for category labels in a series of three studies (Horne, Hughes, & Lowe, 2006; Horne, Lowe, & Randle, 2004; Lowe, Horne, & Hughes, 2005). In this series of studies, typically developing children between the ages of one and four learned category labels for sets of objects (wooden blocks with unfamiliar shapes) by either saying the category label when shown the object (Expressive) or pointing to the block when hearing the category label (Listener). The authors consistently found that those children trained in the Expressive method could also produce the corresponding listener responses for category labels. On the other hand, only some of the children trained in the Listener method could produce expressive responses for category labels.

The patterns of transfer between listener and expressive language are also supported in literature from less behavioral disciplines. For example, Goldin-Meadow,



Seligman, and Gelman (1976) tested 2-year-old children to determine how their listener and expressive repertoires compare. Children were tested to determine if they could select objects upon hearing the name and say the name upon seeing the objects. Objects were familiar and represented by toy items. The authors conclude that children fall into two categories of language development. The Receptive group of children responded with listener responses (pointing or selecting) to almost three times as many nouns as they could name expressively when shown the item. In this study the nouns were actual names of objects. The Productive group of children produced expressive responses for nearly every item that they could produce listener responses for. No child failed to produce listener responses for items that they could correctly identify expressively. Although no attempt was made to teach the children new words in one modality and test performance in the other, these findings support those of the behavioral studies. These findings show that some children can produce listener responses to items that they cannot name expressively – that is, they lack the skill of mutual entailment. However, another group of children can produce expressive responses to items they can identify receptively – that is, they have very strong mutual entailment skill.

The literature suggests that for some children there is an asymmetry between (a) mutual entailment from existing expressive responses to inferred listener responses and (b) mutual entailment from existing listener responses to inferred expressive responses. However, these patterns of mutual entailment between listener and expressive responses have not been widely discussed in the RFT literature. With the exception of one study (Lipkens, Hayes, & Hayes, 1993), researchers have not made it a point to confirm that mutual entailment may be performed at different levels of accuracy dependent on the

type of training implemented. In the study by Lipkens et al., a 17-month-old infant was taught to say a name (nonsense word) when shown a picture. He was then tested for mutual entailment (listener responding) by being shown two pictures and asked “Where is the (name)?” The child pointed to the correct picture with few errors. On a different set of pictures, the child was taught to point to one of the two pictures when asked “Where is the (name)?” The child was then shown a single picture and asked “What is this?” He made numerous mistakes in this condition but improved over time with multiple exposures to the stimuli. This case study is an example of the recognition by RFT researchers that the ability to correctly respond with mutual entailment may depend on the direction of the entailed relation. More researchers in this area should investigate this pattern of responding.

It will be important to begin to explore factors that may influence a child’s ability to perform mutually entailed responses. One potential skill that may be related to ability to respond with mutual entailment is memory. It seems logical that a child who is able to retain more information for periods of time will be better equipped to produce an expressive response after learning listener responses or vice versa. Working memory is described as the ability to “hold in mind relatively meaningless information for short periods of time” (Gathercole, 1999, p. 410). One type of working memory, phonological working memory, is often measured by tests such as digit span, recall of words, or nonword repetition. In each of these tests, the child is asked to listen to a series of digits, words, or nonwords. The child is then asked to vocally repeat the sequence heard. As the child progresses through the test, the sequences of digits, words, or nonwords gets progressively longer until the child is no longer able to repeat the sequences accurately

(Gathercole). Research is needed to determine whether some measure of working memory predicts which children produce expressive responses following listener training, and vice versa, while some children fail to do so.

### Conclusions

In order to build the research foundation needed to develop an intervention to increase preschool children's ability to respond with combinatorial entailment (untrained intraverbals), we must determine how several aspects of language are related to the ability to derive intraverbals. First, we must determine whether each teaching method (Expressive or Listener) is likely to result in the ability to derive mutual entailment and intraverbals. By doing this, we will be able to not only determine the relative effectiveness of each type of training; but more importantly, we will gain better understanding of the degree of asymmetry of mutual entailment for this population of children and learn more about their ability to derive untrained responses based on two training procedures with differing instructional power. This will allow us to begin to explore why some children are able to learn in less than optimal teaching situations. The information gained from these comparisons can then be used to design interventions that will most efficiently increase children's ability to derive untrained responses. Second, we must determine how mutual entailment and derived intraverbals (combinatorial entailment) are related. This will allow us to decide whether an intervention aimed at improving ability to derive intraverbals should include a component to enhance mutual entailment abilities. If it is found that mutual entailment is related to combinatorial entailment ability, then it may be important to continue to explore the exact nature of this

relationship and identify factors that influence a child's ability to perform mutual entailment in each training method.

Therefore, the purpose of the current study is to clarify the relationships among training method, mutual entailment responses, and untrained intraverbal responses. The relationship between these variables and working memory will be evaluated as a secondary analysis. The results from the current study will set the stage for future intervention studies to determine how to improve a child's ability to derive untrained responses. Research questions concerning these relationships follow:

1. When taught expressive responses for item names and category labels, how does preschool children's performance compare to chance on:
  - a. Listener responses (mutual entailment)?
  - b. Untrained intraverbal responses (combinatorial entailment)?
2. When taught listener responses for item names and category labels, how does preschool children's performance compare to chance on:
  - a. Expressive responses (mutual entailment)?
  - b. Untrained intraverbal responses (combinatorial entailment)?
3. How do children receiving each training method compare on their ability to:
  - a. Perform mutual entailment responses?
  - b. Perform untrained intraverbal responses?
4. When preschool children are taught either expressive or listener responses to item names and category labels, what is the nature of the relationship between the ability to perform mutual entailment responses and the ability to perform untrained intraverbal responses?

- a. How is the ability to produce mutually entailed responses related to the ability to produce untrained intraverbal responses?
  - b. How does this relationship differ for each training method?
5. In what ways does working memory relate to preschool children's performance on:
  - a. Mutual entailment
  - b. Untrained intraverbals

## CHAPTER II

### METHOD

The five research questions are addressed using a combination experimental and correlational design. Training method is assigned as an experimental variable and correlations among multiple language outcomes are analyzed. This design is particularly appropriate for this study since there is little research in this areas and this study is intended to be exploratory in nature (Gall, Gall, & Borg, 2003).

We worked with children participating in a local Head Start program. We divided participants into two groups – Listener and Expressive training – through stratified random assignment. Both groups learned names and categories for a set of unfamiliar cartoon characters through their assigned training method. Following training, we assessed all participants for mutual entailment, untrained intraverbals, delayed mastery, and working memory. Results allow us to test: (a) the effectiveness of Expressive training for each dependent measure, (b) the effectiveness of Listener training for each dependent measure, (c) how participants in the Expressive training group perform in comparison to the Listener group on each dependent measure, (d) how mutual entailment is related to untrained intraverbal responses, and (e) how working memory is related to mutual and intraverbal responding.

#### Participants and Setting

Preschool children between the ages of 54 and 67 months ( $M = 60.80$ ,  $SD = 3.70$ ) attending a Head Start program participated in this study. All parents or guardians

provided informed consent prior to children's participation. A total of 87 children participated in this study. Of those participants, 68 completed the procedures within the required two instructional sessions. Specific demographic information concerning age, ELL status, and rates of training completion for each training method will be analyzed in the Results section.

We implemented experimental procedures in a quiet corner of the classroom with a table and two chairs or outside the classroom with the experimenter and child seated on a bench just outside the classroom door. All sessions took place during a self-selected activity time in which children had free choice of activities within their preschool classroom.

One doctoral student at Utah State University served as the primary experimenter. This investigator holds a Texas K-12 Special Education teaching certificate and has experience providing behavioral evaluations and consultation for schools through a local pediatric psychology clinic at the Center for Persons with Disabilities.

### Materials

Intervention materials consisted of four picture cards portraying cartoon characters (i.e. four different faces). Each picture consisted of a black outline with one additional color filling the body of the character. Colors were confined to blue, red, yellow, and green so that they were easily discriminable. Pictures are shown in Table 5. Pictures were 3 x 3 inches in size and printed on a sheet of paper (landscape) in a 2 x 2 array (see Appendix A for an example). We used all possible combinations of character locations within the 2 x 2 array in training and testing. We assigned a nonsense word

name for each picture (e.g. Kip). No physical feature of the picture was predictive of item names. We assigned category labels in the form of familiar actions for each character (e.g. likes to sing/dance). We divided the four pictures in a set into two categories, each with two members. For example, Kip and Bo like to sing and Lou and May like to dance. No physical features were predictive of category membership.

Materials for conducting a test of working memory included a small speaker which was connected to an MP3 Player. We played stimuli for working memory tests, spoken digits, from an electronic file uploaded to the MP3 Player.




### Procedures

Children were divided into groups receiving either Expressive or Listener training. We stratified groups based on English Language Learning (ELL) status and age. ELLs included children whose parents reported speaking only a language other than English in the home. Primary English speakers (PES) included children whose parents reported speaking only English or English along with another language in the home. We first divided children into groups of ELLs and PESs and then further divided these groups into cells based on age. Within each of these cells, we randomly assigned children to training groups, with the requirement that the number of children completing each type of training remain equal. If a child within a cell failed to complete training, then we assigned an additional child to that same training condition to keep the number of children completing each training type equal within each cell.



Table 5

*Examples of Pictures, Names, and Categories*

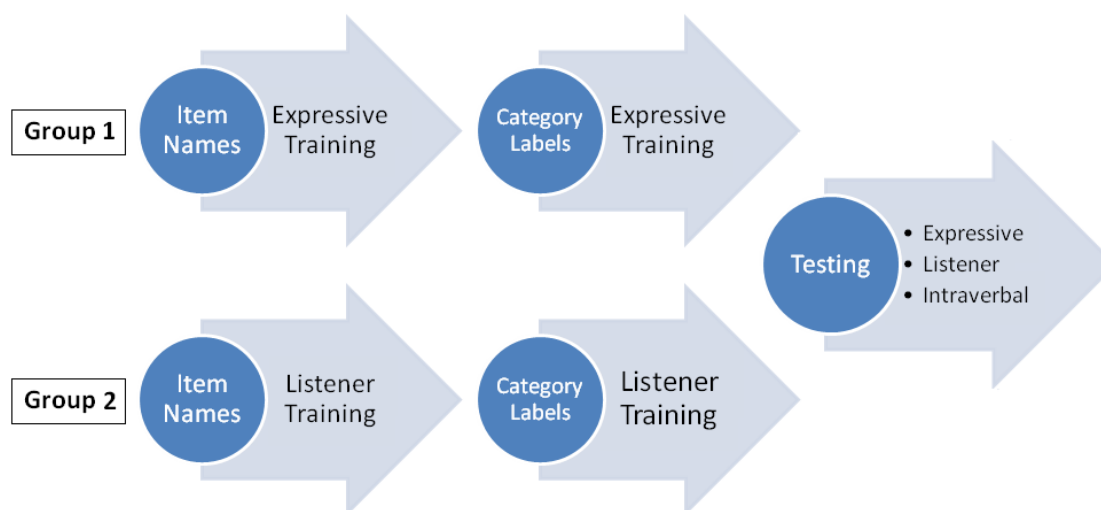
Picture	Item name	Category label
	Kip	Likes to sing
	Lou	Likes to dance
	Bo	Likes to sing
	May	Likes to dance

This experiment took place over the course of twenty-six school days. Each child received training and testing individually. Most participants required a single 10-20 minute session to complete training and testing (excluding the working memory test). If a child failed to achieve mastery of the directly taught material within the 20 minute session, we scheduled a second session to complete training and testing. This second attempt took place an average of 3 days (range 0 – 19) following the first attempt. If the child failed to complete training and testing within two sessions, training was discontinued. Children who failed to complete training in two sessions were administered the working memory test, but not tests of mutual entailment and intraverbals. Participants completed working memory tests in a later session following completion of all other training and testing procedures. The length of time between completion of training and administration of the working memory test was an average of 13 days (range 0 – 48).

Figure 3 shows the basic sequence of training and testing within a single session. Each child received either Expressive or Listener training on item names and category labels for a single set of four pictures. Following training on names and categories, participants completed tests of expressive responses, listener responses, and derived intraverbal responses.

#### *Item Name Training*

All four pictures were visible to the child throughout the training. We taught item names by pointing to a picture and telling the child the name that goes with that picture



*Figure 3.* Sequences of training and testing.

(e.g. “This is Kip”). We then asked the child to respond. In the Listener training, we asked the child to select the picture with a particular name (e.g. “Point to Kip”). In the Expressive training, we pointed to the picture and asked the child to say the name (e.g. “Who is this?”).

Following a correct response, we provided verbal praise, presented a different array of the four pictures and began the next trial. We corrected incorrect responses by providing a model (e.g. “This is Kip”) and repeated the original question (e.g. “Who is this?”) to obtain a correct response.

The child was required to correctly name or select the first picture three consecutive times before we introduced the name for the second picture. Responses immediately following models (either the first trial for each picture or the first trial after an incorrect response) counted toward this requirement. The final two consecutive responses could not immediately follow models. The second picture was introduced by

modeling and we again asked the child to name or select the picture. This time, trials on the first and second pictures were mixed in random order. We required the child to correctly name both the first and second pictures three consecutive times to move on. This pattern of training and testing was carried out with the third and fourth pictures. We modeled each item name and required the child to correctly respond to the newly trained item along with all previously trained items. Figure 4 illustrates this process. Order of trials (i.e., which picture is tested) was randomized within each set. Children met mastery criteria (described below) for item name training before proceeding to category label training.

#### *Category Label Training*

Category label training was similar to name training. We pointed to a picture, said the category label (e.g. “This one likes to sing”), and then asked the child to respond either as a listener (“Point to one who likes to sing”) or expressively (“What does this one like to do?”). We required the child to perform three consecutive correct responses before being trained on the next picture. When categories for each picture had been introduced, the child responded to randomly-ordered trials on all category labels that had been trained. It should be noted that once two pictures have been trained for a single category, the response instruction changes slightly in the Listener training group. We asked the child to point to all the pictures with a category label (“Point to the ones who like to sing”). If the child pointed to two pictures, both responses were recorded. If the child only pointed to one picture, the experimenter prompted the child to point to a second picture by instructing the child to “Point to the other one who likes to sing.” If the child

pointed to the correct picture following the prompt, this response was recorded as correct. The sequence of training for category labels is illustrated in Figure 5. We corrected errors in the same way as described for item name training.

### *Mastery Criteria*

The mastery criterion for training was three consecutive correct responses on each of the four pictures in the mixed set. The first of the three responses could have been in response to a model prompt (following introduction of a new item or an incorrect answer). The final two consecutive responses must have been unprompted correct responses. The shaded test boxes in Figures 4 and 5 indicate mastery tests. If the child produced an incorrect response during mastery testing, we corrected the error and progress toward the mastery criterion was restarted. Testing continued until mastery was reached.

### Dependent Variables

Once the children reached mastery criteria for training, they immediately completed three tests. All children completed tests in the same order. First, the children completed the test of mutual entailment – an Expressive Test for the Listener group or Listener Test for the Expressive group. Next, the children completed the Intraverbal Test. And finally, they completed the delayed mastery test – an Expressive test for the Expressive group or Listener Test for the Listener group. During a later session, each child also completed a test of working memory. For each picture, we asked the child to do the following:

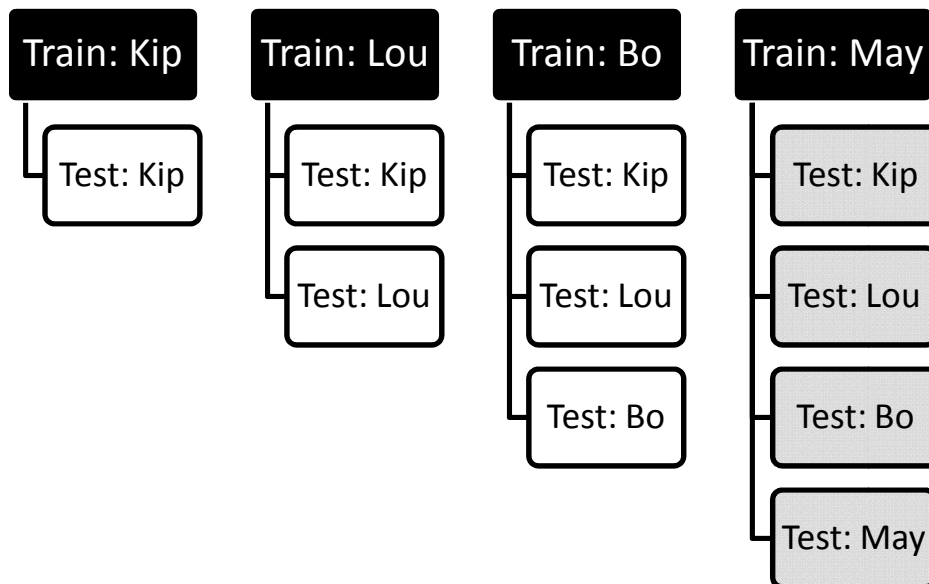


Figure 4. Training and testing for item names. Shaded items indicate mastery tests.

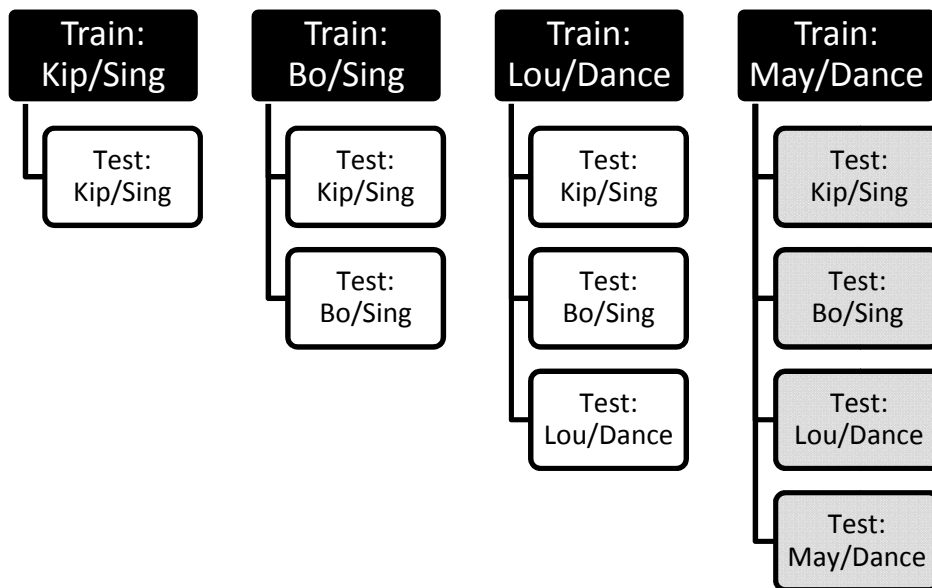


Figure 5. Training and testing for category labels. Shaded items indicate mastery tests.

1. Expressive Test – We pointed to each picture and asked, “Who is this?” The child said the name of the picture. If the child said more than one name, we asked the child to choose one answer (“Which one do you choose?”). Approximations of names were accepted if they were clearly distinguishable from other item names. We also asked the child to state the category of each picture (“What does this one like to do?”). The child said the picture’s category label. A total of eight responses were required (names and categories of each of four pictures). These responses were taught in the Expressive training condition; therefore, for the children in the Expressive condition, this was a delayed mastery test. In the Listener condition, participants were not taught these responses; therefore, this was a test of mutual entailment of item names and category labels.
2. Listener Test – We asked the child to point to a particular picture when given its item name (“Point to Kip”). We also asked the child to point to pictures when given the category (“Point to the ones who like to sing”). If the child only pointed to one picture, we prompted the child to “Point to the other one.” A total of eight responses were required. These responses were taught in the Listener training condition; therefore, for the children in the Listener condition, this was a delayed mastery test. In the Expressive condition, participants were not taught these responses. Therefore, this was a test of mutual entailment of item names and category labels.
3. Intraverbal Test – The child answered questions involving the pictures’ names and category labels with no pictures available. We asked the child to provide the category when given the picture name (“What does Kip like to do?”). We also

asked the child to provide the item name when given the category label (“Who likes to sing?”). If the child gave only one name, s/he was asked, “Who else likes to sing?” Approximations of names were accepted if they were clearly distinguishable from the other item names. A total of eight responses were required. This assessed combinatorial entailment for children in both conditions.

4. During a later session all children also completed a test of working memory – Memory for Digits. This test is a subtest from the *Comprehensive Test of Phonological Processing* (CTOPP) (Wagner, Torgesen, & Rashotte, 1999). During this test, each child listened to a series of digits from an audio file. The series of digits started at two digits in length and increased to eight digits. Once the child listened to a series of digits, we asked the child to repeat the digits in the same order that s/he heard them. We recorded raw scores from this subtest for use during analysis.

During intervention and testing, procedures were used to maintain motivation.

Stamp cards with 12 pictures printed on them were used to reward progress through the intervention (see Appendix B for an example). Prior to each session, we allowed the children to look at three or four small prizes and told them that they could choose one to take home if they earned all of the stamps. We placed a stamp on the card each time the child moved on to training on a new item or reached mastery. We also awarded stamps following completion of each dependent measure regardless of accuracy. Once a child earned all 12 stamps, s/he chose a prize. Children did not earn stamps during the working memory test. Instead, we told the child that s/he could choose a sticker when the test was complete.



