STABILITY OF CONVERSATIONAL LANGUAGE SAMPLES FROM CHILDREN IN KINDERGARTEN: THE EFFECTS OF TIME, SAMPLE LENGTH, AND TOPIC

By

Lindsay DeBrock

March, 2010

Chair: John Heilmann, Ph.D

Major Department: Communication Sciences and Disorders

Abstract: The purpose of this study was to examine the reliability of measures from conversational language samples in the school-aged population. Two 10-minute conversational language samples were collected from 20 at-risk kindergarten children to determine their reliability and feasibility as an assessment and progress monitoring tool. All samples were collected and transcribed by one examiner using an outlined elicitation protocol in which the children were asked open ended questions about school and home activities in various time segments. Test-retest reliability was determined across eleven language measures for each child. Significant reliability correlation coefficients were observed for each language sample measure. The amount of variance due to sample length and topic were analyzed utilizing generalizability theory, which observes various facets accounting for measurement error. Variance was calculated for each language measure. Conversational language samples were found to have strong test-retest reliability across all language measures. The greatest amount of variance in measures was attributed to the child in most language measures with the interaction between the child, session, and length/topic accounting for the second largest amount of variance in most

measures. Length and topic accounted for negligible amounts of variance in language measures. These findings demonstrated that shorter conversational language samples are reliable across time and that the length and topic of the sample do not have a substantial impact on the reliability of language measures.

Stability of Conversational Language Samples from Children in Kindergarten: The Effects of Time, Sample Length, and Topic

A Thesis Presented to the Faculty of the Department of Communication Sciences and Disorders

East Carolina University

In Partial Fulfillment of the Requirements for the Degree Masters of Sciences in Communication Sciences and Disorders

by

Lindsay DeBrock

March, 2010

Stability of Conversational Language Samples from Children in Kindergarten: The Effects of Time, Sample Length, and Topic

by

Lindsay DeBrock

Approved By:

Director of Thesis:

_____

John Heilmann, Ph.D

Committee Member:

_____

Chris Riley-Tillman, Ph.D

Committee Member:

_____

Marianna Walker, Ph.D

Committee Member:

_____

Laura Ball, Ph.D

Committee Member:

_____

Kevin O'Brien, Ph.D

Chair of the Department of Communication Sciences and Disorders:

_____

Gregg D. Givens, Ph.D

Dean of the Graduate School:

_____

Paul J. Gemperline, Ph.D

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: INTRODUCTION

Language Assessment

Accurate assessment is essential for case management of children with language impairment. Language assessments serve many purposes including diagnostics, treatment planning, monitoring progress, and documenting outcomes. While language assessments are most often used to identify whether a child has an impairment, they should also be an integral component during intervention. In order to identify effective therapy techniques, clinicians must be able to document that a change in language abilities is in fact occurring. For this purpose, reliable assessments that can monitor progress and document outcomes are necessary.

Standardized tests and criterion referenced assessments are two methods used to measure language skills in children. Standardized tests are norm-referenced and assess language abilities in a decontextualized context (Paul, 2007). Decontextualized contexts are situations that do not naturally occur in a child's day-to-day life, so standardized tests assess language skills in a way that a child would not normally use language. These tests use contrived measures in order to determine if a child has a language deficit by comparing performance between children with typically developing language abilities and children suspected of having a language deficit. The tests are administered and scored in the same manner each time they are administered. They must be both reliable and valid for the criteria they intend to measure. When constructed properly a standardized test permits clinicians to evaluate the child's language abilities using standard scores, percentile ranks, or age-equivalent scores. Standardized assessments are often used because they are thought to be less time consuming than other methods and offer an objective evaluation of the child's language skills.

While standardized tests are a means of measuring language skills against those who have typically developing language, there are some limitations. First, this type of assessment views language abilities at only one point in time and in a decontextualized format. Therefore it may not be a sufficient representation of the child's functional communication and language skills. Standardized tests do not provide in-depth analyses of deficient language skills; they simply determine if the child's performance is significantly different from other children. They are less effective for establishing baseline function, identifying goals for intervention, or monitoring progress (McCauley, 2000).

