EFFECTS OF A CONVERSATION INTERVENTION ON EXPRESSIVE VOCABULARY OF YOUNG CHILDREN

by

HILARY P. RUSTON

(Under the Direction of Paula J. Schwanenflugel)

ABSTRACT

Entering school with a low level of vocabulary has been associated with poor academic outcomes (Biemiller, 1999). Young children with larger vocabularies typically live with parents who talk more and with greater sophistication. Introducing this kind of talk into early childhood learning environments may be a strategy for increasing children's vocabulary knowledge (Dickinson, 1991). Current observational research has consistently recorded that preschool and elementary teachers rarely talk in depth with children.

This quasi-experimental study investigated the effects of a vocabulary enriched conversation intervention (500 minutes over 3 months) that took place between an adult and pairs of prekindergarten students. Three vocabulary measures were used for pre- and post-testing; a standardized expressive vocabulary test (EVT), a language sample based measure of vocabulary diversity (D) and a definition task (Schwanenflugel et al, 1996). Conversation guidelines included the intentional introduction of novel vocabulary in context, explicit definitions, joint attention tasks with varied prompts and allowance for child initiated topics.

Students in the experimental group had significantly higher post-test scores on the EVT. When scores from children with lower levels of a priori vocabulary were separately analyzed the effect of the intervention was significant for both D and EVT scores. The groups did not differ significantly on the definition task. The results of the present investigation indicated that a school-based practice of enriched conversation could be an effective strategy for increasing the vocabulary levels of children starting out with less developed vocabularies.

The first study examined the validity of D as a measure of lexical diversity. Messick’s (1989) framework for unitary construct validity was used to evaluate the evidence collected from the study. D performed well on internal consistency and test-retest analyses and was moderately related to the EVT. However D scores varied in relation to different prompts used in language sampling. Overall evidence suggested that D is a valid and reliable measure of lexical diversity and may be useful to educators, language therapists and researchers.

INDEX WORDS: Vocabulary, Intervention, Children, Teaching, School, Preschool, Expressive Vocabulary, Vocabulary Diversity, Vocabulary Measurement, Conversation
EFFECTS OF A CONVERSATION INTERVENTION ON EXPRESSIVE VOCABULARY OF
YOUNG CHILDREN

by

HILARY P. RUSTON

B. S., The School for International Training, 1987
M.A., The University of Georgia, 2004

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2007
EFFECTS OF A CONVERSATION INTERVENTION ON EXPRESSIVE VOCABULARY OF YOUNG CHILDREN

by

HILARY P. RUSTON

Major Professor: Paula J. Schwanenflugel
Committee: Stacey Neuharth-Pritchett
Liang Chen
Steve Cramer

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2007
DEDICATION

Thank you - Lorraine, Kate and Max for being such a wonderful family and supporting me in so many countless ways. Thank you - Mom and Dad for all your contributions and encouragement. I love you all.
ACKNOWLEDGEMENTS

Thank you to Paula Schwanenflugel for her guidance, support, encouragement and enthusiasm – every step of the way. I have been so fortunate to work with such a first-rate scientist, who is an expert in so many areas including educational psychology, child development, linguistics, research, writing, etc… Thank you for providing me with so many rich opportunities to learn about reading, teacher education and child development. Even more important and rare, not only do you believe that good things should happen for children who have odds against them, you work in overdrive for those things to happen. I really value your friendship.

Thank you Liang Chen for helping me with all the language sampling analysis problems that continued to arise. You were always knowledgeable and patient.

Thank you Stacey Neuharth-Pritchett for your guidance, support and kindness throughout this study and for your commitment to children in need.

Thank you Steve Cramer for your inspirational and humorous instruction in statistics and assessment.

Thank you to Brian Richards a primary developer of D, and to Brian MacWhinney the developer of the CHILDES project. Both of you were so kind about answering any question (no matter how basic) quickly and completely.

Thank you Chris, John, Jennifer and Elaine for carrying out this research so professionally and to Betsy and Chris for your help in transcribing. A special thank you to the children, teachers and directors of the child care centers who were so cooperative.
# TABLE OF CONTENTS

ACKNOWLEDGEMENTS............................................................................................................. v

LIST OF TABLES....................................................................................................................... viii

LIST OF FIGURES ....................................................................................................................... ix

CHAPTER

1 INTRODUCTION ......................................................................................................... 1
   Statement of Purpose....................................................................................................... 40

II STUDY 1 ..................................................................................................................... 42
   A Validation Study of an Expressive Vocabulary Measure............................................ 42
   Method........................................................................................................................... 52
   Results .......................................................................................................................... 58
   Discussion ..................................................................................................................... 70
   Conclusion..................................................................................................................... 75

III STUDY 2 ..................................................................................................................... 78
   Talk Time, an Intervention for Enhancing Expressive Vocabulary using
   Vocabulary Enriched Talk........................................................................................... 78
   Method........................................................................................................................... 79
   Results .......................................................................................................................... 82
   Discussion ..................................................................................................................... 91
LIST OF TABLES

Table 1.1: Comparison of Adult-Child Conversation at Home and School in Wells Study ..........12
Table 2.1: Demographics of Participants .......................................................................................54
Table 2.2: Correlations between Total D Scores and Split Half Odd and Even Tokens ..........65
Table 2.3: Full Sample Compared to Half Sample .........................................................................66
Table 2.4: Intercorrelations between Measures of Vocabulary .....................................................68
Table 2.5: Means, Standard Deviations of D and TTR Scores ......................................................69
Table 3.1: Means and Standard Deviations on Assessments for Experimental and Control Groups ..............................................................................................................................................................83
Table 3.2: Means and Standard Deviations of EVT Scores by A Priori Vocabulary Levels.......86
Table 3.3: Repeated-Measures ANOVA Results on EVT Scores by A Priori Vocabulary Levels ............................................................................................................................................................................85
Table 3.4: Means and Standard Deviations of D Scores by A Priori Vocabulary Levels .............88
Table 3.5: Repeated-Measures ANOVA Results on D Scores by A Priori Vocabulary Levels....88
Table 3.6: Means and Standard Deviations of Definition Scores by A Priori Vocabulary Levels ............................................................................................................................................................................90
Table 3.7: Repeated-Measures ANOVA Results on Definition Scores by A Priori Vocabulary Levels ............................................................................................................................................................................90
**LIST OF FIGURES**

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>Teacher Use of Rare Words by Activity from Cote (2001)</td>
<td>19</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Children’s Use of Rare Words by Activity from Cote (2001)</td>
<td>19</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>Mean D Scores from This Study and Two Others as a Function of Age</td>
<td>63</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>Effects of the Intervention on EVT Standard Scores for Whole Sample</td>
<td>85</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>Change in EVT Scores as a Function of the Intervention for Children with Low versus Average Initial Vocabulary Skill</td>
<td>86</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>Effects of the Intervention on D Standard Scores for Whole Sample</td>
<td>87</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Change in D Scores as a Function of the Intervention for Children with Low versus Average Initial Vocabulary Skill</td>
<td>89</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>Effects of the Intervention on Definition Standard Scores for Whole Sample</td>
<td>90</td>
</tr>
</tbody>
</table>
CHAPTER I
INTRODUCTION

Children’s vocabulary levels are of considerable interest to education researchers and policymakers because of the influence vocabulary has on reading achievement. Vocabulary knowledge is considered by the National Reading Panel (NICHD, 2000) to be one of the five pillars supporting the process of learning to read. The U.S. Department of Education has considered assessment of progress along each of the pillars as important in diagnosing the problems of reading education in our nation. Research over the past few decades exhibits general agreement on the main factors that contribute to the process of learning to read. During early literacy the most influential factors include the alphabetic principle, phonemic awareness, and exposure to vocabulary through literacy materials and wide-ranging experiences. Later, skill development and experience with phonics, fluency, more vocabulary and text comprehension strategies rise in importance (Adams, 1990; Biemiller, 1999; C. E. Snow, Burns, & Griffin, 1998). Three of the pillars, phonemic awareness, phonics and fluency, are considered to be easier to instruct and measure. The two other pillars, vocabulary and text comprehension, each with long-term importance, are constructs which appear to be more complex to define, measure, and instruct.

The level of vocabulary knowledge of preschool children has a pervasive and persistent impact on both early reading achievement and long-term reading comprehension. Children who begin school with small vocabularies are most likely to have difficulty in learning to read and are at risk for long-term reading problems (Copeland & Edwards, 1990). Early oral language skills,
particularly vocabulary knowledge, play an increasing role during early literacy development, and by second grade become a primary determinant of reading success (National Institute of Child Health and Human Development Early Child Care Research Network, 2005). Early problems can have long lasting effects. In one study vocabulary assessments collected in the first grade predicted over 30% of the variance in reading comprehension ten years later in the eleventh grade (Cunningham and Stanovich, 1977).

Despite its fundamental role in literacy development, vocabulary development has had a neglected position within early childhood education. Some factors which may have contributed to the lack of curricular emphasis include the difficulty in quantifying children’s vocabulary levels and related remediation needs, as well as a historical downplay of academics in early education. In the past few decades early child care and learning environments have been oriented toward providing social support or simply a balanced meal and safe play environment (Ramey & Ramey, 1992). Even programs with an explicit remediation intent, such as Head Start, have typically had minimal training requirements for child care providers, focusing mainly on health and safety procedures.

Current educational policy initiatives, such as No Child Left Behind, have engaged educators and researchers in further examining the effects of setting academic goals in Kindergartens. There is increasingly more confidence among educational researchers that foundational literacy skills can be introduced effectively and in developmentally appropriate ways in preschool and Kindergarten classrooms (Schwanenflugel et al., 2005). Studies have provided support for the effectiveness of targeted early intervention efforts on behalf of reading readiness skill development (Hargrave & Senechal, 2000; Whitehurst et al., 1988). However, prior to the second grade, preschool and elementary school teachers have typically emphasized
phonemic, phonological and print awareness (Dickinson, Anastasopoulos, McCabe, Peisner-Feinberg, & Poe, 2003). The development of these skills is, indeed, helpful in preparing children to decode early texts. Nonetheless oral language skills should not be ignored, as they are integrally important to the development of early literacy processes (Roth, Speece, & Cooper, 2002; Storch & Whitehurst, 2002). Evidence supports their amenability to early intervention programming as well (Peisner-Feinberg et al., 2001; Schweinhart & Weikart, 1997) during the preschool period (Ramey & Ramey, 1998; Bates, 1999).

**Large variation of vocabulary knowledge among young children**

The challenge of supporting language development in preschools grows increasingly complex as the nation’s population of preschool-aged children grows increasingly diverse. A growing number of children, up to 63%, are enrolled in regular child care prior to starting elementary school (U. S. Census Bureau Brief, 2005). An increasing number of children are poor, with up to 19.8% of children under 6 years old in the United States living in low-income households (with incomes less than 200% of the poverty level) (National Center for Children in Poverty, 2004). The proportion of children in care who are learning English as a second language and children born to new immigrant families has also steadily increased (Annie E. Casey Foundation, 2005; Hodgkinson, 2003). The numbers of young children entering schools from homes in which English is not spoken, has increased from 13.9% in 1990 to 18.4% in 2000 (U. S. Census, 2005). Both poverty and lack of English language knowledge most often directly relate to the need for language support and enhancement.

The variation of vocabulary development among young children is significantly correlated with socio-economic status (SES) factors (Duncan & Brooks-Gunn, 1994; Smith & Dixon, 1995). Preschool-aged children growing up in poverty have fewer opportunities to
develop vocabulary, beginning school disadvantaged in early literacy foundational skills. By school entry, the vocabulary gap between children from low-income homes and their peers is approximately 6,000 words (Hart & Risley, 1995). Hart and Risley (1995) in their extensive study of language in homes concluded that conversations between children and parents were the most influential contributors to increasing vocabulary prior to school. They and others found strong correlations between parental use of words and their children’s; including quantity, density of rare words, and ratio and quantity of affirmative comments (Weizman & Snow, 2001; Hart & Risley, 1995). Opportunities for meaningful conversation are constrained by poverty related factors such as extensive time spent at work, increased incidence of depression, parental use of intoxicants, young maternal age and lower education levels of parents (Duncan & Brooks-Gunn, 1994).

Children who grow up in lower SES homes are more likely to lack school readiness skills and to score poorly in academic assessments (Shonkoff & Phillips, 2000; Brooks-Gunn, Duncan, Klebanov, & Sealand, 1993). Poverty related factors continue to have a powerful influence on later measures of reading ability and dropout status (McWayne, Fantuzzo, & McDermott, 2004). The U.S. paradigm of preschool has been 'play centered' and is shifting to 'early learning environment' in an effort to build in readiness skills. It will be important to ensure that preschool teachers have training in developmentally appropriate methods of supporting vocabulary and oral language development. Although the rationale supporting the importance of these skills is established, additional scientifically based research on effective strategies is needed.

*How do home environments influence vocabulary and language development?*

*Early language development.* Caregivers influence the language development of infants from its inception. Babies as early as six weeks begin to *coo* in response to encouragement from
a caregiver (Hoff-Ginsberg, 1997). Cooing becomes more varied in quality and extends into laughing and babbling as a part of pre-lexical development. Throughout this period, in most cultures, adults direct and modify speech to infants in ways intended to engage infants in practicing communication, as well as introducing them to rudimentary grammar (Hirsch-Pasek, Kemler-Nelson, Jusczyk, & Cassidy, 1987). First words are not typically produced until the age of 10 to 15 months (Huttenlocher & Smiley, 1987), but, even as early as this, there is important variation that can be observed in the vocabulary of young children.

The quantity and quality of language interactions in the home contribute to differential vocabulary of young children. When caregivers communicate with toddlers they tend to place emphasis on certain speech features which in turn appears to influence the development of toddler vocabularies. In the United States and cultures speaking Romance languages, adults have been observed to emphasize nouns, and in their children’s early vocabularies nouns outnumber other parts of speech. In Asian cultures, however, mothers have been observed to emphasize verbs with a concurrent verb dominance observed in their children’s speech (Au, Dapretto, & Song, 1994; Choi & Gopnik, 1995; Tardif, 1996; Tardif, Shatz, & Naigles, 1997). American caregivers, particularly, tend to prompt for object labels with great frequency. Among families with higher education and income, there is a higher rate of explicit labeling and a concomitant higher number of nouns known by their children (Goldfield & Reznick, 1990). Another indicator of the rate of vocabulary growth in this early stage between 1 and 2 years is the number of words spoken by mothers to children (Huttenlocher et al., 1991). Differences in vocabulary levels related to SES between children are detectable early on, but become significant by the age of three years (Hart & Risley, 1995; Hoff-Ginsberg, 1997). Evidence has accumulated suggesting that lower incidence of sophisticated language and available reading
material in homes and academic environments is correlated with lower child vocabulary measurements (Duncan & Brooks-Gunn, 1997; Weizman & Snow, 2001).

**SES associations with language development.** Features of the home environment and characteristics of family members exert a major influence on language development. Children born into middle and high SES families tend to engage in more interactive discussions with their parents and may participate in more book reading and trips to libraries, museums and even grocery stories. Such children are usually guided through language by a parent with explicit interests in building children’s language skills and knowledge (Hart & Risley, 1995; Hoff, 2003). Language directed at children in low-income families appears to include more imperatives and prohibitions, with an emphasis on teaching obedience in direct contrast to the dialogue and engagement in information sharing found in middle class homes (Hart & Risley, 1995). These differences between families appear to result in the dramatic differences in the vocabulary levels of preschoolers at school entry.

The qualities of early language inputs have enduring effects on children’s vocabulary. Hart and Risley (1995) conducted an extensive home observation study of 42 children living in families they categorized into welfare, working and professional class. Their analysis of this longitudinal study found SES to be an overall predictor of parental input of language, both of quality and quantity, as well as children’s output. They found that although all families devoted similar amounts of talk to socializing around feeding, dressing and safety, the difference in language skills was related to talk that extended beyond the basic needs. This extended talk accompanied activities such as putting together a puzzle, or describing an event at mealtime. Although Hart & Risley’s study (1995) was very informative and data collection methods were very comprehensive, the numbers of children in each group were small, including only 6 children
from the low SES group. Combined with other studies though, we can build the case that SES acts as a predominant contributor to language differences. However, the trends are not always linear.

The importance of the quantity of words spoken in relation to the quality of words is different relative to the family income level. Pan, Rowe, Singer and Snow (2005) studied the language development of 108 children from low-income families between the age of 1 and 3 years. They found maternal lexical diversity to be a significant predictor of later vocabulary development but they did not find ‘talkativeness’, or simply the number of words used by the mother to be predictive of later vocabulary. In another study of low-income families, Weitzman and Snow (2001) found a very strong relationship between the number of sophisticated words parents had spoken to children during preschool and their second grade vocabulary. Additionally, the nature of the conversation made a difference; helpful and instructive conversations were positively associated with larger vocabularies. Together, the density of sophisticated words and the frequency of instructive and helpful conversations accounted for almost two thirds of the variation in children's second grade scores. These findings differ from Huttenlocher et al.’s (1991) findings involving 22 children of middle class families. In middle class samples both diversity and talkativeness explained significant variance. Why the actual number of words spoken does not seem to be as important an influence for low SES families needs further investigation, but it may be due to the higher incidence of directive or negative speech recorded in interactions (Hoff, Laursen, & Tardif, 2002).

Another feature of familial conversation which influences children’s language skill is the use of decontextualized language, meaning language extending beyond concrete descriptions and directions used to discuss more abstract concepts or ideas. Examples of decontextualized
language may include elaborated explanations (e.g. I think our house is safe because a wolf can’t blow it down), personal narratives (e.g. I think if I’m good, I’ll get some candy, because I saw Ellie did) and pretend playing (e.g. I’m going to be the teacher, now sit down!). Families with higher incomes engage in greater usage of decontextualized language with their children (Curenton & Justice, 2004) and those children tend to score higher on language measures at school (Entwisle et al., 1997; Snow et al., 1991). In a sample of 126 African-American teenage mothers receiving Aid to Families with Dependent Children (AFDC), Britto and Brooks-Gunn (2001) found a low percentage (5%) of maternal utterances to be decontextualized. They found strong associations between maternal decontextual language use during book reading and children’s expressive language scores \( r = 0.72, \ p < .0001 \).

A useful indicator of SES is maternal education and most studies attribute significant weight to the influence that the level of maternal education brings to bear on children’s vocabulary (Ensminger & Fothergill, 2003; Restrepo, Schwanenflugel, Blake, Neuhar thor-Pritchett, Cramer, & Ruston, 2006). Strong associations exist between maternal education levels, the number of books in the home and the belief, by mothers, that they can effectively assist in the education of their children (Gadsden, 1995). Maternal communication delivered with the intention of eliciting conversation was shown to be related to larger vocabulary and syntactic complexity in child speech (Hoff-Ginsberg, 1986; Hoff-Ginsberg, 1991; Huttenlocher et al., 2002). Hoff (2003) found a significant correlation between maternal education and the numbers of different words both heard and produced by two year olds.

Parental warmth and attachment indicators have also been linked to language development. The emotional climate surrounding language interactions has been shown to relate to emerging language abilities of children (Edwards & Pleasants, 1997). In a study of story book
reading interactions, the use of warm encouraging language during reading increased language benefits; whereas the use of disciplining language had negative results (Bus & van Ijzendoorn, 1995; Pianta, 1999). The benefits appear to be enduring. Hart and Risley (1995) found that early parental encouragement to verbally communicate resulted in more elaborate patterns of communication persisting long after the age of two years. Positive literacy outcomes are linked to home environments in which children have frequent and varied language interactions with adults (Beals, De Temple, & Dickinson, 1994).

Finally it may be important in relation to child care center staffing and planning to consider whether dyadic and multi-party language interactions have different language effects on children. One study noted that the focus of mother-child talk is different in dyads, centering more on the language, whereas when other children are included talk becomes directed at activities (Oshima-Takane & Robbins, 2003). Most studies have approached this question through comparing language development related to birth order. The occurrence of one-on-one adult-child discourse interactions appears to benefit the vocabulary development of first-born children within families. Fenson (1994) studied information gathered from parents of 1,803 infants and toddlers over time and found small significant values favoring the vocabulary level and growth of children born first. Hoff-Ginsburg (1997b) gathered spontaneous language samples from a group of 63 two year olds and found first-born children used a more diverse vocabulary. First-born children receive more one-on-one attention from adults than later-born children; therefore it is likely that the increased verbal interaction is the factor that contributes to a larger and more sophisticated vocabulary.

An alternative conclusion was drawn from the Home-School Study when contrasting the benefit of multi-party mealtime conversations with dyadic maternal conversations. Dickinson
and others designed the Home-School Study of Language and Literacy Development so that language patterns of 84 children at school and home could be studied (Dickinson & Tabors, 2001). When Dickinson and Snow (1987) tried to figure out which influences were most important in preparing children for 2nd grade reading and language arts, they found a unique contribution from the mealtime conversation. There have been numerous studies showing correlations between families eating dinner together and language benefits. Unfortunately, though, the incidence of family meals is declining. Only about 50% of families, with children ages 9 years and younger, are likely to eat together most days of the week. Working class families often have additional constraints related to long working hours that place them on the lower end of the continuum for eating together and higher on the continuum for the number of television hours watched (Anderson, Wilson, & Fielding, 1988; Bowden & Zeisz, 1997; Gillman et al., 2000).

A study of mealtime conversations among different families shows it to be a time in which varied language skills can be practiced as children take on different roles. They may co-construct an event that they participated in with other family members, they may have to figure out how to add to a sibling’s tale, or even be involved in two conversations at the same time (Blum-Kalka, 1997; Dickinson & Tabors, 2001). They may also have opportunities to listen to different types of language used in familiar contexts, such as irony, indirectness and various types of humor. This setting provides an opportunity for some children to hear dialogue in the home dialect and code-switching as family members recount an event where they might be imitating another person. These interactions tend to be more challenging than dyadic interactions with a mother or caregiver. Primary caregiver dyadic interactions are often supportive and offer scaffolding opportunities with the intention of correcting vocabulary,
grammar and construct, whereas in a dynamic ‘dinner table’ conversation children may be fair

game for teasing and are much less likely to receive facilitation. One aspect of multi-party

conversations with multi-aged participants is the opportunity to practice roles and listen to others

act as experts or critics, and also to observe different styles of managing conversations (Snow &

Blum-Kulka, 2002).

This brief review illustrates some of the features that appear to be most productive for

language development. Although the context of home environments is different than that of

child care centers, it may be a useful starting place from which to consider productive practices

for preschool teachers. In summary, simply increasing the number of words spoken to children

may benefit middle income children but not those in most need of vocabulary development

(Hoff, 2006). All children benefit from involvement in discourse using more sophisticated

vocabulary and syntactic structures (Hart & Risley, 1995; Hoff, Laursen & Tardiff, 2002). Vocabulary outcomes are improved when discourse is situated in a warm, encouraging setting

(Pianta, 1999) and is engaged in one-on-one (Oshima-Takane & Robbins, 2003). Yet, other

language skills may be enhanced by multiparty dialogue particularly as it occurs around sharing

a meal (Snow & Blum-Kulka, 2002).