## Social Validity

This study does not involve an intervention to be incorporated into a curriculum. Therefore, many of the more conventional measures of social validity such as teacher willingness to continue procedures or teacher rating of benefit to the child would not be appropriate for this study. Instead, we used two somewhat less conventional methods to assess social validity. We assessed importance of the outcomes being measured in this study through a comparison to a typical preschool task involving a categorization skill - matching. The test of matching was administered at the end of the session following completion of tests of mutual entailment, intraverbals, and delayed mastery. In this task, we used the cartoon characters taught during the instructional phase of the procedures. We presented one item (a sample) and asked the child to select the other item (from a set of three) that goes with the sample based on category labels (“Which one likes to do the same thing as this one?”). This was repeated for each of four items. A second measure of social validity compared the types of outcomes evaluated in this study to those deemed important by the Head Start organization. This was accomplished through comparing the outcomes evaluated in this study to the Head Start’s benchmark test (Bear River Head Start, 2008-2009). We reviewed the benchmarks included in the Head Start assessment and selected those related to categorization skills. Then we compared these benchmarks with the outcomes measured in this study to determine if there was alignment between Head Start benchmarks and study outcomes.

## CHAPTER III

### RESULTS

#### Treatment Fidelity and Reliability of Measures

We video recorded all sessions for later observation by an independent observer. Data collection sheets used by both the primary experimenter and the independent observer are located in Appendix C. The independent observer assessed treatment fidelity of the intervention and interobserver agreement (IOA) of the assessments. Treatment fidelity and IOA were measured for both Listener and Expressive training sessions and were distributed evenly throughout the duration of this study.

For treatment fidelity, the independent observer completed a checklist of intervention/assessment steps (see Appendix D) from video recordings for 30% of participants. The mean percentage of intervention/assessment steps followed was 99.9% (range 96% - 100%). Fidelity was measured separately for the test of working memory (see Appendix D for fidelity checklists). The percentage of steps followed for this test was 100% for all participants.

For IOA, the observer independently scored 30% of sessions from video recordings. Point-by-point agreement was calculated for each test. Mean IOA for the test of mutual entailment was 98% (range 75% - 100%). We obtained IOA for mutual entailment below 80% in only one session. In this session, two of the eight responses for the mutual entailment test were inaudible on the video tape. Scores for these two responses were the only disagreements for this test resulting in IOA of 75% for this session. For the intraverbal test, mean IOA was 99% (range 88% - 100%). The IOA for

delayed mastery was 100% for all participants. Mean IOA for the social validity measure, matching, was 99% (range 75% - 100%). In this case, we obtained IOA below 80% in only one session. Only four responses were required in the matching test. Therefore, 75% IOA indicates one disagreement out of the four responses. Finally, for the test of working memory, mean IOA was 99% (range 90% - 100%).

### Descriptive Analysis

Prior to the analysis of results, we present demographic information and highlight notable trends in the demographic data. A total of 87 children were assigned to either Listener or Expressive training groups. Of those, 68 children (78%) completed training with 34 participants completing each type of training. A significantly higher proportion,  $\chi^2(1, N = 87) = 7.11, p = .008, \phi = .286$ , of children completed Listener training (92%) than Expressive training (68%). Although the correlation coefficient between completion and training condition is small, the discrepant proportions of children completing each training protocol are important because additional children were assigned to the Expressive group to ensure equal numbers of participants completing each type of training. This introduces bias since the children who were able to complete Expressive training may be different in some way from those able to complete Listener training. That is, the children in the Expressive group were more highly selected than those in the Listener group; thus, they likely differ in their ability to learn novel names and categories within a short period of time. Therefore, caution should be taken in interpreting comparisons between Listener and Expressive groups.

Also notable are patterns of completion rates among participants who were Primarily English Speakers (PES) and those who were English Language Learners (ELL). Completion rates are reported in Table 6. We calculated chi square statistics to determine whether different proportions of PES and ELL children completed each type of training. It is important to note that one assumption of the chi square statistic is violated in each of these comparisons. Expected frequencies for some cells of the contingency table are less than five. This can increase the likelihood of spurious results for the chi square by adding variability in the obtained  $p$  values. We used the continuity correction in an attempt to prevent spurious effects due to the small expected cell frequencies. However, the continuity correction reduces the power of the chi square statistic. Approximately equal proportions of ELLs and PESs completed Listener training,  $\chi^2(1, N = 37) = .10, p = .747, \phi = .168$ , and Expressive training,  $\chi^2(1, N = 50) = 3.57, p = .059, \phi = .317$ . However, if the continuity correction had not been used, the proportions of PES and ELL students completing the Expressive training would have been significantly different,  $\chi^2(1, N = 50) = 5.03, p = .025, \phi = .317$ , with a higher proportion of PES children completing this training than ELL children.

The ages of children assigned to Expressive and Listener training groups were similar,  $t(85) = .832, p = .408, d = .18$ . Those children who completed each type of training were also very similar in age,  $t(66) = -.099, p = .921, d = -.02$ . Means and standard deviations for these comparisons are reported in Table 7.

Table 6

*Completion Rates of PES and ELL Subgroups*

	Listener		Expressive	
	PES	ELL	PES	ELL
<i>N</i> assigned	28	9	38	12
<i>N</i> completed	25	9	29	5
Pct. Completed	89%	100%	76%	42%

Table 7

*Age of Participants Assigned to and Completing Training*

	<i>N</i>	Mean	<i>SD</i>
Assigned			
Listener	37	61.19	3.68
Expressive	50	60.52	3.73
Completed			
Listener	34	61.03	3.67
Expressive	34	61.12	3.66

Table 8 includes the means and standard deviations for each dependent variable by training group. Performance on mutual entailment was higher on average than performance on intraverbals for both the Listener training group (mutual entailment  $M = 5.15$ , intraverbal  $M = 3.94$ ) and the Expressive training group (mutual entailment  $M = 7.53$ , intraverbal  $M = 4.88$ ). Delayed mastery scores were very similar in Listener ( $M = 7.21$ ) and Expressive ( $M = 7.24$ ) training groups. Statistical comparisons between groups will be reported in a later section.

Table 8

*Mean (SD) of Dependent Variables by Training Group*

	Listener training	Expressive training	Combined
Mutual entailment	5.12 (1.87)	7.53 (1.05)	6.32 (1.94)
Intraverbals	3.94 (2.09)	5.82 (2.25)	4.88 (2.35)
Delayed mastery	7.21 (1.10)	7.24 (.96)	7.22 (1.02)

Dot plot distributions for mutual entailment (Figure 6) and intraverbal results (Figure 7) differ across training groups. Dot plot distributions show that while the data for mutual entailment are roughly normally distributed for the Listener training group, these data are negatively skewed for the Expressive training group. The negative skew is due to a ceiling effect in the Expressive group. The dot plot distributions of intraverbal responses show a similar, but less dramatic, pattern with an approximately normal distribution for the Listener Training group and a ceiling effect for the Expressive



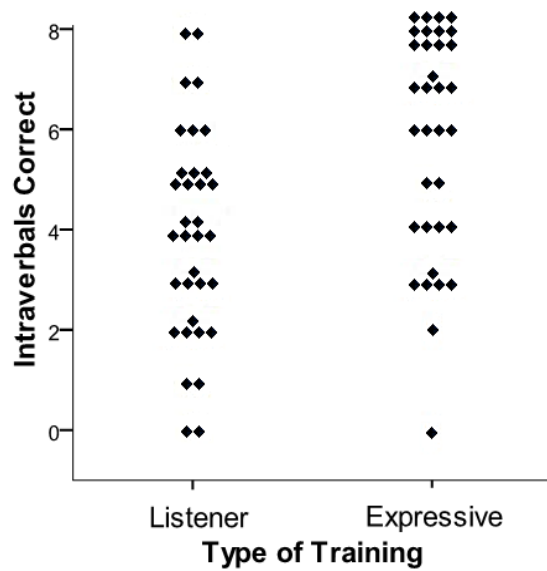


Figure 7. Dot plot distributions for intraverbals.

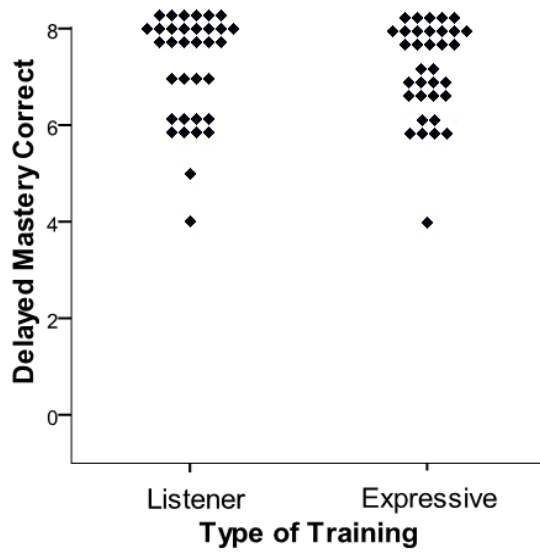


Figure 8. Dot plot distributions for delayed mastery.



We compared performance data for ELL and PES groups to determine whether these groups should be analyzed separately for each research question. Performance on mutual entailment and intraverbals of ELL and PES students is displayed in Figures 9 and 10 and means and standard deviations for each subgroup are reported in Table 9. We calculated effect sizes using Cohen's *d*. Although Cohen's *d* is a parametric measure of effect size, we used it along with visual representations of the effects to describe the magnitude of effect.

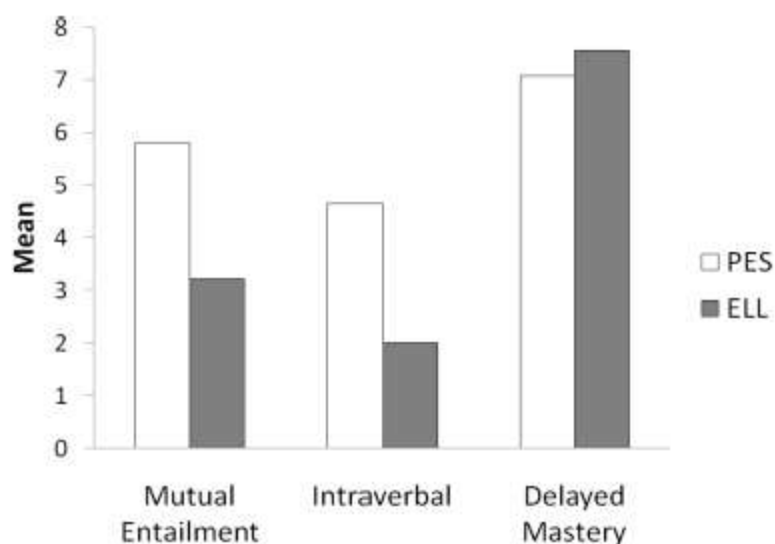


Figure 9. Comparison between ELL and PES means (Listener).

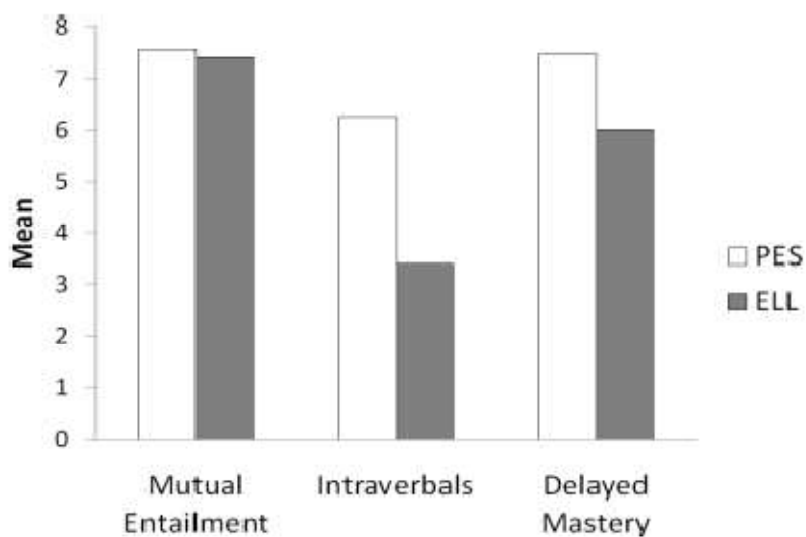


Figure 10. Comparison between ELL and PES means (Expressive).

Table 9

Mean (SD) of Dependent Variables for PES and ELL Subgroups

	PES	ELL
Listener		
Mutual entailment	5.80 (1.29)	3.22 (1.99)
Intraverbals	4.64 (1.80)	2.00 (1.58)
Delayed mastery	7.08 (1.19)	7.56 (.73)
Expressive		
Mutual entailment	7.55 (1.02)	7.40 (1.34)
Intraverbals	6.24 (1.99)	3.40 (2.30)
Delayed mastery	7.45 (.74)	6.00 (1.23)

Within the Listener training group, PES children performed significantly better than ELL children on mutual entailment ( $Z = 1.46, p = .028, d = 1.73$ ) as well as intraverbals ( $Z = 1.67, p = .008, d = 1.51$ ). Effect sizes for these comparisons were large. The ELL children performed slightly better than the PES children on delayed mastery ( $Z = .64, p = .807, d = .44$ ) with a medium effect size. However, this difference was not statistically significant. In the Expressive group, the PES children performed better than the ELL children on mutual entailment ( $Z = .199, p = 1.00, d = .14$ ), intraverbals ( $Z = 1.21, p = .107, d = 1.40$ ), and delayed mastery ( $Z = 1.21, p = .107, d = 1.78$ ). However, these differences also were not statistically significant. While the effect size for mutual entailment was small, the effect size for the intraverbal and delayed mastery comparisons within the Expressive group was large. Since PES and ELL children performed significantly differently on mutual entailment and intraverbals within the Listener group and since effect sizes were large for intraverbal and delayed mastery comparisons within the Expressive group, we analyzed results for each research question for the entire data set with ELLs and PESs combined as well as for PESs and ELLs separately.

We also considered a second subanalysis. Although all children included in analyses demonstrated initial mastery of names and categories prior to testing by responding correctly to names and categories three consecutive times in a mixed set, some participants performed less than perfectly (less than all eight items correct) on the delayed mastery test. When those children who made mistakes on the delayed mastery items were compared to children who answered all eight items correctly there was no statistically significant difference in performance on mutual entailment ( $Z = .52, p = .950, d = .19$ ) or intraverbals ( $Z = 1.11, p = .167, d = .39$ ). Effect sizes for this comparison are

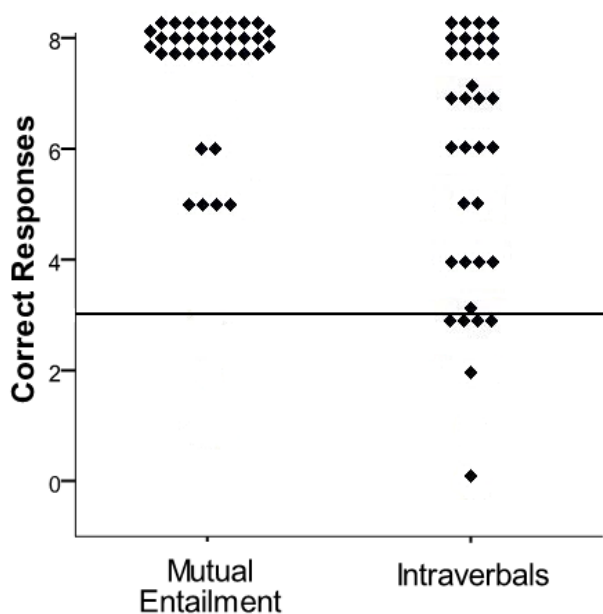
small to medium in strength. Since those who made mistakes on delayed mastery do not perform substantially differently on mutual entailment or intraverbals when compared to those who answered all eight items correctly, a separate subanalysis with data modified to correct for errors due to lack of delayed mastery is not indicated. However, if the reader is interested in the results of a subanalysis modified to account for errors in delayed mastery, this subanalysis for each research question can be found in Appendix F.

Research Question 1 -- When taught Expressive responses how does preschool children's performance compare to chance on mutual entailment and untrained intraverbals?

We report results for this research question as a visual analysis of dot plot distributions in addition to statistical tests. We calculated chance by assuming that the children guessed from the pool of four character names and two categories. Therefore, chance-level performance would average three correct answers on the mutual entailment and untrained intraverbal tests. This is a very conservative estimate of chance (see Appendix G for a detailed description of chance calculation).

The results of this analysis are visually displayed in Figures 11-13. Figure 11 shows the distribution of mutual entailment and intraverbal scores in comparison to chance (indicated by a horizontal line). For the Expressive training group, all 34 participants scored above chance on mutual entailment and 27 of them scored above chance (with 5 scoring at chance level) on intraverbals. We repeated the same analysis for the subgroups of PES (Figure 12) and ELL (Figure 13) children. The results of this subanalysis for PES parallel those in the full analysis of all participants. Within the

expressive training group for PES children, all 29 participants scored above chance on mutual entailment and 25 of them scored above chance (with 3 scoring at chance level and 1 below chance) on intraverbals. For the ELL children, all 5 participants scored above chance on mutual entailment. On intraverbals, 2 children scored above chance (with 2 scoring at chance level and 1 scoring below chance).



*Figure 11.* Comparisons to chance for all participants (Expressive). Chance is indicated by a horizontal line.

Due to the negatively skewed distribution in the Expressive training group, a nonparametric test is most appropriate for statistical comparisons to chance within the Expressive group. We used the Wilcoxon Signed Ranks Test to compare mutual

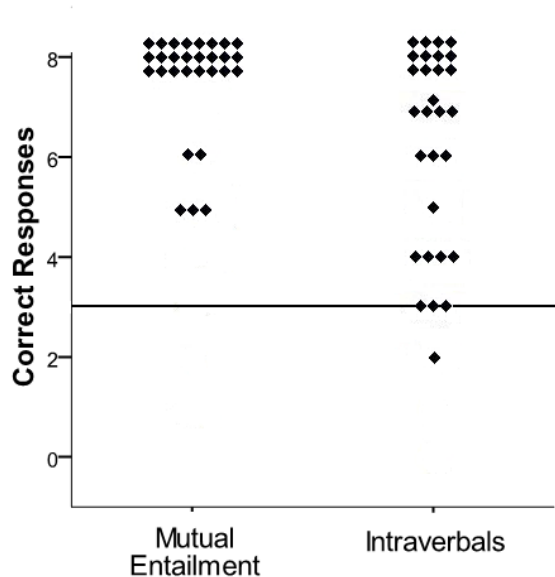


Figure 12. Comparisons to chance for PES (Expressive). Chance is indicated by a horizontal line.

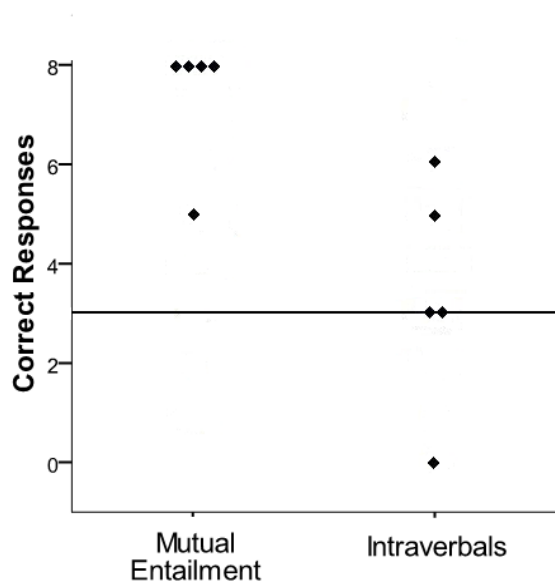


Figure 13. Comparisons to chance for ELL (Expressive). Chance is indicated by a horizontal line.

entailment and untrained intraverbal scores to chance. Results for the Wilcoxon Signed Ranks Test are displayed in Table 10. With the exception of intraverbals for the ELL subgroup, all comparisons were statistically significant with scores better than chance. Similarly, all effect sizes were large (range  $d = 1.25 - 4.31$ ) except for the effect size for intraverbals in the ELL subgroup, which was small ( $d = .17$ )

Table 10

*Wilcoxon Signed Ranks Test Comparisons to Chance for Expressive Training*

		<i>N</i>	Mean rank	Sum of ranks	<i>Z</i>	<i>p</i>	<i>d</i>
All participants							
Mutual entailment	Negative ranks	0	0	0	-5.47	<.001	4.31
	Positive ranks	34	17.5	595			
	Ties	0					
	Total	34					
Intraverbals	Negative ranks	2	6.50	13	-4.47	<.001	1.25
	Positive ranks	27	15.63	422			
	Ties	5					
	Total	34					
PES							
Mutual entailment	Negative ranks	0	0	0	-5.06	<.001	4.46

(table continues)

		<i>N</i>	Mean rank	Sum of ranks	<i>Z</i>	<i>p</i>	<i>d</i>
	Positive ranks	29	15	435			
	Ties	0					
	Total	29					
Intraverbals	Negative ranks	1	3	3	-4.44	<.001	1.63
	Positive ranks	25	2	348			
	Ties	3					
	Total	29					
ELL							
Mutual entailment	Negative ranks	0	0	0	-2.12	.034	3.28
	Positive ranks	5	3	15			
	Ties	0					
	Total	5					
Intraverbals	Negative ranks	1	2.50	2.50	-.27	.785	.17
	Positive ranks	2	1.75	3.50			
	Ties	2					
	Total	5					



Research Question 2 -- When taught Listener responses how does preschool children's performance compare to chance on mutual entailment and untrained intraverbals?