Criterion referenced assessments provide alternative methods of assessing children's language skills. Criterion referenced assessments do not compare a child's language abilities to a normative sample. Rather, this type of assessment examines a child's ability to reach a certain level of performance on a specified task. Through observation or by eliciting responses from the child, the examiner documents whether the skill was present. The examiner usually creates these assessments and can tailor them to the language areas of interest. Specific language or communicative behaviors can be viewed more thoroughly when using criterion referenced assessments. The examiner uses established criteria and developmental norms to create criterion referenced assessments (Paul, 2007). Using appropriate developmental levels allows the examiner to observe areas in which the child exhibits language deficits. Given that criterion referenced assessments provide more detailed descriptions of specific skill sets from naturalistic contexts, they can be effective in monitoring children's progress and documenting treatment outcomes.

Language Sample Analysis

Language sample analysis is a type of criterion referenced assessment that is effective at monitoring progress and documenting outcomes among children with communication disorders (Tilstra & McMaster, 2007). It involves collecting a naturalistic language sample from children through conversation or a narrative. The sample is recorded and transcribed. Language samples can measure various domains of language, including form, content, and use. Additionally language sample analysis provides robust measures that are sensitive to developmental change (Leadholm & Miller, 1992). However, research is needed to determine the psychometrics of language sample analysis and its clinical feasibility.

Test Reliability

Regardless of the chosen assessment method, a clinician must evaluate all assessment measures for reliability. Measurement error is inevitable even on the same task due to the variability in human performance. Error may occur because of several different factors including the examinee's behaviors, the actions of the examiner, the content of the tests, time aspects, and situation factors (Shiavetti & Metz, 1997).

Reliability refers to the stability and consistency of a measurement. Language assessment measures should be stable and result in little change in outcome across multiple administrations. Three categories are used to estimate the reliability of a test: stability, equivalence, and internal consistency. Test-retest analyses are used to determine the coefficient of stability. In this case, the measurement under consideration is administered repeatedly and results are compared. In equivalence, an alternative or parallel form of the measurement is administered. A correlation coefficient is then calculated, which helps estimate the reliability of the measure. The last

category of reliability is internal consistency, which compares two halves of the assessment with a correlation coefficient (Shiavetti & Metz, 1997).

Reliability of Language Sample Measures

The stability, or test-retest reliability, of language sample measures was examined in a study completed by Gavin and Giles (1996). Conversational samples in the context of freeplay were collected from preschool children and four quantitative measurements were obtained across different sized language samples. Four language measures assessed included total number of words (TNW), number of different words (NDW), mean length of utterance in morphemes (MLU-m), and mean syntactic length (MSL) which measure semantic and syntactic abilities. The four measures were calculated for two different time based lengths and all measures except TNW were calculated for multiple utterance lengths. Results from this study indicated greater stability as the language sample increased in length (based on time or number of utterances). NDW, MLU-m, and MSL showed significant correlation for the larger sample sizes indicating acceptable reliability of these measurements in language samples. TNW did not reach a significant correlation between sessions in both time-based sample sizes, signifying low reliability for this measure. Stability was found to be adequate when the sample size was 20 minutes in length or at least 100 complete and intelligible utterances based on a minimum reliability coefficient corresponding to the level in which less than 50% of the variability is due to measurement error (Gavin & Giles, 1996).

Another study focusing on the reliability of play-based language samples from young children found no significant difference in the information obtained within language samples and across language samples (Cole, Mills, & Dale, 1989). Language sample measures were evaluated

to determine the test-retest and internal consistency reliability of language delayed children. MLU and categories of lexical and morphological production were examined using language samples taken at different times and of various utterance lengths. Split-half reliability was conducted on the second language sample to evaluate the degree of variability within language samples. Results revealed a high correlation for MLU and a lower correlation for the proportion of questions used by the child, but no significant difference in either measure. The lexical items and bound morphemes used in the language samples were not significantly different between the two halves of the sample. There was greater correlation between language measures on the split-half reliability comparison than the test-retest comparisons. The study also found little variability in MLU between the first and second half of a language sample and greater lexical diversity in the first half of a sample compared to the second half of the sample. The results of test-retest and split-half reliability indicated greater variability across samples than within samples. (Cole, Mills, & Dale, 1989).