The effect of preschool settings on language development

Characterizing and quantifying the language interactions between teachers and students

in child care centers is a complex task. Further complicating this line of inquiry is the need to
discern whether highly verbal children are disproportionately engaged by teachers as

conversational partners. Most studies of early childhood education focus on observing overall

instruction or specific activities such as storybook reading. Some studies of child care quality

include a measure of verbal responsiveness between teachers and children, but they rarely
include in-depth distinctions in the types of talk that take place at child care centers (Kontos & Wilcoxon-Herzog, 2002; Dickinson, 1994). These studies find overall that preschool teachers talk relatively infrequently to children, and that some less verbal children receive little of no individual attention and are likely to be conversationally neglected (Kontos & Wilcoxon-Herzog, 1997). In one study of University affiliated pre-schools with low teacher-student ratios (1:4) and high teacher education levels, Wilcoxon-Herzog and Kontos (1998) found that 81% of the time teachers did not talk to children even when they were within three feet of them. This might suggest that teachers place themselves near children for whom behavior intervention may be needed, but they do not engage in conversation. There is a need for more detailed observations of language interactions in today’s preschools. Quality, complexity and frequency of language exposure are important factors measured in home environments and should be carefully examined in the preschool environment.

In Bristol, England, Gordon Wells designed a study to directly compare the conversations at home and school in a preschool intended to be language rich. As part of the study, language samples were collected from home and school, and different aspects of talk were compared as shown in Table 1.1:

Table 1.1
Comparison of Adult-Child Conversation at Home and School in Wells Study (N =32)

<table>
<thead>
<tr>
<th>Characteristics of Talk</th>
<th>Home</th>
<th>School</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean no. of child utterances to adults</td>
<td>122.0</td>
<td>45.0</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Mean no. of adult utterances to child</td>
<td>152.7</td>
<td>128.7</td>
<td>n.s.</td>
</tr>
<tr>
<td>Mean no. categories of semantic content in child speech*</td>
<td>15.5</td>
<td>7.9</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Mean no. of child turns per interaction</td>
<td>4.1</td>
<td>2.5</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Questions (from child)</td>
<td>12.7</td>
<td>4.0</td>
<td>p &lt; .001</td>
</tr>
<tr>
<td>Extending child meaning (from adult)</td>
<td>33.5</td>
<td>17.1</td>
<td>p &lt; .001</td>
</tr>
</tbody>
</table>

* N =16 (p. 193)
The Bristol sample included families from a diverse socioeconomic base, yet overall the homes appeared to provide better linguistic environments than the high quality preschool (Wells & Wells, 1984). This study has not been replicated in the United States, although it would be useful to do so in order to assist teachers in better analyzing classroom language use.

Preschool teachers spend much of their time facilitating play and attending to management. In some settings, over half of teachers' verbalizations center around providing children with assistance in obtaining items, managing behavior, supporting children in peer relationships, praising children for appropriate behavior and providing instructions ((Dickinson & Smith, 1991; Dickinson, de Temple, Hirschler, & Smith, 1992; Kontos & Keyes, 1999). One study of 119 classrooms of 3 and 4 years olds found that children interacted with other children about 50% of the time, and that teachers spent 30% of their time interacting with the class as a whole and only 10% with individual children (Layzer, Goodson, & Moss, 1993). In a study of 63 early childhood education teachers, researchers found that children were notably less engaged in productive activities in classrooms with teachers who were more controlling and used more directive and less interactive language. On the other end of the spectrum, in classrooms where teachers engaged in elaborative and responsive interactions, children sustained longer engagements in activities. Differences in teacher behavior were directly related to education and training (de Kruif, McWilliam, Ridley & Wakely, 2000).

Education levels among child care center teachers are notoriously low. Most children in poverty are cared for in child care centers or family day cares that typically employ caregivers with low levels of education in addition to having high ratios of children to caregivers. Funding levels at these centers provide for no more than minimum wage based salaries that target
unskilled workers, minimally with a General Equivalency Degree or high school education (Laverty, Burton, Whitebook, & Bellm, 2001). These employees however may use less challenging language constructions associated with higher education levels, and higher salaries. Even federal intervention programs such as Head Start, which theoretically emphasize enrichment, have narrowed potential language enrichment by hiring the parents of children at the center, who are not likely to have high levels of education. A lesser degree of education does not always result in the use of less enriched language, but more limited knowledge of development and appropriate child care practice is likely to result in fewer positive teacher-child interactions (Burchinal, Roberts, Nabors & Bryant, 1996; Pianta et al., 2005).

Although there is no research directly relating vocabulary levels of the child care provider and the children in care, Ferguson (1991) found that 25% of the variance in achievement in a study of Texas public schools was linked to the teachers’ scores on the verbal section of the Teacher’s Examination of Current Administrators and Teachers. Current funding levels may serve to further replicate limited discourse experiences by minimizing the involvement of trained and educated teachers. Observational studies find the majority of teacher talk to be related to routine management concerns. This trend increases in centers serving a majority of low-income children (Tabors, Snow, & Dickinson, 2001).

Children’s varied experiences with multiple languages and dialects in preschool. A student's vocabulary, comprehension and ability to converse, has a direct influence on the degree to which they can profit from school. Along with varied and flexible language skills comes the improved ability to form questions and build on existing knowledge (Naude, Pretorius, & Viljoen, 2003). Consistency between home language and school language patterns may facilitate the journey between the two environments (Liontos, 1991). For generations, vocabulary and
comprehension have been taught through storybook reading and discussion in the traditional style and language of European Americans (EUAs) (McCarthey, 1997). Yet, the best literacy outcomes appear to occur when reading and speaking practices at school are consistent with home practices (Moll & Gonzalez, 1994; Neuman & Roskos, 1994). As the child care population continues to ethnically diversify, more flexibly responding to multiple discourse patterns may promote better outcomes.

Children may hear and speak different languages in a variety of ways. Parents may speak a language other than English at home; they may speak a native language and English at home; and teachers and peers may speak only English, or multiple languages. Some children hear two languages spoken from birth and others are exposed later to a second language. Overall there does not appear to be a significant difference in learning one or two languages as long as the level of input from each is adequate and consistent (Genesee, Paradis, & Crago, 2004; Petitto et al., 2001). A monolingual child’s vocabulary is likely to be larger than that of a child learning two languages due to more limited exposure to a single language (Pearson, Fernández, Lewedeg, & Oller, 1997). The same effects appear in the grammar development where the effect of input in each language largely determines the course of development.

Given the persistent achievement gap between African American (AA) and EUA children, additional attention to early differences is merited. Banks and McGee Banks (1997) suggest that part of the achievement gap for AA children may result from the unfamiliarity that EUA teachers’ have with African American English (AAE) learning habits, as well as teachers’ negative attitudes toward cultural and language behaviors of incoming AA children. One study showed that pre-service teachers often enter college with a minimal understanding of multicultural education and typically do not progress very far during their education. In fact
some students leave programs feeling resentful of having to learn about traditions such as Kwanza (Neuharth-Pritchett, Reiff, & Pearson, 2001).

In the early part of this century, scholars approached the study of different dialects as a deficit in the language ability of AA children. Not until the 1960s did linguists, such as William Labov (1972) and Walt Wolfram (1969), establish the validity of AAE as having a logical, interdependent structure with roots as a linguistic Creole. If teachers do not recognize AAE as a separate dialect, they may mistakenly assess its use as indicative of a speech disorder or learning disability (Champion, 2003; Washington & Craig, 1994). Because of the similarities in vocabulary and some elements of grammar, some of the systematic differences between AAE and General American English (GAE) are subtle and difficult to differentiate especially because there exists a wide variety of AAE forms, both regionally and within communities. Increased attention to language features by teacher education programs may help address mis-referrals to special education that are based on dialect (Washington & Craig, 1994; Washington & Craig, 1998).

Language acquisition processes are the same for all typically developing children. Many children, however, enter school from environments in which their only experience has been speaking and listening to AAE. Consequently they must learn a new dialect in which the differences can be difficult to navigate in the midst of communication and beginning school. Wyatt (2001) proposes that children exposed to GAE and AAE by the age of four demonstrate a better ability to "code-switch" in different situations than those who have not had interactive language exposure to GAE. Piestrup (1973) found a very strong negative correlation between the degree of vernacular, which is highly related to SES, spoken and reading achievement. When she studied 200 first grade students, Piestrup found that for children who spoke with a high
degree of dialect, the teacher’s reaction to the child’s language had a significant impact on achievement scores. Teachers who responded positively to language and pointed out differences between GAE and AAE produced students with much higher achievement scores than those who used the interrupting approach and corrected speech, evaluating dialect use solely as being incorrect.

Middle class EUA children are more prone to try to provide the ‘correct’ word in response to questions about objects, or to repeat the expected answer (Champion, 2003), whereas often members of AA culture place a high value on linguistic playfulness and creativity (Smitherman & van Dijk, 1988). These skills may not traditionally be valued (or assessed) in the Pre-K and Kindergarten classrooms and may not be interpreted by teachers as productive in a large group setting. Setting aside of individual time for teacher-child conversation may provide an appropriate forum for talking about words and languages directly and encourage language development.

*Strategies used by preschool teachers to foster vocabulary*

Skills learned prior to Kindergarten have a stable and enduring effect on later school success (Shonkoff & Phillips, 2000). However, teacher surveys reveal that explicit programs directed at vocabulary have not been typically introduced until the 2nd grade, when most children can read (Morrison, Williams & Massetti, 1998). In preschool classrooms, most vocabulary growth occurs incidentally through participation in conversations, listening to storybooks, song lyrics, and even watching educational television (Rice & Woodsmall, 1988; Rice et. al., 1990). However in most preschools, the likelihood of teachers engaging children in conversation appears to be more haphazard than as a strategy to support language development goals. Nonetheless, conversation as a teacher practice shows promise. Wells (1986) found “a
clear relationship between children’s rate of language learning and the amount of conversations they experienced” (p. 44). Of particular importance in Wells analysis is that “the child’s experience of conversation should be in a one-to-one situation in which the adult is talking about matters that are of interest and concern to the child” (p. 44). Research shows that, not only is the quality of teacher talk important, the quantity or frequency of adult-child verbal interactions is important for encouraging children’s oral language development (Hart & Risley, 1995; Huttenlocher, Haight, Bryk, Seltzer, & Lyons, 1991; Wells, 1986). Specifically, the number of utterances directed to children by caregivers and the number of times children initiate conversations with caregivers predicts children’s performance on language measures (McCartney, 1984). There is ample support that children are especially amenable to vocabulary instruction and that methods are differentially effective (Elley, 1989; Hargrave & Senechal, 2000; Whitehurst et al., 1988).

A useful study conducted by Cote (2001) evaluated the differences in language use according to preschool activity by children in Head Start classrooms (N=76). The results suggested that children’s expressive language development can be supported through a variety of activities, but that times when teachers are most likely to intentionally introduce new vocabulary (such as during large group activities) is when children are also most likely to use more rare words. She found that children used the most words during free play, and used the most rare words during large group time and the same pattern was seen in teacher rare word use (shown in Figures 1.1 and 2.1). She also discovered that mealtimes can be organized to be more productive if teachers sit with children as opposed to circulating. In this case, children tend to talk more and the types of conversations are more likely to include decontextualized and playful talk (Cote, 2001).
There is evidence that current pre-school teaching practices do not typically provide high quality language support. One study of 61 lead Head Start teachers found that they discussed the
meaning of words less than 1% of the time (Champion, 2003; Champion, Hyter, McCabe, & Bland-Stewart, 2003). In these same classrooms, the children’s low receptive vocabulary measurements would recommend explicit vocabulary education in addition to other strategies. However, verbal interactions between preschool teachers and children currently tend to be related to routine matters (L. Dunn, Beach, & Kontos, 1994), typically involving concrete talk. Concrete talk engages children in a low-level of challenge, rather than exposing them to cognitively challenging talk. Typically, this concrete talk takes the form of questions that test their knowledge of basic facts and information. The typical language routine known as initiation-response-evaluation (IRE) (Cazden, 2001) occurs when a teacher asks a question (e.g., “What shape is this?”), expects a short response (e.g., “Square.”), and provides a quick evaluation (e.g., “Yes; or No, it is a circle.”). Asking display or concrete questions, particularly to children from low SES homes, tends to limit children’s talk (Simmons, 1976). In contrast to concrete talk, cognitively challenging talk involves activities such as analyzing, predicting, discussion of vocabulary, summarizing, clarifying and evaluating (Dickinson & Smith, 1994).

In the longitudinal Home-School Study based in the northeastern United States, Dickinson and his colleagues were interested in a number of similar quality issues in addition to quantity figures related to talk. He also found that teachers used more rare words during large group time although not significantly more than during free play time. However the majority of cognitively challenging conversations took place with EUA children. In the Home-School Study, researchers coded teacher-talk in categories of 1) cognitive abstraction (e.g., what might a dog think?), 2) non-present talk (e.g., What was your favorite holiday?), 3) explaining, 4) extending (e.g. repetitive calendar routines), 5) pretend, and 6) no talk. Children who were in clusters with teachers who practiced the first three language behaviors showed positive
associations on vocabulary and other language measures, and demonstrated higher amounts of
talk and varied words in Kindergarten. Children in groups whose teachers used more extending
talk than explaining talk had lower language scores, especially in receptive vocabulary (Beals,
De Temple & Dickinson, 1994). A common factor, teacher education level, was found to be
significantly related to engagement in cognitively challenging talk and use of rare words (Smith
and Dickinson, 1994).

The relationship between teachers’ use of cognitively challenging talk and children’s
growth in literacy, specifically reading achievement, continues on through elementary school
(Taylor, Pearson, Peterson, & Rodriguez, 2003). Thus it is important to find ways to engage
teachers in thinking about how to provide children with cognitively challenging and abstract talk
as a way to build vocabulary and to help them better participate in the decontextualized use of
language within the classroom (Rosemary & Roskos, 2002). It is also apparent that these
opportunities for language practice need not be limited to large group time, but may be
productively pursued in multiple settings including small group, family style meals, one-on-one
and free-play.

Limitations of these approaches

This review of studies so far has presented a patchwork of observations of child care both
in the U.S and the U.K. From these observational studies, strong common associations emerge
related to language practices. Environments, whether they are homes or child care centers, vary
in the quality and quantity of language interactions. Early language experiences shape the
amount and complexity of children’s talk and initiate a trajectory for the future (Entwisle,
Alexander, & Olson, 1997; Farkas & Beron, 2004; Hart & Risley, 1995). Over all, children
perform better on language and cognitive measures when they have more and varied experience
conversing with adults. Limitations are inherent, though, related to what we can infer from relationships drawn from observational studies particularly in environments as complex as child care centers. For example, it is unclear whether teachers engage in more cognitively challenging conversations in child care centers serving middle income children because the children are better able to partner with them or because the teachers have higher educational levels.

For many years teachers in early education programs have been instructed that developmentally appropriate practice is to not interfere in children’s play time or center activity (Golbeck, 2001). A conversational invitation could be interpreted as an interruption of a child’s learning process.

Currently there are few studies that have provided an intensive and long-term study of teacher-child talk in preschool. Unfortunately, many of the studies used only one observation point for only part of a day. In some cases children were observed on a different day than teachers. A notable exception is the longitudinal Dickinson Home-School Study during which data was collected from homes, centers, schools and teachers, over a period of four years. Even so, only one home and center visit per year was conducted per child.

An important consideration in research is whether conclusions are generalizable to a target population or the general population. Many studies reflect practices in child care centers serving either low income children or University affiliated populations. To some degree this represents two ends of an existing continuum in which children are distributed in centers according to income. Inferences made from these studies can only extend to the specific population studied. In few cases (such as the Georgia Pre-Kindergarten program free to all 4-year-olds is one) is there an integration of economic groups in child care. More research is needed in such settings. Given the broad diversity of languages, cultures and ethnicities
represented in child care centers it is important that the differential effects of practices are considered. It is prudent to study and follow up findings that appear contradictory such as different associations of more talk for children from different economic groups. This information should directly inform teacher training. As the US determines how to invest resources in child care it is important that effective practices are promoted.

Teacher-child relationships are complex and influenced by biases and personalities. Previous preschool paradigms may have more closely accommodated a one-size-fits-most curriculum management goal. New goals related to early literacy education call for a different set of teaching strategies and perhaps a different philosophical underpinning. Some children may be less able to take advantage of learning opportunities in the discourse environment of the preschool classroom. Students from communities engaging in different discourse patterns may be less adept at meeting typical expectations. It is important that the training which teachers receive should include information on different discourse patterns and methods of engaging all children in conversation.

*Interventions intended to support vocabulary development through adult-child talk in preschools*

Investing in improving the linguistic environments of preschools may be the most direct route to a literate society. In 2001, the National Center on Education and the Economy (NCEE) developed classroom guidelines based on research to promote ways that teachers can increase opportunities for children to improve their oral language skills. These guidelines recommend child-teacher conversations, support of narrative development, and the use of elaborated explanation. Despite policy and curriculum statements such as the NCEE’s, the actual efficacy of implementing programs designed to enhance interactive communication between teachers and children is understudied.
Teacher communication thought to be most useful in building child communication is based on principles of shared interest, expanding and elaborative questions and comments, and multiple turn taking (Snow, 1983). Much of school-based communication has been observed to follow a different format in which teachers ask more interrogative questions, in an IRE format, during which they assess children’s knowledge (de Kruif, et al, 2000). Girolametto and colleagues (Girolametto, Weitzman, Lieshout, Duff, 2000; Girolametto, Weitzman, & Greenberg, 2006) conducted in-service training with Canadian preschool teachers to determine whether training could shift the language behaviors of teachers, specifically increasing responsive talk and decreasing the number of directives given to children. They studied teachers of toddlers and preschoolers between the ages of 1 ½ and 5 ½ years of age and were particularly interested in children’s talk in small groups (4 children or less). Prior to training they found there were no significant age group effects. Teachers controlled topic choice and turn-taking at similar levels across age groups. In their observations prior to the intervention, they found teachers used good language modeling strategies, but they scored at lower rates for interaction-promoting strategies (encouraging turn-taking).

The Canadian intervention consisted of a teacher training in the curriculum *Learning Language and Loving It*. The training provided eight evening group sessions and seven individual classroom based sessions. They found that the eight teachers in the training group increased their interactive behaviors, their overall talkativeness and engaged in promoting more turn-taking and verbal interaction. The new strategies were maintained for a period of nine months at which time follow up testing was conducted. As the teacher’s language behavior changed, so did the children’s; those in the experimental group talked more and used more multiword combinations. As part of the curriculum, the researchers, who specialized in speech
and language pathologies, encouraged teachers to simplify their speech in order to be maximally understood by the children. Interestingly, the teachers did not modify their speech and continued to use complex language and sentence structure, suggesting that modifying speech levels may be a more difficult adaptation (Girolametto et al., 2000; Girolametto et al., 2003). A large body of observational research would suggest that children who have greater exposure to more complex language and rare words have improved language development, therefore simplifying speech might limit opportunities for normally developing children (Weizman & Snow, 2000). Regardless, following the intervention, children in the experimental group talked more, produced more combinations, and talked to peers more often than the control group (Girolametto et al., 2003).

One study of narrative development included the opportunity for one-on-one and group communication between a teacher and preschool children in a diverse set of classrooms (Nicolopoulou, 2002). During the day a child could choose to tell a story to a teacher who would interact responsively and transcribe it. At the end of the day the teacher would read the story to the whole class, giving the author and his or her peers a chance to act it out. Nicolopoulou collected over 3,500 stories from middle class children and 166 from Head Start classrooms. Although the Head Start children began the year with fewer narrative skills than their middle-class counterparts, the same trends were evident between the two groups, with narrative complexity and sophistication increasing throughout the year. Both groups added detail to character development and increased the amount and coherency of the action as a function of the intervention. Predictably the intervention group showed a significant improvement over the control group on a measure of narrative tasks. The Expressive Vocabulary Test (Williams, 1997) scores
showed a significant increase for the intervention group between September and May, (from M=92 to M=95.33), whereas the control group decreased (M= 96.26 to M=92.42). However, although Nicolopoulou attributed the changes in children’s narrative skills to the practice they gained as a function of the intervention, it is also possible that the practice time the children had alone with the teacher engaging in conversation shaped and developed their language.

Intentional teacher-child conversations were a component of a multifaceted early literacy intervention, PAVEd for Success (Schwanenflugel, et al, 2005). This study was designed especially to evaluate the relative effectiveness of different programmatic emphases related to explicit vocabulary and phonological awareness. Each of the teachers in the experimental conditions participated in training that was designed to support a foundation of quality early literacy practices in children from diverse economic and social backgrounds. One practice included in this training was Building Bridges, a program designed to systematize a teacher’s engagement in conversation with each child 3 times a week. The conversations could occur in small group or dyadically. The children in classrooms who received the quality early literacy practices training performed better on both expressive and receptive vocabulary language measures, but the effectiveness of the Building Bridges program was not separately evaluated. Teacher interviews conducted at the conclusion of the intervention indicated that teachers thought this practice was valuable, but that the methods suggested for monitoring the inclusion of each child were cumbersome. Teachers also perceived the practice to especially benefit children who were withdrawn or who were English language learners.

The question of differential effects for children related to teacher practice is important to consider in light of the feedback provided by the teachers in the PAVEd for Success study.
Teachers do informally evaluate the effectiveness of practices. If they do not perceive a benefit they are unlikely to invest their time in continuing it. This may have been the case with *Building Bridges*, where some children arrived with good verbal skills and conversed regularly with the teachers and children in multiple settings. There is reason to question whether this practice is as valuable for children with good verbal competency as for those who have less language practice at home. Teachers cite time and feasibility constraints, given the number of children versus teachers; however, children at centers who are observed to spend more time dyadically with teachers have better language outcomes (McCartney & Scarr, 1984). A parent training study conducted in Bermuda found that a verbal interaction intervention did not add benefits to the children of parents who were motivated toward their children’s success (Scarr & McCartney, 1988). Providing enhanced language practice may be most effective and efficient when targeted to children scoring on the lower end of expressive language competencies.

The benefits of exchanging communication in responsive and interactive ways are likely to have positive effects on teacher-child relationships as well as language skills. Verbal risk taking and opportunities for oral language growth are most likely to occur in the context of individual conversations about meaningful topics. In a study of Head Start teachers, Brody, Stoneman and McCoy (1994) found that teachers who were more affectionate and responsive toward their students produced students who scored at a higher level on language and literacy measures. Similarly, preschool children in another study were observed to use more complex language when they have a positive relationship with their teacher (Pianta, 1995). The importance of positive language use and affirmations are also present in the Hart and Risley (1995) home study. Positive teacher talk also appears related to the development of secure child-
teacher attachments which has broad benefits. Children who have secure, positive relationships with teachers are more successful at mastering new skills and have higher academic achievement and school adjustment (Birch & Ladd, 1997).