For the second research question, we used the same data analysis methods as in the first research question. Results for all participants are displayed in Figure 14. Similar to the results in the first research question, 28 of 34 participants scored above chance (with 2 scoring at chance level) on mutual entailment and 20 scored above chance (with 5 scoring at chance level and 9 below chance) on intraverbals. When only the PES children are evaluated (Figure 15), the same pattern of results is seen with 23 of 25 participants scoring above chance (with 2 scoring at chance level) on mutual entailment and 19 scoring above chance (with 3 scoring at chance level and 3 below chance) on intraverbals. In the ELL subgroup (Figure 16) five of 9 participants scored above chance on mutual entailment (with 4 below chance), and on intraverbals, one participant scored above chance (with 2 scoring at chance level and 6 below chance).

We used the Wilcoxon Signed Ranks Test to determine whether the Listener training group performed statistically significantly better than chance on mutual entailment and intraverbal variables. Although mutual entailment and intraverbals are approximately normally distributed for the Listener training group, we use this nonparametric test in the interest of consistency and to produce comparable statistical results across analyses. Results for this analysis are displayed in Table 11. All comparisons were statistically significant with scores better than chance except for mutual entailment and intraverbals for the ELL subgroup. Effect sizes for the statistically significant comparisons are medium to large (range  $d = .45 - 2.17$ ). Effect sizes for the

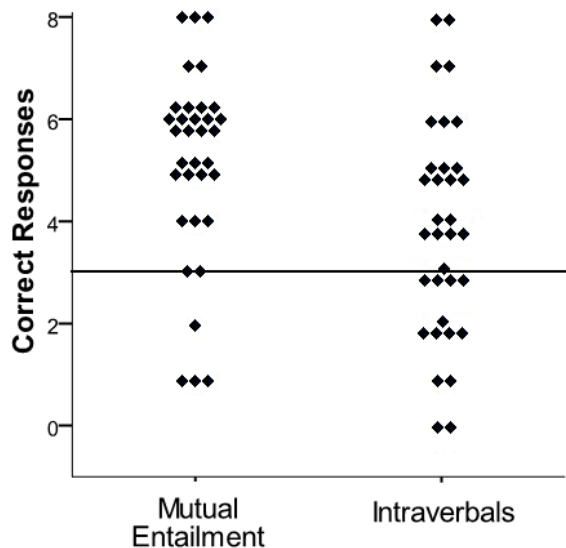


Figure 14. Comparison to chance for all participants (Listener). Chance is indicated by a horizontal line.

ELL subgroup are small for mutual entailment ( $d = .11$ ) and medium in a negative direction for intraverbals ( $d = -.63$ )

Research Question 3 -- How do children receiving each type of training compare on mutual entailment and untrained intraverbals?

In the third research question, we compare the two training groups. This analysis also required a nonparametric statistic since the Expressive training group's scores are not normally distributed. We used the Kolmogorov-Smirnov test to compare groups. However, one assumption of this test is violated. While nonparametric statistical tests do not require normal distributions, both parametric and nonparametric tests assume that the

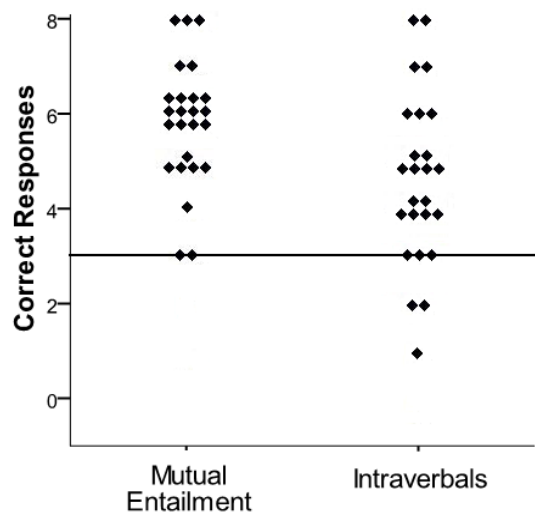


Figure 15. Comparisons to chance for PES (Listener). Chance is indicated by a horizontal line.

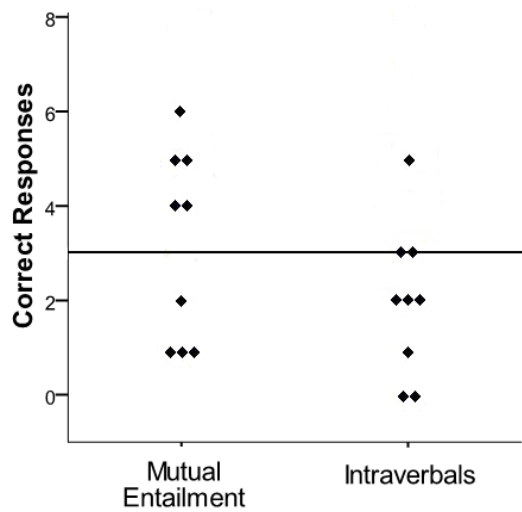


Figure 16. Comparisons to chance for ELL (Listener). Chance is indicated by a horizontal line.

Table 11

*Wilcoxon Signed Ranks Test Comparisons to Chance for Listener Training*

		<i>N</i>	Mean rank	Sum of ranks	<i>Z</i>	<i>p</i>	<i>d</i>
All participants							
Mutual entailment	Negative ranks	4	7.75	31	-4.41	<.001	1.13
	Positive ranks	28	17.75	497			
	Ties	2					
	Total	34					
Intraverbals	Negative ranks	9	12	108	-2.39	.017	.45
	Positive ranks	20	16.35	327			
	Ties	5					
	Total	34					
PES							
Mutual entailment	Negative ranks	0	0	0	-4.28	<.001	2.17
	Positive ranks	23	12	276			
	Ties	2					
	Total	25					
Intraverbals	Negative ranks	3	7	21	-3.46	.001	.91
	Positive ranks	19	12.21	232			

(table continues)

		<i>N</i>	Mean rank	Sum of ranks	<i>Z</i>	<i>p</i>	<i>d</i>
Ties		3					
Total		25					
		ELL					
Mutual entailment	Negative ranks	4	5	20	-.30	.762	.11
	Positive ranks	5	5	25			
	Ties	0					
	Total	9					
Intraverbals	Negative ranks	6	3.92	23.50	-1.62	.105	-.63
	Positive ranks	1	4.50	4.50			
	Ties	2					
	Total	9					

distributions being compared have similar shapes (Blalock, 1972). The distributions of the Listener and Expressive groups for both mutual entailment and intraverbals are not similarly shaped due to the ceiling effect in the Expressive group compared to the approximately normal distribution of the Listener group. Therefore, we place more emphasis on the visual analysis of data than on statistical significance.

The dot plot distribution displayed in Figure 17 shows the differences in performance between groups on mutual entailment and intraverbals for all participants. For mutual entailment, the Expressive group ( $M = 7.53$ ,  $SD = 1.05$ ) performed better than

the Listener group ( $M = 5.12$ ,  $SD = 1.87$ ). The Expressive group ( $M = 5.82$ ,  $SD = 2.25$ ) also performed better than the Listener group ( $M = 3.94$ ,  $SD = 2.09$ ) on intraverbals. Figure 18 shows the results for the PES subgroup, which parallel the overall analysis. For mutual entailment and intraverbals, the Expressive group (mutual entailment  $M = 7.55$ ,  $SD = 1.02$ , intraverbal  $M = 6.24$ ,  $SD = 1.99$ ) performed better than the Listener group (mutual entailment  $M = 5.80$ ,  $SD = 1.29$ , intraverbal  $M = 4.64$ ,  $SD = 1.80$ ). The patterns of comparison for the ELL analysis (Figure 19) also parallel the overall analysis with the Expressive group (mutual entailment  $M = 7.40$ ,  $SD = 1.34$ , intraverbal  $M = 3.40$ ,  $SD = 2.30$ ) performing better than the Listener group (mutual entailment  $M = 3.22$ ,  $SD = 1.99$ , intraverbal  $M = 2.00$ ,  $SD = 1.58$ ).

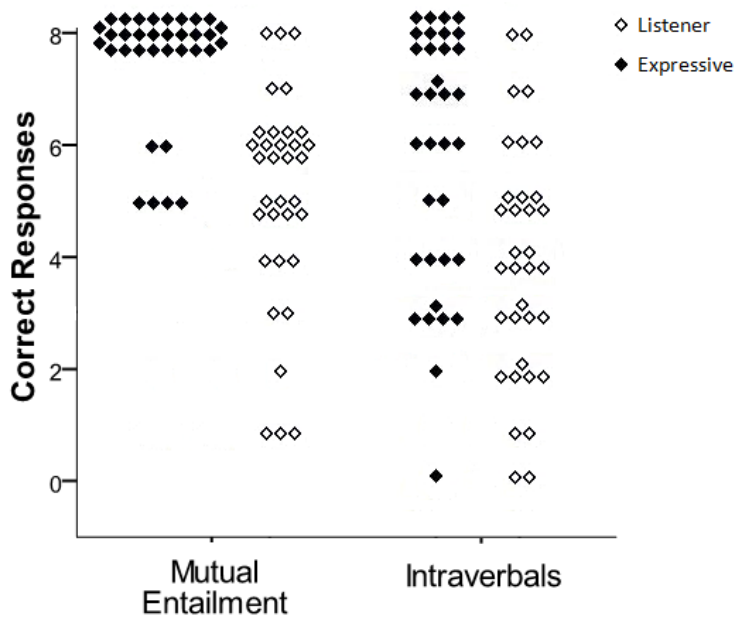


Figure 17. Dot plot distribution of mutual entailment and intraverbals for all participants (Listener and Expressive).

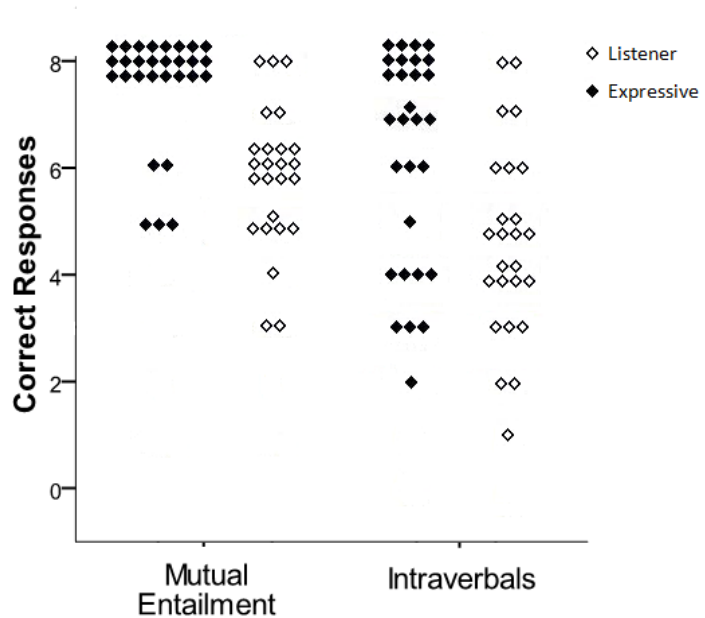


Figure 18. Dot plot distributions of mutual entailment and intraverbals for PES (Listener and Expressive).

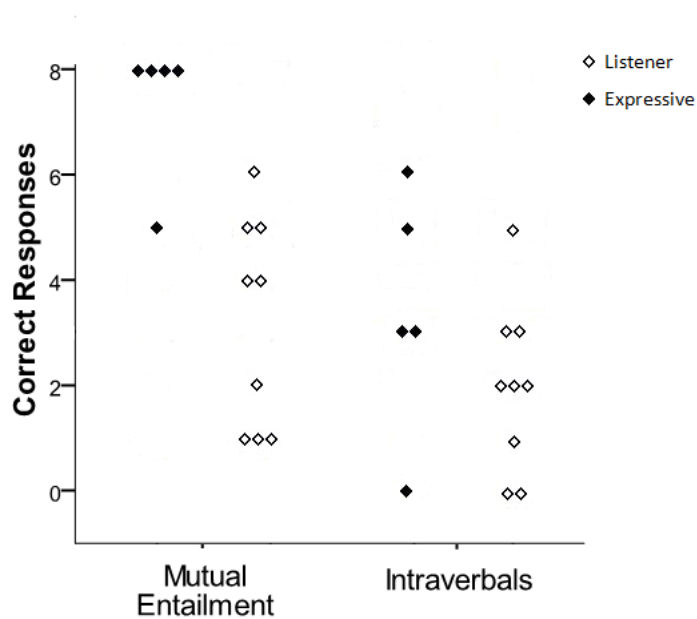


Figure 19. Dot plot distributions of mutual entailment and intraverbals for ELL (Listener and Expressive).

The results from the Kolmogorov-Smirnov test are displayed in Table 12. When we include all participants in the analysis, the Listener and Expressive training groups differ significantly on both mutual entailment and intraverbals with the Expressive group demonstrating better scores on both variables with large effect sizes. When we include only the PES children in the analysis, the results parallel that of the full analysis. For the PES children, the Listener and Expressive training groups differ significantly on both mutual entailment and intraverbals also in favor of the Expressive group with large effect sizes. For the ELL children the Listener and Expressive groups differ significantly on mutual entailment in favor of the Expressive group with a large effect size. The ELL children in Listener and Expressive groups do not differ significantly on intraverbals with a medium effect size. However, the statistical comparison of training groups for ELL children has low power due to the small number of participants ( $n = 14$ ).

Research Question 4 -- What is the nature of the relationship between the ability to perform mutual entailment and untrained intraverbals?

This research question explores the nature of the relationship between mutual entailment and combinatorial entailment (intraverbal) performance. We initially planned a contingency table analysis along with a chi square test of independence for this analysis. This test would have assessed whether mastery (all eight items correct) of mutual entailment was independent of mastery of intraverbals. But, due to the low proportion of participants mastering mutual entailment or intraverbals, especially in the



Table 12

*Kolmogorov-Smirnov Test Comparing Listener and Expressive Training*

	<i>N</i>	Max. absolute difference	<i>Z</i>	<i>p</i>	<i>d</i>
All participants					
Mutual entailment	68	.735	3.03	<.001	1.59
Intraverbals	68	.412	1.70	.006	.87
PES					
Mutual entailment	54	.708	2.59	<.001	1.52
Intraverbals	54	.426	1.56	.015	.84
ELL					
Mutual entailment	14	.800	1.43	.033	2.33
Intraverbals	14	.467	.837	.486	.76

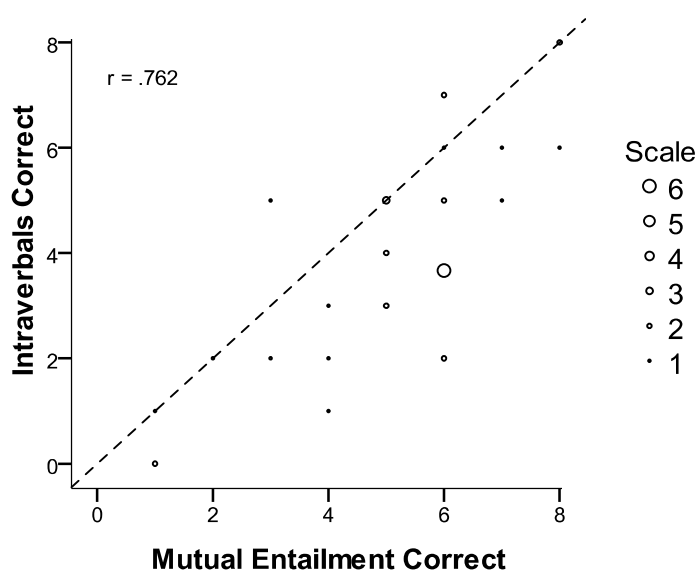
Listener training group, this analysis would have been insensitive. Instead, we analyzed the nature of the relationship between mutual entailment and intraverbals through scatterplots and correlation coefficients. If the reader is interested in the outcomes of the contingency analysis and chi square test of independence, this analysis is included in Appendix H.

Figures 20 and 21 show scatterplots of mutual entailment and intraverbal performance for the two groups. The broken line displayed in each scatterplot (located at

$x = y$ ) divides the plot into three sections: (a) those who scored better on mutual entailment than on intraverbals (below the line), (b) those who scores lower on mutual entailment than on intraverbals (above the line), and (c) those who scored the same on both mutual entailment and intraverbals (on the line). In each Figure, the number of overlapping data points is indicated by the size of the dots. Larger dots indicate a larger number of overlapping data points.

In Figure 20, the scatterplot for the Listener group, 23 of the 34 participants (68%) scored in the lower portion of the scatterplot. These children scored better on mutual entailment than on intraverbals. And seven children (21%) scored on the line, which means that these children earned the same score on both mutual entailment and intraverbals. Overall, the majority (88%) of children in the Listener group performed mutual entailment at levels equal to or better than performance on intraverbals. Only four participants (12%) scored better on intraverbals than on mutual entailment. The correlation between the number of correct intraverbals and mutual entailment responses was .762 which was statistically significant ( $p < .001$ ).

The scatterplot for the Expressive group is displayed in Figure 21. All children in this group performed mutual entailment at levels equal to or better than intraverbals. Twenty-one of the 34 participants (62%) scored better on mutual entailment than on intraverbals. Thirteen children (38%) scored the same on mutual entailment and intraverbals and none of the children scored better on intraverbals than on mutual entailment. In the Expressive group, no child scored below five items correct on mutual



*Figure 20.* Scatterplot of mutual entailment and intraverbals for all participants (Listener).

entailment, but scores for intraverbals range from 0 to 8 items correct. The correlation between mutual entailment and intraverbals is  $.592$  ( $p < .001$ ). However, the correlation for the Expressive group is weaker than for the Listener group. It is important to note that the strength of this correlation may have been reduced by the ceiling effect for mutual entailment.

Scatterplots of mutual entailment and intraverbals for the PES subgroup are shown in Figures 22 – 23. In Figure 22, the scatterplot for the Listener group, 17 of the 25 participants (68%) scored better on mutual entailment than on intraverbals. And 4 children (16%) earned the same score on both mutual entailment and intraverbals. Four

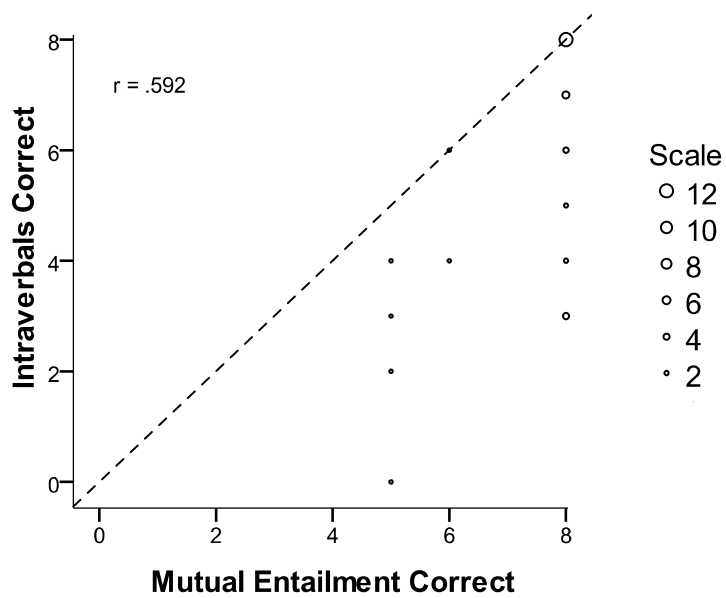


Figure 21. Scatterplot of mutual entailment and intraverbals for all participants (Expressive).

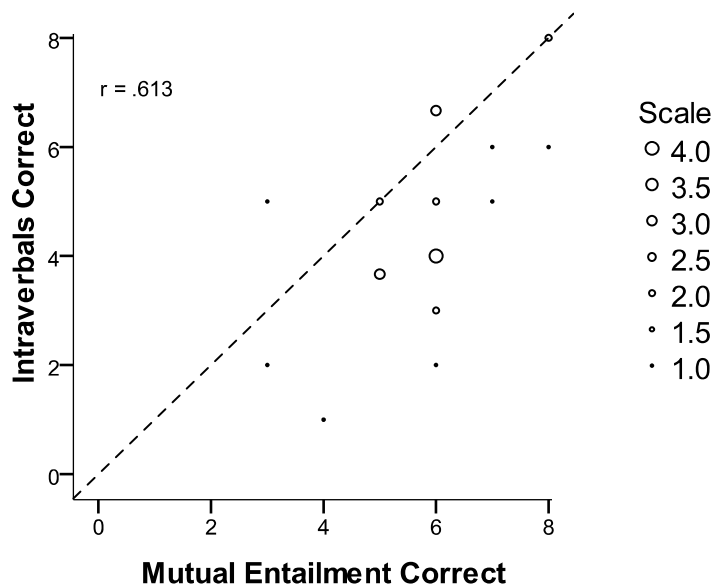


Figure 22. Scatterplot of mutual entailment and intraverbals for PES (Listener).

participants (16%) scored better on intraverbals than on mutual entailment. These four participants were the only children in either training group to score better on intraverbals than on mutual entailment. For the Listener group, the correlation ( $r = .613$ ) was statistically significant and positive ( $p = .001$ ).