The studies completed by Gavin and Giles (1996) and Cole et al. (1989) found that conversational language sample measures are reliable in preschool-age children when a relatively long language sample is collected (i.e., 100 + utterances). Heilmann, Miller, Iglesias, Fabiano-Smith, Nockerts, & Digney-Andriacchi (2008) documented significant reliability coefficients for language sample measures collected from young school-age children who completed a narrative retell task. The participants were school-age English Language Learners (ELL) who produced oral narratives in both English and Spanish. The samples were transcribed using the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2007) using parallel transcription conventions and comparable language measures in each language. After equivalent measures were determined, Pearson correlation coefficients were calculated for four

narrative measures from two different sessions approximately two months apart. The language sample measures obtained from the ELLs' narratives had significant correlations with moderate strength. These correlations indicated that reliable language measures can be generated from short narrative retells (i.e., approximately 25 utterances) collected from school-age children (Heilmann et al., 2008). Even though the samples were substantially shorter than those from the previous studies, Heilmann et al. suggested that strong test-retest correlations were facilitated by the consistent sampling context (all children told the same story after following an initial model) and the children's age (school-age have stronger and more stable language skills than preschool-age children).

Use of Language Samples for Progress Monitoring

Language samples can be used as a tool to monitor progress as they have good test-retest reliability, sample language in a naturalistic manner, and are repeatable (Leadholm & Miller, 2005). Identification of the best sampling context for language needs to be determined in order to elicit an accurate sample of a child's language abilities. While conversational samples are not as developmentally sensitive as other contexts, such as narratives, in school-age children (Leadholm & Miller, 1992), they are still good indicators of language impairment in older children (Heilmann, Miller, & Nockerts, In Press). Narrative language samples may not be the ideal choice for monitoring outcomes, however, due to the possible influence of story or context. Such effects, which could positively or negatively skew the language sample data. In the present study we will be evaluating the properties of conversational language measures collected from young school-age children.

Properties of Conversational Language Samples

Because conversations are more interactive and have less intrinsic structure than narratives, large amounts of variability can be introduced to the samples based on the interactive style of the communication partners, conversational topic, and activities. Evans and Craig (1992) compared measures collected from an interview context to those collected from a play-based conversation in a group of school-age children with specific language impairment. The children produced longer utterances, a greater number of utterances, and more complex syntactic and semantic forms in the interview context. Evans and Craig concluded that the structure of their interview procedure provided a more reliable assessment context when compared to a freeplay context.

In addition to differences in measures as a function of elicitation procedures, measures from conversations could vary as a function of the conversational topic. Conversational protocols in the literature have combined the conversational topics of school activities and out-of school activities in their elicitations of conversational samples, assuming that the specific conversational topic has a modest influence on the language sample measures when using the same elicitation procedures (Evans & Craig, 1992; Leadholm & Miller, 1992). Literature examining the effects of sampling context on language measures has focused on differences across genres (e.g., narrative versus conversation; Evans & Craig, 1992; Westerveld, Gillon, & Miller, 2004; Abbeduto, Benson, Short, & Dolish, 1995), but has not looked at more detailed distinctions within a specific genre (e.g., conversational topics).

There are three plausible outcomes when comparing measures across story topics: both topics could yield similar language, the topic of school activities could elicit more complex

language, or the topic of home activities could elicit language abilities that are more sophisticated. First, the topic may not influence the language that is collected because both topics have commonly been used in eliciting language samples and are topics that children can easily discuss. The same elicitation protocol is used for both school and out-of school activities, which decreases variability. The topics should have similar discourse rules, so the pragmatic conventions will essentially be the same. However the language skills elicited from the children could be different, depending on the chosen topic. Discussion of school activities may require the children to use more decontextualized language. Decontextualized language requires use of complex grammar and vocabulary to express meaning because information must be shared about abstract objects, events, and situations. It requires more complex language such as more precise vocabulary and formal syntactical structures. Also known as literate language, these more elaborate language features are seen in the academic setting and are needed for success (Curenton & Justice, 2004). Conversations about school activities may elicit more literate language because of the topic's decontextualized context. Conversely, school-based conversational topics may elicit less complex language use. Sturm and Nelson (1997) found that children's use of language in school is influenced by a set of implicit rules that direct the way in which they speak while in the classroom. They found that children use shorter and fewer utterances in the classroom setting. Children may follow these classroom discourse rules when discussing activities related to school, resulting in higher values on the language sample measures when the children are talking about home activities.

In addition to documenting the effect of conversational topic, additional research is needed to determine if the language sample analysis procedure can become more efficient. On average, it takes five minutes to transcribe each minute of a language sample. A 15 minute sample would

take well over one hour to transcribe. Given the busy schedule of speech-language pathologists, taking an hour to transcribe a sample is not feasible; therefore, there is a need to determine if shorter samples still provide reliable data. One parameter to investigate that would aid in the reliability of shorter language samples is the consistency and standardization of elicitation procedures. Evans and Craig (1992) found that reliable measures could be obtained from a more structured elicitation procedure (interview context) while still preserving the naturalness of conversation, which in turn creates a more efficient sampling method.