Limitations of studies and challenges inherent in the study of language development

The literature directly linking parental talk to vocabulary and language development in children (Huttenlocker, Haight, Bryk, Seltzer & Lyons, 1991; Weizman & Snow, 2001; Hart & Risley, 1995) is relatively abundant. A reasonable inference resulting from these data might be that teacher talk would have similar direct benefits for children’s language development, but it is difficult to find studies that have tried to isolate the effects of increased language practice in a preschool setting. The reasons for this paucity of research may be due to the difficulty of determining causal factors in the complex environment of the child care center, even in experimental studies. How can researchers factor out or disentangle the many variables that may be simultaneously contributing to improved language skills? For example, social talk between children naturally occurs in several settings throughout the preschool day including outside play and center time. Pellegrini and others studying the preschool setting have suggested that talk between children is a critical component of social development and should be valued (Pellegrini, Galda, Dresden, & Cox, 1991). Centers vary in the level of accommodation and facilitation of social talk between children encouraged, which may be a factor in the level of improvement in language skills. Another influential factor may be a teacher’s attitude toward the non-standard communication style of a particular incoming student which might impact a child’s motivation to speak in the classroom. At the present time, we don’t have methods to measure such attitudes.

The dynamic nature of conversation also involves the varying language and pedagogical skills of individual teachers, factors which are difficult to measure and factor out.
One way to address the variability of teacher language skills is to use examiners as conversation partners; however this investigative strategy has not been pursued. The Nicoupoulou (2002) study provides some support on behalf of the beneficial effects of language interaction between the teacher and student, but this study included an additional component of group interaction. In addition, the student-teacher talk was voluntary and may not have represented a range of children. The PAVEd for Success (Schwanenflugel, et al, 2005) study provided valuable information related to teachers’ reactions to implementing a conversation program, but data as to whether or not the practice improved language skills were not available.

Another set of challenges lies in the area of measurement. It is time consuming for both children and researchers to conduct reliable and valid measurements of different expressive language skills. Even so, for a number of commonly used language assessments, concerns exist whether skills of non-EUA children are underestimated (Restrepo, et al., 2006; Washington & Craig, 1998).

Despite these real challenges, learning whether enriched conversational interactions between teachers and students is a useful method of supporting language development is important information for early childhood center curriculum design. Recent attention at the national level to raising academic standards in elementary education has exerted a strong influence toward higher expectations from early childhood curricula. As the system undergoes redesign, there is a current emphasis on scientifically based research as a guide for curriculum. Although the scientific community has determined that oral language skills, especially vocabulary, appear to be key predictors of later academic success, the field has not yielded many actual methods to guide teachers’ practice toward its development. Most researched methods are anchored on read-alouds (Arnold, Lonigan, Whitehurst, & Epstein, 1994; Elley, 1989; Hargrave
The present danger is that early childhood programs may focus on phonemic and alphabetic awareness because of teaching practices and assessments which have undergone rigorous review to the exclusion of vocabulary development programs which are under researched.

*Can the systematic practice of vocabulary enriched challenging conversation between an adult and student improve vocabulary and language development in preschool children?*

Two theoretical perspectives are used to consider this oral language development question. First, oral language is considered to be a critical component of early literacy (Dickinson et al., 2003; Schatschneider, Francis, Carlson, Fletcher, & Foorman, 2004). Literacy development is conceived of as a set of cognitive processes that work together to build successful reading development. These processes can be roughly organized in two interdependent strands that interact to support reading, the *outside in* and *inside out*, (Whitehurst & Lonigan, 1998). The *outside in* skills, vocabulary, print awareness, broad oral language experience and life experience, are highly related to home environment, SES, and parent education. The *inside out* skills are related to the coding aspects of reading, phonemic awareness, phonics and decoding, and are related to instructional environment and phonological processing abilities (Maclean, Bryant, & Bradley, 1987; R. K. Wagner, Torgeson, & Rashotte, 1994). These cognitive processes have been scientifically studied in an effort to understand the relationships between individuals, their environment and reading progress and difficulties. The relative importance of oral language to reading continues to be studied, but methods of intervention in cases of deficits have not been adequately developed (Catts, Fey, Zhang, & Tomblin, 1999; Scarborough, 1998; Storch & Whitehurst, 2002).
Oral language development within the preschool classroom is approached through a social-interactionist theory of language acquisition (Dickinson & McCabe, 1991; Girolametto, Weitzman, Lieshout, & Duff, 2000). In the early childhood environment, discourse patterns between children and adults are likely to have a strong influence on children’s developing language. Whether or not children are engaged in conversational interactions responsive to their interests is likely to influence motivation and ongoing initiation of talk. The presence of semantically appropriate and syntactically diverse language that provides the opportunity for modeling is likely to positively affect children’s developing language (Cicognani & Zani, 1992; Girolametto, Weitzman, & Greenberg, 2003; Justice & Ezell, 2000; Pelligrino & Scopesi, 1990; Polyzoi, 1997).

The purpose of this investigation is to examine a specific method of increasing vocabulary and oral language skills in preschool classrooms. The plan is to provide systematic, cognitively challenging conversation (involving interpreting, hypothesizing and clarifying) between the teacher and students. This strategy is derived from observations of the normal ways that children learn vocabulary throughout language development. Language input by parents and teachers can have a significant effect on child output (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). Multiple aspects of language appear to influence children’s vocabulary development in a positive direction, including the total number of words, density of rare words and the ratio of approving words (Hart & Risley, 1995; Weizman & Snow, 2001). In addition, different types of construction and complexity, such as the use of complex clauses and passive structures by teachers might influence children’s language production (Huttenlocher et al, 2002; Reese & Fivush, 1993; Vasilyeva, Huttenlocher & Waterfall, 2006). There is little known, however, of the degree of intervention necessary to affect vocabulary growth significantly. To
learn more about different ways that expressive vocabulary skills develop, I will measure expressive vocabulary in children following an intervention of regular small group systematic vocabulary enriched conversations.

The intervention reported here was designed to be short term and intensive. With the assistance of three UGA undergraduate assistants, I attempted to control the key features of conversation, quality and quantity. En lieu of using the children’s actual preschool teachers, using these trained experimenters, I hoped to control for many factors unrelated to the intervention. Teachers vary in education and experience, they have varying interest in individual interaction, and they have competing work related duties. Although researcher-conducted conversation cannot be considered as an exact proxy for teacher-child conversation in early childcare settings, this practice could be carried out by teachers in a number of possible ways; from my previous work with PAVED for Success (Schwanenflugel, et al, 2005), I found that treating talk as a center activity or ‘talk center,’ or as part of ‘family style meals’ where the teacher or paraprofessional might converse with a small number of children on a rotating basis. Although researcher conducted conversation does not automatically generalize to the preschool classroom, this study seeks to fill a gap in the literature regarding whether regular oral language practice in the context of child care might result in positive, measurable effects on children’s language development. I focused on preschools that served high poverty communities and was particularly interested in the effects of the intervention on children whose vocabulary pretest scores indicated that they fell in the lower third of the normal curve in vocabulary skill.

Measuring expressive vocabulary development

In this section I outline some issues in the quantification of language elements and support the use of both a standardized assessment and language sampling. I begin with theory,
followed by introductory information about a standardized vocabulary assessment the Expressive Vocabulary Test (EVT) (Williams, 1997), a definition assessment instrument, and the calculation of lexical diversity from a language sample. I then covered several aspects of interest related to the elicitation and analysis of language samples. In each case I provided support for the methods used in my study.

Theoretical basis. Measuring oral language development, particularly vocabulary, using reliable and valid methods has posed a challenge to researchers and educators for several decades. Three dimensions of vocabulary development are important in this research: expressive vocabulary, lexical diversity, and definitional skill. Abilities related to vocabulary can be measured in multiple ways. Two aspects of vocabulary commonly measured are: expressive vocabulary which refers to words used in speech or writing, and receptive vocabulary which refers to words which can be read or recognized and understood in a spoken or written format. Lexical diversity refers to the range and variation of vocabulary deployed in speech. Definitional skill measures provides information on a child’s level of knowledge of given words and their ability to express their understanding of the meaning.

Measuring these different oral language skills in preschool children presents a dilemma. The use of standardized assessment and the analysis of language samples represent two very different approaches to this task. Standardized assessments can indicate whether a subject can produce specific oral language skills in separate tests, or subtests (e.g., vocabulary, syntax, narrative structure, expository comprehension). These tests are more likely to be used by school systems to gain a baseline understanding of their students’ language skills and to identify children who may be in need of special services. By contrast, a language sample is intended to
represent expressive language as deployed in genuine speech. Language sample analysis has been used by researchers in the study of expressive vocabulary, language development, communicative features, second language ability, language impairment and numerous other linguistic areas. This method is often employed by speech and language pathologists. Analyses derived from samples can also yield quantitative scores useful in comparing children in terms of development and diagnosis of language disorders.

Conceptualizing expressive vocabulary as a construct can be confusing and complex. The first question is what to consider a word. For instance the word run can be a noun or a verb, inflected forms include ran, running or adjectival runner. Do we count these as individual words or as one root word? The Oxford English Dictionary (2nd edition) includes about 500,000 head words, whereas the number of word families is estimated to be closer to 88,700 (Nagy & Anderson, 1984). Estimates of how many words a first grader is likely to know vary greatly from 2,500 to 26,000 (Lorge & Chall, 1963) depending on the inclusion criteria. Even at the low end of the estimate, it would not be possible for teachers to elicit 2,500 words from each student (bearing in mind that the words themselves would vary according to student). Therefore some type of sampling is required when estimating children’s expressive vocabularies.

Children learn many commonly used words in roughly the same pattern of order between Kindergarten and fifth grade. However, the variance between children at the end of second grade can range between 3,200 and 6,200 root words depending on their previous home and educational environments (Biemiller & Slonim, 2001). Counting words in this way a child would need to have learned, on average, two root words per day between the ages of one and eight in order to be at a normative level of 5,000 root words. The degree to which a child truly knows a word is not transparent either. Schwanenflugel, Stahl and McFalls (1997) classified the varying
degrees of ‘knowledge’ as none or incorrect knowledge, partial knowledge or full knowledge in a study of learning words from context. The partial knowledge category was useful in describing a range of responses that showed familiarity with the target word, perhaps through a root association, but not full knowledge.

**A standardized expressive vocabulary test, the EVT.** The EVT was developed to measure vocabulary through one-word responses. Subjects up to the age of 5 are asked to provide a label for a picture. From age 5 through adult a picture and label are provided to the subject who is asked to provide a synonym. Both labels and synonyms may be nouns, verbs or adjectives. The developers of the EVT selected words from frequency lists including A Spoken Word Count (Jones & Wepman, 1996); Computational Analysis of Present Day American English (Kucera & Francis, 1967); Word Frequencies of Spoken American English (Dahl, 1979) and sources of commonly used language from children’s literature and the Merriam Webster’s Collegiate Thesaurus (1988). By requiring subjects to actively label the picture provided the EVT, developers propose that two constructs are tested, vocabulary knowledge and word retrieval (Williams, 1997).

Although there are certain cost and ease benefits of standardized one-word answers, retrieving a synonym may measure expressive vocabulary more validly for children who are quick and succinct versus those who may be more careful and deliberate. Further, bias concerns continue to surface about incidences of lower standardized vocabulary scores of African American children even when matched with European American children for SES and education (Washington, 2001), but the EVT may provide a more accurate assessment of vocabulary than the PPVT-III receptive vocabulary test (Restrepo et al., 2006) to which it can be directly compared.
Several compelling features of the EVT are reported by the developers. The test manual reports convincing discriminate and convergent validity evidence when compared with concurrent testing through other language batteries. The EVT scores had moderately high correlations with oral expression and listening comprehension measures, but not so high as to be measuring the same skill. Evidence is presented by the developers in support of the test’s utility in adequately discriminating between children with expressive language delays or impairments in order to direct services. However in a study comparing four vocabulary tests as to their utility in discriminating SLI, Gray, Plante, Vance and Henrichsen (1999) found the EVT to be the least predictive of the four, resulting in 15% over-referrals and 16% under-referrals. Although the EVT is a measure that is relatively easy to use in schools and has reasonable reliability values and validity support, validity concerns related to its accuracy for children with less educated mothers, or children with SLI are worrying (Gray, Plante, Vance and Henrichsen, 1999).

Definition assessment. Expressive vocabulary can also be measured through assessments which ask children to provide definitions of words. This oral language skill represents another dimension of vocabulary knowledge. The ability to define words varies according to age, metalinguistic development, and educational level (Benelli, Belacchi, Gini & Lucangeli. 2006; Kurland & Snow, 1997; Snow 1990). An initial level of cognitive development enabling the perception of language as a sign-referent system is necessary to providing a definition. Linguistic skills are also necessary to be able to reflect on a word and provide a meaningful definition.

The use of definition assessment, however, enables the determination of partial word knowledge, something that simple one word answers do not permit. The process of learning and sharing the meaning of individual words also follows a development direction described and
categorized by Schwanenflugel, Stahl and McFalls (1997) as unknown, partially known and known. Answers scored as partially known may also include partly incorrect answers and incomplete associations. As a child builds more associations from hearing a word used in multiple contexts they are able to develop a more refined and detached definition. Some factors that facilitate definition learning are presenting words in rich contexts, building familiarity with word components, increasing frequency of word exposure and saliency of meaning for practical usage (Schwanenflugel, Stahl & McFalls, 1997).

The development of definition skills follows age growth (Nippold, 1995). Initially it is difficult for children to differentiate the word from the object and early attempts at definitions are characterized by has followed by descriptive features. “Dog has black ears” typically refers to the one dog a child may know well. Early definitions may include a synonym, but often the relationships between descriptives, synonyms and the target word are not meaningfully presented. This stage usually shifts around age 7 to is responses which general include more abstract conceptions (Benelli, et al, 2006). Later in pre-adolescence a student may be able to provide even more formal definitions, often referred to as Aristotelian in form. These definitions may include information about category memberships as well as descriptive features (Benelli, 1988; Litowitz, 1997). In this case a synonym is often included and is presented according to grammatical rules that link the concepts together appropriately (Benelli et al, 2006). The practice of asking for definitions and scoring answers on a three point scale has been used by researchers to study vocabulary development in early education settings (Coyne, Simmons, Kame’eniu & Stoolmiller, 2004; Justice, Meier & Walpole, 2005; Leung, 1992; Snow, Tabor, Nicholson & Kurland, 1995).
The ability to provide a more formal definition appears to be highly related to education-based practice. In an Italian study evaluating the ability to provide a definition between subjects ranging from 5 to 11 years and a group of adults with higher education and one without, researchers found that the group of adults without higher education scored similarly to 11 year olds whereas more educated adults scored significantly higher (Benelli et al, 2006). A significant relationship between definition skills and educational practice was also found in a low-income sample of children and parents in the U.S. (Kurland & Snow, 1997). In an intensive home study Hart and Risley (1999) documented the regularity with which higher educated parents introduce new words in context and follow them with definitions, e.g. “Look at that conductor driving the train. Who is driving the train?” Children in homes with less language support lack practice in hearing this critical skill modeled and in having opportunities to use it.

Language sampling. Analyzing features of a language sample of words actually used by a child in speech is a more common method for speech and language pathologists, but it is a method unlikely to be used by educators and school psychologists, who tend to prefer standardized assessments. Language samples can provide information on actual vocabulary use and seem less amenable to charges of racial or ethnic bias. A positive feature of this method is that once a transcription is prepared, other language features can be examined such as syntax, phonological, and discourse-features. However, with current technology limitations, this method is time intensive, as the language must be taped and then manually transcribed. Measuring vocabulary knowledge from a language sample is not entirely indirect. One increasingly favored measure of vocabulary obtained from language samples is, D, a measure of lexical diversity, but other measures exist. However, the reliability and validity of these measures have not been
clearly established. Thus, difficult decisions have to be made by researchers as to how to assess vocabulary levels and language skills.

*D, a measure of lexical diversity calculated from a language sample.* D is based on a different approach to measuring expressive language. Instead of sampling commonly known words, D relies on collecting a sample of speech for analysis of characteristics of expressive language a child actually uses. The strongest validity support relates to the content aspect of construct validity. Consistent with the theory underlying lexical diversity, D reflects vocabulary deployment independent of NDW and MLU as well as increased diversity of vocabulary use as age increases. Initial evidence related to criteria show reasonable support for D as a diagnostic tool for language impairments and more modest support as a vocabulary measure. There is very little evidence thus far that the measure is generalizable because the corpora are small and relatively homogenous in relation to SES, race/ethnicity and age. The authors analyzed for sex differences in each of the reported studies and found no differences; however, insufficient evidence is provided for other population differences.

Approaching assessment in three different ways should provide a more complete picture of a child’s expressive vocabulary level than any single assessment alone. A standardized test, such as the EVT, can effectively scan for vocabulary knowledge through assessing vocabulary that has been carefully arranged hierarchically by difficulty. D provides a measure of lexical diversity from language deployment in a relatively authentic setting. Finally, the definition test narrows in on a smaller set of words, but measures the ability to share information about the meaning of words.
Statement of Purpose

The primary purpose of this series of studies is to investigate whether a somewhat modest amount of adult-child complex conversation emphasizing vocabulary-enriched talk carried out for 500 minutes over the course of 10 weeks has a positive effect on expressive vocabulary development. The second study focuses on the intervention itself. Important practical implications arise in the event of positive evidence. Although preschool curricula may note the importance of oral language practice, they do not typically include a systematic method for teachers to engage in and account for conversation with children. If this conversational intervention has a positive effect for children with low vocabularies, this would indicate a need to include vocabulary-enriched talk between classroom-adults and children (hereafter referred to as Talk Time) as part of early literacy instructional time in classrooms. This study is conducted with 4 year olds because of the particular importance of enhancing vocabulary prior to learning to read. Further, age four is increasingly a normative age for school entry around the country because of the existence of prekindergarten programs for children at risk.

A secondary purpose is to conduct a validity study of D, a measure of lexical diversity and, in particular, vocabulary knowledge, as measured by the computer program vocd (MacWhinney, 2000). I used Messick’s (1989) framework for unified construct validity to investigate this measure focusing attention on the structural aspect, specifically reliability issues of internal consistency and test-retest; and the external aspect including the relationships that D demonstrates with the EVT and the definition test. A fuller description of this study is presented in Study 1.
In short, the present series of studies address the following questions:

Study 1:

1. When evaluating D, using Messick’s (1995) framework for unitary construct validity, is there evidence to following validity aspects: content: substantive, structural, generalizable and external?

2. To what extent does D, when collected during a standard protocol that queries children for personal narratives and narration of a wordless picture book, correlate with a standardized measure of expressive vocabulary and a definitional task in a diverse sample of preschool children?

3. To what extent is D a relatively unbiased measure of vocabulary?

Study 2:

1. Will an intervention of vocabulary-enriched talk result in improved vocabulary levels as measured by a standardized vocabulary assessment, a definition assessment and D?

2. Will these effects be larger for children who begin the intervention with lower assessed standardized vocabulary skill?
CHAPTER II

STUDY 1

A Validation Study of an Expressive Vocabulary Measure

The purpose of this study is to examine the validity of D, calculated through the vocd program, as a measure of lexical diversity. Currently, it is not clear whether lexical diversity tabulated in this way relates to standardized assessments of vocabulary, nor is it clear whether standard procedure I developed for collecting language samples provide an acceptable measure of lexical diversity. I used Messick’s (1989, 1995) framework for unitary construct validity, to evaluate the evidence collected from the study. The following aspects were considered:

1. **Content:** To what extent does the test sample the domain of interest as defined? Is there consistency between the theoretical and empirical bases of the definition of the construct and the structure and boundaries of the assessed domain?

2. **Substantive:** Is there empirical evidence that the assessment tasks aligned with the theoretical rationale?

3. **Structural:** Does the task and scoring structure reflect the construct’s domain?

4. **Generalizability:** Are the score properties and interpretations replicable and consistent across other population groups and time?

5. **External:** Is there evidence of convergent and discriminant relationships with other measures, and are those relationships consistent with the theory of the construct being assessed?
6. **Consequential:** In relation to the issues of bias, fairness and distributive justice, do the score interpretations hold value to support the actual and potential consequences?

To support the validity of a measure a combination of evidence is needed from all six aspects, which are considered interrelated (Messick, 1995). Validity is best considered holistically as an overall judgment that integrates evidence from the six aspects, so the validity of a measure is always considered to be a matter of degree rather than absolute. Because D is calculated from a transcription of sampled speech or a sample of writing, special consideration is due to the method of elicitation and the degree of standardization of practices. The following section is a review of issues related the standardization of sample elicitations and concludes with a rationale for the method proposed for use in Study 1 and Study 2 (Appendix A).

**Language sample elicitations**

For researchers using language sampling to note language growth among typically developing children, one of the most salient concerns is the lack of standardized elicitation methods. Researchers may use one or several methods including narration of a wordless pictures book, retelling the story from a previously-read picture book, spontaneous speech during free play with toys or play dough, structured interview questions or personal narrative with photo or picture prompts. A reason for variation of methods among researchers in this area may be that many language samples are carried out with clinical concerns in mind. Speech language pathologists are often interested in gathering language sampling targeting information regarding suspected language impairment, and some elicitation tasks are likely to provide more focused data.

Variation can be evident even within the results of a single method such as story retelling. Story retelling has been used by speech language pathologists as a means of predicting language
delay, in part because variants of oral storytelling are common to many cultures in the Americas (Fazio, Ardmore, & Connell, 1996). However a recent study by Gazella and Stockman (2003) uncovered a wide range of elicitation methods within the storytelling method alone. Story retelling might include reading a story or showing a video followed by asking the child to retell the story as remembered, with or without the pictures as prompts. In some studies, children were asked to discuss any familiar story or movie. Other studies blur the distinction between story retelling and story generation, by encouraging the child to add their own ideas (Dollaghan, Campbell, & Tomlin, 1990; Liles, 1993). A fairly common method is asking a child to generate a story from a wordless picture book, such as the Frog stories by Mercer Mayer. Some researchers present the book to the child pointing out the pictures throughout and then ask the child to tell the story using the pictures (Berman et al., 1994; Curenton & Justice, 2004). In other contexts, the examiner tells the story using the pictures and then asks the child to tell the story, which might target slightly different competencies.