The scatterplot for the PES Expressive group is displayed in Figure 23. For the Expressive group, all children performed mutual entailment at levels equal to or better than intraverbals. Sixteen of the 29 participants (55%) scored better on mutual entailment than on intraverbals. Thirteen children (45%) scored the same on mutual entailment and intraverbals. For the Expressive training group, the correlation is also significant and positive ( $r = .599$ ,  $p = .001$ ).

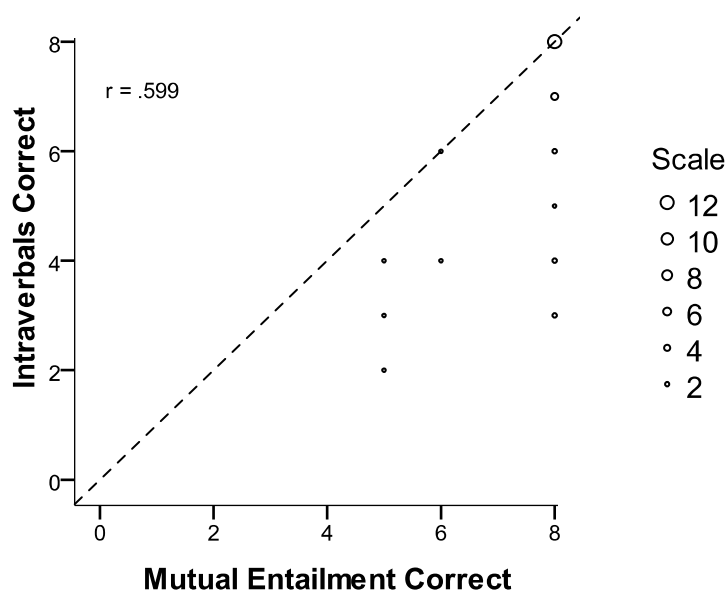
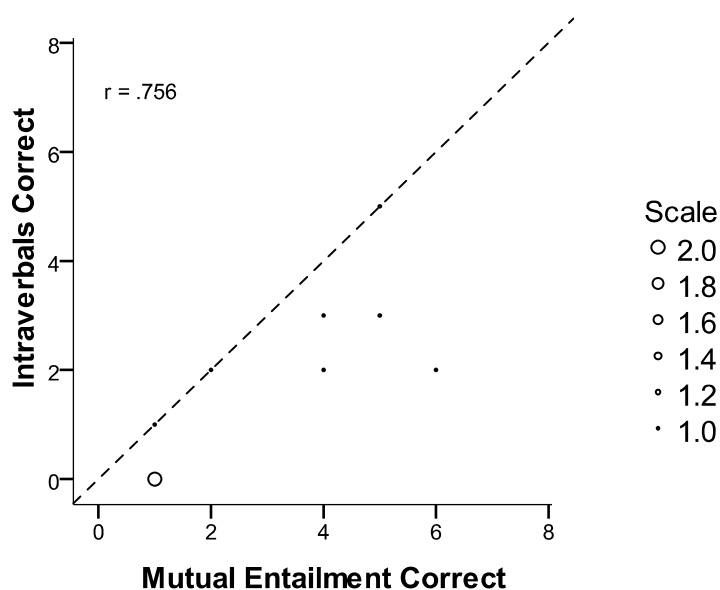


Figure 23. Scatterplot of mutual entailment and intraverbals for PES (Expressive).

Scatterplots of mutual entailment and intraverbals for the ELL subgroup are shown in Figures 24 – 25. For the ELL subgroup, all children performed mutual entailment at levels equal to or better than intraverbals in both training groups. In Figure 24, the scatterplot for the ELL Listener group, 6 of the 9 participants (67%) scored better on mutual entailment than on intraverbals. And 3 children (33%) earned the same score on both mutual entailment and intraverbals. All ELL children answered 6 or fewer items correctly on mutual entailment and 5 or fewer on intraverbals. The correlation between mutual entailment and intraverbals was positive and statistically significant ( $r = .756$ ,  $p = .018$ ).



**Figure 24. Scatterplot of mutual entailment and intraverbals for ELL (Listener).**

The scatterplot for the ELL Expressive group is displayed in Figure 25. For the Expressive group, all 5 of the participants scored better on mutual entailment than on

intraverbals. Although 4 of the 5 children answered all 8 mutual entailment items correctly, only two children answered more than three intraverbal items correctly. For the Expressive training group, the correlation based on these 5 children is strongly positive, but not statistically significant ( $r = .826$ ,  $p = .085$ ). Given the very small number of data points, the statistical significance test is not relevant.

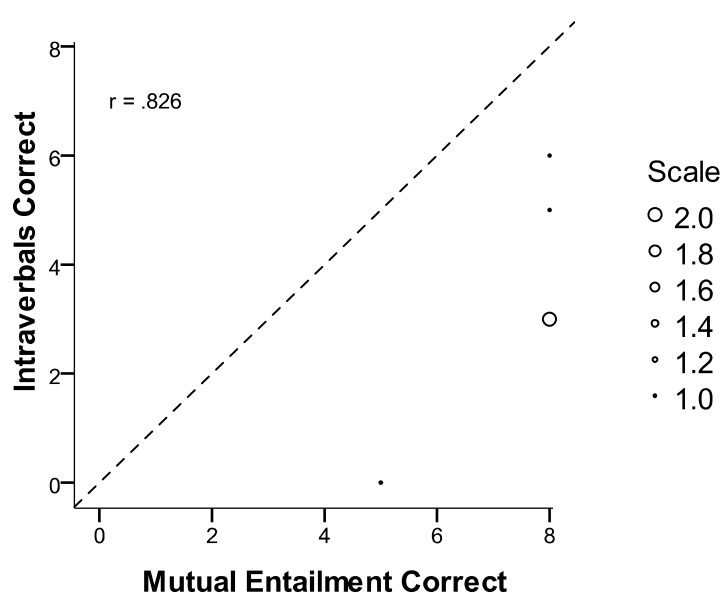


Figure 25. Scatterplot of mutual entailment and intraverbals for ELL (Expressive).

Research Question 5 -- What is the nature of the relationship between working memory and mutual entailment and untrained intraverbals?

Variation in working memory may explain some of the variation in performance on mutual entailment and intraverbals. We measured working memory through the Memory for Digits subtest of the CTOPP (Wagner et al., 1999). Of the 68 children who completed training, 67 also completed the Memory for Digits subtest. The one child who

did not complete this subtest was not available for testing due to extended and frequent absences from school. As seen in Table 13, most notable finding is that working memory is significantly correlated with mutual entailment ( $r = .368, p = .035$ ) and intraverbals ( $r = .444, p = .010$ ) in the Listener training condition. However, the magnitude of correlation coefficients for these relationships is small to medium. Correlations in the Expressive group were near zero and not significant. The visual analysis of scatterplots within the Listener training condition shows an unclear relationship between working memory and mutual entailment (Figure 26) and untrained intraverbals (Figure 27). There are two outlying data points in which two children scored 0 on working memory but scored 2 and 6 on mutual entailment and scored 2 and 4 on intraverbals.

Correlations for the PES subgroup in the Listener training group are not statistically significant for either mutual entailment ( $r = .094, p = .670$ ) or intraverbals ( $r = .332, p = .122$ ). Similarly, in the ELL subgroup, correlations for Listener training are not statistically significant for mutual entailment ( $r = .250, p = .517$ ) or intraverbals ( $r = -.103, p = .729$ ).

### Social Validity

We assessed social validity of the outcomes measured in this study through a comparison of study outcomes to performance on a typical preschool categorization task. To accomplish this, we presented a categorization matching task to each child following completion of the other tests. Because children selected one picture out of an array of three, random responding would result in an average of 33% correct. Therefore, scores of 1 or 2 (out of 4) correct should not be interpreted as an indication of proficiency.



Table 13

*Correlations Between Working Memory and other Dependent Variables*

Training group	Mutual entailment	Intraverbals
All participants		
Listener	.368*	.444*
Expressive	.017	.081
PES		
Listener	.094	.332
Expressive	-.040	-.005
ELL		
Listener	.250	-.103
Expressive	.232	-.203

\* $p < .05$

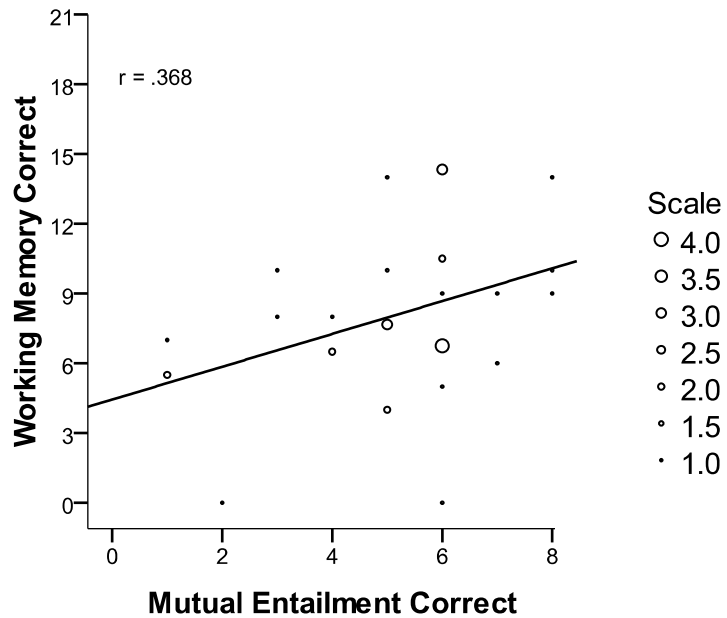


Figure 26. Scatterplot of mutual entailment and working memory (Listener).

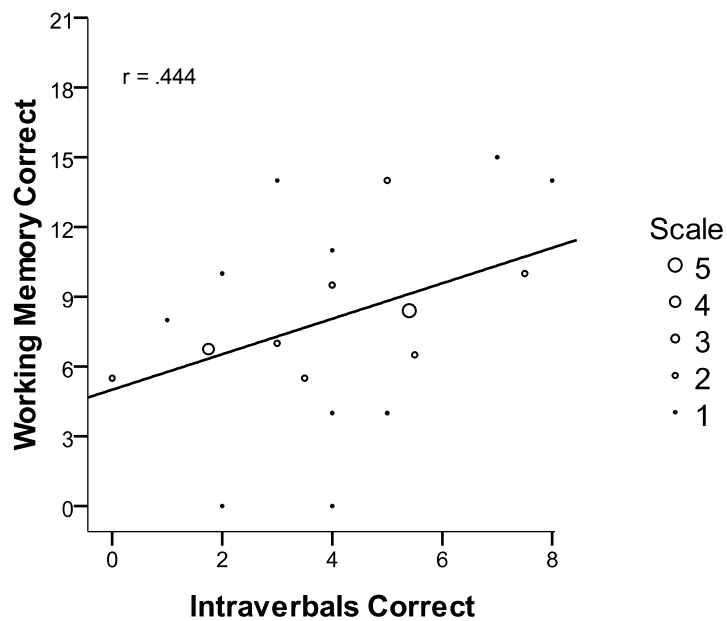


Figure 27. Scatterplot of intraverbals and working memory (Listener).

Dot plot distributions (Figure 28) for Listener and Expressive training groups on matching are negatively skewed due to a ceiling effect in both groups. In fact, 23 of 34 children (67%) in the Listener group made no errors on this task. In the Expressive group, 24 of 34 children (71%) made no errors on the matching task. When the Listener and Expressive groups are compared on matching scores, the groups do not perform significantly differently ( $Z = .121, p = 1.00$ ). However, this comparison between groups is likely affected by the fact that most children produced perfect performances on this measure. This ceiling effect may prevent differences in matching ability from being detected within this study.

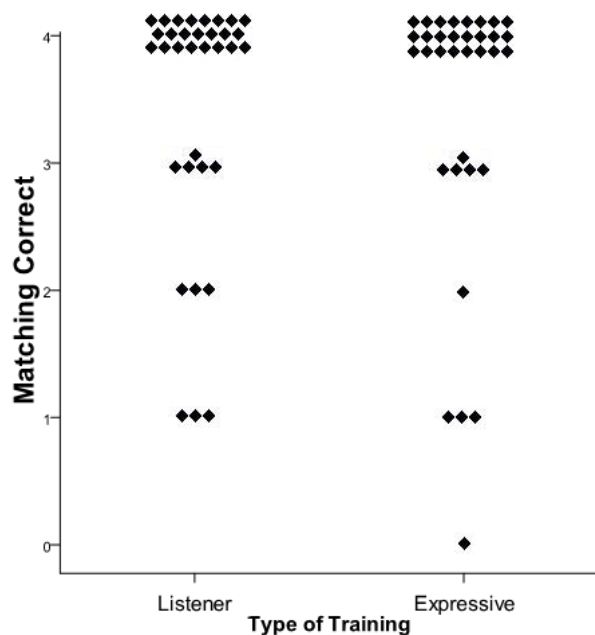


Figure 28. Dot plot distributions for matching (Listener and Expressive).

Scatterplots for matching and each dependent variable in the Listener and Expressive groups are displayed in Figures 29 – 32. These scatterplots clearly show the ceiling effect in the matching variable with many scores clustered at the top of the graph. Correlation coefficients for the relationship between matching and each dependent variable are included in Table 14. Due to the prominent ceiling effect in both training groups, correlation coefficients are less meaningful measures of strength of relationship between matching and other dependent variables. Caution should be taken in interpreting these results since few children scored below the maximum score on matching.

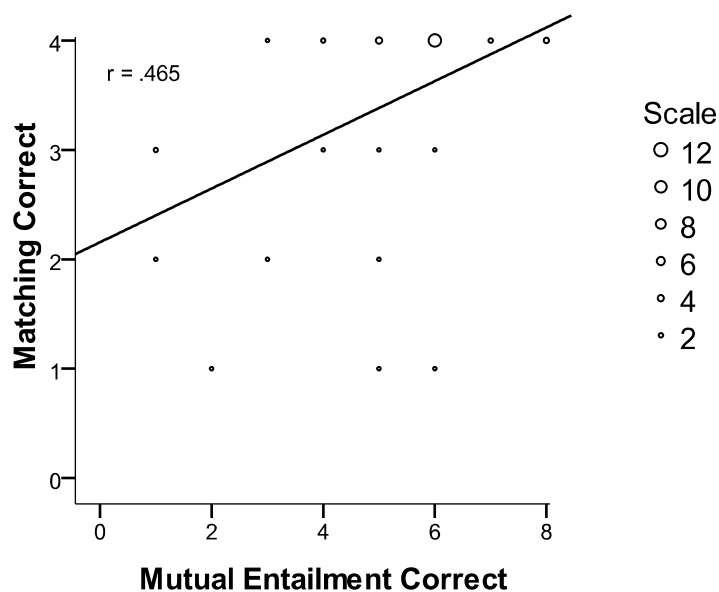


Figure 29. Scatterplot of matching and mutual entailment (Listener).

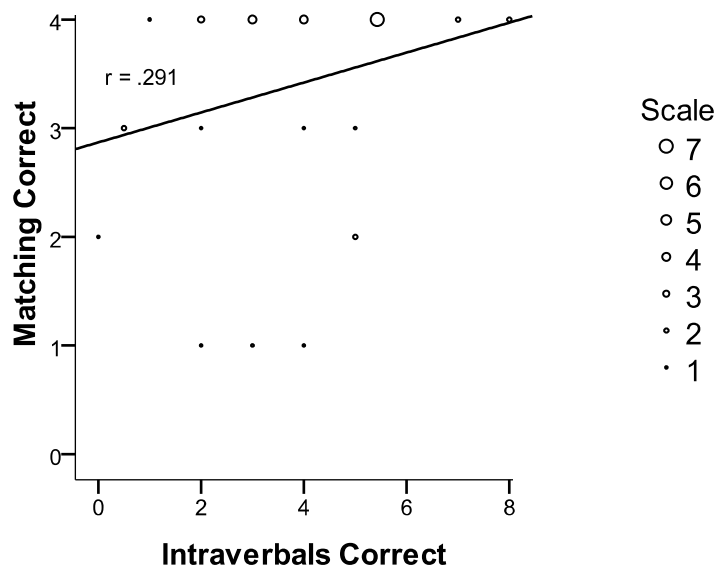


Figure 30. Scatterplot of matching and intraverbals (Listener).

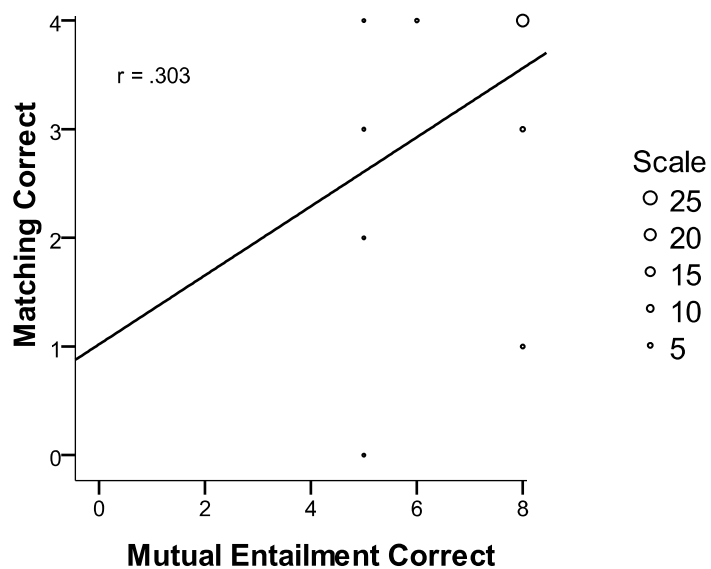
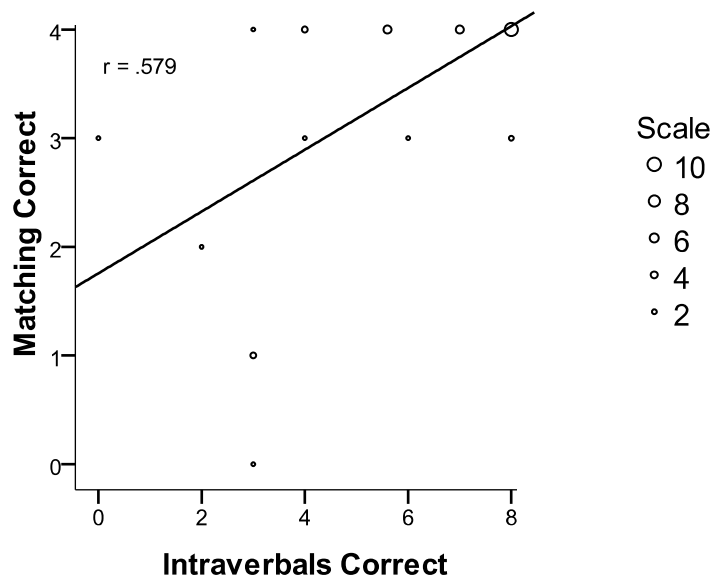


Figure 31. Scatterplot of matching and mutual entailment (Expressive).



*Figure 32.* Scatterplot of matching and intraverbals (Expressive).

We also assessed social validity of these measures by examining the benchmark test used by the Head Start program (Bear River Head Start, 2008-2009). This assessment included several items concerning children's ability to categorize objects as well as items requiring the child to verbally respond to categories of objects. Table 15 displays several Head Start benchmarks along with the outcomes from the current study that address similar skills.

Table 14




*Correlations Between Matching to Sample and other Dependent Variables*

Training group	Mutual entailment	Intraverbals
All participants		
Listener	.465*	.291
Expressive	.303	.579*
PES		
Listener	.245	.046
Expressive	.432*	.548*
ELL		
Listener	.161	-.140
Expressive	-.147	.415

\* $p < .05$

Table 15

*Comparison of Head Start Benchmarks to Current Study Outcomes*

Head Start benchmark	Outcome from current study
<ul style="list-style-type: none"> <li>• “Identifies objects based on category”</li> </ul>	 Points to picture when presented with category label
<ul style="list-style-type: none"> <li>• “Names classes/categories of sorted objects”</li> </ul>	 Says category label when presented with picture
<ul style="list-style-type: none"> <li>• “States the characteristics and attributes of an object or place”</li> </ul>	
<ul style="list-style-type: none"> <li>• “Compares items by category and function”</li> </ul>	 When presented with one picture from a category, selects the other picture from the same category
<ul style="list-style-type: none"> <li>• “Classifies animals into groups”</li> </ul>	
<ul style="list-style-type: none"> <li>• “Classifies objects by their state”</li> </ul>	



## CHAPTER IV

### DISCUSSION

In this chapter we present an overview of the rationale and purpose of this study, summarize study findings, consider alternative explanations, and finally discuss directions for future research.