The purpose of the present study was to evaluate the measures of conversational language samples collected from young school-age children for test-retest reliability. Before language sample analysis can be used to reliably document change in language skills, it is critical to determine if reliable language measures can be obtained when one is not tracking change in a child's language. If the measures do not demonstrate test-retest reliability when no change is expected, the measures will not be reliable when a clinician wants to track change and use language sampling as a progress monitoring tool. In order to address some of the clinical concerns associated with collecting language samples, an analysis targeting two factors was completed to determine their influence on the language measures. The study examined if sample length had a substantial influence on conversational language sample measures in young school-age children. It also evaluated the impact of conversational topic on language output in this population. This study addressed the following questions:

1) Are significant test-retest reliability coefficients observed for measures from two 10-minute conversational language samples produced by at-risk children in kindergarten?

2) Does sample length account for a substantial amount of variability in measures from conversational language samples produced by at-risk children in kindergarten?

3) Does conversational topic account for a substantial amount of variability in measures from conversational language samples produced by at-risk children in kindergarten?

CHAPTER 2: METHODS

Participants

Twenty kindergarten children at-risk for speech and/or language problems were recruited for this study. Participants were recruited from an elementary school in Greenville, NC following the approved protocols from the East Carolina University and Pitt County Schools Institutional Review Boards. Both the principal of the participating elementary school and Associate Superintendent for Instruction agreed to allow the examiner to work with the school psychologist, who identified the potential pool of participants and assisted with the consent process. Parental consent forms were signed for each child prior to collecting the language samples as well as a verbal assent from each child (refer to Appendix A).

The participants were enrolled in tier 2 of the response to intervention program at their school since these are the children that may be evaluated and monitored through language sampling. To be included in tier 2, the child's teacher and/or parent must have identified that the child is struggling in language use, reading, and/or academics and may qualify for special education services. The children in tier 2 have not previously received special education services. Tier 2 requires the involvement of parents, teachers, and supports to help the child succeed. Academic and/or reading skills are addressed and interventions are developed and implemented within the classroom to address the specific needs. Records were reviewed by the school psychologist and confirmed that all participants passed a hearing screening at 20dB for the frequencies of 1000Hz, 2000Hz, and 4000Hz at the beginning of the school year; certified audiologist or a certified speech-language pathologist completed the hearing screenings.

Tympanograms were also completed on all children and revealed that they all had normal middle ear function. In the present study 10 participants were male and 10 were female.

Elicitation Procedures

Two 10-minute structured conversational language samples were collected from each participant. The sessions were approximately one week apart and all samples were digitally audio recorded. The procedure to elicit the two 10-minute conversations was adapted from the protocol used by Evans and Craig (1992), which found that structured conversations facilitated consistent and reliable samples. In this study, the conversations focused on two separate topics: activities completed in school and activities completed out of school. Questions such as, "What can you tell me about your family?", "What kinds of things do you like to do in class?", or "What do you like to do when you're not in school?", were asked to introduce a topic. Because these samples were relatively short, the examiner provided an initial model of the activity to reduce the "warm-up" effect. When eliciting conversational samples, it can take children time to adapt to a new conversational partner and new activity, resulting in less productive language during the beginning of a language sample (Leadholm & Miller, 1992). To provide an initial model of the task, the examiner modeled an example of things she likes to do in school or out of school. The initial model lasted approximately one minute and occurred during the beginning of each session so the participants could become more comfortable with the examiner.

After the examiner modeled the activity, she began asking the target questions to initiate the conversation. A predetermined list of questions about school activities and out of school activities was used in each sample to introduce a topic (see Appendix B for the complete list of questions). If the child failed to respond to the question for five seconds after hearing the initial

question, the examiner rephrased the question. If the child did not provide a response within the next five seconds, the question was rephrased a second time. After the participant provided a description of the target activity, the examiner cued the child to expand on his/her description by providing four open-ended prompts for each question based on the child's response (e.g., "Tell me more about circle time,"). The examiner used open-ended prompts and then allowed up to three seconds for the child to respond to the prompt in order to elicit the most representative sample of the child's language abilities. Such prompts have been shown to be useful in stimulating greater language usage and eliciting language when a child is reluctant to speak (Evans & Craig, 1992). Once the prompts were provided by the experimenter and a topic was introduced, the conversation was directed by the child. If the child departed from the target conversational topic (i.e., conversation about school or out of school activities), the examiner used prompts to redirect the conversation back to the target topic. The same examiner was used in all sessions to create the least amount of variability in adult-child interaction. Data for each participant was collected over two sessions. Sessions were at least 4 days apart but no more than 1 week apart.