Gazella and Stockman (2003) studied story retelling in two modalities. They found no modality difference between an initial audio-only and an audio-visual story presentation (see also, Schneider & Dube, 1997). They found that both methods fell short of producing the minimum 50 utterance sample length typically necessary for most analyses. They had also included a direct questioning condition (9 questions following the retell) which, when added, produced enough language. By combining two methods, broader language skills appeared to be elicited. Direct questioning elicited a larger number of utterances and number of different words (NDW), whereas story retelling produced longer more complex grammatical structures. However, issues related to the lack of standardization for this one method alone illustrate the difficulty in comparing studies.
A number of studies have found differences when comparing the elicitation of personal narratives with storytelling, however. Westerveld, Gillon and Miller (2004) developed a normative database of speech samples from 268 New Zealand (N.Z.) children between the ages of 4; 5 and 7; 6 years. They used three elicitation methods and were able to compare results between them. After a review of the literature, they developed a method designed to gather the most comprehensive data: personal narrative, conversation and story retelling. In this case, photo prompts were used for the personal narrative in a method developed by Peterson and McCabe (1983); for instance the examiner might show a photo of a young child receiving care from a physician with the following prompt, “This little girl had a really bad sore throat, so the doctor examined her. Have you ever been to the doctor?” The conversation protocol was based on a method developed by Evans and Craig (1992), in which the child was instructed to choose an object from the classroom and bring it to the session. The examiner asks the child to talk about the object followed by semi-scripted questions about family members, school, and after school activities. In the story retelling setting, the child and researcher listened to the tape of a children’s book in English. The book was written in a South Pacific language unknown to the children. After listening to the tape twice, the child was asked to retell the story without the picture book and with minimal prompting. In each context, the role of the examiner was to provide supportive listening and to try to limit the introduction of new vocabulary or topics.

Westerveld et al. (2004) used quantitative language measures to calculate syntactic competence, as calculated by mean length of utterance in morphemes (MLU-M) and percentage of word errors per utterance (PcEW); semantic diversity also referred to as expressive vocabulary, calculated by number of different words (NDW), and verbal productivity a calculation of total number of words (TNW). Significant differences were found between
different elicitation contexts for all age groups, except for semantic diversity which for 7 year olds were the only group showing significant differences, in this case, in favor of the personal narrative context. For the whole sample, verbal productivity was higher for the personal narrative and syntactic competence was highest in the story retell condition. The researchers concluded that collecting a language sample using both personal narrative and story retelling was the best way to represent a broad range of oral language abilities (Hadley, 1998; Westerveld et al., 2004). The N.Z. sample was compared to U.S. sample and there were non-significant differences between the groups. The elicitation prompts were very similar. The scores lagged a little for U.S. children, between ages 5 and 7 years, probably because most U.S. children start school at a later age than N.Z. children. This study provided useful guidance for language sampling and useful normative data.

Wagner, Nettelbladt, Sahlen and Nilholm (2000) studied Swedish preschoolers with language impairments and used multiple methods of language elicitations. They compared an interview ‘conversation’ (questions about family and interests followed by interview questions) called NELLI with ‘narration’ which combined story retelling and story generation. During the interview task, children were more fluent and intelligible which are concerns especially related to language impairment, whereas during the narration context, children produced more MLU (by word), more phrasal expansions and grammatical morphemes.

Nippold, Hesketh, Duthie and Mansfield, (2005) compared a kind of conversational interview consisting of questions about a subject’s interests, family and pets; with an expository condition, during which the subject was asked to describe their favorite sport. The examiner asked the subject to explain how the sport works, from beginning to end, followed by a final prompt asking for an explanation of the sport is won. The participants were 120 people ranging
from 7 to 49 years old. The researchers found that for all age groups the expository context elicted greater syntactic complexity, which they measured in a number of ways, concluding that the most effective were mean length of T-units (an independent clause with a subordinate clause (Hunt, 1970) and relative clause use.

There is a wide variety of protocols used to elicit what are referred to as ‘conversations’ ranging from child initiated talk during play to structured interviews. To gather more spontaneous speech, researchers usually reduce the quantity and qualities of speech provided by the adult conversational partner. However the different levels of encouragement and different prompts such as play dough, pictures, and toys add variability. Owen and Leonard (2002) collected speech samples obtained during free play and were able to discriminate specific language impairment (SLI) using vocd (D) scores, but noted concerns in their conclusion that for some analyses the lack of constraints in a free play structure may not yield representative speech as it relates to having to answer or respond to questions. Typically there is a high degree of variability in examiner encouragements that could contribute unintended variance in unstructured free play settings. Evans and Craig (1992) compared conversations elicited during free play with an interview and found that children produced an increased number of words and grammatical complexity during the interview. More structured conversations are easier to standardize and may prompt a greater variety of language. For this study I used the Westerveld, Gillon, and Miller (2004) methods which presented a useful set of tools with which to elicit a broad sample of expressive language use.

The development of the calculation D. Valid and reliable methods of calculating oral language measures from language samples have been sought after by researchers and clinicians. Theoretically, a useful measure should reflect language variation and sophistication that develops
over time and can identify the presence of speech and language disabilities. Initially, researchers believed that measuring the number of different words (NDW) in a language sample would provide a quantitative measure of range of vocabulary. However, results were found to be highly dependent on sample size, and attempts at trying different ways of truncating samples (e.g. random sampling or first 100 words) led to different sets of results, evidence of unreliability.

For some time, the Type Token Ratio (TTR), which is the result of dividing the number of types (or different words in the language sample) by the number of tokens (or the sheer number of words spoken in the sample), was thought to have solved the sample size problem. TTRs provide a more robust statistic than NDW but are still troubled by the size of the sample. The longer the sample, the more likelihood that words will be repeated and the ratio begins to decrease. The expectation was that the ratio would stabilize and plateau, but because of the higher probability of repeating high frequency words given a longer text a downward trend follows a short plateau. The same problems of standardizing NDW samples arose for TTRs. However a number of researchers find NDW a better index of lexical diversity than TTR (Klee, 1992; Watkins, Kelly, Harbers, & Hollis, 1995). A study comparing TNW, NDW and the EVT found in a significant correlation between EVT and NDW (r = .48, p<.01, N=28) but not with TNW (Ukrainetz & Blomquist, 2002).

Malvern and Richards (2002) carefully studied the promises and mistakes of different attempts over time to measure and standardize lexical diversity. They developed D, a quantitative measure of lexical diversity, intended to adjust for sample size. Lexical diversity includes vocabulary range, and in addition, researchers have provided other qualitative descriptions such as vocabulary richness (Read & Chapelle, 2001) and verbal creativity (Fradis, Mihaiilescu, & Jipescu, 1992). These descriptions correspond to ‘expressive’ in a dynamic way.
“There is a general underlying assumption among those concerned with educational
development, language learning and acquisition, and language impairment that a high lexical
diversity is ‘a good thing’, an indication of a combination of vocabulary size and the ability to
use it effectively.” (Malvern et al., 2004, p.6). The theoretical underpinnings of D, in addition to
measures of range such as NDW and TTR, include the early insights of Thompson and
Thompson (1915) that the deployment of vocabulary is critically important, including the
patterns of repetitions and the frequency of token groupings.

The model that Malvern and Richards (2002) developed represents a curve that takes into
account features of TTR, NDW and patterns of repetitions and frequencies, it is:

“…a particular value for best-fit between the ideal curves and those derived from
real transcripts over the standard range of points of the TTR versus N curve drawn
by a standardized procedure. As we have pointed out, this gives it an advantage
over any method of measuring diversity which depends on one point. We have
also shown because the random sampling uses all the data in the full transcript, it
reflects the whole manuscript.” (p.59).

The analysis for D uses a speech sample consisting of a minimum of 50 tokens, although 250
tokens have been recommended for reliability by Owen and Leonard (2002), and samples sets of
words selected randomly. An additional advantage of D is that teacher speech can also be
analyzed using the same methods.

Reliability estimates provide information on the stability and consistency of scores from
a test. Test-retest and internal consistency are two preferred methods of estimation. Malvern,
Richards, Chipere and Durán (2004) report several reliability statistics in their description of the
development of D. Internal consistency was estimated by a split half reliability coefficient
collected using even versus odd numbered words from transcripts. A Spearman-Brown derived estimate of \( .866 (df=36; p<.001) \) was obtained from 38 transcripts from corpora from the Child Language Data Exchange System (CHILDES) (MacWhinney, 2000; Dale, Bates, Reznick, & Morisset, 1989; Wells et al, 1985). A review of literature did not find reported test-retest estimates which would be helpful information related to this measure.

Questions related to whether or not a measure adequately and accurately captures the construct of interest fall into the framework of validity inquiry. Theoretically, the dynamic nature of language sampling could provide substantive validity compared with most standardized test formats. Speech and language clinicians commonly use language sampling over standardized tests because of the rich information gathered on how the child functionally uses language, including vocabulary (Furey & Watkins, 2002). A number of researchers have suggested that language sampling might provide a better assessment for speakers of African American English (Seymour & Bland, 1991) because of their concerns related to the validity of standardized tests for non-majority children.

Using the New England and Bristol corpora correlations between D and a number of oral language measures have been significant, including NDW, oral examination and fluency, but not with others such as the English Picture Vocabulary Test, words per minute, complexity of structure, or accuracy. Several studies have supported the use of D as a diagnostic measure for language disorders. Silverman and Ratner (2002) collected transcripts and administered the Expressive One-Word Picture Vocabulary Test – Revised (Gardner, 1990) to 30 children, 15 of who stuttered. Unlike previous studies comparing D to standardized vocabulary tests, they found a significant correlation between D and the Expressive One-Word Vocabulary Test, which is similar to the EVT, \( (r=.48; p=.01) \). One of the markers of Specific Language Impairment (SLI)
is delayed learning of vocabulary and D was able to differentiate children with SLI from age matched and MLU matched peers without SLI in a study conducted by Owen and Leonard (2002). Another study found that D along with MLU could clearly discriminate Cantonese children with SLI (Stokes & Fletcher, 2000). Collecting language samples might be less threatening to young children (especially with language disorders) than standardized tests and have shown utility in English, French, Cantonese and Spanish (Klee et al., 2004).

Consistent with the theory underlying lexical diversity, D reflects vocabulary deployment independent of NDW and MLU as well as increased diversity of vocabulary use as age increases. Initial evidence related to criteria show reasonable support for D as a diagnostic tool for language impairments and more modest support as a vocabulary measure. There is very little evidence thus far that the measure is generalizable because the corpora are small and relatively homogenous in relation to SES and race/ethnicity. The authors and colleagues (Malvern, Richards, Chipere & Durán, 2004) analyzed for sex differences in each of the reported studies and found no differences; however, insufficient evidence is provided for other population differences.

Within the content of an individual’s sample of speech, is there information that can be quantified and be used to discriminate the ability to effectively use vocabulary or identify impairment? One caution mentioned related to the mathematical model underpinning D is that a skilled speaker may intentionally use repetition for rhetorical effect and this would cause a decreased value of D because of the decreased ratio of new tokens and types, a concern affecting validity and reliability. In contrast a child could have “collocational diversity” (Stokes & Fletcher, 2000) represented by less productive combining of verbs and aspect markers and end up with a higher D.
The purpose of this validity study is to examine the evidence collected and determine whether it provides support for the kind of inferences and interpretations that are made about vocabulary diversity; “One validates not a test, but an interpretation of data arising from a specified procedure” (Cronbach, 1971, p.477). The present study used the unitary construct validity theoretical framework proposed by Messick (1989). Methods for considering the content and substance aspects of validity were also informed by Kane’s interpretive argument approach (Kane, Crooks & Cohen, 1999) which suggests the inclusion of both confirmatory and falsification evidence (Popper, 1962). The integral importance of consequence is addressed at the outset, in terms of test usefulness, as well as in the sixth aspect, consequential, (Messick, 1995) as suggested by Bachman and Shepard (Bachman & Palmer, 1996; Shepard, 1993).

Method

Participants

Children. Seventy-three four-year-old children participated who were part of a larger study examining the effects of additional linguistically complex talk on children’s language development. Children were attending universal lottery-funded, center-based, full-day, prekindergarten programs in a metropolitan community in Georgia. Recruitment letters (Appendix B) and consent forms (Appendix C) were distributed to all 4 year old children attending six lottery-funded prekindergarten classrooms in private child care centers. All had parental permission to participate and children assented to their own participation.

EVT testing was conducted with all children who received consent to participate. Students with scores higher than one standard deviation above the mean of the test were excluded because their vocabulary scores were already at a such a high level that additional language experience was unlikely to have much effect. Children who scored lower than one and
one-half standard deviations below the mean were excluded because they were beginning English language learners receiving different language services according to their school and county. Five students were excluded from the sample because of these exclusionary conditions. Selected were seventy three children (30 girls and 43 boys) with a mean age of 4 years and 4 months (SD = 3 months).

Because I sampled child care centers known to serve low income children, the overall SES of the population was expected to be low. To provide general information regarding the socioeconomic environments of the children in the study, two questions were added to the parental permission form. The parent filling out the form was requested to note their occupation and last year of schooling. Occupation is considered to a stable and reliable indicator of socioeconomic status (Zimmerman, 1992). Occupation was coded according to the Nakao and Treas (1992) index. Each respondent was assigned a status score between 0 and 100. The rationale for collecting information on the education of one of the parents is based on the high collinearity of maternal and paternal education levels (Kalmijn, 1991). Educational capital was scored using the following system suggested by Entwisle and Anton (1994): less than high school (0), high school graduate (1), some college (2), Bachelor’s degree (3) and higher degree (4). Although social scientists differ on specific details, they broadly agree that a combination of education, income and occupational status are three valuable indicators of financial capital (Entwisle & Astone, 1994). In reconfiguring the SEI index used to score occupational status, Nakao and Treas (1992) consistently found that education level played a greater role than income in the socio-economic index ratings. For this study mother’s educational level and current occupation were collected and each variable was standardized. Scores were averaged to create a SES score, or when only question was addressed that value was used (Geoffrey, Cote, Borge,
Larouche, Seguin & Rutter, 2007). The range of values for the fifty eight mothers who answered at least one question was between -1.73 and 1.96 with a mean of -.02 (SD=.90). For most mothers with an SES of -.02 or less, high school was the highest level of education completed and their work entailed service in food, sales or child care industries.

Ethnicity information was not collected on the parent permission form. The ethnicity of children in the study approximately reflected the counties’ demographic distribution for children of that age which is about 67% European American, 28% African American and 5% Hispanic, Latino and other (U.S. Census, 2005). Demographic features of the subject sample are displayed in Table 2.1.

Table 2.1

<table>
<thead>
<tr>
<th>Demographics of Participants (N = 73)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item</td>
</tr>
<tr>
<td>Age in months</td>
</tr>
<tr>
<td>SES</td>
</tr>
<tr>
<td>Gender (M, F)</td>
</tr>
</tbody>
</table>

For this validity study, only subjects with complete datasets were used. One child during post-testing refused to participate and his scores (including the pretest) were eliminated from both studies. Three subjects were unavailable during the post-testing session and two pretest recordings did not yield the adequate number of tokens for analysis, resulting in a sample of 68 participants. Among these subjects, there were 26 girls and 42 boys with a mean age of 54.5 months ($SD=3.60$) and a mean SES of $.08 (SD=.30.)

**Experimenter characteristics and training.** Three senior undergraduate psychology majors (two females, one male) and one male beginning graduate student attending the University of Georgia carried out the intervention. I also served as an experimenter. Each experimenter had a minimum of one semester of supervised tutoring, teaching or intervention
experience with young children. All were participating for research course credit. Two experimenters were assigned 5 pairs of children, one was assigned 4 pairs, one 3 pairs and I met with one pair. Further, I served as a substitute when one was needed.

Experimenters attended two two-hour training sessions prior to the intervention (Appendix F). In the first two hours, training included instruction in techniques intended to foster children’s joint interest in an activity (e.g. allowing adequate wait time for children to initiate conversation and following the lead of the child during conversation.) In addition, emphasis was placed on introducing vocabulary naturalistically into conversation through vocabulary recasting (with elaboration) and expansions of children’s utterances. Training included strategies designed to encourage children to balance turn taking and to build upon ongoing subjects in their conversation pairs. In the second two hours, experimenters practiced language modeling techniques with two pilot children and received ongoing feedback and coaching on their performance from me.

The experimenter team communicated via e-mail or in weekly meetings to share and update information about the progress of the intervention and receive information for the upcoming week. Throughout the intervention, I monitored the implementation of the intervention by the experimenters and met with them individually on a regular basis to discuss progress, questions and concerns and to provide feedback. A mid-intervention conversation was taped from each pair of children to determine fidelity with trained intervention practices.

Procedure

Testing was carried out by the experimenter team. Three types of vocabulary assessment were administered: (a) the EVT; (b) a definitional assessment; and (c) a language sample in order to obtain a measure of lexical diversity. The EVT and definition assessment were carried out
and counterbalanced in one session and the language sample was given in a second session that took place within two weeks of the first. The first session took approximately 20 minutes and the second session took approximately 15 minutes. A standard protocol was used which included open ended questions about play activities, scripted questions related to picture prompts and the narration of a wordless picture book (see Appendix A for detailed protocol).

The same set of tests was conducted in the same manner 11-12 weeks later. For half the children, this assessment followed a 10-week intervention designed to elicit conversation and talk in four-year olds. (See Chapter 3 for more details about this intervention.) In no instance was the person conducting the post-test the same person who carried out the intervention. Children were provided a similar, but different wordless picture book from the pretest for language sampling.

**EVT assessment.** The EVT is an individually administered assessment of expressive single-word vocabulary designed for children 2 ½ years and onward. The basic format of the test requires the examiner to point to a picture and ask the child for a word label. It is composed of 190 items, including nouns, verbs and adjectives, the first 38 requiring labels and the remaining 152 requiring synonyms. For most items a picture is presented and either a label or synonym is sought from the subject. For others, the examiner points to a part of his or her body and asks for a label. Both labels and synonyms may be nouns, verbs or adjectives. The test, which is not timed, takes about 10 to 15 minutes to complete for preschool children. The EVT was conducted and scored according to guidelines presented in the manual.

The EVT is designed to be used primarily by classroom teachers and school personnel in order to help screen for language impairments, monitor expressive vocabulary growth in English (for native speakers and English language learners) and by researchers in areas of reading,
language development and speech disorders. Specifications allow for the test to be given by a wide range of people, including teachers, undergraduates or paraprofessionals under supervision. According to the test manual, measures of internal consistency ranged from .90 to .98 (median, .95; N=100 per age group, 25 age groups).

Test-retest reliability information on the EVT ranged from .77 to .90 according to the test manual. I scored all EVT pre- and post-tests and another researcher re-scored 20% of the tests. There was 100% agreement.

Definition Assessment. The definitional assessment (Appendix D) was comprised of fifteen single-word items. All of the words were concrete nouns. Prior research had indicated that, for young children, nouns are easier than verbs or adjectives to define and concrete nouns are easier than abstract ones (Benelli et al, 2006; Johnson & Anglin, 1995). Some words were selected from items used in a study by Benelli et al. (2006) who focused on words from intelligence tests, and others were selected from a vocabulary activity developed as training activities for PAVEd for Success (Schwanenflugel, et al., 2005). Children were asked to describe the meaning of each word. In addition, a model definition was provided for pencil, a word believed to be in most children’s vocabulary at age 4. Items were scored for partial knowledge using 0 = no knowledge or incorrect knowledge, 1 = partial knowledge and 2 = full knowledge as described by Schwanenflugel et al. (1996). Interrater reliability was carried out by a second rater on 37% of the sample and the resulting Cohen’s Kappa was .84.

Language sample. Language samples were collected using a SONY digital audio tape recorder. The examiner escorted the child to a quiet area in a hall or room immediately outside of the classroom. The assent script was read and the digital tape recorder and microphone were shown to the children. Children assented in all cases. The examiner then engaged the child in an
interview following the Westerveld, Gillon, and Miller (2004) protocol (Appendix A). The first prompts were conversational and included open-ended questions about family and favored play activities. The second set of prompts was a series of photos intended to elicit personal narratives with a goal of three narratives. The final prompt was a wordless picture book *Good Dog Carl* (Day, 1991) or *Carl Goes Shopping* (Day, 1989) which the child was asked to narrate. This book was counterbalanced so that if the child received one book at pretest, he or she received the other at post-test.

For all prompts, the experimenters asked questions that helped extend the conversation but they tried to avoid introducing new vocabulary. When introducing the wordless picture book, the experimenters explained that there was no right or wrong story, but the child might make up any story to go along with the pictures. Language samples were transcribed according to CHAT specifications for analysis using the *vocd* program. See Appendix E for sample CLAN commands and output. Transcriptions were carried out by one of three researchers trained in language transcription and were rechecked by me to ensure transcription accuracy.

DAT recordings were converted to .wav files using Audacity ([http://audacity.sourceforge.net](http://audacity.sourceforge.net), 12/1/2006) and downloaded onto CDs. The .wav file was uploaded into Transcriber ([http://trans.sourceforge.net](http://trans.sourceforge.net), 1/12/2007), a program developed to assist in transcribing digital audio files. For a complete description of the transfer and data preparation process, refer to Appendix G.

**Results**

The results are analyzed according to Messick’s (1995) six aspects of construct validation. The analyses of these various forms of validity are taken together to form a
comprehensive picture of a measure’s validity. In what follows, each aspect of validity is evaluated in turn.

**Content Aspect of Construct Validity**

The methods, materials and results of obtaining, transcribing and analyzing the language samples and D scores were reviewed and evaluated related to theoretical fit. Attention was focused on the results gathered from the study as to whether they demonstrated characteristics discussed in the literature related to expressive vocabulary. D directly represents the relationship of the number of different words to the total number of words with a correction for variation related to sample size. There is no subjective element in the calculation of the D value. D provides a measure of diversity of vocabulary without regard to rare word use or accurate word choice, in other words the level of difficulty of the words used does not factor into the calculation. A teacher or researcher seeking a measure of vocabulary or vocabulary diversity is typically interested in the degree of a child’s vocabulary that the child puts to productive and communicative use. Although this is not identical to expressive vocabulary as assessed by psychometric tests such as the Expressive Vocabulary test, it should be moderately related to it.

In general a speaker with a more sophisticated vocabulary uses a more diverse set of words in communication (Malvern et al., 2004). However, some concerns arise related to expressive language variations that affect the ratio of unique words to total words. A child may omit common pronouns, conjunctions and prepositions and receive a higher D score than a child who speaks using more conventional grammar. To illustrate this concern, representative excerpts from three different transcripts are presented below. The first is an example from a child, who spoke using typical 4-year-old grammar with an average D score; the second is an example from a child using typical grammar but more sophisticated language who received a
higher D score. Last is an example from a transcript of a child who used a number of one-word unelaborated answers who obtained what could be considered to be an inflated D score:

Average D scores, average speaker, partial transcript, ID #25, (D = 41.93)

CHI: play.
CHI: with keeley in housekeeping.
CHI: play mommy.
CHI: playing kitty too.
CHI: play with my mommy at the park.
CHI: play and feed the duckies.
CHI: brother.
CHI: two.
CHI: and he's always feeling mean.

High D score, sophisticated speaker, partial transcript, ID #30, (D = 80.46)

CHI: i like to do blocks housekeeping and table toys.
CHI: you play with toys?
CHI: yeah but watch when you squeeze the legs.
CHI: it's i don't remember his name.
CHI: yeah actually his his [/] name is you know the name like when you know the name.
CHI: i don't really know the name but actually you know like one of those things that crawl on the ground and they can make a ball of theirselves that's what i'm talking about their names they make a ball out of theirselves and they have a big nose hey guess what me
and my daddy saw a big old black thing with a big old nose and eyes
and you know those things we saw one of those bad things that eat a
lot you remember the things that eat everything and they're mean.