#### Overview of Study

This study was designed to explore the relationship between learning naming responses and producing untrained intraverbals. Behavioral theories of language development have predicted patterns of language learning concerning these types of verbal behavior. Skinner (1957) suggested that different types of verbal behavior (e.g. tacts and intraverbals) are learned independently, at least early in language development. For example, if a child learns to say “cat” or “animal” upon seeing a cat, the child would not necessarily be able to answer a question in the absence of the object such as “What is a cat?” by saying “animal.” However, there are environmental circumstances and learning histories that make transfer among types of verbal behavior more likely. RFT (Hayes et al., 2001) also predicts that these responses are learned independently only very early in language development. RFT adds to Skinner’s predictions by specifying how naming and intraverbal responses become dependent. Over many exposures to words used as names and also as intraverbals, the child learns the underlying relationship among responses. After learning this general relationship, the child is able to answer the question “What is a cat?” by saying “animal” even though s/he has only been taught to say the

words “animal” and “cat” upon seeing a cat. Although several studies (Luciano, 1986; Miguel et al., 2005; Partington & Bailey, 1993; Petursdottir et al., 2008) have been conducted to test whether young children and older children with mental retardation learn these responses independently, results have been inconsistent.

This study was conducted to clarify the nature of the relationship between training method, mutual entailment, and intraverbals in a categorization task for preschool children. We evaluated these relationships along with related language outcomes such as delayed mastery and working memory. The findings from this study will be useful in clarifying behavioral theories of language development. The findings will also serve as a foundation for practical applications of these theories in a preschool setting.

## Findings

We summarize findings for completion rates as well as each research question in this section. We discuss applications and present implications for future research in a later section.

### *Completion Rates*

A statistically significantly smaller proportion of children completed Expressive training (68%) than Listener training (92%). While there are many possible explanations for why this is the case, it is clear that learning to say four names and two categories in Expressive training is more difficult than learning to point to characters when hearing these four names and two categories in Listener training. This may be because one form of training (Expressive or Listener) is a more powerful form of instruction for directly

teaching names and categories. That is, Listener training may be more powerful in that it resulted in a higher proportion of children reaching mastery criterion for both names and categories within one or two training sessions. On the other hand, it is also possible that the tasks required in the two types of training are not equally difficult. Saying names and categories for each of four characters in the Expressive condition may be more difficult than pointing to characters when hearing names and categories in the Listener condition. The latter explanation is likely more accurate. Learning to say names and categories is likely more difficult than learning to point to pictures upon hearing names. This is reflected in prior literature in which very young children were shown to learn listener responses to words (selecting) before learning to say those same words (e.g., Goldin-Meadow et al., 1976). Therefore, it is most likely that the differences in completion rate are at least partially attributable to differences in the difficulty of the learned responses (saying names is more difficult than pointing to items).

Regardless of whether the difference in completion rates is a result of the power of instruction or the difficulty of tasks, the children who completed Expressive training likely entered the study with different skills than those who completed Listener training. Those completing the Expressive training learned to say responses upon seeing pictures while those in Listener training learned to select pictures upon hearing names.. Essentially, the Expressive group passed a more difficult 'entry test' prior to the posttesting. This difference introduces bias in group membership. It is not known whether the bias in group membership had any affect on the results of this study. However, if the children in the Expressive group do have higher language skills than the Listener group, then all comparisons between groups may be affected. That is, superior

performance of children in the Expressive group as compared to the Listener group on dependent variables could be attributed to higher general language skills instead of (or in addition to) group membership.

Future researchers could address this issue in several ways. For example, all participants in future studies could first complete an Expressive pre-training in which they learn to say names and categories for a set of characters not used in the study. Only those children who complete mastery of pre-training within a time period equal to or less than that allowed during experimental training (e.g. two twenty minute sessions) would be included in the research sample. Thus, all children in both groups would have been shown to be capable of learning a set of names from this type of instruction. This would increase the likelihood that all children are equally capable of completing training in either training group. Another approach to reducing the possible selection bias may be to implement a more powerful or more extensive training program – one that would enable a larger proportion of children to attain mastery. For example, training could be continued for multiple days until each child reached mastery.

*Research Question 1. When taught Expressive responses how does preschool children's performance compare to chance on mutual entailment and untrained intraverbals?*

Children in the Expressive training group performed significantly better than chance on both mutual entailment and intraverbal tasks. In fact the mean mutual entailment score for the Expressive group ( $M = 7.53$ ) was near the maximum of eight correct responses. PES children in the Expressive training group also performed significantly better than chance on both mutual entailment and intraverbals. However, the

ELL group of children performed significantly better than chance only on mutual entailment. Performance on intraverbals was similar to chance for this subgroup of children. This analysis suggests that for many preschool children taught to say names and categories for a set of pictures (Expressive training), both mutually entailed responses and intraverbal responses can be derived without direct training. However, it is important to notice that although performance on mutual entailment and intraverbals was better than chance, not all children answered all items correctly. In fact about 18% of participants made at least one error on mutual entailment items and 65% of children made at least one error on intraverbals. Therefore, the Expressive training procedures did not result in a firm ability to produce mutual entailment and untrained intraverbals in all children. In a later section, we argue that Expressive training is a stronger form of instruction than Listener training. It is important to note that preschool children, given a relatively strong form of instruction, still produce some incorrect responses, especially in deriving intraverbals.

*Research Question 2. When taught Listener responses, how does preschool children's performance compare to chance on mutual entailment and untrained intraverbals?*

In the Listener training group, children also performed at levels significantly better than chance on both mutual entailment and intraverbal responses. While these scores were better than chance, they were not close to the maximum score for each dependent variable (mutual entailment  $M = 5.12$ , intraverbal  $M = 3.94$ ). The PES subgroup of children also performed at levels significantly better than chance on both mutual entailment and intraverbals while the ELL subgroup of children performed at

levels similar to chance on both mutual entailment and intraverbals. This analysis suggests that some preschool children receiving Listener training can also derive mutually entailed and intraverbal responses. However, similar to the Expressive training group, it is important to notice that although performance on mutual entailment and intraverbals was better than chance, not all children answered all items correctly. About 91% of participants made at least one error on mutual entailment items and 94% made at least one error on intraverbals. The ability to produce untrained mutual entailment and intraverbals is not firm for most children when they receive Listener training. As noted in the summary of the Expressive group's comparisons to chance, comparisons between training groups (discussed in a later section) will show that Listener training is a relatively weak form of instruction. Given that Listener training is weak, it is important to note that a small percentage of children in the Listener group (9% for mutual entailment and 6% for intraverbals) derived all possible responses and most children derived some of the untaught responses. This shows that some preschoolers are able to benefit from relatively weak forms of instruction when the goal of instruction is deriving mutual entailment and intraverbal responses.

*Research Question 3. How do children receiving each type of training compare on mutual entailment and untrained intraverbals?*

The purpose of the third research question was to examine the differential effects of Expressive and Listener training to determine the relative effectiveness of each form of training on derived mutual entailment and intraverbal responses. The Expressive training group performed significantly better than the Listener training group on both mutual

entailment and intraverbals with large effect sizes in both comparisons. This was also true of the PES subgroup of children. However, within the ELL subgroup of children, the Expressive group performed significantly better than the Listener group only on mutual entailment. In fact, the ELL subgroup in Expressive training scored well above chance on mutual entailment ( $M = 7.40$ ,  $d = 3.28$ ) while the Listener group scored at levels similar to chance ( $M = 3.22$ ,  $d = .11$ ). For ELLs, the difference between intraverbal performance in Expressive and Listener groups was not statistically significant. However, the effect size for the difference between Listener and Expressive groups on intraverbal performance indicates a large effect ( $d = .76$ ). The lack of statistical significance for intraverbals may be due to the small number of participants ( $n = 14$ ) in the ELL subgroup.

This analysis shows that teaching an expressive response (saying) may be more effective than teaching a listener response (selecting) in producing mutual entailment and untrained intraverbals. However, this finding is tentative due to the selection bias in group membership. A statistically significant smaller proportion of children completed Expressive training than Listener training. Thus, the children who completed training in the two groups likely differ in their language learning skills. This difference may confound results in the comparison between training groups. We will discuss additional aspects of the training and testing that may have affected performance of children in each training group in the following paragraphs.

The finding that children in the Expressive group perform better on mutual entailment compared to children in the Listener group is likely not surprising to many researchers in the disciplines related to child development and language. There is already

a large amount of research (e.g. Horne et al., 2006; Wynn & Smith, 2003) showing that children who are trained in expressive naming responses (saying object names upon seeing the object) then produce listener responses (pointing to the object upon hearing the name) at high levels of accuracy without additional training. Those trained in listener responses are less likely to produce the expressive naming response without additional training. However, it is unclear whether the relative effectiveness of Expressive training on producing mutual entailment is a result of the instruction or the difficulty of the mutual entailment tasks. It could be that Expressive training is a more powerful form of instruction and thus results in higher levels of mutual entailment performance. On the other hand, the mutual entailment task for Expressive training (selecting) may be easier than the mutual entailment task for Listener training (saying). Thus, children may produce mutual entailment in Expressive training (selecting) better than mutual entailment in Listener training (saying) for reasons unrelated to the form of instruction – the differential ease of the mutual entailment task.

While there are multiple research studies addressing the ability to perform mutual entailment under Expressive and Listener training conditions, there was little previous evidence available to show that a similar pattern of responding is present for intraverbals. While mutual entailment performance may differ due to differential difficulty of the assessment tasks, for intraverbals the responses are identical in both training groups (saying). It seems logical that those children taught to say names (expressive) would likely perform better on intraverbals than those taught listener responses to names since the intraverbal response also requires them to say the names (that is, they have the same response modality). Those children in the Expressive group practice saying names and



categories during training, while those in the Listener group merely point to items. This study provides empirical evidence to suggest that children who are trained in expressive names and categories produce higher levels of intraverbals than children trained in listener responses. Given the relative strength of Expressive training in producing both mutual entailment as well as derived intraverbals, it appears that Listener training is a weaker form of instruction when mutual entailment and intraverbal outcomes are the goal of instruction.

*Research Question 4. What is the nature of the relationship between the ability to perform mutual entailment and untrained intraverbals?*

To further explore the relationship between performance on mutual entailment and intraverbals, we analyzed the patterns in scatterplots depicting these variables. Overall, the large majority of children (Listener = 88%, Expressive = 100%) produced mutual entailment scores equal to or higher than their intraverbal scores. This general pattern of responses holds across PES and ELL subgroups of children. Only four out of 68 children scored higher on intraverbals than on mutual entailment. These four children were in the Listener training PES subgroup.

Correlations were also calculated for each comparison. Correlations between mutual entailment and intraverbal performance were statistically significant and positive within each training group. The correlation for the Listener group indicated a strong relationship between mutual entailment and intraverbals ( $r = .762$ ) while the correlation for the Expressive group indicated a medium strength of relationship ( $r = .592$ ). For the PES subgroup, correlations were also statistically significant and positive indicating

medium strength relationships among variables. Finally, within the ELL subgroup, the correlation for the Listener group was statistically significant and positive indicating a strong relationship between mutual entailment and intraverbals. However, the correlation in the Expressive group indicated a strong, positive relationship but was not statistically significant – this was likely due to a very small number of participants in this subgroup ( $n = 5$ ).

In combination, the visual analyses of scatterplots and the correlations, suggest that performance on intraverbals is related to mutual entailment performance. The nature of this relationship, however, is unclear. While performance on mutual entailment is *predictive* of performance on intraverbals, several types of relationships are possible. Very few children in this study (6%) produced a higher number of correct intraverbal responses than they did mutual entailment responses with most children (65%) producing fewer intraverbals than mutual entailment. This suggests that the mutual entailment response is learned earlier than the intraverbal response. Earlier acquisition of mutual entailment may signify that mutual entailment is a simpler skill than intraverbals and develops earlier due to the relative ease of learning to derive mutual entailment compared to intraverbals. It is also possible that mutual entailment is acquired earlier than intraverbals because it is prerequisite to intraverbal performance. Children may be unable to learn to derive intraverbals until after they have learned to derive mutual entailment responses. The exact nature of the relationship between mutual entailment and intraverbal performances is an empirical question and should be explored in future research.

*Research Question 5. What is the nature of the relationship between working memory and mutual entailment and untrained intraverbals?*

We evaluated the relationship between working memory and other dependent variables using a correlational analysis. Correlations and examination of scatterplots revealed that within the Listener training group, working memory was significantly but weakly correlated with both mutual entailment ( $r = .368$ ) and intraverbals ( $r = .444$ ). Correlations within the Expressive group were not statistically significant. Within the PES and ELL subgroups, correlations between working memory and the two main dependent variables were small and not statistically significant.

It is notable that correlations between working memory and mutual entailment and intraverbals are statistically significant only in the Listener training group, not in the Expressive group. It is possible that working memory is correlated with mutual entailment and intraverbals in the Listener training group due to the nature of the training and the modality of tested responses. That is, children in the Listener group practiced selecting pictures upon hearing names during training. But, they were required to say those names in tests of both mutual entailment and intraverbals. It is possible that producing this unpracticed response during testing requires relatively high demands on working memory ability. On the other hand, it is possible that correlations in the Expressive group were suppressed by the ceiling effects present in the mutual entailment and intraverbal responses.

More research is needed to clarify how working memory is related to mutual entailment and intraverbals in each type of training as well as how this relationship may be translated into practical applications. For example, future researchers could examine

whether an intervention targeted at increasing a child's ability to recall words has a positive effect on mutual entailment and intraverbal performances.

### *Social Validity*

We assessed social validity in two ways in this study. Because this study does not involve an intervention that would be incorporated into a curriculum, many of the more conventional measures of social validity such as teacher willingness to continue procedures or teacher rating of benefit to the child would not be appropriate for this study. Therefore, two somewhat less conventional methods were used to assess social validity. We assessed the importance of the skills being evaluated in this study through correlation with a typical preschool task involving a categorization skill – matching. A second measure of social validity compared the types of skills evaluated in this study to skills considered important by the Head Start organization. This was accomplished through comparing the skills evaluated in this study to skills assessed by items on the Head Start's benchmark test.

We compared performance on the dependent measures in this study to performance on a typical preschool matching task. Because most children answered all matching items correctly, little can be said about the relationship between mutual entailment or intraverbals and matching. However, it appears that the ability to perform the untaught matching task is present before the ability to produce mutual entailment and intraverbal responses. When training groups are compared on their matching scores, children in the Listener and Expressive training groups perform similarly – both groups show near ceiling level performance. In the current study, the matching task required

children to make a derived (in the case, combinatorial entailment) response, but this task seemed to be easier than mutual entailment and intraverbals as measured in this study. The lower level of difficulty may be due to the smaller amount of information required to correctly match pictures. In this task, the child only needed to remember, or derive, the categories for each picture. A child who did not remember, or derive, any of the picture's names could still perform the matching task accurately. It may be due to the low difficulty level of the matching task that most children performed without error.

A second indicator of social validity compared the dependent measures in this study to skills assessed by items on an assessment used regularly at the Head Start preschool. Several of the outcomes on that preschool assessment were very similar to dependent variables being measured in this study. The assessment included items requiring both listener and expressive categorization responses. These measures of social validity suggest that the dependent measures in this study assess important skills that are reflected in the preschool curriculum. Therefore it is important to evaluate the dependent variables in this study to determine how children learn to categorize and to build a foundation for future interventions to increase children's ability to perform these skills.

### Alternative Explanations

We describe alternative explanations for the results as well as limitations in this section. Alternative explanations and limitations include native language and delayed mastery. We also discuss child and task characteristics important for generalization. These include (a) age, (b) risk status, and (c) type of relation. Recommendations for

future research are included in this section for each alternative explanation and area of generalization.

### *Native Language*

The current study included students who spoke primarily English at home (PES) as well as those whose primary home language was other than English (ELL). The PES subgroup in Listener training performed significantly better on mutual entailment ( $d = 1.73$ ) and intraverbals ( $d = 1.51$ ) when compared to the ELL subgroup. Although differences between PES and ELL groups on each dependent variable in the Expressive group were not statistically significant, the effect size for intraverbals ( $d = 1.40$ ) was large with PES children scoring better than ELL children. It should be noted that both mutual entailment and intraverbals in the Listener group and intraverbals in the Expressive group are expressive (saying) responses. This difference in performance between PES and ELL subgroups can be explained in several ways including (a) differences in ability to learn new words, (b) differences in ability to retain directly taught relations, (c) differences in ability to derive mutual entailment and intraverbals across languages, or (d) differences in ability of to understand testing instructions or questions.

First, differences in PES and ELL performance on mutual entailment and intraverbals may be due to differences in ability to learn new words. However, all children who completed training met the initial mastery criterion for names and categories prior to completing posttests. This ensured that all children who completed training had at least a minimum ability to learn new words. Therefore, differences in

performance on posttests do not indicate a difference in ability to learn directly trained words.

ELL performance may also have been affected by a lower ability to maintain either listener or expressive responses once learned. While the ELL children performed delayed mastery better than PES children in the Listener group ( $d = .44$ ), they performed delayed mastery at a lower level of accuracy in the Expressive group ( $d = 1.78$ ). However, neither of these differences were statistically significant. Although ELL children's ability to maintain learned responses over time does not significantly differ from PES children, the large effect size for delayed mastery in the Expressive group suggests that ELL children may maintain expressive (saying) responses at different levels than PES children. This suggests that ELL children may have a lower ability to maintain directly taught expressive responses than PES children. Lower performance on mutual entailment and intraverbals by ELL children may be partially explained by a lower ability to maintain directly taught expressive responses.

A third explanation for the disparity in ELL and PES performance is that children who speak a language other than English are less capable of performing mutual entailment and intraverbal responses regardless of the language of instruction. However, this is unlikely. Theoretically, the ability to derive intraverbal responses based on names and categories should not differ across languages. The existing literature evaluating the ability of students to derive intraverbals has been conducted in multiple languages. The study by Luciano (1986) evaluating whether teenagers with mental retardation could derive intraverbals after being taught to name objects was conducted in Spanish, the native language of the participants. Yet, these results are very similar to the replication of

the study (Partington & Bailey, 1993) which was conducted in English, the native language of the participants in that study. Thus, it is unlikely that the ability to perform mutual entailment and intraverbals differs across native languages.

A final explanation for why the ELL students performed differently than the PES students in the current study appears to be more probable. The current study was conducted entirely in English. It is likely that ELL children had a lower level of English proficiency than those in the PES group (although no measure of English language proficiency was administered in the current study). ELL children may not have understood the instructions for testing or they may not have understood the test questions. Several of the ELL children displayed signs of confusion during testing such as failing to answer the question or repeating portions of the question. For example, when asked the question “What does Bo do?” one child responded, “Bo do.” It is plausible that if the testing instructions and questions had been delivered in the child’s native language, the ELL group of children would have answered more questions correctly. Therefore, it is likely that differences in mutual entailment and intraverbals of the ELL group when compared to the PES group are attributable to a lack of understanding of testing instructions and questions as well as a lower ability to maintain learned expressive responses rather than differences in ability to learn new names and categories, or differences in ability to derive mutual entailment and intraverbal responses across languages.

The limitations introduced into this study due to low English proficiency could be used to help refine future research methods in several ways. First, the research could be conducted in each child’s native language. However, if this is not feasible, then a test of



English proficiency could be used to identify potential participants, excluding those with English proficiency scores below a specified level. This would ensure that all children included in the study have at least a minimum level of English proficiency to understand the instructions and test questions. Implications for future research concerning low performance on delayed mastery tests will be addressed in the next section.

### *Delayed Mastery*

Following other testing procedures, children completed a delayed mastery test to assess whether they had forgotten any of the directly trained responses. Not all children performed perfectly (all eight items correct) on the delayed mastery test. Mistakes on the delayed mastery test could be explained in two ways. First, the children may have forgotten the trained responses – this might suggest that the mastery criteria used to terminate training was not stringent enough to ensure that children could maintain all of the directly trained responses through the entire assessment period. On the other hand, the delayed mastery test took place late in the session. Mistakes on the delayed mastery test could also be attributable to student fatigue.

While less than perfect performance on delayed mastery tests did not seem to substantially impact performance on mutual entailment and intraverbal items, it does point to a limitation in the procedures of this study. Future research could be refined in a way to ensure that inadequate delayed mastery performance does not influence performance on dependent variables. This could be accomplished in several ways. First, while training could be delivered in one or two sessions similar to the current study, the mastery criterion could be raised. For example, instead of three consecutive correct

responses in a mixed set, the children could be required to produce a larger number of consecutive correct responses. Another way to strengthen the mastery criteria would be to add a fluency component. In this case, the children would have to produce a specified number of consecutive correct responses in a mixed set and within a specified time limit. Yet another option to increase the chances of better performance on the delayed mastery test may be to require a mastery criterion to be met on each of a specified number of consecutive sessions. Whether any or all of these mastery criteria would actually result in improved performance on delayed mastery is an empirical question and a topic for future research.