High D score, unsophisticated speech, partial transcript, ID #40, (D = 77.42)

CHI: play.
CHI: housekeeping.
CHI: library.
CHI: blocks art science.
CHI: table toys.
CHI: football pirates of carribean.
CHI: oh you had a boat.
CHI: i got me two little sisters.
CHI: i'm the brother.
CHI: mcdonalds.

The final transcription presented illustrates a potential problem in calculating vocabulary diversity. If the child does not use typical grammatical conventions with its common high frequency function words, the D score is likely to have an inflated value over a child who used the same words but presented them with common conjunctions, prepositions and pronouns repeated throughout. When D was used to measure transcripts of student learners of a second language, this same pattern was observed when students omitted high frequency function words (Meara & Bell, 2001).

A concern on the other side of the spectrum relates to a speech style that may be considered sophisticated but which results a deflated D score. A speaker may intentionally
repeat a phrase to add emphasis, such as “It was huge as a house, not just huge as a house, but gigantic.” When calculating the ratio of unique tokens to total tokens, the repetition of huge as a house would diminish D, in this case, the phrase is repeated stylistically but not necessarily for a lack of alternative vocabulary. However, in the set of language samples collected in this study, children did not tend to repeat speech for emphasis and only in a few cases did they consistently omit conjunctions. Moreover, the two children who did respond primarily with one word answers did not yield the 50 tokens necessary for the program to run and subsequently those transcripts were not included in the analysis. Consequently, although this issue might be a concern in theory, in practice it did not appear to have had much affect on the validity of D for the current study.

A measure of expressive vocabulary should reflect typical growth that occurs over time. Two studies included as part of their evaluation a consideration of whether D scores reflect language development over a span of years. The Bristol, England study (Malvern et al, 2004) and a Hong Kong based study of children speaking Cantonese (Klee, Stokes et al. 2004) gathered transcripts and evaluated scores of typically developing children. The present study focused on collecting data from children who were at moderate-to-high risk for expressive vocabulary deficits. The scores from these three studies are plotted on Figure 2.1. When plotted against the values reported by these other studies, the values obtained in the present study fit broadly within the range of the expected scores. They represent a slightly lower limit which is to be expected because the present sample was selected from the lower end of the expressive vocabulary normal curve as determined by a standardized assessment.
Figure 2.1

*Mean D Scores from This Study and Two Others as a Function of Age*

**Substantive Aspect of Construct Validity**

Consideration of the substantive aspect included an analysis of whether different tasks included in the sample elicitation resulted in different D scores. Young children were expected to display higher D scores in the personnel narrative section because they are most likely to have practiced speaking about those episodes (Hudson & Shapiro, 1991). Although the scenarios pictured in the wordless picture books were intended to include familiar objects and events, it was predicted that more cognitive resources would be dedicated to trying to figure out the story and fewer on expression (Westerveld et al., 2004). Further, maintaining the main discourse characters across the storytelling would have the effect of reducing D as well. The transcripts were evaluated with regard to whether expected differences related to elicitation method.

Transcripts were divided according to elicitation method (personal narrative versus wordless picture book) and D scores were compared for each method. Only transcripts with 500
tokens or more were used, and each subdivided transcript was evaluated to ensure an adequate number of tokens for analysis. A D score was calculated for each elicitation type. A paired-samples $t$ test was conducted using the two sets of D scores from each sub-transcript to evaluate whether their means were significantly different according to elicitation method. As predicted, the results indicated that the mean for the personal narrative prompts ($M = 55.46, SD = 10.39$) was significantly greater than the mean for the wordless picture book narrative ($M = 41.15, SD = 15.68$), $t(40) = 5.75, p < .001$. The standardized effect size, $d$, was .90, indicating a large effect. The 95% confidence interval for the mean difference between the D scores was 9.27 to 19.30. Among the 41 transcripts with 500 tokens or more, 33 were attributable to different children. Eight of the transcripts were posttests of children with a pretest used in the analysis. A second paired-samples $t$ test was conducted without those 8 posttest scores to ensure that the assumption of independence was in no way violated and the result remained significant, $t(32) = 4.51, p < .001, d = .80$. Overall, children spoke about their personal experiences using much more diverse vocabulary.

The internal consistency of transcripts was another measure of interest in relation to the task. Internal consistency was computed using all the transcripts: a split-half coefficient expressed as a Spearman-Brown corrected correlation and coefficient alpha. Every other word was selected from each transcript resulting in an even words transcript and an odd words transcript. The decision to divide transcripts in this way took into account the concern that variation would occur within a transcript according to topic and elicitation, or even warming up versus focused time. The correlations between the full transcript and the two half transcripts: odd words and even words are presented in Table 2.2.
Table 2.2

*Correlations between Total D Scores and Split Half Odd and Even Tokens (N=136)*

<table>
<thead>
<tr>
<th></th>
<th>Even Words</th>
<th>Odd Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total D score</td>
<td>.884**</td>
<td>.902**</td>
</tr>
<tr>
<td>Even words</td>
<td>.645**</td>
<td></td>
</tr>
</tbody>
</table>

*p < .01

The internal consistency estimate between the even and odd word transcripts, using the Spearman Brown correction (Crocker & Algina, 1986), was .78 indicating satisfactory reliability. This procedure is based on the assumption that the two half-transcripts are parallel. Analyzing every other word throughout the transcript is preferable to dividing the transcripts into sections.

The protocol for language sampling in this study had only one form, but it is of interest how consistently children responded to this protocol at different times. A coefficient of stability was computed using the pre- and posttest scores of the control group. To date, test-retest reliability scores have not been reported in the literature for D. This type of reliability evidence is of concern particularly because of the lack of standardization in elicitation methods across researchers in other studies (Klee et al., 2004; Owen & Leonard, 2002; Silverman & Ratner, 2002). In the current study, the language samples were collected within 12 weeks of each other. The test-retest value was computed for the 32 children in the control group using pre- and posttest scores. The resulting coefficient of stability $r = .68$ ($p < .01$) provides support that D has substantial test-retest reliability.

Another potential source of evidence for the substantive aspect of validity is whether D is sensitive enough to detect a difference of scores in response to an intervention. A paired-samples $t$ test conducted using the means for the pre- and post-test scores of the control group
did not find a significant difference, \( t (31) = -1.37, p = .18 \). However for the experimental group the paired-samples \( t \) test indicated a significant difference between the pre- and post-test scores, \( t (35) = -2.63, p = .01, d = -.44 \). This difference is consistent with the hypothesis that the intervention would effectively increase vocabulary diversity. What is important to the validity study is that this difference is detectable following an intervention of such short duration.

**Structural Aspect of Construct Validity**

A persistent concern in the study of vocabulary diversity measures is the influence of sample size, overly small samples are unstable and yet scores of increasingly large samples tend to plateau and even decline. A comparison of D scores was conducted by analyzing different lengths of transcripts. According to Malvern et al. (2004), D scores should not be dependent on sample size; therefore if the sample is truncated at 250 tokens (words), the D score should not differ much from the same transcript original which may have 350 tokens. In the CLAN system the command line: `vocd +t*CHI +r6 +DS0 (and DS1) -s@exclude.cut sample.cha` produces a new version of the transcript comprised of half the tokens, including only every other word. This command initiates the following process: `vocd` is the program that calculates D; `t*CHI` restricts the analysis to the child’s tier of speech; `+r6` excludes self repetitions; `+DS0` restricts the analysis to every odd word of the child; `-s@exclude.cut` filters a pre-specified list of items from the analysis (eg. “mmhmm” and “uhhuh”); and `sample.cha` refers to the individual child’s file. This method of reducing the number of tokens was preferable to using only the first, middle or last 50 or 100 tokens, because the nature of the prompt has an influence on D.

In order to evaluate the effect of sample size, I compared the even and odd word transcripts prepared for the split half consistency analysis with the full pretest. This resulted in an overall comparison of a half transcript with a whole transcript regardless of original token size. The token range for the pretest was \( M = 423, SD = 238, \) range = 1161 and for the posttest
was $M = 411$, $SD = 214$, range = 968. The results of the paired-samples $t$ tests are shown in Table 2.3. The lack of significant difference between the samples is one indication that token size did not influence the D scores overall across a range of transcript sizes.

Table 2.3

*Full Sample Compared to Half Sample*

<table>
<thead>
<tr>
<th>Group</th>
<th>$M$</th>
<th>$SD$</th>
<th>$df$</th>
<th>$t$</th>
<th>$p$</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D Pretest – Odd Words</td>
<td>-1.49</td>
<td>8.2</td>
<td>67</td>
<td>-1.490</td>
<td>.141</td>
<td>6.73</td>
</tr>
<tr>
<td>D Pretest – Even Words</td>
<td>-1.14</td>
<td>9.6</td>
<td>67</td>
<td>-.979</td>
<td>.331</td>
<td>8.45</td>
</tr>
<tr>
<td>D Posttest – Odd Words</td>
<td>-0.63</td>
<td>8.7</td>
<td>67</td>
<td>-.595</td>
<td>.554</td>
<td>8.10</td>
</tr>
<tr>
<td>D Posttest – Even Words</td>
<td>-2.11</td>
<td>10.1</td>
<td>67</td>
<td>-1.721</td>
<td>.090</td>
<td>7.98</td>
</tr>
</tbody>
</table>

**Generalizability Aspect of Construct Validity**

One dimension of generalizability is whether D produces scores which would be biased against members of certain groups and in favor of others. The language sampling protocol intentionally included topics likely to be familiar to most children such as talking about family and common experiences; and one that is likely to be unfamiliar to most; narrating a wordless picture book.

To examine the generalizability of score meaning, scores across groups based on gender and SES were compared. A one-way multivariate analysis of covariance (MANCOVA) was conducted to evaluate the relationship between D and each factor. The D scores for the pretest and the posttest were entered as the dependent variables, while gender served as a fixed factor and SES was entered as a covariate. No significant differences were found for the main effects of gender, $F (2, 50) = .455$, $p = .637$, partial eta squared = .018, or SES $F (2, 50) = 1.438$, $p = .248$, partial eta squared = .054. The effects of gender and SES did not significantly influence the results in the sample suggesting substantial generalizability across groups.
**External Aspect of Construct Validity**

To assess the external aspect of D’s validity, the relationship between the D scores and the other two vocabulary measures, the EVT and definition assessment, was evaluated. Correlations are reported in Table 2.4.

Table 2.4

*Intercorrelations between Measures of Vocabulary (N=68)*

<table>
<thead>
<tr>
<th>Measures</th>
<th>Pre EVT</th>
<th>Pre Def</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre D</td>
<td>.522**</td>
<td>.426**</td>
</tr>
<tr>
<td>Pre EVT</td>
<td></td>
<td>.444**</td>
</tr>
</tbody>
</table>

** p < .01 (2-tailed), p < .05 level (2-tailed)

D measures an aspect of expressive vocabulary deployment, whereas the EVT introduces additional language elements by involving synonym knowledge, picture recognition, a one-word condition and includes words likely to be outside most children’s vocabulary. We would expect a moderate correlation between these two measures. The definition assessment included words with varied difficulty, similar to the EVT, but was designed to assess somewhat different vocabulary skills than expressive vocabulary, albeit an important skill for school. As a result one might expect a somewhat smaller correlation between definition assessment scores and D than between EVT and D. In this analysis I focused on pre-test scores because they were unaffected by differential treatment effects. As can be seen in Table 2.4, all correlations were moderate, significant, and in the expected direction.

Measurement statistics for vocabulary diversity are based on the ratio of types to tokens (or TTR), but previous ratio formulae were found to be significantly influenced by sample size and therefore could not be compared. Selecting a standard number of tokens from transcripts did not solve this problem because ratios taken from different parts of samples varied significantly.
Regardless of the unsatisfactory characteristics of TTR, D should be moderately correlated with TTR at least. D, a measure with a sample size correction, and TTR, the traditional ratio uncorrected for sample size, were correlated to see if this expected moderate correlation held.

Table 2.5 presents the means, standard deviations of the D and TTR pre-test scores. A correlation of .30 ($p < .05$) was obtained between the D and TTR scores, which was somewhat smaller than anticipated, but still statistically significant.

Table 2.5

<table>
<thead>
<tr>
<th>Group</th>
<th>$M$</th>
<th>$SD$</th>
<th>$N$</th>
</tr>
</thead>
<tbody>
<tr>
<td>D Pretest</td>
<td>54.26</td>
<td>16.88</td>
<td>68</td>
</tr>
<tr>
<td>TTR Pretest</td>
<td>.409</td>
<td>.093</td>
<td>68</td>
</tr>
</tbody>
</table>

Consequential Aspect of Construct Validity

Language assessments have the power to influence a child’s placement in leveled classes, referral to therapy and their relationship with their teacher. Messick (1980) and others have included the consideration of consequence as an integral part of validity study. Particularly important is to ensure that negative consequences of misdiagnosis may be prevented. In an educational setting a speech and language therapist could use D to discern change in response to various clinical strategies. However with the current level of development of D, it is not yet a standardized normative measure so scores cannot yet be used to attribute significant language problems and therefore cannot yet be determined to have consequential validity. However, the findings presented above suggest that it may, indeed, eventually be expanded to have those implications.
Discussion

The purpose of Study 1 was to examine the evidence collected and determine whether it provided support for the kind of inferences and actions that are evoked in relation to a vocabulary diversity score. Using Messick’s (1989) framework, evidence related to the validity of D as a vocabulary diversity measure was gathered, evaluated and presented.

Content Aspect of Construct Validity

In discussing the content aspect of validity it is necessary to define the domain boundaries of lexical diversity as well as including relevant social and cultural values and assumptions related to children’s expressive vocabulary skills. The first step of domain analysis was to define what is being measured as lexical diversity and for whom those measurements could be useful and important. Measures of vocabulary, both expressive and receptive are of interest to teachers, speech and language professionals and parents because of their researched relation to reading and academic success (NICHD 2000; NICHD ECCRN 2005; Shonkoff & Phillips, 2005.) These studies have used vocabulary measures based on survey tests such as the EVT and the PPVT. The domain of vocabulary sampled in these tests is culled from texts used in many schools and word frequency lists. To a large degree school success depends on a student’s understanding of the texts they are assigned to read, so the basis of a survey test is sound. For preliterate children though, it is difficult to remove the differential effects of literature exposure which is highly related to SES, from their score. D provides an alternate measure of expressive vocabulary deployment that is not linked to a child’s knowledge of rare vocabulary words.

Aspects of vocabulary not measured by survey tests, but related to achievement are: 1) communicative use of different words in speech and writing and 2) level of knowledge of meaning of words. D assesses an aspect of expressive vocabulary related to communicative
speech, measuring a person’s use of different words in relation to total words. Interest in vocabulary diversity includes the study of expressive vocabulary development and related disorders and extends to the study of communicative style, readability of texts, teacher and student language and is even used in forensic linguistics for authentification of a speaker’s words or writing. Some researchers (Hirsch-Pasek, Kochanoff, Newcombe & de Villiers, 2005; Laufer, 2003) have expanded the definition of vocabulary diversity to include variation in language sophistication, but D as measured through vocd does not capture degrees of difficulty, sophistication or accuracy in vocabulary use. This study was restricted to evaluating D as a measure of children’s expressive vocabulary development for use in linguistic investigation that could occur in homes, schools or laboratories.

One concern noted in the results was that an overestimated D score could result when a child used limited grammar, mostly nouns and verbs. Sample transcripts were shown in which children only provided one or two words answers for which an ‘inflated’ D value was assigned. Presently there is neither a protocol for language sampling nor rules for exclusion and inclusion of samples. If transcripts were restricted to those in which normal grammar was used, this problem would be solved.

Substantive Aspect of Construct Validity

Messick (1998) warns of two threats to construct validity, construct irrelevance and construct underrepresentation. The promise of the D measure is that it may circumvent construct irrelevance that is almost unavoidable in traditional tests involving surveys of vocabulary which run the risk of differentially favoring children who have had a wider array of experiences usually directly related to SES. By collecting samples of natural speech, there is a greater likelihood of measuring positive qualities of expressive language regardless of SES and other concerns related to bias in testing.
The threat of construct underrepresentation in relation to D was studied by evaluating the consistency of D with regards to two methods of language sampling. At this time there is no standard protocol associated with sample collection. In this study we have demonstrated that how a language sample is collected affects the size of D. For 4 year olds, language elicited from narrating a wordless picture book produced D scores indicating that children’s language was less diverse than when they were assessed while creating personal narratives. Until there is a consistent protocol associated with obtaining D, scores cannot be interpreted and compared across studies without study and evaluation of the particular protocol prompt used. The full protocol used for this study shows promise in providing a broadly representative sample of speech.

Using the present protocol, internal consistency estimates were satisfactorily high. A version of the split-half method was used, creating two half transcripts out of every other word. The lack of independence between the two split transcripts of conversational speech is likely to boost the reliability estimate, however the estimate from this study and others (McKee, Malvern & Richards, 2000; Malvern et al., 2004) provide confidence in the area of internal consistency. The substantial test-retest reliability estimate obtaining from the control group pre- and post-tests contributes new validity evidence to the literature and confidence of stability.

Finally, a limitation of standardized survey based vocabulary tests is that they are not designed to be effective instruments for measuring vocabulary gain from short term (less than one year) interventions. Speech and language therapists need measures that can distinguish vocabulary-related growth in shorter time periods so as to evaluate their methods. Education researchers seeking to evaluate the effectiveness of expressive language interventions also need
sensitive measures. In this study D was shown to be sensitive to a vocabulary centered intervention, a useful quality for research and evaluation of teaching strategies.

*Structural Aspect of Construct Validity*

Previous measures of lexical diversity, NDW and TTR, have been considered invalid because of their dependence on sample size. D was developed specifically to address the sample size problem. In this study, I compared split-half samples with the whole sample and found no significant differences in mean D scores. Evaluation of the structural aspect of a measure also must consider the task involved. The task used for D is one of the measure’s strong points. D is calculated from a sample of natural speech that represents the kind of talk that occurs during the academic day. A speech sample is a fairly direct representation of the vocabulary they are able to use. One concern related to using language samples should be noted however. It is time consuming to transcribe language samples, yet it is necessary preparation for the analysis program.

*Generalizability Aspect of Construct Validity*

The generalizability aspect of validity is concerned with whether the measure is consistent across population groups and time. A multivariate analysis was conducted and the results did not reveal any significant differences due to gender or SES. The lack of a significant effect for gender is consistent with previous analyses conducted by Malvern et al. (2004) using the Bristol and New England Corpuses. In the sample used for this study the mean SES was lower than average and not highly variable. Malvern et al. (2004) similarly, found no significant effect for SES in the Bristol Corpus, but they did find a weak one in the New England Corpus as did Wells (1985) in his English sample with higher D scores associated with children from a higher SES. Unfortunately, information regarding race/ethnicity was not collected for this study.
However, in another study I conducted (Ruston, Schwanenflugel, & MacLean, 2006) using similar language sample elicitation methods, D scores were collected from 19 children (10 African American and 9 European American) who attended either Head Start or 4-year-old prekindergarten program similar to the one reported here. That study also did not find differences related to race/ethnicity.

The children in the control group were measured at the beginning of the study and then again 10 weeks later without a significant difference between the scores. Thus this test-retest reliability suggests that D provided a stable measure over a short time period, as might be expected if children were not part of an intervention to improve their language development. Taking all of these findings together, it can be inferred that this measure shows substantial generalizability across groups.

*External Aspect of Construct Validity*

Evidence of convergent and discriminant relationships between D and other measures of vocabulary were analyzed to evaluate the external aspect of validity. We expected a moderate correlation with EVT and definition test scores because each assessed a slightly different aspect of expressive vocabulary. The correlations were significant and moderate and slightly higher between D and EVT than between D and the definition test. The EVT requires the subject to produce vocabulary words which is part of what is assessed in D; the definition task measures the more metacognitive ability of being to articulate the meanings of words and would thus be expected to have the least association. These results were consistent with Silverman and Ratner’s (2002) study. They found a significant correlation between D and the standardized expressive measure used (nearly identical to the one presented here, .48), but nonsignificant relationships with a receptive measure.
There was a low, but statistically significant correlation between D and TTR. This was somewhat surprising given the variability of sample size in this study and the dependence of TTR on sample size. Silverman and Ratner (2002) did not find D and TTR to be significantly related in two separate analyses. First they used the entire sample, as I did, and then modified the TTR transcripts to include only the middle 50 utterances from each sample. In the present study the scores were related but not robustly.

Consequential Aspect of Construct Validity

In the context of teaching vocabulary in the elementary school classroom, D is not likely to become a widely used measure by teachers or speech pathologists because of the time it takes to collect and transcribe a sample of children’s language. However, when automatic transcription technology becomes available and financially feasible, this measure could be useful to speech pathologists. New advances in automated transcribing software such as Dragon Naturally Speaking Preferred suggest that research and development is moving forward in this area. For teachers as well as speech and language pathologists, D measures a vocabulary dimension that is different than what is measured in the Expressive Vocabulary Test. A low score on D would in most cases indicate a need for additional speech and language support with an emphasis on vocabulary. A high score on D could indicate an aptitude for word learning that is independent of SES.

Conclusion

In sum, Messick’s unitary construct validity framework was used to evaluate the use of D as a measure of vocabulary diversity. When the D scores from this study were plotted with scores from other studies, it was seen that D was able to measure development in vocabulary diversity over time. Psychometric qualities of the measure were demonstrated to be strong, both
internal consistency and test-retest reliability estimates were significant. Patterns of correlation suggested that D measures an aspect of expressive vocabulary related to the construct measured in a standardized expressive vocabulary test but was not exactly the same. The evidence collected in this study showed D to be generalizable across gender and SES. In addition, D was successful in detecting effects of a vocabulary intervention in the short period of 10 weeks.

The limitations of D as a measure encountered in this study were more related to the preparation of the transcripts and the interpretation of the D scores than to the psychometric aspects. A first limitation is that a standard protocol is necessary before comparisons of D scores between studies can be made because it was shown that D is influenced by the type of prompts involved in the sample collection. However, as of yet, a standard protocol has not been associated with the measure, severely limiting the kind of judgments and inferences that can be made across studies. A second limitation is a concern regarding practicality. The time taken to prepare transcripts, given current technology, is considerable in contrast to the relatively short time it takes to administer a standardized vocabulary test. Finally, care should be taken when interpreting D values. D calculates a score for the diversity of vocabulary used in speech, but does not include any evaluation of the density of rare or sophisticated words or the accuracy of use. Therefore very simple speech could receive a high score if high frequency words were omitted and very sophisticated speech runs the risk of devaluation if speech conventions using repetition are intentionally used to create emphasis and suspense.

Overall, D scores provided a useful measure of vocabulary deployment that reflected a relative degree of language development and was reasonably correlated with another expressive vocabulary measure. Although D scores differed according to the type of language prompts used, they did not vary as a function of sample size. In this study D was successful in
discriminating differences attributable to an intervention, a promising feature for future research and therapeutic uses. Expressive vocabulary knowledge is a construct of great interest to educators and researchers, and D offers a relatively valid and reliable measure of one aspect; lexical diversity.
CHAPTER III

STUDY 2

Talk Time, an Intervention for Enhancing Expressive Vocabulary using Vocabulary Enriched Talk

This study evaluated the effects of an intervention designed to enhance the vocabulary levels of prekindergarten children by increasing the quantity and quality of oral language experiences of children within a preschool setting. I studied whether or not a measurable increase in vocabulary level can result from a simple preschool-based intervention involving systematic engagement of children in conversation carried out over a 10-week period which totalled an additional 500 minutes of talk. The conversations took place between a researcher trained to communicate responsively and pairs of children. Conversation techniques were used that emphasized cognitively challenging talk (interpreting, hypothesizing, and clarifying), a factor linked to reading achievement. The researcher modeled the use of vocabulary words, and used extensions, expansions, and recasts. I evaluated whether this level of intervention improved children’s expressive vocabulary through both the standardized assessment, the Expressive Vocabulary Test, and unstandardized assessments, definition assessment and language sampling, described in Study 1.

It was anticipated that this intervention would be particularly beneficial for children with low levels of vocabulary and language development. As noted in the introduction, one major cause for low levels of language development and vocabulary in children with limited vocabulary skills was the relative lack of exposure to systematic cognitively challenging speech
from caretakers. This intervention was directly intended to support that need. Thus, while it might be that all children might benefit from this type of intervention, I predict that children who begin the intervention with low levels of vocabulary might benefit particularly from this intervention compared to children with average or good vocabulary skills.

Method

Participants

Children were the same as those from Study 1. For this study, missing scores were estimated. Two D pretest scores were missing because of recording difficulties and three children from the control group were unavailable for post-testing. The total sample for this study was 73 children and was described in Table 2.1.

Experimenter characteristics and training. Three senior undergraduate psychology majors (two females, one male) and one male beginning graduate student attending the University of Georgia carried out the intervention. I also served as an experimenter. Each experimenter had a minimum of one semester of supervised tutoring or intervention experience with prekindergarten children and/or school teaching experience. All were participating for research course credit. Two experimenters were assigned 5 pairs of children, one was assigned 4 pairs, one 3 pairs and I met with one pair. Further, I served as a substitute when one was needed.

Experimenters attended two two-hour training sessions prior to the intervention (Appendix F). In the first two hours, training included instruction in techniques intended to foster children’s joint interest in an activity (e.g. allowing adequate wait time for children to initiate conversation and following the lead of the child during conversation.) In addition, emphasis was placed on introducing vocabulary naturalistically into conversation through vocabulary recasting (with elaboration) and expansions of children’s utterances. Training
included strategies designed to encourage children to balance turn taking and to build upon ongoing subjects in their conversation pairs. In the second two hours, experimenters practiced language modeling techniques with a pilot child and practiced delivering the assessments. Further, throughout the intervention period, experimenters received ongoing feedback and coaching on their performance from me.

The experimenter team communicated via e-mail or in weekly meetings to share and update information about the progress of the intervention and receive information for the upcoming week. I monitored the implementation of the intervention by the experimenters by talking with them on the telephone or in person biweekly to discuss progress, questions and concerns and to provide feedback. A mid-intervention conversation was taped from each pair of children to determine fidelity with trained intervention practices. I set criteria that each experimenter needed to reach in order to be considered to have met the program goals for the conversations. For each experimenter, I tabulated six conversation behaviors from the first 15 minutes of recorded conversation: (a) number of new vocabulary words introduced in context (b) number of new vocabulary words explicitly defined, (c) number of new vocabulary words presented by recasting, (d) number of open-ended questions, (e) number of forced choice questions and (f) number of yes/no questions. To be considered to meet with program vocabulary goals, experimenters would have needed to introduce 20 new words through a combination of context, definition or recasting presentations. To meet the program conversation goals, the experimenter would have needed to use at least 15 open-ended questions and fewer than 40 forced choice or yes/no questions. No significant differences were found between experimenters while they engaged in conversation.
Finally, an informal short post-intervention interview was conducted with each experimenter by telephone. Three questions were asked: what worked, what didn’t; and what surprised you?

*Intervention Procedures*

Prior to the intervention, teacher input was gathered to assist in arranging compatible child pairs. Each experimenter met with the children for 25 minutes twice a week over a period of 10 weeks for a total of 500 minutes of conversation per pair of children. This same adult conversational partner (who we referred to with the children as *Talking Buddies*) met regularly with the same children to maintain consistency of the pairs and to build rapport. However, I served as a substitute for one conversation when an experimenter was ill.

Conversations took place in a quiet area of the center. At two centers the hall was used, at one center the cafeteria was used and at the remaining center the staff room was used. Conversation topics were suggested for the initial sessions so that the experimenters had a starting point for conversations. These stems included talk about family members and pets, holiday activities and toy prompts such as wordless picture books, toy phones and plastic figures. Conversation topics for later sessions were based on experimenter’s knowledge of children’s interests and experimenters continued to bring a variety of different props and toys designed to stimulate conversation. Children received a book at the end of the intervention for their participation.

*Assessments.* Detailed information on the pre- and post-test assessment protocol is provided in Study 1. Pre- and post-test assessments included the EVT, a definition task, and the elicitation of an oral language sample including the narration of a wordless picture book and guided conversation. Examiners conducted EVT and definitional assessments during one 15-
minute session, with the order of these assessments counterbalanced within the session. At a second session, a language sample was collected. These two sessions were not counterbalanced because EVT scores were used to eliminate subjects that performed more than one standard deviation above or one and one-half standard deviations below the mean. All pre-test assessments took place within a period of two weeks prior to the intervention. Post-test assessments were conducted within two weeks following the intervention. Children’s post-tests were not conducted by the same adult who provided the intervention to control for the children’s familiarity with the tester at pre- and post-test.

Design

A pre-test post-test control group design was used. Subjects were initially matched on a standardized vocabulary test score and one member of each pair was randomly assigned to the intervention group. Thus, the study could be considered quasi-experimental because the subjects were matched on vocabulary pre-tests. The same assessments were used for pre-testing as for post-testing. Group (experimental versus control) served as a between-subjects factor. Time (pre-test versus post-test) served as a within-subjects factor. Sample size was determined by using Cohen’s (1977) power chart, which recommended a sample size of 34 for a power of 80 using an effect size (i.e. $d$) of .3 for a one-tailed alpha level of .1, given a directional hypothesis. Dependent variables were EVT standard scores, lexical diversity or $D$, and definition assessment sum scores using partial knowledge scoring.

Results

Pre-test and post-test comparisons were examined so as to identify any differences between the experimental and control groups on any of the three dependent variables; $D$, EVT
and definition assessment on the pre- and post-tests. The mean scores are presented in Table 3.1.

Table 3.1

*Means and Standard Deviations on Assessments for Experimental and Control Groups*

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th></th>
<th>Experimental</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>N</td>
<td>M</td>
</tr>
<tr>
<td>EVT pre-test</td>
<td>96.53</td>
<td>9.10</td>
<td>38</td>
<td>94.83</td>
</tr>
<tr>
<td>D pre-test</td>
<td>52.73</td>
<td>14.97</td>
<td>38</td>
<td>53.87</td>
</tr>
<tr>
<td>Definition pre-test</td>
<td>11.6</td>
<td>6.5</td>
<td>38</td>
<td>12.7</td>
</tr>
<tr>
<td>EVT post-test</td>
<td>97.61</td>
<td>10.99</td>
<td>38</td>
<td>99.63</td>
</tr>
<tr>
<td>D post-test</td>
<td>56.74</td>
<td>15.60</td>
<td>38</td>
<td>60.95</td>
</tr>
<tr>
<td>Definition post-test</td>
<td>11.87</td>
<td>5.9</td>
<td>38</td>
<td>13.5</td>
</tr>
</tbody>
</table>

The analyses presented including descriptive and inferential statistics were chosen to respond to the primary research question investigated in this study which asked whether an intervention of vocabulary enriched talk would result in improved vocabulary levels as measured by a standardized vocabulary assessment, a definitional task and D. For each measure, I also tested the prediction that children who began the intervention with low levels of vocabulary would show significant growth compared to children with simply average or good vocabulary skills who may or may not benefit from this type of intervention. To determine the extent to which *a priori* vocabulary skills differentially influenced the post-test outcomes, the subjects were further grouped into a low vocabulary group and an average group based on EVT pre-test scores. The low group included children scoring below at or below one half of a standard deviation below the mean of the pre-test EVT. Children in the low group had a standard score of 93 or lower which placed them at or below the 32nd percentile and was intended to represent children with moderate to high risk for low language skills. The average group was comprised of children who achieved a standard score of 94 through 114, which placed them between the 35th and 84th percentile.
Subject matching and the use of the same test battery are two important elements necessary for the use of within-subjects, repeated-measures ANOVA (Gliner, Morgan & Harmon, 2002). Subject matching, based on expressive vocabulary skill, was intended to reduce the variability among the participants. Under these conditions, data is best analyzed using a statistic that accounts for the reduction in participant error variance which repeated measures ANOVA does. Using the same instruments for pre- and post-testing reduced error based on effects of different instrumentation and met the second requirement for the use of repeated-measures ANOVA.

To investigate the difference between and within groups, repeated-measures ANOVAs were performed comparing the means for the pre- and post-test scores on the three vocabulary measures. Children’s natural development of language skills during the first months of prekindergarten is a likely outcome of educational programs; therefore a statistical procedure was used that compared pre-test and post-test scores both between and within the groups to minimize the effect of maturation on tests for intervention outcomes. The assumption of sphericity necessary for this statistical procedure did not present a concern because there were not more than two levels of a within subjects factor; Mauchley’s W statistic of 1.00 indicted that the levels of within-subject variables were equal. The following analyses investigated the effect of the intervention and all were conducted using SPSS Version 14.0 for Windows.

A 2 condition (intervention versus control) X 2 time (pre- versus post-test) ANOVA was carried out on EVT standard scores with group as a between-subjects variable and time as a repeated-measures variable. There was a significant main effect of time, $F(1, 71) = 11.37, p = .001$, but a nonsignificant main effect of group, $F(1, 71) < 1$. As predicted, there was a significant interaction between these two factors, Wilk’s $\Lambda = .94, F(1, 71) = 4.56, p = .036,$
partial eta squared = .06 (which related to a Cohen (1988) “medium effect size”). The results are displayed in Figure 3.1.

![Graph showing EVT Standard Scores for Whole Sample](image)

**Figure 3.1. Effects of the Intervention on EVT Standard Scores for Whole Sample**

A second analysis was carried out to test the prediction that the effects of the intervention would be greater for initially low vocabulary children than for children with average to good vocabulary skills. The means for each group are displayed in Table 3.2. A 2 condition X 2 time X 2 skill ANOVA did not yield the anticipated 3-way interaction, $F (1, 68) = 2.57$, $p = .114$, partial eta squared = .036. Still, given that I had a strong prediction that children who initially had low vocabulary skills were more likely to benefit from the intervention, I carried out a 2 Condition X 2 Time partial ANOVAs on each skill group’s data separately. The results of these tests for the interaction can be found in Table 3.3 and Figure 3.2 which shows that there was a significant benefit of the intervention for children starting the intervention with low vocabulary skills, but not for children with average vocabulary skills.
Table 3.2

Means and Standard Deviations of EVT Scores by A Priori Vocabulary Levels

<table>
<thead>
<tr>
<th></th>
<th>Pre-test</th>
<th></th>
<th>Post-test</th>
<th></th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low Group Control</td>
<td>88.00</td>
<td>5.06</td>
<td>88.93</td>
<td>4.93</td>
<td>15</td>
</tr>
<tr>
<td>Low Group Intervention</td>
<td>87.59</td>
<td>4.85</td>
<td>95.12</td>
<td>7.82</td>
<td>17</td>
</tr>
<tr>
<td>Average Group Control</td>
<td>102.09</td>
<td>6.42</td>
<td>103.26</td>
<td>10.15</td>
<td>23</td>
</tr>
<tr>
<td>Average Group Intervention</td>
<td>101.67</td>
<td>5.61</td>
<td>103.89</td>
<td>9.72</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.3

Repeated-Measures ANOVA Results on EVT Scores by A Priori Vocabulary Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1, 30</td>
<td>7.721</td>
<td>.009</td>
<td>.205</td>
</tr>
<tr>
<td>Average</td>
<td>1, 39</td>
<td>.185</td>
<td>.669</td>
<td>.005</td>
</tr>
</tbody>
</table>

Figure 3.2. Change in EVT Scores as a Function of the Intervention for Children with Low versus Average Initial Vocabulary Skill
The same set of analyses were then carried out on D scores. The first ANOVA was a 2 condition (intervention versus control) X 2 time (pre- versus post-test) with group as a between-subjects variable and time as a repeated-measures variable. There was a significant main effect of time, $F(1, 71) = 11.113, p = .001$, but a nonsignificant main effect of condition, $F(1,71) = .623, p = .433$. Different than predicted, there a nonsignificant interaction between these two factors, Wilk’s $\Lambda = .988, F(1, 71) = .848, p = .360$, partial eta squared = .012 (which related to a Cohen (1988) “small effect size”). The results are displayed in Figure 3.3.

A second set of analyses were carried out to test the prediction that the effects of the intervention would be greater for initially low vocabulary children than for children with average to good vocabulary skills using the criteria for determining average and low levels of vocabulary. The means for each group are displayed in Table 3.4. As predicted, significant results were found from a 2 condition X 2 time X 2 skill ANOVA 3-way interaction $F(1, 68) = 8.39, p = .005,$
partial eta squared = .110 (which related to Cohen’s (1988) “large effect size”). I then carried out 2 Condition X 2 Time partial ANOVAs on each skill group’s data separately. The results of these tests for the interaction are provided in Tables 3.4 and 3.5 and displayed in Figure 3.4. They indicated a significant benefit of the intervention for children starting the intervention with low vocabulary skills, but not for children with average or good vocabulary skills.

Table 3.4

Means and Standard Deviations of D Scores by A Priori Vocabulary Levels

<table>
<thead>
<tr>
<th></th>
<th>Pre-test M</th>
<th>Pre-test SD</th>
<th>Post-test M</th>
<th>Post-test SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Group Control</td>
<td>45.0</td>
<td>12.31</td>
<td>48.1</td>
<td>12.31</td>
<td>15</td>
</tr>
<tr>
<td>Low Group Intervention</td>
<td>44.58</td>
<td>14.82</td>
<td>60.37</td>
<td>14.16</td>
<td>17</td>
</tr>
<tr>
<td>Average Group Control</td>
<td>57.77</td>
<td>15.31</td>
<td>62.38</td>
<td>15.30</td>
<td>23</td>
</tr>
<tr>
<td>Average Group Intervention</td>
<td>62.65</td>
<td>16.54</td>
<td>61.49</td>
<td>17.45</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.5

Repeated-Measures ANOVA Results on D Scores by A Priori Vocabulary Levels

<table>
<thead>
<tr>
<th>Level</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1, 30</td>
<td>10.366</td>
<td>.003</td>
<td>.257</td>
</tr>
<tr>
<td>Average</td>
<td>1, 39</td>
<td>1.617</td>
<td>.198</td>
<td>.040</td>
</tr>
</tbody>
</table>
The final set of analyses using the definition scores were carried out. First, a 2 condition (intervention versus control) X 2 time (pre- versus post-test) ANOVA with group as a between-subjects variable and time as a repeated-measures variable was computed. There was not a significant main effect of time, $F(1, 71) = 1.44$, $p = .234$, partial eta squared .020, nor a significant main effect of condition, $F(1,71) = .992$, $p = .323$. The interaction between these two factors, was also nonsignificant, Wilk’s $\Lambda = .995$, $F(1, 71) = .386$, $p = .537$, partial eta sq .005.. The results are displayed in Figure 3.6. Similarly when the data from the tests were analyzed by a priori vocabulary skill level, a nonsignificant result from the 2 condition X 2 time X 2 skill ANOVA 3-way interaction, $F(1, 68) = .103$, $p =.749$, partial eta squared = .002. The results are displayed in Tables 3.6 and 3.7.
Figure 3.5. *Effects of the Intervention on Definition Standard Scores for Whole Sample*

Table 3.6

*Means and Standard Deviations of Definition Scores by A Priori Vocabulary Levels*

<table>
<thead>
<tr>
<th>A Priori Vocabulary Levels</th>
<th>Pre-test M</th>
<th>Pre-test SD</th>
<th>Post-test M</th>
<th>Post-test SD</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low Group Control</td>
<td>8.33</td>
<td>4.32</td>
<td>9.00</td>
<td>3.53</td>
<td>15</td>
</tr>
<tr>
<td>Low Group Intervention</td>
<td>10.29</td>
<td>5.07</td>
<td>11.84</td>
<td>5.60</td>
<td>17</td>
</tr>
<tr>
<td>Average Group Control</td>
<td>13.74</td>
<td>6.90</td>
<td>13.74</td>
<td>6.9</td>
<td>23</td>
</tr>
<tr>
<td>Average Group Intervention</td>
<td>14.89</td>
<td>5.57</td>
<td>15.22</td>
<td>5.89</td>
<td>18</td>
</tr>
</tbody>
</table>

Table 3.7

*Repeated-Measures ANOVA Results on Definition Scores by A Priori Vocabulary Levels*

<table>
<thead>
<tr>
<th>Level</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>partial eta squared</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>1, 30</td>
<td>.344</td>
<td>.562</td>
<td>.011</td>
</tr>
<tr>
<td>Average</td>
<td>1, 39</td>
<td>.06</td>
<td>.808</td>
<td>.002</td>
</tr>
</tbody>
</table>
During the informal post-interviews with experimenters several common themes emerged. When asked what aspects of Talk Time worked best, each of the examiners expressed that the sessions with children were fun and enjoyable. The props considered most successful were pictures from books and magazines, play dough and small figures; and the most productive topics were friends, parties and holidays. When asked what aspects didn’t work, experimenters commented that 25 minutes was uncomfortably long at first and it took several weeks to adjust to the length of time. Two experimenters discussed behavior problems they felt adversely affected certain groups. In one situation a child consistently tried to monopolize the attention of the experimenter and in another pair, one child was very physically active and difficult to keep engaged. When asked what surprised them, two common themes were expressed. One was that children didn’t begin to converse with each other until the end of the intervention, if at all. Most conversations were dyadic between a child and experimenter, with each child taking turns. Notably experimenters mentioned that for each pair during a session within the first month the relationship between the three of them shifted to become more comfortable. Following this shift it appeared that children became much more risk-taking in their language use.

Discussion

The findings of the study reported here provide evidence in support of vocabulary enriched conversation between adults and children held in school settings as an effective strategy for improving vocabulary levels, particularly for children with poor vocabularies. Although the findings were not robust, the concentration of the intervention, 500 minutes of cognitively challenging conversation distributed over ten weeks, was very modest. Five hundred minutes spread throughout a normal school year of thirty five weeks is roughly equivalent to 15 minutes of conversation per week per pair of children. This study provided evidence that this practice
could significantly improve the expressive vocabulary levels of four year olds. Furthermore, for children with low levels of vocabulary, improvement was robust. Given our understanding that children fall behind in school achievement when they begin their education with impoverished vocabularies, this study provides useful information on a strategy for augmenting preliterate children’s vocabulary levels.

Vocabulary growth as measured by standardized assessments is typically slow and steady. The EVT is intended to be used in school settings to survey vocabulary knowledge and growth on a yearly basis. In studies such as Peterson and McCabe’s (1994) and in PAVEd for Success (2005), expressive vocabulary grew as a result of the intervention, but that growth was not measured by standardized assessments such as the EVT until 6 months or more following an intervention. Finding improvement through this measure after the relatively short duration of this study is notable.

The results from D were somewhat different. Whereas no significant difference in lexical diversity emerged for the group taken as a whole, children in the intervention group with low a priori vocabulary levels were measured to have significantly higher lexical diversity on post-tests. This evidence when combined with the EVT results supports the use of sustained cognitively challenging conversation with enriched vocabulary as an intervention for children at the lower end of the vocabulary spectrum. Both lexical diversity and standardized expressive vocabulary measures improved for the low group. In Study 1 it was demonstrated that the lexical diversity measures a different aspect of expressive vocabulary than the standardized vocabulary measure. That lexical diversity did not improve for the overall group may be attributable to factors related to the construct. For lexical diversity to increase, a speaker must use a more diverse vocabulary in their natural speech. Measures of lexical diversity calculated
from samples of average speakers may grow more gradually than what can be measured in the short, ten-week span of this study.

Improvement on the definition task scores was in a positive direction but was not statistically significant. However, this task does not measure expressive vocabulary directly, but, instead, is a related skill that becomes increasingly important to reading achievement and curriculum-based assessment of vocabulary knowledge as children get older. The ability to reflect on and describe the meaning of a word involves metalinguistic development that likely has a very high variation at this young age. In the Hart and Risley (1995) study, it was noted that the parents of the high vocabulary group tended to introduce new vocabulary and follow with a definition. That practice was incorporated by the experimenters in this study. However the intervention did not include ask children to engage in this practice themselves by defining the new words that they used. The ability of the child to verbalize a meaning for deployed vocabulary was not tested in the Hart and Risley (1995) study and for this young age group existing research on that skill is limited. Thus, in terms of having immediate importance for the evaluation of this intervention, I consider the EVT and D to be more direct measures of the skills I have attempted to affect.

Beyond the obvious language practice and modeling of language, there are several other reasons that the particular choices I made may have enabled the intervention to be successful. In setting up the intervention, I had to decide what the optimal group size might be that would be both feasible for teachers to carry out and for children to benefit from. The decision to hold conversations with pairs of children, as opposed to individual children or with larger groups, was driven both by practical considerations and for its consistency with social-interactions theory of language acquisition (Dickinson & McCabe, 1991; Girolametto, Weitzman, Lieshout, & Duff,
Groups containing larger numbers of children are more likely to present the problem that more talkative children might dominate the conversation, thereby replicating a problem that already exists in preschool classrooms. Although conversing with children individually might more closely resemble optimal parent/child home settings, it would be less feasible for a teacher to sustain. By conversing with children in pairs, some of the intensity of the individual conversation is achieved, while also providing an opportunity to build the beneficial conversation skills related to talking with adults and peers in small groups. In the early childhood environment, the discourse patterns that children and adults engage in are likely to have a strong influence on children’s developing language as well as on their motivation to engage in future conversational interactions. In setting up conversations with pairs of children, skills important to talking with adults as well as peers were modeled and practiced. In addition, children reluctant to speak had the benefit of listening to another child converse with the adult and then the opportunity to practice in a low-threat environment.