We have explored several issues that may impact the interpretations of the current study, but it is also important to begin to understand how the findings from this study can be generalized to other populations and other language tasks. The next few topics addressed have to do with the degree to which the findings from the current study can be generalized across child characteristics such as age and risk-status or task characteristics such as the specific type of relationships taught.

### *Age*

The children who participated in this study were older (age range = 54 – 67 months) than the participants in previous research in which inconsistent results were reported. Three children in a previous study (Petursdottir et al., 2008) between the ages of 41 to 50 months failed to produce untrained intraverbals following Expressive and Listener training in names and Listener training in categories. Children between the ages of 39 and 60 months in two previous studies (Partington & Bailey, 1993; Miguel et al.,

2005) produced intraverbals inconsistently following Expressive training in names and categories. Based on the findings in previous studies as well as the current study, younger children perform derived intraverbals less accurately than older children. However, due to the small number of participants in the previous research, it is difficult to make generalizations about the age at which this skill develops. Therefore, the extent to which the findings in the current study can be generalized to children of different ages is not known. More research should be conducted to determine how both younger and older children perform on the dependent variables measured in the current study.

Research in a variety of age groups will allow the identification of developmental trends for mutual entailment and untrained intraverbals. Researchers in one study have suggested that mutual entailment develops as early as 17 months and combinatorial entailment develops as early as 24 months in an equivalence (sameness or identity) relation (Lipkens et al., 1993). Yet, these ages may be specific to the equivalence relation and perhaps specific features of the stimuli and task. In the current study, while almost all children in the Expressive group produced mutual entailment responses, 65% of these children made at least one error on intraverbals. For the Listener group, 91% and 94% of children made at least one error on mutual entailment and intraverbals, respectively. This study suggests that many typically developing children between 54 and 67 months are still developing both mutual entailment and intraverbals in a categorization relation since they failed to produce correct responses on all items.

*Risk Status*

The preschool children who participated in this study were considered at-risk for school failure. They attended a Head Start preschool program for children from low income families. According to Hart and Risley (1995), children from economically disadvantaged backgrounds tend to differ on various language skills compared to children from more economically advantaged backgrounds. In a longitudinal study of 42 families from a variety of socioeconomic backgrounds, the Hart and Risley identified patterns of language development. They found that children from economically disadvantaged backgrounds had smaller vocabularies than children from economically advantaged backgrounds. The children from economically disadvantaged backgrounds also had flatter vocabulary growth curves. That is, they were not learning words as rapidly as children from economically advantaged backgrounds. Based on this information, it is important to understand how children in the current study may differ from groups of children from different socioeconomic backgrounds. This is especially true of children receiving Listener training since children with better mutual entailment and intraverbal skills are more likely to be able to say words learned as listener responses and thus increase the size of their vocabularies. Future researchers should determine if children from different socioeconomic backgrounds perform differently on derived mutual entailment and intraverbal responses within a categorization task. If children from low income backgrounds perform at lower levels of accuracy on these dependent variables, and if these skills are important for future language development, then preschool interventions targeted at teaching children to produce these responses are warranted.

*Families of Relations*

In the current study, mutual entailment and intraverbals were tested within a categorization task. It is possible that the findings from this study are specific to the categorization relationships among names and intraverbals. Yet, preschool children are asked to respond to many types of relations other than categorization such as equivalence, difference, part-whole, opposites, cause-effect, and temporal relations among many others. Other relations that involve naming and intraverbal responses should be researched to determine whether the patterns found in the current study also apply to other relations. That is, having learned to derive mutual entailment and intraverbals in a categorization relation, to what degree can the child then produce mutual entailment and intraverbals in a different relation without direct training? For example, in a part-whole relationship, children can be taught to name whole objects and parts of objects. Mutual entailment could be tested for whole object names and object part names by testing listener skills if training was expressive or testing expressive skills if names were trained as listener responses. Intraverbals could be tested by having the child answer questions about the whole object and its parts in the absence of the object. If the child learned to name a “car” and also learned to name its parts such as “wheels,” “doors,” and “hood” the intraverbal could be tested by asking the child “What are the parts of a car?” while car is not present. This would be important to RFT in that it would show mutual entailment and combinatorial entailment (intraverbals) to be skills that children can generalize across types of relations. Whether children can generalize these skills would also be important when developing a language instruction program. It would be important to know whether

mutual entailment and intraverbals need to be specifically taught in each type of relation or whether these skills are generalizable across types of relations.

In addition to research on relations that involve names and intraverbals, future research should also be conducted to determine if the findings in the current study also apply to types of combinatorial entailment other than intraverbals. Intraverbals are only one example of combinatorially entailed responses. Other types or responses such as naming and matching could be arranged to test combinatorial entailment. For example, an equivalence task could involve three stimuli that are the same in some respect. A child could be taught to say a name when presented with an actual object. The child could also be taught to say the same name when presented with a picture of the object.

Combinatorial entailment could be tested through a matching task (a form of listener responding). The child would select the picture when presented with the object or select the object when presented with the picture. This is similar to the matching task used in the current study to assess social validity. However, almost all children in this study performed perfectly on the matching task. To assess the relationship between performance on mutual and combinatorial entailment using a matching task, researchers would need to recruit younger children than those in the current study or children with developmental delays. These populations of children would be more likely to produce more variable results on a matching task allowing analysis of relationships among mutual entailment and combinatorial entailment skills.

### Implications for future research

In this section, we explore theoretical and practical implications of the findings concerning each research question.

In the first two research questions, performance of preschool children trained in either listener or expressive responses was compared to chance. Performance was significantly better than chance for both mutual entailment and intraverbals in all comparisons except for the ELL subgroup. These findings are important both theoretically and practically. First, these findings clarify some of the inconsistencies in previous studies (Luciano, 1986; Miguel et al., 2005; Partington & Bailey, 1993; Petursdottir et al., 2008). In previous studies some, but not all, children produced untrained intraverbal responses following training in names and categories. Since these studies used single subject designs and included a small number of participants, broad conclusions about the ability of preschool age children (as a population) to perform untrained intraverbal skills were extremely tentative.

The current study shows that many preschool age children are capable of producing both mutual entailment and intraverbals having been trained in only item names and categories. This finding adds support to the assertions of RFT. It seems that by preschool age, many children have learned the underlying relationships among names and intraverbals and are able to produce untrained intraverbals after learning only names and categories. It is also important to note that there is variability in performance of mutual entailment and untrained intraverbals in both training groups. Although *some* children performed all 8 mutual entailment and intraverbal responses flawlessly, some

children performed at lower levels of accuracy. For mutual entailment, 18% of children in the Expressive group and 91% of children in the Listener group made errors. For intraverbals, errors were made by 65% of children in the Expressive group and 94% of children in the Listener group. This suggests that the skills of deriving mutual entailment and intraverbal responses are not firm in all preschool age children. Future interventions could be focused on teaching these skills for children who are not firm in mutual entailment and intraverbal performances.

Within the Listener training group, it is notable that a small number of children (3 out of 34) were able to perform both mutual entailment and intraverbals perfectly (all eight items correct). This is notable because children in Listener training were not as highly selected as those in Expressive training and because Listener training appears to be a relatively weak form of instruction. These children would likely have an advantage in learning new names and intraverbals compared to other children since they learn more from activities that do not require expressive responding. This could be important since not all children can be expected to produce expressive responses in all activities. Some children are less talkative than others and do not engage in expressive responses regularly. In other cases, large class sizes may prevent teachers from frequently eliciting expressive responses from all students. Many potential learning experiences also take place outside of the classroom (e.g. home, community, etc.) and may not require expressive responding. It would be beneficial to learn why some children were able to learn so much from Listener training. This information could support future interventions to increase the amount of information gained from less than optimal teaching situations



and natural learning experiences outside of the classroom in which not all children are given the chance to practice expressive skills.

The fourth research question concerns the relationship between mutual entailment and intraverbal responses. This relationship has been largely unexplored in the existing literature. In the current study, most children performed mutual entailment at levels equal to or higher than intraverbals. Only 4 out of 68 children scored higher on intraverbals than on mutual entailment. This demonstrates that mutual entailment develops earlier than intraverbals, at least with the tasks used in this study. This pattern of acquisition of mutual entailment and intraverbal responses adds clarity to RFT. According to Hayes et al. (2001, p. 31) “it seems highly likely that combinatorial entailment usually emerges slightly later in language training than mutual entailment due to its complexity and training history. In principle, however, combinatorial entailment need not be linearly related to mutual entailment.” The study provides empirical evidence for this pattern of acquisition in a categorization task. However, additional research is needed to further explore the nature of the relationship between mutual entailment and intraverbals. For example, it is unknown whether mutual entailment develops earlier than intraverbals because it is a simpler relation or whether it is a prerequisite skill. It is still unknown whether *improvements* in mutual entailment skills have a parallel effect on performance of intraverbal skills. It should be determined whether targeting an intervention at mutual entailment will also have effects in intraverbal performance even though intraverbals are not directly taught in the intervention.

The ability to produce mutual entailment and untrained intraverbals may be important targets for early language interventions. Strand, Barnes-Holmes, and Barnes-

Holmes (2003) have suggested that the ability to derive relational responses (such as mutual entailment and combinatorial entailment) may underlie performance in academic domains such as general intelligence, problem solving, vocabulary, and analogies.

Research has also shown that the ability to perform relational tasks is correlated with vocabulary and arithmetic performance in adults (O’Hora, Paleaz, & Barnes-Holmes, 2005). Currently, little research is available concerning the effect of improving mutual entailment and untrained intraverbal skills on other academic domains. It is an empirical question whether targeting a child’s ability to produce mutual entailment and untrained intraverbals will result in improved academic performance such as vocabulary or arithmetic abilities or other related cognitive abilities such as problem solving and analogy. Future researchers should attempt to increase mutual entailment and untrained intraverbals through interventions and test whether parallel increases are also found in other academic skills.

Several studies have already been conducted to explore the kinds of specific instruction that might improve mutual entailment skill. One study (Greer, Stolfi, Chavez-Brown, & Rivera-Valdes, 2005) focused specifically on improving children’s ability to perform mutual entailment when trained in listener responses. In this study, children between two and four years old were taught a listener response for a set of pictures. The listener response required the child to match pictures after hearing the instruction to “Match (name).” The child placed a picture on top of the corresponding picture out of an array of two pictures. The ability to then say the name of the picture when shown the picture (mutual entailment) was evaluated. Following pretests on mutual entailment, the children received multiple exemplar training which included teaching the children to

produce both listener and expressive responses for several pictures. Mutual entailment improved for all participants following multiple exemplar training. This study shows that it is possible to increase mutual entailment for a naming response by teaching both listener and expressive responses for multiple exemplars.

Taken together, the results of the current study and the study by Greer et al. (2005) suggested that children at the preschool age level may benefit from an intervention to improve mutual entailment, which can be taught through multiple exemplar training. The current study suggests that within the Head Start preschool population some children, especially those in the Listener group, do not perform mutual entailment or intraverbals reliably and may benefit from training in these skills. Greer et al. suggested that children who do not perform mutual entailment reliably may benefit from multiple exemplar training. The next study in this line of research should assess the effects of multiple exemplar training on intraverbals. If intraverbals do not improve when multiple exemplar training is implemented on mutual entailment, then researchers should assess whether multiple exemplar training directed specifically at intraverbals will improve intraverbal performance in a categorization task.

In conclusion, this study provides a foundation for a line of research on features of RFT that have direct applications in preschool settings. Ideally a line of research following the current study would include (a) studies to determine whether teaching the skill of mutual entailment has a positive effect on intraverbals, (b) studies to determine whether we can reliably teach children in a way that will enable them to derive relations (mutual entailment and intraverbals) from relatively weak Listener training, and (c) studies to determine if teaching mutual entailment and intraverbals has positive effects on

future academic performance. The lines of research just described would be aimed at producing a set of practical interventions for improving language skills that are important for language and academic performance.

The lines of research described will be particularly important for children who are at-risk for academic failure. If children can be taught the underlying relationships in language such as the relationship between naming and intraverbals, interventions can be used to increase the efficiency of language learning for children at-risk for school failure, and the chances of academic failure may be decreased. This might be accomplished by structuring preschool curriculum to include interventions based on the underlying relationships of language specified by RFT. Whether this possibility will become reality and what these interventions will include will be dependent on the outcomes of future research.

## REFERENCES

- Bear River Head Start. (2008-2009). *HELP*. (Available from the Bear River Preschool Head Start, 95 West 100 South #240, Logan, UT 84321)
- Blalock, H. M., Jr. (1972). *Social statistics*. New York: McGraw-Hill.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research: An introduction*. Boston: Pearson Education.
- Gathercole, S. E. (1999). Cognitive approaches to the development of short-term memory. *Trends in Cognitive Sciences*, 3(11), 410-419.
- Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year-old. *Cognition*, 4(2), 189-202.
- Greer, R. D., Stolfi, L., Chavez-Brown, M., & Rivera-Valdes, C. (2005). The emergence of the listener to speaker component of naming in children as a function of multiple exemplar instruction. *The Analysis of Verbal Behavior*, 21, 123-134.
- Hall, G., & Sundberg, M. L. (1987). Teaching mands by manipulating conditioned establishing operations. *The Analysis of Verbal Behavior*, 5, 41-53.
- Hanley, G. P., Iwata, B. A., & Lindberg, J. S. (1999). Analysis of activity preferences as a function of differential consequences. *Journal of Applied Behavior Analysis*, 32(4), 419-435.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experiences of young American children*. Baltimore: Paul H. Brookes.

- Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition*. New York: Kluwer Academic/Plenum.
- Hayes, S. C., Zettle, R. D., & Rosenfarb, I. (1989). Rule-following. In S.C. Hayes (Ed.), *Rule-governed behavior: cognition, contingencies, and instructional control* (pp. 191-220). New York: Plenum Press.
- Horne, P. J., Hughes, J. C., & Lowe, C. F. (2006) Naming and categorization in young children: IV: Listener behavior training and transfer of function. *Journal of the Experimental Analysis of Behavior*, 85(2), 247-273.
- Horne, P. J., Lowe, C. F., & Randle, V. R. (2004). Naming and categorization in young children: II. Listener behavior training. *Journal of the Experimental Analysis of Behavior*, 81(3), 267-288.
- Kelley, M. E., Shillingsburg, M. A., Castro, M. J., Addison, L. R., & LaRue, R. H., Jr. (2007). Further evaluation of emerging speech in children with developmental disabilities: Training verbal behavior. *Journal of Applied Behavior Analysis*, 40(3), 431-445.
- Lamarre, J., & Holland, J. G. (1985). The functional independence of mands and tacts. *Journal of the Experimental Analysis of Behavior*, 43(1), 5-19.
- Lipkens, R., Hayes, S. C., & Hayes, L. J. (1993). Longitudinal study of the development of derived relations in an infant. *Journal of Experimental Child Psychology*, 56, 201-239.

- Lowe, C. F., Horne, P. J., & Hughes, J. C. (2005). Naming and categorization in young children: III. Vocal tact training and transfer of function. *Journal of the Experimental Analysis of Behavior, 83*(1), 47-65.
- Luciano, M. C. (1986). Acquisition, maintenance, and generalization of productive intraverbal behavior through transfer of stimulus control procedures. *Applied Research in Mental Retardation, 7*, 1-20.
- Miguel, C. F., Petursdottir, A. I., & Carr, J. E. (2005). The effects of multiple-tact and receptive-discrimination training on the acquisition of intraverbal behavior. *The Analysis of Verbal Behavior, 21*, 27-41.
- O'Hora, D., Pelaez, M., & Barnes-Holmes, D. (2005). Derived relational responding and performance on verbal subtests of the WAIS-III. *The Psychological Record, 55*, 155-175.
- Partington, J. W., & Bailey, J. S. (1993). Teaching intraverbal behavior to preschool children. *The Analysis of Verbal Behavior, 11*, 9-18.
- Petursdottir, A. I., Carr, J. E., Lechago, S. A., & Almason, S. M. (2008). An evaluation of intraverbal training and listener training for teaching categorization skills. *Journal of Applied Behavior Analysis, 41*(1), 53-68.
- Petursdottir, A. I., Carr, J. E., & Michael, J. (2005). Emergence of mands and tacts of novel objects among preschool children. *The Analysis of Verbal Behavior, 21*, 59-74.
- Skinner, B. F. (1957) *Verbal behavior*. Acton, MA: Copley.
- Skinner, B. F. (1966) An operant analysis of problem solving. In B Kleinmuntz (Ed.), *Problem-solving: Research, method, and theory* (pp. 225-257). New York: Wiley.

- Strand, P. S., Barnes-Holmes, Y., & Barnes-Holmes, D. (2003). Educating the whole child: Implications of Behaviorism as a science of meaning. *Journal of Behavioral Education, 12*(2), 105-117.
- Wagner, R. K., Torgesen, J. K., & Rashotte, C. A. (1999). *Comprehensive test of phonological processing*. Austin, TX: PRO-ED.
- Wynn, J. W., & Smith, T. (2003). Generalization between receptive and expressive language in young children with autism. *Behavioral Interventions, 18*, 245-266.



APPENDICES

Appendix A

Example of Test Stimuli Page

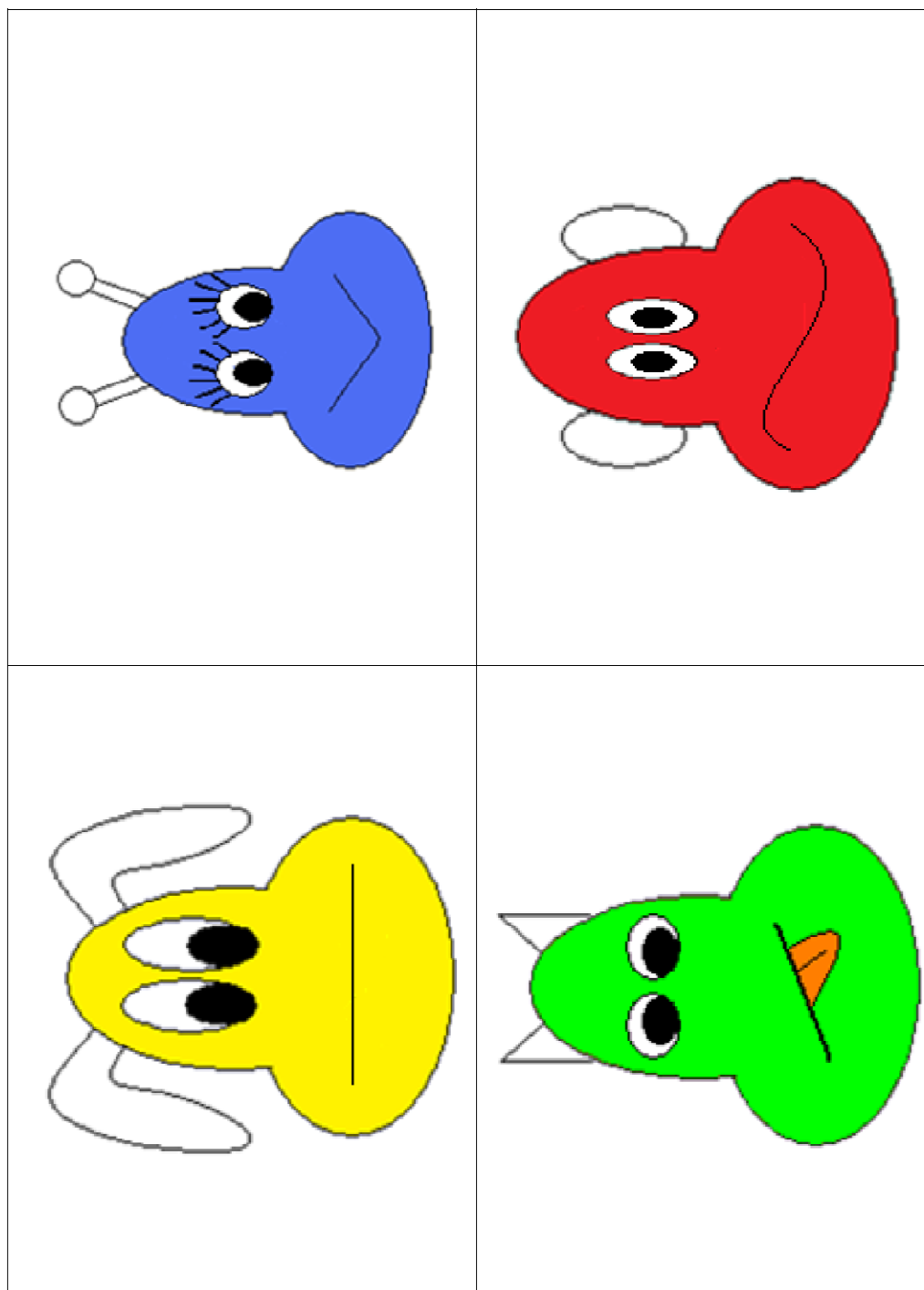


Figure A1 Example stimuli page used in training and testing procedures.