A second ancillary factor that may have supported the effectiveness of the intervention for children was the fact that the children could develop a positive attachment to their Talking Buddy. A strong association exists between language competence and a secure attachment according to a meta-analytic study conducted by van Izendoorn, Dijkstra and Bus (1995). Although their study measured associations and could not determine causation, it is likely that the factors of language development and attachment share a bootstrapping relationship. As a child becomes a more competent and engaging communicator, adults are increasingly likely to verbally interact, in turn allowing for further exposure to increased and possibly complex language (McCabe, Peterson & Connor, 2006). Studies of children’s attachment to child care providers have found a direct relationship between quality of care factors and multiple child
development indicators (Howes and Hamilton, 1993; Peisner-Feinburg & Burchinal, 1997). It is reasonable to believe that the skills involved in building and maintaining positive relationships with care providers can be carried into Kindergarten and the first years of school. Early relationships with teachers strongly contribute to students’ achievement of academic and social competence in their early years of school (Pianta & Stuhlman, 2004; Pianta, LaParo, Payne, Cox, & Bradley, 2002). O’Connor and McCartney (2007) were able to discern from the National Institute of Child Health and Human Development Study of Early Care and Education (NICHD) that relationships between children and teachers vary and impact achievement accordingly. After controlling for child and family factors, they found that the quality of teacher-child relationship was a significant predictor of third grade achievement.

A foundational plank of the teacher-child relationship rests on effective communication. However many instructional strategies emphasize developing academic achievement using pedagogical methods that do not include building relationships with children who may have difficulty communicating. Building conversation skills could be a crucial piece of developing better relationships with students and thus could become a key component for increasing achievement (O’Connor & McCartney 2007.)

In this study, attachment indicators were not measured systematically and studied as a variable. However during post-interviews with the experimenters, each of the experimenters spontaneously noted a perceptible shift occurred toward improved ease and quality of conversation when a level of trust, comfort and connection with the child had been established. Each of us noticed that when that occurred children became more risk-taking in their language use and willingness to try out new words. The amount of time necessary to establish this level of comfort varied between children and was likely influenced by a number of factors. These factors
might include the personality match between the child and the experimenter, prior experience of
the child at school and home, temperament and personality characteristics of the child as well as
dynamics related to the pairing of children. Overall though, it was impressive how quickly this
sense of relationship developed given the short duration of this intervention. In a classroom,
there is tremendous variation between children’s readiness to develop a comfortable relationship
from which to practice their language skills. I would suspect that the quietest children who may
not have engaged in any teacher-child communication would be the most at risk for never
developing that relationship. Implementing a systematic conversation program could help the
quietest children advance in their language skills and perhaps in their relationships with their
teachers.

How might this intervention be carried out in a real classroom rather than through using
outside experimenters as was done here? In the PAVEd for Success intervention
(Schwanenflugel et al., 2006), some teachers successfully carried out conversation time with
children as a center activity or ‘talk center,’ as part of ‘family style meals’, during playtime
either at a bench or on a walk, at naptime with non-sleepers, and during pickup and drop-off
time. ‘Walk Time’ can be carried out during recess where a teacher can take a small group of
students on a walk around the school yard, conversing as they go. Over time both the students
and the teacher will look forward to these conversations, a factor likely to contribute to the
teacher sustaining the practice.

Limitations

Several limitations in the methodology must be noted when interpreting the findings of
this study. First, the intervention was not carried out by the children’s teachers. The
experimenters were University students with prior experience working with young children,
professionally or in research contexts. However, the overall experience level of this group was considerably less than the typical population of preschool and kindergarten teachers. It is possible that the effects of the intervention might have been even stronger and pervasive if the children’s actual teachers were used.

Second, the conversations did not take place within the classroom although they did take place within the school day. Conversations held during the intervention took place in a setting just outside the classroom. Should a conversation component be added to the kindergarten curriculum, it would likely have to take place within the classroom, at a lunch or snack table in the cafeteria, or possibly on the playground. If sessions were conducted by a speech and language professional it is possible they might take place outside the classroom. Both teachers and center directors agreed that, as experimenters, if we tried to hold conversations in the classroom we would have been distracting to and distracted by the other children. However a teacher, assistant or regular volunteer should be able to establish behavior expectations for other children related to a conversation center or program.

Teachers, unlike the experimenters, have already begun to establish relationships with children, a factor which might reduce the amount of time needed in pairs. Experimenters for the most part reported good relationships with their pairs of children with the exception of one case in which the child became very needy of attention and in another where the researcher had difficulty controlling the behavior of one child. Some variation would also be expected to occur in classrooms related to the degree of bonding between teachers and students.

Because teachers did not carry out the conversations, there is a need to examine the feasibility of managing a systematic conversation program within the limitations of time and space of the preschool or kindergarten setting. A pullout model such as adopted by the current
intervention might also be feasible in which another adult in the school, such as a paraprofessional, a volunteer, or a speech and language professional met with children who were at risk of low achievement due to poor vocabulary. The effectiveness of training teachers to increase their open-ended questions and conversation skills while decreasing their directive language, related to behavior management and response control, has been shown to be possible in the classroom (Girolametto, Weitzman & Greenberg, 2003) as well as during book reading (Valdez-Menchaz and Whitehurst 1994; Whitehurst et al. 1988).

Conclusion

This study provides evidence supporting the inclusion of Talk Time as a valuable strategy available to teachers for enhancing the vocabulary of young children with poor vocabularies. Teacher-child conversation appears to differ in character from the kind of dialogue that typically accompanies interactive book reading. Studies set in child care centers comparing language produced by children elicited from a book reading context with that from interactions in which a teacher sits beside a student during a free play context have found that more complex speech and more overall talk is produced by children during free play. Unquestionably children’s literature provides a rich context to learn useful vocabulary central to many different domains. However, in the book reading context, teachers tend to ask more concrete and yes/no answered questions, or questions that are constrained by the topic of the narrative (O’Brien & Bi, 1995; Giralometto, Weitzman, van Lieshout & Duff, 2000). The kind of language practice arising from interactive storybook reading is valuable for later academic skills and is likely a key source of receptive vocabulary learning. Talk Time would be expected to enhance children’s ability to interactively participate in book talk as well as in other settings.
Prior to this study, it was well established that language input both from parents and teachers can have a significant effect on child output (Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002). From home-based and center-based studies, we know that there is a strong positive relationship between the total number of words, the use of rare words and the increased complexity of phrasing that adults use to and the resultant quality of theirs children’s language (Hart & Risley, 1995; Weizman & Snow, 2001; Huttenlocher et al, 2002; Reese & Fivush, 1993; Vasilyeva, Huttenlocher & Waterfall, 2006). This study is one of the first to try to discuss whether relatively small amounts of interventive language can affect vocabulary growth. Of significant interest is that children beginning the intervention with low vocabularies finished with a 7 point standard score increase and were well on their way toward achieving an average vocabulary. This suggests that even relatively small amounts of additional complex language input can have dramatic effects on vocabulary development.
CHAPTER IV
GENERAL DISCUSSION

This research has addressed two different areas of inquiry. Of primary interest was the evaluation of an intervention intended to increase the vocabulary knowledge of young children. In order to measure the effects of the intervention, however, a relatively new method of calculating vocabulary diversity, D, was used, in addition to a standardized test, the EVT, and a curriculum-based assessment of the type often used in school, definition assessment. Evidence was collected and used to conduct a validity study of D.

The Validity of D as a Measure of Vocabulary Knowledge

The results of the validity study suggest that D acts as a valid and reliable measure of lexical diversity. The five aspects of Messick’s (1989) framework of unitary construct validity were used to systematically evaluate how D functioned as a measure when used with young children. D scores represent a measure of lexical diversity calculated from speech or writing samples. The ratio of different words used is weighed against the total number of words and the curve estimation method accounts for sample size dependency that has plagued previous attempts of calculating the diversity of words used in everyday speech.

Analyses of internal consistency suggested that D had good reliability as had been demonstrated in previous research (Malvern et al. 2004; Owen & Leonard, 2002; Silverman & Ratner, 2002). This study was the first to estimate test-retest reliability and the results showed excellent score stability between the two test administrations. Sample size was not found to be an influence on D in this study; however methods of sample collection did reveal significantly
different D scores. Language sampled from personal narratives was more diverse than language sampled from a wordless picture narration. This kind of difference suggests that a standard protocol should be included as part of the reporting of the measure. A strength of D appears to be its generalizability across economic, gender and race/ethnicity groups. Results from both this study and a study comparing Head Start to non-Head Start children (Ruston, Schwanenflugel, & MacLean, 2006) show it to be relatively free from bias that has been associated with survey-based vocabulary tests (Restrepo et al., 2006; Washington & Craig, 1998).

The relationships between D and the standardized measure of the EVT and the definition test were moderate as expected. In other studies, D scores have not been significantly related to receptive vocabulary test scores (Richards & Malvern, 2004; Silverman & Ratner, 2002) which secures its placement more squarely in the expressive language domain. Evidence suggests that D measures an expressive vocabulary skill that is somewhat different from traditional vocabulary tests yet is within the domain of interest for researchers, teachers, and speech therapists.

Limitations encountered with the assessment of D were concentrated in the areas of data preparation and interpretation of the meaning of various levels of D. The process of gathering and analyzing a language sample for D may take an impractical amount of time unless additional information available from the transcripts is also of interest. To prepare for the vocd program, children’s speech must be recorded and transcribed. The time required per language sample from the beginning of recording to running the vocd program was a minimum of two hours, which is significantly more than the 30 minutes needed to administer and score the EVT. The second concern related to the language sample was one of reliability. Because there is not a standard language sampling protocol associated with D and because different prompts can yield different D scores, scores should not be compared between studies (unless the protocols were the
same). Using D scores to infer normal or abnormal development cannot be made at this time due to the lack of a normative sample and a standardized protocol for collecting D.

My findings indicate, however, that when the same protocol is used within a study, it was reasonable to interpret changes in D scores as an indicator of growth. In this study the same protocol was used for pre- and post-testing and D scores distinguished differential growth between the intervention and control groups. This difference was detected following a fairly short intervention of 10 weeks. This sensitivity to the effects of intervention is a promising feature for future research and therapeutic uses.

The results of the intervention study suggested that a relatively minimal amount of vocabulary enriched conversation between an adult and pairs of children could result in improved vocabulary levels for the children. Statistically significant differences in expressive vocabulary scores as measured by a standardized vocabulary test were found between the experimental and control groups of children.

The use of vocabulary- and sentence-structure enriched adult-child conversation as a school-based vocabulary enhancement strategy was informed by study of the strong relationships measured between families that regularly engage in more enriched talk and their children’s vocabulary. The idea that a teacher could promote the development of vocabulary of children who arrive at school with impoverished vocabularies through highly elaborative, vocabulary enriched, and topically sensitive talk is not foreign to preschool teacher training programs related to early literacy (NICHD, 2000; Schwanenflugel et al., 2005). However, no research had been conducted to date as to whether the amounts of such talk practical to carry out in preschool classrooms might actually effect positive change in vocabulary levels. For this intervention, pairs of children received 25 minutes of conversation twice a week for 10 weeks. This total
amount of linguistically complex talk might be carried out by preschool teachers over the course of the school year in a more distributed way or by volunteers or therapists much like it was carried out in the current study.

The need for vocabulary augmentation is heightened for children entering school with very low levels of vocabulary. A strong relationship exists between low vocabulary scores in kindergarten through second grade and low achievement scores in high school (Biemiller, 1999). The initial gap in vocabulary widens over time in school and for the children at the low end can lead to difficulties with comprehension and general reading skills later (Cunningham & Stanovich, 1998). Therefore children who pre-tested with low to very low vocabularies (in the bottom third of the normal distribution on the EVT) were of particular interest in this study. When the scores of this low vocabulary group were analyzed separately, the effects of the intervention were highly significant on both the EVT and D scores. The group of children beginning the intervention with average vocabularies did not benefit nearly to the same degree. Given these results, I believe that this type of intervention is best targeted to children with low vocabularies.

Evaluators of vocabulary programs have concluded that presentation of new vocabulary in rich and multiple contexts is most effective (Juel, Biancarosa, Coker, & Deffes, 2003; Stahl, 1986). Most vocabulary interventions are centered around words found in texts that will be read by the teacher or student and can be referenced in multiple times and ways (Beck, McKeown & Kucan; Schwanenflugel et al., 2005; Wasik & Bond, 2001). This study used very small group conversation as the instructional medium. Adults conducting the conversations were trained to allow for child initiation of topics and thus conversations held high interest levels for the children. Topics chosen by children, such as events that happened the night before, favorite
games or special interests (such as dinosaurs and babies) provided personal and meaningful contexts for the inclusion of new and rare vocabulary.

**Implications for Educational Practice**

The results of the present investigation indicated that a school-based practice of enriched conversation could be an effective strategy for increasing the vocabulary levels of children starting out with less developed vocabularies. Previous research conducted by Giralometto et al. (2003) demonstrated that teachers can be trained to engage in conversation that is likely to maximize the expressive language opportunities of children. Skills such as establishing joint attention, increasing wait time, asking open-ended questions and using rare words in context as well as defining them enhance the language learning opportunities of students.

The *Talk Time* program was based on *Building Bridges*, a component activity in the *PAVEd for Success* (Schwanenflugel et al., 2005) early literacy intervention. For *Building Bridges* teachers were asked to hold 5 minute conversations in small groups so that each child was included three times a week. However, *Building Bridges* was not faithfully sustained during the over the course of the intervention year by teachers. In interviews conducted as part of the intervention, teachers stated that although they believed small group conversation to be a worthwhile and important practice, scheduling was a deterrent to maintaining the practice. Teachers had difficulty both finding time and keeping track of these conversation sessions. *Talk Time* and *Building Bridges* share the same underlying structure and method but differ in their deployment. *Talk Time* targets the specific linguistic problems possessed by children with low vocabularies and general language skills. Scheduling regular and consistent conversations with these children can ensure that they receive adequate expressive language practice to improve their vocabulary. In addition, it is recommended that conversation time be protected from
interruption as much as possible, last for a minimum of 10 minutes per pair of children, and include linguistically complex talk. Teachers in PAVEd for Success have contributed useful ideas for conversation settings. These range from talking centers, snack times, lunch times, pre- and post-nap times, long walks during recess, to arrival and departure times.

Limitations

There are a number of limitations of the study that may not allow its findings to be generalized to typical preschool classroom settings. One limitation of the study is that the current study used a fairly high concentration of additional enriched conversation. This research represents the first attempt to quantify a minimum amount of enriched conversation necessary to effect a measurable increase on children’s vocabulary. Although the amount of conversation time engaged in during the intervention was modest (500 minutes), it was fairly concentrated (25 minutes twice a week over 10 weeks). Over a year, 500 minutes per targeted child may be fairly easily achieved. However, more research is needed to determine whether the concentration of dosage in this intervention was an influential aspect necessary to effect the positive change. Further, it is necessary to determine whether such results could be obtained as part of a general classroom strategy, rather than as a pull-out program such as the one presented here.

A second limitation of the study is the practicality within the classroom of carrying out the intervention using pairs of children as was done here. The model for this study was to emulate some of the characteristics of family environments most associated with high vocabularies. Having pairs of children in each group most closely resembles the level of shared input found in a family setting. More importantly, providing input in pairs addresses the concern that the children most in need of practice are often neglected in larger groups because of their general lack of conversational skills. However, meeting with only two children at a time may be
difficult for teachers to schedule. Due to the constraints of time that teachers face and the likelihood of high concentrations of children with poor vocabularies in some childcare settings, it would be worthwhile to study whether the same effects would hold in groups larger than pairs.

A third limitation is that conversations took place outside of the classroom with an experimenter. In this setting distracters such as other children were eliminated, interruptions were minimized and the focus was directed to the children’s speech. Whether or not teachers could maintain the necessary focus in the context of a busy classroom needs to be researched. That the experimenter was a stranger may have had a positive or negative impact. On one hand, conversations may have started slowly as experimenters established the trust necessary for children to take speaking ‘risks’. On the other hand, the novelty of the experimenter and the ‘special’ attention may have increased children’s willingness to engage in a challenging activity. An experimental study of this practice carried out by teachers would help answer these important questions.

The effects of the intervention were encouraging. Children in the low group started with a mean standard score of 88 on the EVT and finished with a mean score of 95. In terms of the normative age equivalencies reported in the EVT manual, on average this is an increase from 3 years and 7 months to 4 years and 2 months. Although the gap between the national mean and the children in the study was not closed completely, it came close. It is important to know if these gains could be further increased and sustained over time. Moreover, whether vocabulary gains obtained in this way would have long-term consequences for later reading comprehension would also need to be evaluated. A longitudinal study would be necessary to study these important questions.
REFERENCES


Goldfield, B. A., & Reznick, J. S. (1990). Early lexical acquisition: Rate, content, and the

vocabulary tests administered to preschool-age children. *Language, Speech, and Hearing
Services in Schools, 30*, 196-206.

Guttfreund, M., Harrison, M., & Wells, G. (1989). *Bristol language development scales*


who have limited vocabularies: The benefits of regular reading and dialogic reading. *Early
Childhood Research Quarterly, 15*(1), 75-90.

Brookes.

American children*. Baltimore; Paul Brookes.

Hess, C. W., & And Others. (1986). Sample size and type-token ratios for oral language of
preschool children. *Journal of Speech and Hearing Research, 29*(1), 129-134.


APPENDIX A

LANGUAGE SAMPLE PROTOCOL

Conversational context (5 min) (Evans & Craig, 1991)
Interview with the child. Respond to child with rewording of child’s comments or “that’s interesting, tell me more about that.” Try to avoid leading question. Allow the child to take the lead.

- Tell me about the sorts of things you do in the classroom? What do you like to play with?
- What do you like to do when you’re not in school?
- Do you have any brothers or sisters?

Personal narrative (goal is to elicit at least 3 personal narratives) (Peterson & McCabe, 1983)
- “I also brought some photos to show you” (Talk about the photos as outlined below. If the child responds “no”, go to the next photo. If the child says “yes”, ask him/her “Can you tell me about it?”)

Prompts
- Oh look, this girl fell off the bars and hurt her knee. She had to go to the emergency room and they put a cast on. Have you ever broken anything? Did you ever hurt yourself on the playground?
- Can you see the bee on the flower? I got stung by a bee once. Did a bee ever sting you?
- This little girl had to go to the doctor, because she had a bad cough. Have you ever been to the doctor?
- The dentist visited my daughter’s (sister’s) school last year. All the children had to go for a check-up. Have you ever had a toothache? Have you ever been to the dentist?
- Look, this is Santa. He visited my daughters (sister’s) school last year. Have you ever seen Santa anywhere?
- These children went on a school-trip. They all went on a bus to the Zoo. Have you even been on a school trip?
- Oh look who’s this? (Ronald McDonald). I went to a birthday party at McDonald’s last year. Have you ever been to McDonalds?

Story narration:
Books: Good Dog Carl and Carl Goes Shopping
- I would like to hear you tell a story using the pictures in this book, but the story can be from your own imagination. Would you like to do that? There is no right or wrong story, you can make up any story to go along with the pictures.
Prompts:
- Repeat what the child said. Can you tell me more? What’s happening? What else do you see?
- Remember, try not to introduce any new vocabulary.


APPENDIX B

RECRUITMENT LETTER

Dear ‘Center’ families,

The attached permission form is for a research study “Can conversation in preschool classrooms improve vocabulary levels?” The researchers are interested in oral language development of preschool children and how different strategies might be supportive and in different ways of assessing language development.

In this study, children will be given several vocabulary assessments that ask him or her to name objects in pictures. In another session, they will participate in creating a story from a wordless picture book and answering some questions designed to produce vocabulary. This session will be tape recorded. Over a period of twelve weeks, some children will participate in conversations based on a variety of subjects designed to help develop vocabulary and conversation skills. These conversations will take place just outside the classroom, or in a quiet area in the classroom, with pairs of children for 25 minutes, 2 times a week. Conversations will be conducted by trained U.G.A. students. After that, all children will be given vocabulary assessments that ask him or her to name objects in pictures, narrate a wordless picture book and answer a set of questions.

We would also like to collect information regarding your occupation and education to make sure that these ways of collecting oral language do not present an erroneous picture of any group of children’s language development. We have included these questions on the attached consent form.

All the information will be held confidential unless otherwise required by law. All children’s identifies will be referred to by codes (not names) and all data will be kept in a secured location in Aderhold Hall at U.G.A. Audio tapes will be erased by December 31, 2008. Attached to this letter is a more detailed permission form. Please feel free to give the researchers a telephone call if you have any question about this research.

Thank you,

Hilary Ruston, MA
Telephone (706) 614-8660
Email: hilaryr@uga.edu

Paula Schwanenflugel, PhD
Professor of Educational Psychology
Email: pschwan@uga.edu

325R Aderhold Hall
University of Georgia
Athens, GA 30602
APPENDIX C

PARENT CONSENT FORM FOR CHILD

Parent Permission Form for Child’s Participation in Research

I, __________________, agree to allow my child: __________________ who was born on ______________ (Month/Date/Year), to take part in a research study titled “Can conversation improve vocabulary?”

This research is being conducted by Hilary Ruston, MA (706-614-8660) and Dr. Paula Schwanenflugel (706-542-4273), both of the Department of Educational Psychology, University of Georgia. I do not have to allow my child to take part in this study. My child can stop taking part or choose not to take part at any time without giving a reason and without penalty. I can ask to have information related to my child returned to me, removed from the research records, or destroyed.

The researchers are interested in developing our knowledge of both how classroom activities may help children develop their oral language skills and child assessments of vocabulary. The researchers are interested in whether regular conversation in a preschool setting can have immediate effects on vocabulary and grammar and how typical vocabulary tests may not capture child language growth as well as sampling children’s speech does. In this study the researchers will compare children’s speech to performance on standardized tests of vocabulary. The researchers are not interested in how my child does on this test in particular, but in how children my child’s age in general perform. My child will have an opportunity to practice expressive vocabulary skills with an individual adult. My child is expected to experience no harm or discomfort from his/her involvement in this research. If I do not want my child to take part then s/he will remain in the classroom.

This study will include fours sessions of assessment, each 15 minutes. In two, my child will be provided a wordless picture book and tape recorded while telling the story in his/her own words, and respond to a set of scripted questions. and have his/her speech tape recorded. In the other two sessions my child will look at a series of pictures and be asked for a label to go with the picture (Expressive Vocabulary Test.) My child will also be asked to provide definitions for words that are likely to be at the edge of his or her vocabulary.

In addition, for ten weeks my child may participate in two conversations per week, each 25 minutes, with a trained U.G.A. student and another child from the classroom. These conversations will be prompted from a set of questions designed to help build language skills. Three of the conversations will be recorded.
In each session my child will be accompanied by a trained researcher to a quiet table outside the classroom in the hall. The process will be explained by the researcher who will not proceed without the assent of the child. My child will be praised for his/her performance.

The only people who will see the results are the researchers. The researchers are not interested in my child’s performance specifically, but in how vocabulary develops and how different measures capture expressive vocabulary. Any information collected about my child will be held confidential unless otherwise required by law. My child’s identity will be coded, and all data will be kept in a secured location. Audio tapes will be erased by December 31, 2008.

If I have any questions, I can call the researchers at the phone number above.