Appendix B  
Example of Stamp Cards



*Figure B1.* Example of stamp card used in motivation procedures.

Appendix C  
Data Sheets

Teaching Method: Expressive/Listener

Picture Set: Monsters

Instructions for teaching portion of session:

- If the child says/points to the correct name/category – place a check mark (✓) in the column for that character
- If the child says/points to the wrong name/category – place an (X) in the column
- If the teacher models the correct answer following a wrong response – place an (M) beside the (X)
- The first response following a model will count as a separate response and will be marked as either correct or wrong in the next space on the data sheet.
- When a stamp is delivered, circle the last response marked before the stamp was delivered.

Instructions for mutual entailment test:

Place a check mark (✓) in the blank beside each name/category if the child says/points to the correct picture when told the name/category. Place an (X) in the blank if wrong.

Note: For each category, there are two correct answers.

Instructions for intraverbal test:

Place a check mark (✓) in the blank beside each name if correct. Place an (X) in the blank if wrong.

Note: Each category has two correct answers.

Instructions for delayed mastery test:

Place a check mark (✓) in the blank beside each name/category if the child correctly says/points to the picture name/category. Place an (X) in the blank if wrong.

Note: For each category, there are two correct answers.

Instructions for matching test:

If the child points to the picture in the same category as the sample put a check mark (✓) in the blank next to the item name. If the child points to a picture in a different category than the sample, place an (X) in the blank next to the item name.

## Item Names

	Kip	Lou	Bo	May
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				



## Category Labels

	Sing		Dance	
	Kip	Bo	Lou	May
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				
17				
18				
19				
20				
21				
22				
23				
24				
25				
26				
27				
28				
29				
30				

**Mutual Entailment**

Does the child correctly say/point to:

Lou \_\_\_\_\_

Bo \_\_\_\_\_

Kip \_\_\_\_\_

May \_\_\_\_\_

Does the child correctly say/point to:

Dance \_\_\_\_\_ (Lou) \_\_\_\_\_ (May)

Sing \_\_\_\_\_ (Bo) \_\_\_\_\_ (Kip)

Does the teacher provide corrective feedback (e.g. model) for any responses? \_\_\_\_ yes \_\_\_\_ no

Is a stamp provided upon completion of this test? \_\_\_\_ yes \_\_\_\_ no

**Intraverbals**

Does the child say the correct category for each of the item names?

Bo \_\_\_\_\_ (Sing)

May \_\_\_\_\_ (Dance)

Kip \_\_\_\_\_ (Sing)

Lou \_\_\_\_\_ (Dance)

Does the child say the correct name for each of the categories?

Sing \_\_\_\_\_ (Bo) \_\_\_\_\_ (Kip)

Dance \_\_\_\_\_ (May) \_\_\_\_\_ (Lou)

Does the teacher provide corrective feedback for any responses? \_\_\_\_ yes \_\_\_\_ no

Is a stamp provided upon completion of this test? \_\_\_\_ yes \_\_\_\_ no

**Delayed Mastery**

Does the child correctly say/point to:

Lou \_\_\_\_\_

Kip \_\_\_\_\_

May \_\_\_\_\_

Bo \_\_\_\_\_

Does the child correctly say/point to:

Dance \_\_\_\_\_ (May) \_\_\_\_\_ (Lou)

Sing \_\_\_\_\_ (Bo) \_\_\_\_\_ (Kip)

Does the teacher provide corrective feedback for any responses? \_\_\_\_ yes \_\_\_\_ no

Is a stamp provided upon completion of this test? \_\_\_\_ yes \_\_\_\_ no

**Category Matching**

Does the child point to the correct picture?

Kip \_\_\_\_\_

Lou \_\_\_\_\_

Bo \_\_\_\_\_

May \_\_\_\_\_

Does the teacher provide corrective feedback for any responses? \_\_\_\_ yes \_\_\_\_ no

Is a stamp provided upon completion of this test? \_\_\_\_ yes \_\_\_\_ no

Appendix D  
Treatment Fidelity Checklists

___ out of 6	New items introduced after 3 consecutive corrects on previous item.
___ out of 1	Name mastery reached before categories introduced.
___ out of 1	Category mastery reached before tests
___ out of 1	All mistakes corrected with a model
___ out of 8	Stamps awarded for each item learned
___ out of 4	No corrective feedback during tests
___ out of 4	Stamps awarded after each test
___ out of 25	Percentage Fidelity = _____

*Figure D1.* Assessment and training steps fidelity checklist.

___ out of 1	Instruction read accurately
___ out of 1	Feedback provided on practice items
___ out of 1	No feedback provided on test items
___ out of 1	Ceiling reached (3 consecutive incorrects)
___ out of 4	Percentage Fidelity = _____

*Figure D2.* Working memory fidelity checklist.

Appendix E  
Statistical Procedures

*Kolmogorov-Smirnov Test.* Several nonparametric tests are designed to evaluate whether two groups perform differently on a dependent variable. The Mann-Whitney test is a nonparametric test designed for this purpose. The Mann-Whitney test depends on ranking the scores of participants in each group. Ranks are assigned based on the magnitude of difference scores. We determined this test to be inappropriate for this data set due to the large number of tied ranks. Since only nine values are possible in each dependent variable, many participants will have equal difference scores and thus equal ranks. For the Mann-Whitney test, a large number of tied ranks can bias the test statistic. This bias is consistently in the conservative direction making significant differences between groups more difficult to identify. Blalock (1972) recommends the Kolmogorov-Smirnov test in situations with numerous tied ranks because this test is less affected by tied ranks. Although the Kolmogorov-Smirnov test is also biased in the conservative direction by a large number of tied ranks, this bias is less pronounced than in the Mann-Whitney test (Blalock, 1972).

The Kolmogorov-Smirnov test is used to detect differences between scores of two groups on a single variable. The null hypothesis assumes that the cumulative frequency distributions for the two samples are equal, which indicates that the groups are from the same population. In a cumulative frequency distribution, instead of representing the number of cases within an interval as in a typical frequency distribution, this distribution represents the proportion of cases equal to or below an interval. The Kolmogorov-Smirnov test results in a statistic that is the maximum difference between the cumulative frequency distributions for each group. The maximum difference is the largest difference between proportions of participants falling at or below each possible rank for each group.

It is calculated by subtracting the lower proportion from the higher proportion at each value of rank in a cumulative frequency distribution and selecting the largest difference. The  $Z$  statistic is calculated to compare the maximum difference to a normal approximation of maximum differences. If the maximum difference is larger than would be expected by chance, then the null hypothesis can be rejected.

*Wilcoxon Signed Ranks Test.* The Wilcoxon Signed Ranks test does not assume a normal distribution. In this test, participants are assigned a rank based on a comparison to chance. Those scoring less than chance are assigned a negative rank and those scoring higher than chance are assigned a positive rank. Difference scores are calculated by comparing each child's obtained score to chance. The specific value of the rank is dependent on the size of the difference score. A participant scoring three points above chance would receive a higher rank than a participant scoring two points above chance. Those scoring equal to chance are ignored in this calculation. Ranks are then summed and the null hypothesis is tested. This test evaluates the null hypothesis that the sum of negative ranks (using absolute values) equals the sum of positive ranks. If these sums are significantly different, then the null hypothesis can be rejected. The Wilcoxon Signed Ranks Test results in a  $Z$  statistic which is evaluated for significance based on a normal approximation of sums of ranks.

Appendix F  
Delayed Mastery Subanalysis



We conducted the delayed mastery subanalysis by first analyzing each child's scores on delayed mastery, mutual entailment, and intraverbal tests. If a child made an error on a delayed mastery item, all questions on the mutual entailment and intraverbal tests requiring knowledge of that response were excluded from the analysis. For example, if a child made an error on the delayed mastery item for the character named Bo, then the mutual entailment item for the character named Bo would be excluded. The two intraverbal items in reference to Bo would also be excluded. Following identification of items to be excluded from each child's tests, we calculated percentages of correct answers from the remaining items for each dependent variable. Percentages were necessary since the number of questions on each test differed across children. These percentages were used to analyze this modified data set on each of the research questions.

*Research Questions 1 and 2.* We did not calculate comparisons to chance in this subanalysis. This is because the number of questions for each participant on each dependent variable is determined by the number of delayed mastery questions answered correctly. This makes chance calculation specific to each child and each dependent variable.

*Research Question 3.* We repeated the comparison between Expressive and Listener training groups using a subset of mutual entailment and intraverbal items determined by performance on the delayed mastery test. The results parallel that of the analysis within the full set of data. Figures F1 – F3 show the dot plot distributions for mutual entailment and intraverbals in each training group. As can be seen in these graphs, the Expressive group scored better than the Listener group on both mutual entailment and

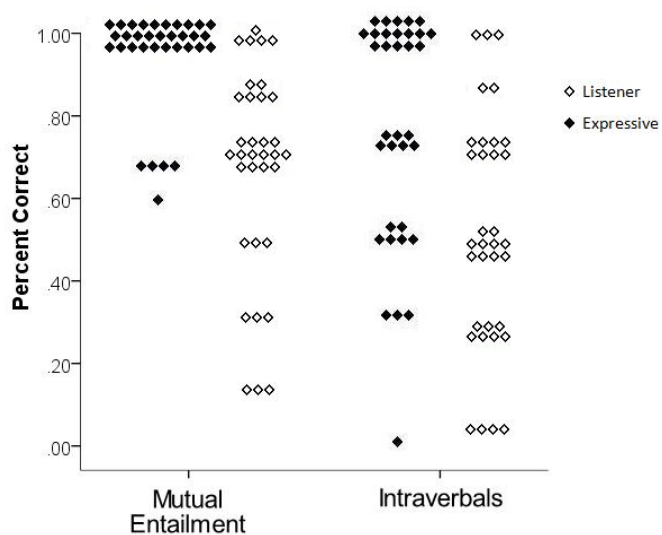


Figure F1. Dot plot distributions of mutual entailment and intraverbals for all participants (Listener and Expressive).

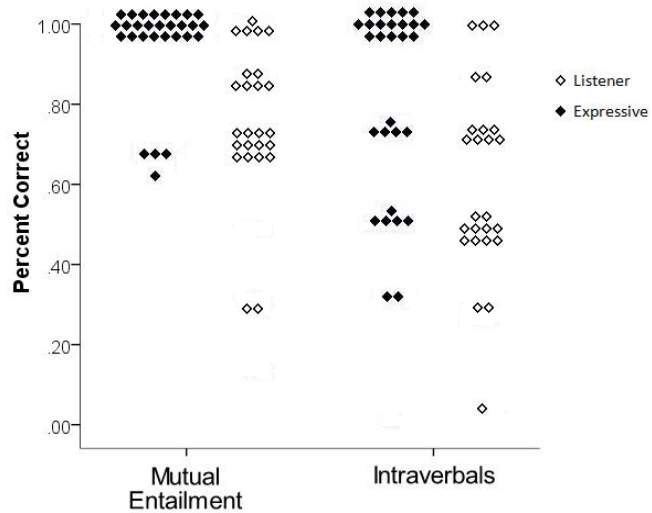


Figure F2. Dot plot distributions of mutual entailment and intraverbals for PES (Listener and Expressive).

intraverbals. This is true of the PES and ELL subgroups as well. In fact, many children in the Expressive group obtained the highest score possible on each dependent variable while few children did so in the Listener group. The results of the statistical analysis are reported in Table F1. The Kolmogorov-Smirnov test was used to test for statistical significance. The Expressive group performed significantly better than the Listener group on both mutual entailment and intraverbals with large effect sizes in both comparisons. This is also true of the PES subgroup. In the ELL subgroup, the Expressive group scored significantly better than the Listener group in only mutual entailment. The difference in intraverbal performance between the Expressive and Listener groups for ELLs was not significant. However, the power of this statistical comparison may have been reduced by the small number of participants in the ELL group ( $n = 14$ ).

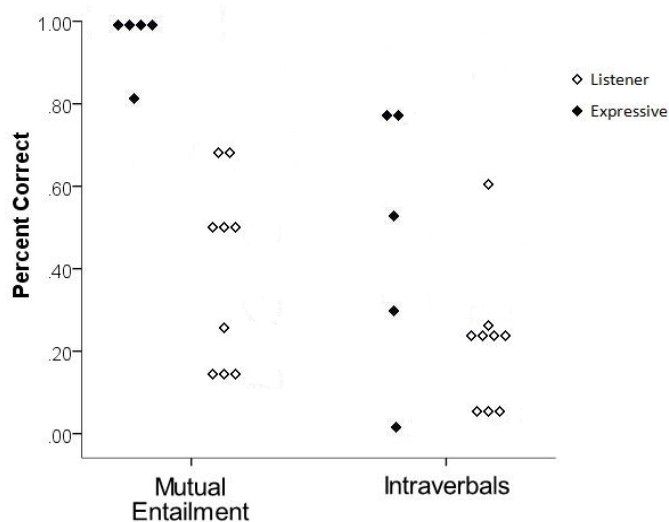


Figure F3. Dot plot distributions of mutual entailment and intraverbals for ELL (Listener and Expressive).

Table F1

*Kolmogorov-Smirnov Test Comparing Expressive and Listener Groups*

	<i>N</i>	Max. absolute difference	<i>Z</i>	<i>p</i>	<i>d</i>
All participants					
Mutual entailment	68	.676	2.79	<.001	1.43
Intraverbals	68	.412	1.70	.006	.89
PES					
Mutual entailment	54	.628	2.30	<.001	1.21
Intraverbals	54	.426	1.56	.015	.83
ELL					
Mutual entailment	14	.889	1.59	.012	2.76
Intraverbals	14	.489	.88	.426	.93

*Research Question 4.* Scatterplots in Figures F4 – F9 are divided into three sections: (a) children who scored better on intraverbals than on mutual entailment (upper portion of graph), (b) children who scored better on mutual entailment than on intraverbals (lower portion), and (c) children who scored the same on mutual entailment and intraverbals (along the broken line at  $x = y$ ). In the Listener training group (Figure F4), most children (94%) produced mutual entailment scores that were equal to or better than intraverbal scores. Only 2 (6%) scored in the upper portion of the graph indicating

better scores on intraverbals than on mutual entailment. The correlation for this comparison is statistically significant and positive ( $r = .812, p < .001$ )

For the Expressive group (Figure F5), all children produced mutual entailment scores either equal to or better than intraverbals scores. Twenty (59%) of children scored in the lower portion of the graph with better mutual entailment scores and 14 (41%) of children scored along the line with equal mutual entailment and intraverbal scores. The correlation for the Expressive group is also statistically significant and positive ( $r = .554, p = .001$ ).

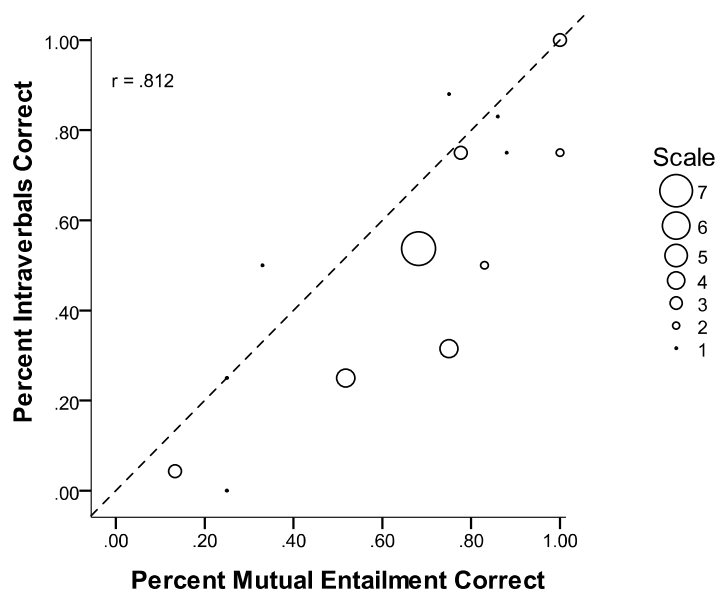
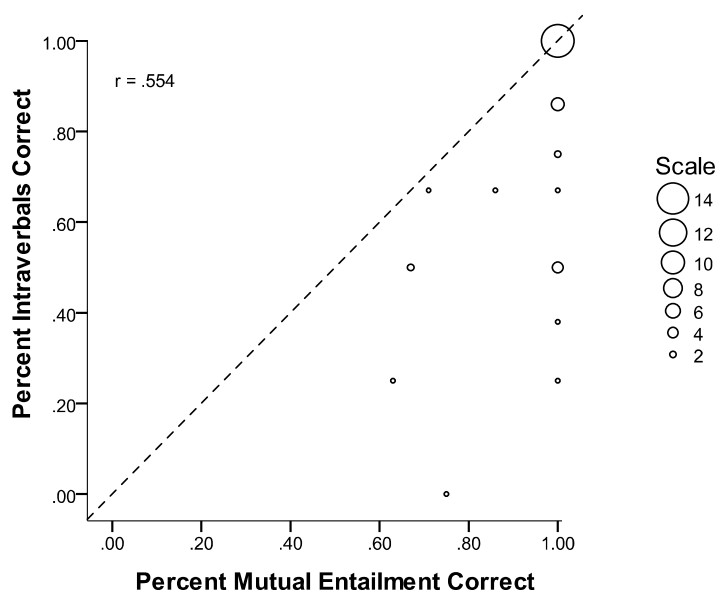


Figure F4. Scatterplot of mutual entailment and intraverbals for all participants (Listener).



*Figure F5.* Scatterplot of mutual entailment and intraverbals for all participants (Expressive).

The PES subgroup results parallel the analysis with all participants. In the Listener group (Figure F6), only two children (8%) scored better on intraverbals than on mutual entailment. All other PES children in the Listener group (92%) scored better on mutual entailment than on intraverbals. In the Expressive group (Figure F7), all PES children scored better on mutual entailment than on intraverbals. Correlations for both Listener ( $r = .726, p < .001$ ) and Expressive ( $r = .600, p = .001$ ) groups are significant and positive.

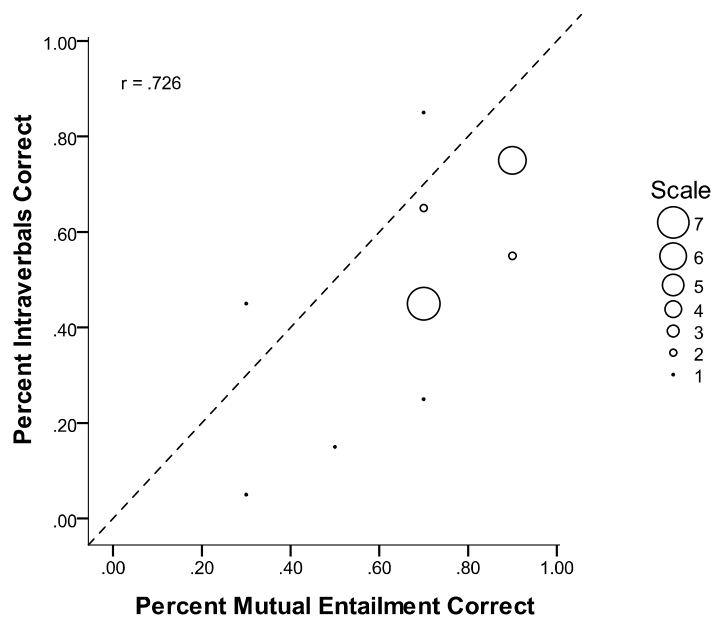


Figure F6. Scatterplot of mutual entailment and intraverbals for PES (Listener).

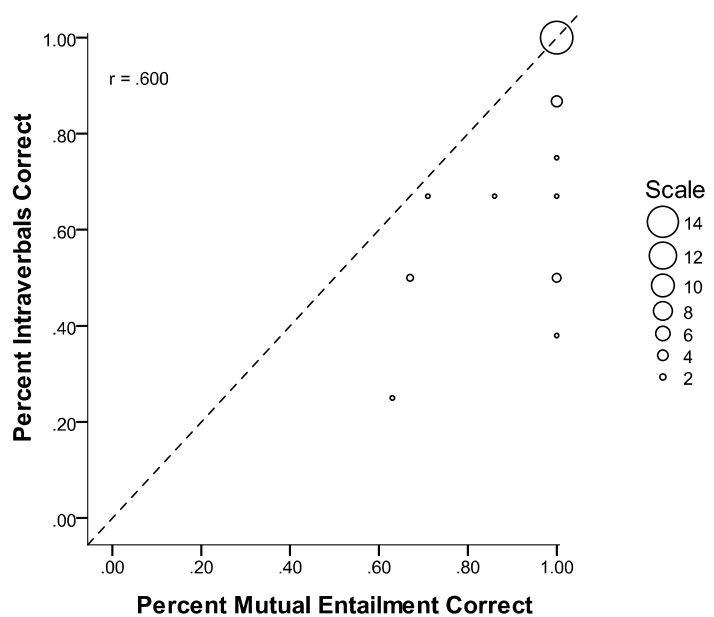


Figure F7. Scatterplot of mutual entailment and intraverbals for PES (Expressive).

All ELL children in both Listener (Figure F8) and Expressive (Figure F9) groups scored equal to or better on mutual entailment than on intraverbals. The correlations between mutual entailment and intraverbals for the Listener group ( $r = .728$ ,  $p = .026$ ) is positive and statistically significant. The correlation for the Expressive group ( $r = .753$ ,  $p = .141$ ) is strong and positive but not statistically significant due to the small number of participants ( $n = 5$ ) in this group.

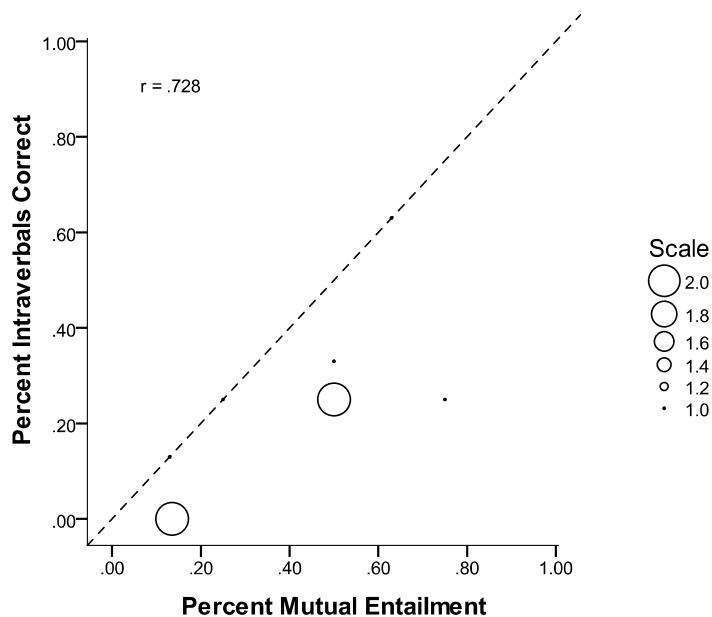


Figure F8. Scatterplot of mutual entailment and intraverbals for ELL (Listener).

*Question 5.* We calculated correlations between working memory and mutual entailment and intraverbals for the delayed mastery subanalysis. These correlations are displayed in Table F2. The results are similar to the analysis using the full data set. Correlations are significant and positive only for the Listener training group. As can be



seen in Figures F10 – F11, these correlations are weak and the relationship between working memory and mutual entailment and intraverbals is somewhat unclear.

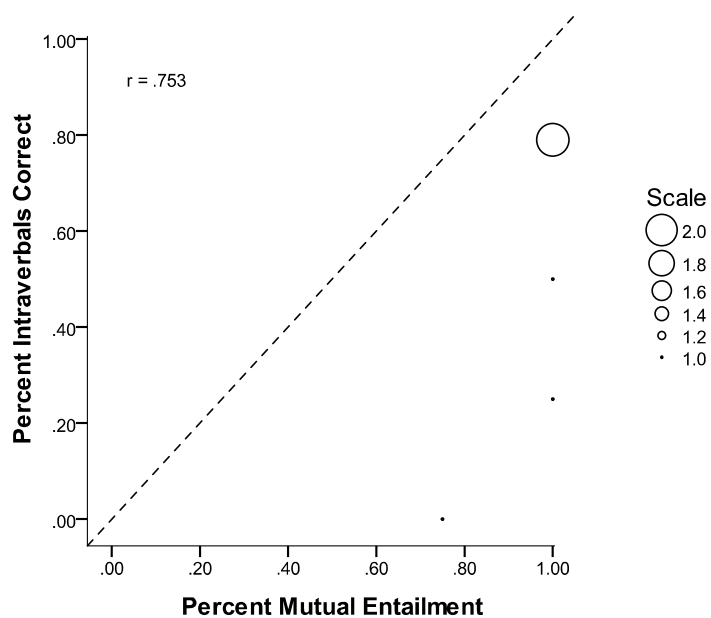


Figure F9. Scatterplot of mutual entailment and intraverbals for ELL (Expressive).

Table F2

*Correlations Between Working Memory and Dependent Variables*

Training group	Mutual entailment	Intraverbals
Listener	.416*	.463*
Expressive	.045	.105

$p < .05$

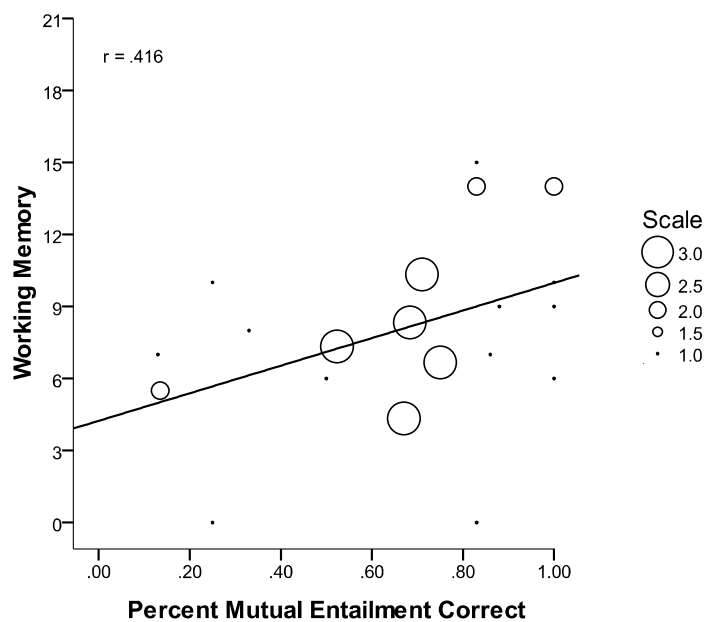


Figure F10. Scatterplot of mutual entailment and working memory (Listener).

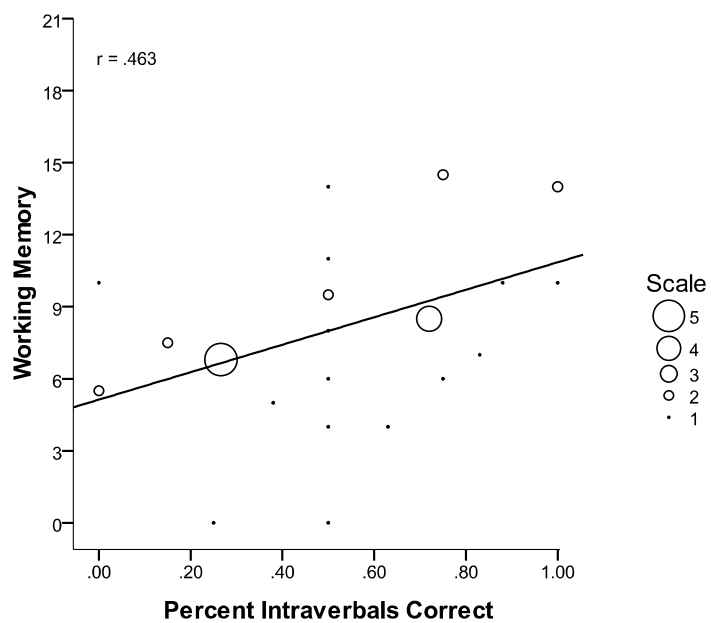
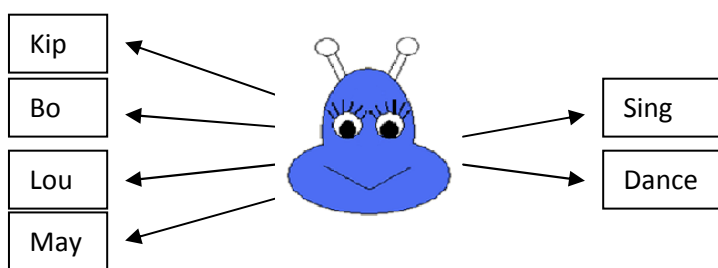


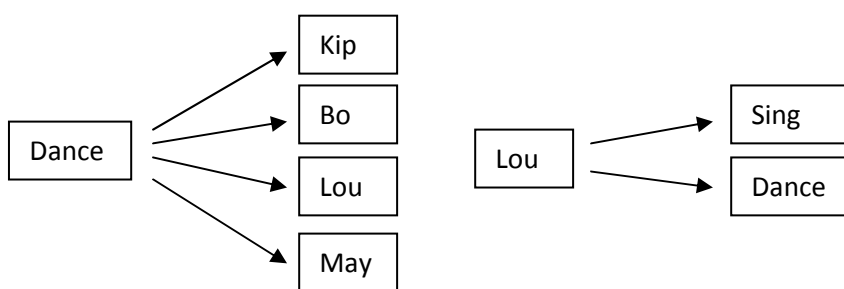
Figure F11. Scatterplot of intraverbals and working memory (Listener).

Appendix G  
Calculation of Chance

We calculated chance for measures of mutual entailment and intraverbals. In chance calculations, we assumed that the children guessed from the pool of four item names and two category labels. Figure G1 shows the possible name and category responses available for each picture in the mutual entailment test. Figure G2 shows the possible name and category responses for the intraverbal test. This is a conservative estimate of chance since it does not take into account the possibility of a child responding with “I don’t know” or similar responses or by providing a name or category outside of those taught in this study.



*Figure G1.* Possible mutual entailment responses for chance calculation.



*Figure G2.* Possible intraverbal responses for chance calculation.

We calculated chance separately for name and category responses. Since there are four choices for item names, the chance of guessing the correct name is .25. Given four

items, the children are likely to guess one answer correctly by chance. For category responses, since there are two choices for category labels, the chance of guessing the correct category is .5. Given four items, the children are likely to guess two answers correctly by chance. Adding name and category responses together results in three as the number of responses equal to chance.

Appendix H  
Contingency Table Analysis

We also conducted a contingency table analysis to explore the relationship between mastery of mutual entailment and mastery of untrained intraverbal responses. Those children who answered all eight mutual entailment or intraverbal (combinatorial entailment) responses correctly were considered to have mastered each response. Those with less than eight correct did not master each response. We conducted this analysis separately for the groups of children who received Expressive and Listener training.

The analysis begins by categorizing participants into those who mastered each type of derived response and those who did not. A contingency table and chi square statistic (test of independence) were used to analyze whether mastery of mutual entailment was independent of mastery of combinatorial (intraverbal) entailment. If these types of responses are independent, then the performance of participants will fall into the cells of the contingency table in proportions similar to those expected by chance (indicated by numbers in parentheses). However, one assumption of the chi square statistic is violated. Expected values in some of the cells of the contingency table are fewer than five. This can increase the likelihood of spurious results in the chi square statistic in either direction. We used the continuity correction in an attempt to prevent spurious effects due to the small expected cell frequencies.

In Table H1 the contingency tables for Listener and Expressive groups are displayed. Almost all of the data for the Listener group fall into the cell for failure to mastery either mutual entailment or intraverbals. Very few children in the Listener group mastered either mutual entailment or intraverbals. In the Listener group mastery of intraverbals is significantly dependent on mastery of mutual entailment,  $\chi^2 (N = 34, 1) = 11.57, p = .001, \text{phi} = .804$ , when the continuity correction is used. In the Expressive

group, almost all children mastered mutual entailment. The degree of relation between mastery of intraverbals and mastery of mutual entailment in the Expressive group falls short of statistical significance when the continuity correction is used,  $\chi^2 (N = 34, 1) = 2.32, p = .128, \phi = .342$ . However, if the chi square statistic had been calculated without the continuity correction, this analysis would also have been considered statistically significant,  $\chi^2 (N = 34, 1) = 3.97, p = .046$ .

Table H1

*Contingency Table Comparing Mastery of Mutual Entailment and Intraverbals*

		Intraverbal		Row sums
		Mastered	Not mastered	
		Listener		
Mutual entailment	Mastered	2 (.2)	1 (2.8)	3
	Not mastered	0 (1.8)	31 (29.2)	31
	Column sums	2	32	34
		Expressive		
Mutual entailment	Mastered	12 (9.9)	16 (18.1)	28
	Not mastered	0 (2.1)	6 (3.9)	6
	Column sums	12	22	34

When this analysis is repeated for only PES children, similar results are evident. Contingency tables for separate training groups are displayed in Table H2. In the Listener



training analysis, mastery of intraverbals is significantly dependent on mastery of mutual entailment,  $\chi^2(N = 34, 1) = 8.17, p = .004, \phi = .799$ , when the continuity correction is used. In the Expressive training condition mastery of intraverbals and mutual entailment responses are not significantly dependent when the continuity correction is used,  $\chi^2(N = 34, 1) = 2.45, p = .117, \phi = .383$ . Similar to the contingency table analysis with the full set of participants, the results for the Expressive group are significant if the continuity analysis is not used,  $\chi^2(N = 34, 1) = 4.27, p = .039$ .

We did not repeat the contingency table analysis for the ELL subset of participants. This is because none of these students mastered the intraverbal response, thus a contingency table and chi square statistic would not indicate the degree of independence of mutual entailment and intraverbals.

Table H2

*Contingency Table Comparing Mastery of Mutual Entailment and Intraverbals (PES)*

		Intraverbal		
		Mastered	Not mastered	Row sums
		Listener		
Mutual entailment	Mastered	2 (.2)	1 (2.8)	3
	Not mastered	0 (1.8)	22 (20.2)	22
	Column sums	2	23	25
		Expressive		
Mutual entailment	Mastered	12 (9.9)	12 (14.1)	24
	Not mastered	0 (2.1)	5 (2.9)	5
	Column sums	12	17	29

## CURRICULUM VITAE

**Ginger Kelso**  
**34 USU Aggie Village Apt. K**  
**Logan, UT 84341**

**Home: (435) 757-3613**  
**Work: (435) 797-3608**  
**Email: ginger.kelso@aggiemail.usu.edu**

---

**Education**

Anticipated August 2009	Ph.D., Disability Disciplines Utah State University, Logan UT Specialization: Special Education Dissertation title: Combinatorial Entailment as a Bridge between Tact and Intraverbal Repertoires
August 2005	M. Ed., Special Education Stephen F. Austin State University, Nacogdoches TX Specialization: Autism Direct Instruction
May 2004	B.A.A.S. Applied Arts and Sciences Stephen F. Austin State University, Nacogdoches TX Specializations: Special Education Human Services

**Awards and Fellowships**

August 2005 – May 2006	Presidential Fellowship - \$15,000 School of Graduate Studies Utah State University
---------------------------	-------------------------------------------------------------------------------------------

**Certifications**

Texas	Special Education EC – 12
-------	---------------------------

## Current Position

- August 2005 – Present
- Graduate Research Fellow**  
 Dept. of Special Education and Rehabilitation  
 Utah State University
- Assist faculty in research projects
  - Assist faculty in course management
  - Teach undergraduate and graduate courses
  - Prepare conference presentations
  - Prepare publishable manuscripts including a comprehensive review of literature
  - Supervise undergraduate or graduate practicum students
- August 2008 – Present
- Utah Regional Leadership Education in Neurodevelopmental Disabilities – Trainee**  
 Center for Persons with Disabilities  
 Utah State University  
 Supervisor: Barbara Fiechtl
- Attend weekly seminars by professionals in fields related to neurodevelopmental disabilities
  - Coordinate and participate in parent directed clinics to provide consultation to parents of children with developmental delays
  - Consult with trainees from various fields in development of leadership projects
- August 2008 – Present
- Technology Advisor – Virtual Home Visits for Children with Developmental Delay**  
 Center for Persons with Disabilities  
 Utah State University  
 Supervisor: Barbara Fiechtl
- Instruct and support families in videoconferencing software usage
  - Develop training manual for virtual home visits
  - Develop online training for virtual home visits
  - Train parents and staff to use software
  - Develop assessment measures
- November 2008 – Present
- Senior Staff Assistant**  
 Autism Support Services: Education, Research, and Training (ASSERT)  
 Center for Persons with Disabilities  
 Utah State University  
 Supervisor: Thomas Higbee

- Direct service provision – discreet trial training and Direct Instruction
- Supervision of staff
- Staff evaluations
- Case coordination
- Curriculum management

## **Professional Experience**

August 2006 –  
Present

### **Clinical Services Graduate Assistant**

Center for Persons with Disabilities

Utah State University

Supervisor: Dr. Robert Cook

- Provide educational expertise in clinical assessments of children with disabilities including developmental disabilities and autism
- Observe children in school and home settings
- Record data concerning child, teacher, and parent behaviors
- Provide summary report of educational and behavioral outcomes of observations
- Provide educational and behavioral recommendations
- Provide care coordination for families of children with disabilities
- Participate in IEP meetings in school for children with disabilities

June 2004 –  
July 2005

### **NCATE Graduate Assistant**

Stephen F. Austin State University

College of Education

Supervisor: Dr. Melanie Jephson, Associate Dean of Education

- Coordinate NCATE Committee meetings
- Organize NCATE documentation
- Train faculty in e-portfolio system
- Train students in e-portfolio system
- Develop PowerPoint presentation for conferences
- Develop Teacher Certification website
- Maintain database for educational certificate plans

## **Publications**

- Kelso, G.L., Fiechtl, B.J., Olsen, S., & Rule, S. (In review). The feasibility of virtual home visits to provide early intervention: A pilot study.
- Kelso, G.L., & Lignugaris-Kraft, B. (In preparation). Viewing mentoring programs through an evidence-based practice lens: Research quality, program outcomes and future directions.
- Kelso, G.L., Slocum, T.A. (In preparation). Language for Learning: Alignment with relational frame theory?

## **Presentations**

- Kelso, G.L., & Slocum, T.A. (May 2009). *Combinatorial entailment as a bridge between tact and intraverbal repertoires*. Paper to be presented at the Association of Behavior Analysis, Phoenix, AZ
- Kelso, G.L., & Lignugaris/Kraft, B. (2008). *Is mentoring evidence based?* Poster presented at the annual meeting of the Association for Behavior Analysis International Education Conference, Reston, VA.
- Lignugaris/Kraft, B., & Kelso, G.L. (2008). *Community and School-Based Mentoring Programs: Perils and Pitfalls in Evaluating the Evidence*. In K.D. Hager (Chair), Evidence-Based Practice Reviews: Applying EBP Standards to Reviews of School-Based Practices. Symposium conducted at the Association of Behavior Analysis Convention, San Diego, CA.
- Fiechtl, B, Olsen, S., & Kelso, G.L. (2007, October). *Virtual home visits: Using technology to provide early intervention services*. International Conference for the Division for Early Childhood, Niagara Falls, Ontario.
- Kelso, G.L. (2007, May). *Language for Learning: A Relational Frame Theory perspective*. In C. Watkins (Chair), Language for Learning, Children with Autism, and Relational Frame Theory. Symposium conducted at the Association of Behavior Analysis Convention, San Diego, CA.
- Stenhoff, D., Kelso, G.L., Alexander, M., & Lignugaris/Kraft, B. (2006, November). *Teacher observation instrument validation: The relationship between teacher behavior and student outcome*. Teacher Education Division of the Council for Exceptional Children Conference, San Diego, CA.
- Kelso, G. L. & Forbush, D. (2006, March). *100% Proficient by 2014: Special Education Students Current and Projected Performance on AYP Measures*. American Council on Rural Special Education, Lexington, KY.

## **Grants (funded)**

Fiechtl, B.J., Olsen, S., & Kelso, G.L. Phase I project: Virtual home visits. Office of Special Education and Rehabilitative Services – Steppingstones of Technology Innovation for Children with Disabilities. \$200,000 per year for two years; 2008-2010.

## **Courses Taught/Assisted**

### **Journal Reading Group – Relational Frame Theory (SPED 7940)**

This is a doctoral level class in which students explore and discuss the literature pertaining to a specific topic. The purpose of this Journal Reading Group is to gain a better understanding of the theories and applications of Relational Frame Theory, with a particular focus on educational applications.

### **Consulting with Parents and Teachers (SPED 5060)**

This course provides preservice professionals with strategies for communicating with parents and other teachers as members of the multidisciplinary team. Students learn to assist parents in communicating feelings and needs, in accessing resources, and in advocacy. Students learn strategies for collaborative problem solving with other teachers including monitoring student progress in the regular classroom and assisting the regular teacher in adapting material.

### **Applied Behavior Analysis 1: Principles, Assessment and Analysis (SPED 5010)**

This course introduces basic principles by which humans learn social, academic, and other skills. Emphasis is placed on defining behaviors, measuring behaviors through direct observation, evaluating interventions via graphic data analysis, and communicating intervention effects to others.

### **Applied Behavior Analysis 2: Applications (SPED 5050)**

This course provides in-depth examination of functional assessment (interview and observation procedures), functional analysis (development of hypotheses based on functional assessment data), behavioral intervention based on functional equivalence, punishment and ethical issues related to implementation, policies governing use of punishment and other behavioral interventions, and assessment/intervention for students with autism and emotional disturbance/behavior disorders. The overarching goal of the course is for the

student to implement and analyze a behavior intervention plan based on a functional assessment with a student in an applied setting.

### **Education of Exceptional Individuals (SPED 4000)**

This course provides basic information on the definitions, causes, characteristics, and educational approaches concerning persons with a variety of disabilities. Students will also become familiar with legal issues surrounding special education.

### **Professional Memberships**

August 2004 – Council for Exceptional Children  
Present

March 2006 – Association for Behavior Analysis  
Present