I understand the procedures described above. My questions have been answered to my satisfaction and I agree to allow my child to participate. I have received a copy of this form.

My occupation: __________________________

Signature of parent or guardian __________________________ Date ______ Last year of school completed: ______

Signature of Researchers __________________________ Date ______
Telephone contact: (706) 542-4273 Email: hilaryr@uga.edu

Please sign both copies, keep one and return one to the researcher.

Additional questions or problems regarding your child’s rights as a research participant should be addressed to The Chairperson, Institutional Review Board, University of Georgia, 612 Boyd Graduate Studies Research Center, Athens, Georgia 30602-7411; Telephone (706) 542-3199; E-Mail Address IRB@uga.edu
DEFINITION TASK

I’m going to tell you some words and I want you to tell me what they mean and describe them for me. For instance if I say pencil, you might say “You write with it, it’s long and thin and has lead at one end and an eraser at the other.” If you don’t know some of the words, that’s okay, just tell me you don’t know that one. Tell me what you know about each word. Are you ready? Do you know what we’re going to do? (Listen for the child to describe one thing about the task).

Let’s practice one: tree

If still not understanding try this: “Let’s imagine that someone from space, from a completely different planet, has come to earth and it’s your job to describe some things to him. He doesn’t know anything about these things so you need to carefully describe them.” Practice: tree

Can you tell me about_________________?

1. house
2. pretzel
3. cave
4. cauliflower
5. aquarium
6. igloo
7. castle
8. hive
9. lobster
10. cocoon
11. donkey
12. umbrella
13. clown
14. thief
15. handkerchief

Write out what the child says next to the word.
APPENDIX E

SAMPLE FREQUENCY AND VOCD CHART

> check +g3 sample.cha
depfile.cut being used
from:
"C:\childes\Clan\lib\depfile.cut"
check +g3 sample.cha
check (24-Apr-2006) is conducting analyses on:
ALL speaker tiers
   and those speakers' ALL dependent tiers
   and ALL header tiers
*********************************
***************
From file <sample.cha>
First pass DONE.
Second pass DONE.
Success! No errors found.

> freq +t"*CHI" +s"*-
  %\%" sample.cha
freq +t*CHI +s*-\%
sample.cha
Wed Jul 12 18:56:20 2006
freq (24-Apr-2006) is conducting analyses on:
   ONLY speaker main tiers matching: *CHI;
*********************************
***************
From file <zee2.cha>
4 a
1 about
16 and
1 another
1 are
1 at
7 baby
1 bad
6 be
check +g3 sample.cha
4 bed
1 bite
1 bread
1 break
1 can
2 can't
1 cd
1 climb
1 canning
1 crying
1 crying
3 did
4 do
8 dog
6 doing
2 don't
1 door
2 down
1 fall
1 fishy
1 flour
1 food
1 foot
1 game
8 get
4 go
9 going
2 got
2 grapes
2 have
23 he
1 head
4 hear
1 her
2 here
15 him
2 his
1 home
9 i
9 in
1 into
2 is
9 it
2 jump
2 know
1 let
1 like
3 look
2 make
2 me
1 might
1 mom
3 momma
2 monster
1 music
2 my
1 Nemo
2 no
1 nose
4 oh
9 on
1 out
1 over
1 player
5 put
6 quiet
2 right
1 sack
4 say
2 see
1 shoes 5 this 9 what
1 sleeves 16 to 1 what's
1 stank 1 toes 1 with
2 sticky 1 tore 2 yeah
2 stop 1 trying 5 you
16 that 3 trouble ------------------------
17 the 6 trying 107 Total number of
data
1 them 2 turn different word types used
7 there 1 up 367 Total number of
3 they 1 want words (tokens)
2 thing 2 water 0.292 Type/Token ratio
2 think 1 wearing

> VOCD +t"*CHI" +r6 -s@exclude.cut
sample.cha

UTTERANCES: (vocd<sample.cha>)
like my music
here dog and a baby and a and a bed baby in
the bed and the dog and the momma dog and
momma in the bed
what's this
what this what you think about me wearing
this
let me see
the baby and the dog
they him put foot on the dog and jump on
the bed he go he go get in trouble him him
go get in trouble dog and him get sleeves on
him and him toes be stank what they doing
what that baby doing on his head
yeah
they are bad
he put it on his nose
it going to break he going to be crying
him put that thing right there and he going to
get in trouble he going to cry oh i know him
doing going in the baby he going to fall the
monster down there
the monster down there
look be quiet he going
quiet
quiet be quiet
he trying to do that he trying to do that get in
that water he trying to get in that water he
think it bite

the fishy oh i see Nemo right there
i did it
he trying to climb up there and turn the cd
player on him trying to go out the door and
he don't got no shoes on what him mom
going to do with him
oh he tore the bread sack what him doing
that dog food
it sticky it sticky
he them grapes my momma got grapes at
home what him doing
he have he might have be quiet be quiet
there it is
put him into
i did that
he trying to put that baby in that i don't
know
i did that
he trying to do that and make and jump on
the thing and make that flour he her say stop
another dog going to get over there he get
the baby what him doing
he say stop
look
oh what is that
yeah
look here this a game
i want to hear this
you have to turn it on
i can't hear it
you say you can't hear that
no
you say i can hear that
### VOCD RESULTS SUMMARY

```
Command line: vocd +t*CHI +r6 -s@exclude.cut sample.cha
File name: sample.cha
Types, Tokens, TTR: <107,367,0.291553>
D_optimum values: <38.71, 39.07, 39.02>
D_optimum average: 38.94
```

<table>
<thead>
<tr>
<th>tokens</th>
<th>samples</th>
<th>ttr</th>
<th>st.dev</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>100</td>
<td>0.7540</td>
<td>0.060</td>
<td>40.443</td>
</tr>
<tr>
<td>36</td>
<td>100</td>
<td>0.7569</td>
<td>0.065</td>
<td>42.432</td>
</tr>
<tr>
<td>37</td>
<td>100</td>
<td>0.7468</td>
<td>0.065</td>
<td>40.737</td>
</tr>
<tr>
<td>38</td>
<td>100</td>
<td>0.7379</td>
<td>0.061</td>
<td>39.470</td>
</tr>
<tr>
<td>39</td>
<td>100</td>
<td>0.7328</td>
<td>0.065</td>
<td>39.195</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>0.7315</td>
<td>0.067</td>
<td>39.858</td>
</tr>
<tr>
<td>41</td>
<td>100</td>
<td>0.7222</td>
<td>0.059</td>
<td>38.488</td>
</tr>
<tr>
<td>42</td>
<td>100</td>
<td>0.7255</td>
<td>0.056</td>
<td>40.261</td>
</tr>
<tr>
<td>43</td>
<td>100</td>
<td>0.7172</td>
<td>0.058</td>
<td>39.108</td>
</tr>
<tr>
<td>44</td>
<td>100</td>
<td>0.7215</td>
<td>0.057</td>
<td>38.847</td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>0.6933</td>
<td>0.055</td>
<td>35.270</td>
</tr>
<tr>
<td>46</td>
<td>100</td>
<td>0.7074</td>
<td>0.055</td>
<td>39.333</td>
</tr>
<tr>
<td>47</td>
<td>100</td>
<td>0.6900</td>
<td>0.058</td>
<td>36.091</td>
</tr>
<tr>
<td>48</td>
<td>100</td>
<td>0.6898</td>
<td>0.056</td>
<td>36.812</td>
</tr>
<tr>
<td>49</td>
<td>100</td>
<td>0.6912</td>
<td>0.053</td>
<td>37.911</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.6840</td>
<td>0.052</td>
<td>37.014</td>
</tr>
</tbody>
</table>

D: average = 38.829; std dev. = 1.798
D_optimum <38.71; min least sq val = 0.001>

<table>
<thead>
<tr>
<th>tokens</th>
<th>samples</th>
<th>ttr</th>
<th>st.dev</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>100</td>
<td>0.7429</td>
<td>0.071</td>
<td>37.556</td>
</tr>
<tr>
<td>36</td>
<td>100</td>
<td>0.7453</td>
<td>0.058</td>
<td>39.250</td>
</tr>
<tr>
<td>37</td>
<td>100</td>
<td>0.7435</td>
<td>0.055</td>
<td>39.874</td>
</tr>
<tr>
<td>38</td>
<td>100</td>
<td>0.7489</td>
<td>0.060</td>
<td>42.451</td>
</tr>
<tr>
<td>39</td>
<td>100</td>
<td>0.7346</td>
<td>0.062</td>
<td>39.653</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>0.7378</td>
<td>0.062</td>
<td>41.508</td>
</tr>
<tr>
<td>41</td>
<td>100</td>
<td>0.7251</td>
<td>0.063</td>
<td>39.214</td>
</tr>
<tr>
<td>42</td>
<td>100</td>
<td>0.7164</td>
<td>0.056</td>
<td>38.010</td>
</tr>
<tr>
<td>43</td>
<td>100</td>
<td>0.7219</td>
<td>0.054</td>
<td>40.279</td>
</tr>
<tr>
<td>44</td>
<td>100</td>
<td>0.7232</td>
<td>0.053</td>
<td>41.565</td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>0.7111</td>
<td>0.057</td>
<td>39.385</td>
</tr>
<tr>
<td>46</td>
<td>100</td>
<td>0.7033</td>
<td>0.063</td>
<td>38.334</td>
</tr>
<tr>
<td>47</td>
<td>100</td>
<td>0.6906</td>
<td>0.059</td>
<td>36.233</td>
</tr>
<tr>
<td>48</td>
<td>100</td>
<td>0.6838</td>
<td>0.051</td>
<td>35.479</td>
</tr>
<tr>
<td>49</td>
<td>100</td>
<td>0.6980</td>
<td>0.051</td>
<td>39.515</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.6860</td>
<td>0.050</td>
<td>37.468</td>
</tr>
</tbody>
</table>

D: average = 39.111; std dev. = 1.834
D_optimum <39.02; min least sq val = 0.001>

---

<table>
<thead>
<tr>
<th>tokens</th>
<th>samples</th>
<th>ttr</th>
<th>st.dev</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>35</td>
<td>100</td>
<td>0.7540</td>
<td>0.060</td>
<td>40.443</td>
</tr>
<tr>
<td>36</td>
<td>100</td>
<td>0.7569</td>
<td>0.065</td>
<td>42.432</td>
</tr>
<tr>
<td>37</td>
<td>100</td>
<td>0.7468</td>
<td>0.065</td>
<td>40.737</td>
</tr>
<tr>
<td>38</td>
<td>100</td>
<td>0.7379</td>
<td>0.061</td>
<td>39.470</td>
</tr>
<tr>
<td>39</td>
<td>100</td>
<td>0.7328</td>
<td>0.065</td>
<td>39.195</td>
</tr>
<tr>
<td>40</td>
<td>100</td>
<td>0.7315</td>
<td>0.067</td>
<td>39.858</td>
</tr>
<tr>
<td>41</td>
<td>100</td>
<td>0.7222</td>
<td>0.059</td>
<td>38.488</td>
</tr>
<tr>
<td>42</td>
<td>100</td>
<td>0.7255</td>
<td>0.056</td>
<td>40.261</td>
</tr>
<tr>
<td>43</td>
<td>100</td>
<td>0.7172</td>
<td>0.058</td>
<td>39.108</td>
</tr>
<tr>
<td>44</td>
<td>100</td>
<td>0.7215</td>
<td>0.057</td>
<td>38.847</td>
</tr>
<tr>
<td>45</td>
<td>100</td>
<td>0.6933</td>
<td>0.055</td>
<td>35.270</td>
</tr>
<tr>
<td>46</td>
<td>100</td>
<td>0.7074</td>
<td>0.055</td>
<td>39.333</td>
</tr>
<tr>
<td>47</td>
<td>100</td>
<td>0.6900</td>
<td>0.058</td>
<td>36.091</td>
</tr>
<tr>
<td>48</td>
<td>100</td>
<td>0.6898</td>
<td>0.056</td>
<td>36.812</td>
</tr>
<tr>
<td>49</td>
<td>100</td>
<td>0.6912</td>
<td>0.053</td>
<td>37.911</td>
</tr>
<tr>
<td>50</td>
<td>100</td>
<td>0.6840</td>
<td>0.052</td>
<td>37.014</td>
</tr>
</tbody>
</table>

D: average = 38.193; std dev. = 1.926
D_optimum <39.07; min least sq val = 0.001>
Guidelines for conversations

Let the child lead:

- Provide pause time to allow the child to initiate talk. Typically adults give children one second to respond before jumping in. Many children need much longer to process questions or information. Increase your wait time (count to ten slowly). Wait for the child to initiate. The initiation may not always be language; it might be pointing at a toy or picture.
- Follow the child’s lead in conversation and respond to their focus of interest.
- Try not to change the subject.
- Listen when the child is talking; try not to interrupt their flow. Show interest through your facial expression or encouraging sounds.
- Try to be sitting face-to-face, or side-by-side showing joint attention.

Responsiveness:

- **Imitate** – Repeat what the child says using similar facial expressions and gestures. This lets the child know that you are interested, engaged and listening. Nod to indicate that you are understanding.
- Model language by using extensions, expansions and recasting.
  - Child: ‘My sister Jennie there’ (pointing to classroom)
  - **Extension**: Your sister Jennie is there in the other classroom, what is she doing?
  - **Expansion**: Your sister Jennie is here in the same school, what can you tell me about Jennie?
  - **Recast**: Your sister Jennie is in the three year old classroom?
- Introducing vocabulary naturalistically into conversation through recasting and expansions of children’s utterances.
  - Child: There aren’t very many trikes.
- **Vocabulary Recast:** So *tricycles* are *rare* at your school? Are any other toys rare?

- Child: She ain’t got a book.
- **Repetition** with question: I wonder why she doesn’t have a *storybook*.
- Ask **open ended** questions: ‘Tell me about this’; or ‘How do you do that?’ or, ‘Why do you think that happened?’
- Ask **clarifying** questions,
  - Instead of ‘That’s such a great picture’ ask ‘Tell me about that picture you drew.’
- You don’t need to correct language or grammar directly. This can be done implicitly by repeating what was said with more context or in a standard form.
- Introduce topics in a vocabulary rich way: ‘Look out the window at the hawk on the *oak* tree. I wonder what it is searching for, what kind of prey, maybe a mouse. What do you think it is doing?’

**Turn-taking in conversation:**

- Children may a set of conversation rules at home that is different than those commonly used at school.
- Encourage children to balance turn taking and build upon ongoing subjects in their conversation pairs.
- Encourage more than one turn on each subject.
- Be clear that you are encouraging the other child to take a turn when one is dominating.
- Urge them to clarify misunderstandings about the subject.

**With children who show some reluctance:**

- Try to establish joint attention with a toy, a picture or an activity.
- Make sure that you don’t communicate pressure to talk.
- Slow the pace.
- Adapt your language to be simpler at first.
Ask simple questions to keep the child part of the conversation. Even a question that will elicit a one word answer might be okay to keep things moving.

Ask questions that demonstrate that you are listening and checking in that you are understanding the story.

Ask questions with a forced choice.
  - On the playground do you like to play on the swings or the slide?

Avoid questions that sound like a test with a ‘right or wrong’ answer.

Avoid questions that are very concrete.
  - Ask ‘What was snack like?’ instead of ‘What did you eat for snack?’

A welcoming environment:

Introduce topics about real-life situations.

Talk about the environment, ‘It’s freezing in the hall today. What temperature do you like?’

‘Look at that drawing of a dog on the wall. My dog escaped into the yard today and I had trouble getting him in. Have …’

Becoming a better story teller:

By your active listening and responsiveness you can help children become more skilled at telling an engaging story. Prompt for missing information, including background details, specific vocabulary and relevant information

Model story telling.

Encourage a story, ‘I bet you did something fun over the weekend.’

Encourage imitations, ‘What does your father sound like when he’s trying to … (whatever the story is about i.e., ‘get everyone in the car.’)

In the case of a sensitive topic:

As we are not counselors; try not to bring up a sensitive topic yourself. If a child initiates one (i.e. about divorce), listen responsively and be supportive. If a child discusses abuse, listen, take notes and contact me immediately, (cell) or Dr. Schwanenflugel (cell). If you are unable to reach either of us, talk to the Center Director, provide her with the details
Don’t leave the Center without sharing the information with us or the person at the highest level.

**Props (discussion starters):** Play telephone, photos from magazines and catalogues, pictures of your family or your animals, toy figures, phone, books, play dough, or something of particular interest to the child that might be found in a child care center.

**Initial Conversation Stems:**

1. Family
2. Extended family/pets
3. Injuries, going to the doctor, visiting the hospital
4. Friends, play, sports
5. Birthdays, holidays, celebrations
6. Photo prompts of places in Athens/nature
7. Home/ neighborhood
8. School, activities, teachers
9. Art, Music, Building
APPENDIX G

DIRECTIONS FOR OBTAINING A D SCORE

PREPARATION:

Download two programs: Transcriber and CLAN. Both programs are free to use.

**Transcription Program:** Transcriber\(^1\) can be found at: http://trans.sourceforge.net/en/presentation.php. An audio file (wav or mp3) can be directly transcribed and linked into the Transcriber program. The User Manual is available for downloading from the presentation page. Transcriber can be downloaded from this page and is available for Windows, MacOS X and Linux.

**Vocabulary Diversity Analysis Program (VocD):** VocD is a language analysis tool included in the CLAN (Computerized Language Analysis) program. This tool, a component of the CHILDES (Child Language Data Exchange System)\(^2\), is available at: http://childes.psy.cmu.edu. Links at this website enable you to download the CLAN program and two manuals available in PDF format; the CHAT transcription system manual, and the CLAN programs manual.

CLAN can be found at http://childes.psy.cmu.edu/clan. It is supported by Windows XP, Mac OS X 10.4 and up and Unix.

**Audio files:** Audio files should be downloaded as wav files ideally or as mp3 recordings. If you are restricted to another format, check that the Transcriber program will accept the recording.

TRANSCRIPTION TASKS:

Open transcriber to a new file and pull up the audio file.

For the purpose of analyzing for vocabulary diversity use, transcribe each speaker as one section. Continue alternating between speakers. Once the passage is complete add the speakers (we use Interviewer and the child’s name, e.g. Sample).

Both interviewer’s questions and child’s responses should be transcribed for the purpose of evaluating interviewer reliability.

Sample:

\(^1\) Transcriber is hosted by SourceForge.net a large Open Source software development web site.
\(^2\) The CHILDES system was started up through the John D. and Catherine T. MacArthur foundation and has been funded by the National Science Foundation since 1999. CHILDES and the TalkBank project are maintained by Brain MacWhinney, Professor of Psychology at Carnegie Mellon University.
Interviewer: what do you like to play at school.
Sample: dramatic play when mister Jack brings his ties and hats well I used to play with Giovanni but now he has gone to another school.
Interviewer: okay what do you like to play at home.

Transcription conventions for VocD: Do not use any capitals except for proper names. Do not use abbreviations such as Mr. or St. Write out the word (mister or street). The only punctuation should be at the end of the speaker’s turn, even if there are several sentences together. Transcribe words according to Merriam Webster Dictionary conventions.

Partial words were not transcribed, only completed words, eg. he jumped on the tram trampoline (tram is not transcribed.)

Transcriptions followed dictionary spelling versus phonetic conventions, eg, ‘the pawtwidge in da peaw twee’ was transcribed as ‘the partridge in the pear tree’.)

In cases of dialect use such as ‘finna’, we would type ‘fixing to’, so as to avoid variations of spelling such as fina, fiinna etc…

Transcripts were double checked for consistency, for instance tv was used consistently instead of teevee.

Retraced words or phrases should be transcribed with the following notation so they will not add to the token calculations: and a <and a> [,], or he < he he he he he he he > [,]

We added speaker names and punctuation after completing a first draft of the transcription. First create the two speakers and then starting at the top of the file use Control T, Enter, down arrow, Control T, Enter, down arrow. This results in alternately assigning the speakers to turns.

Each subject’s file was saved under his or her name, eg. Sample.trs. and postSample.trs.

TRANSFERRING TRANSCRIBER FILES TO .CHA FILES FOR CLAN

Open up transcriber and pull up the file

Export to text
Enter

This will save your transcriber file (Sample.trs) as a text file (Sample.txt).

Open the text file: Sample.txt and
Go to File
Save as
In the file name box change Sample.txt to Sample.cha
CONDUCT CLAN ANALYSIS TO OBTAIN D SCORE:

Open Clan
Click on file and find sample.cha
Save sample.cha to CLAN lib

Now prepare your file for VocD

Edit
Replace: Interviewer: space space
With: *INT: insert tab (box)
Replace all

Edit
Replace: Sample: space space
With: *CHI: insert tab (box)

The CLAN program requires certain information included at the beginning of each file. Type the following code. In this example I am using the child’s name Sample and the Interviewer’s name Hilary.

@Begin
@Languages: en
@Participants: CHI Sample Child, INT Hilary Examiner
@ID: en.sample1hilary.4.11.bb=CHI (This stands for en(English).sample1(timepoint one)hilary .4.11 (age four years eleven months.). bb (school code)=CHI

At the end of each .cha file type
@End

Now your file should be ready for analysis.
First run check on the file until it is error free. If there is an error, this function will identify the error and the line # in which it occurs.

Mode
Check opened file from the Mode menu or (Esc-l)
This will identify the errors until you get the message Success no errors found.

Next run the freq program to conduct a frequency count for the child. This will produce a list of all words used, frequency counts, and a type/token ratio

freq +t*CHI sample1.cha

freq +t*CHI sample1.cha
Mon Feb 26 20:47:54 2007
freq (24-Jan-2006) is conducting analyses on:
The results of the FREQ command provide a helpful list that provides an opportunity to scan the words used from the transcript from typographical or other errors.

Now run the vocd program.

vocd +t*CHI +r6 +f -s@include.cut sample1.cha

In the command is +t*CHI which constrains the analysis to the child’s utterances, +r6 which eliminates retracing of words or phrases from the analysis. A cut file was created for expressions that were to be excluded from the analysis. The cut file list consisted of: xxx, mmm, mmhmm, uhhuh, shhh and individual letters of the alphabet with the exception of a and i. xxx was typed for words we could not understand (sometimes the difficulty was due to pronunciation and other times due to recorder or microphone problems) we typed one xxx even if it was clear that a string of words was lost or unintelligible.

Sample Portion of Output:

UTTERANCES: (vocd<sample1.cha>)

i don't know
yeah
no
my daddy's at the bone doctor he works at the bone doctor
he if people bones are broken he puts a cast on it
and one time someone when someone was trying to get something out of the vacuum cleaner
when it was on they stuck their hand in the vacuum cleaner
no it broke his bone
i know of me and it didn't it stinged a little bit
they have to check his heart
i don't know

VOCD RESULTS SUMMARY
====================
Command line: vocd +t*CHI +r6 +f -s@exclude.cut sample1.cha
File name: sample1.cha
Types,Tokens,TTR: <146,356,0.410112>
D_optimum values: <46.42, 46.84, 47.38>
D_optimum average: 46.88