Because one of the goals of the study was to document variability due to session length (3 versus 7 minute segments) and conversational topic (school versus out of school), the participants were randomly assigned to four groups allowing the length and topic factors to be counterbalanced. The arrangement of the four groups can be viewed in Figure 1. Language samples from four children were collected in each grouping.

Figure 1. Groupings that facilitated counterbalancing of time and topic facets

	Session 1		Session 2	
Group 1	School	Out-of-school	School	Out-of-school
Group 2	Out-of-school	School	Out-of-school	School
Group 3	School	Out-of-school	Out-of-school	School
Group 4	Out-of-school	School	School	Out-of-school

Note: Large blocks represent 7-minute samples and small blocks represent 3-minute samples.

Transcription

All language samples were digitally recorded and were later transcribed and analyzed with SALT (Miller & Iglesias, 2008) using C-units for segmentation. C-units are classified as an independent clause and all of its modifiers or subordinating clauses (Loban, 1976). Each minute was coded in the transcript. When the time segment for the first topic ended (at either 3 or 7 minutes) the subsequent utterances that continued on the previous topic were not analyzed. The next minute was coded when the examiner asked the new topic's question. When the 3 or 7 minute marker occurred in the middle of a child's utterance, the time was coded according to the number of words before and after the 3/7 minute marker; if more words in the utterance occurred before the time marker, the minute was coded after the completion of the utterance and vice versa.

To ensure transcription accuracy 10% of the samples were retranscribed by the initial researcher to document intra-rater agreement. A second research assistant also retranscribed 10%

of the samples to document inter-rater agreement. The selected samples were compared to the original transcripts to document agreement at the word and morpheme level, placement of mazes, and utterance segmentation, which are areas where the transcriber must make critical decisions in order to create an accurate transcription. Percent agreement values were generated by calculating the total number of agreements divided by the total number of agreements and disagreements. Intra-rater agreement was 99.0% at the word and morpheme level, 95.8% for placement of mazes, and 95.6% for utterance segmentation. Inter-rater agreement was 98.3% at the word and morpheme level, 93.3% for placement of mazes, and 96.9% for utterance segmentation.

Language Sample Measures

The SALT program was used to calculate several measures of language production to document the children's form, content, and use of language. The first class of measures examined children's overall productivity (i.e., how much they talked). While productivity measures have received limited study, they may be an important clinical marker of general language performance. Miller (1991) documented that measures of productivity were some of the best measures of general development in 3 – 13 year old children. In their study of short narrative samples, Tilstra and McMaster (2007) documented that productivity measures were some of the most reliable language sample measures. Three productivity measures were examined in the present study. Number of total utterances (NTU) is the total number of utterances spoken by the child during the entire sample. Number of total words (NTW) is the total number of main body words the child used throughout the entire language sample. Words per minute (WPM) is a measure of verbal fluency that documents the rate of the child's speech.

To compute WPM, the number of words, including those in mazes, are summed and then divided by the sample's elapsed time.

The second class of measures documented children's grammatical and lexical skills. Mean length of utterance (MLU) was calculated by dividing the number of words and bound morphemes by the total number of c-units. MLU is a general index of children's grammatical skills and is related to overall language abilities; as children develop language, their utterances become longer and the structures used also increase (Brown, 1973; Miller, 1982). Aram, Morris, & Hall (1993) identified MLU as one of the best tools to identify children with language impairment. Number of different words (NDW) documents children's vocabulary skills by calculating the lexical diversity of a sample (Miller, 1987; Miller & Klee, 1995). NDW is determined by counting the number of different word types used in a language sample.

The third class of language sample measures documented children's discourse skills. In natural discourse, speakers commonly use false starts, repetitions, revisions, and filled pauses, collectively classified as *mazes* (Loban, 1976). Mazes were coded according to the SALT transcription conventions. The percentage of words in mazes to total main body words was calculated to document the child's discourse formulation skills. Mean turn length, which was calculated in words, is an additional index of discourse formulation skills as well, and was found by calculating the average number of main body words produced during a speaker's turn. A turn was defined as all the words contained in consecutive utterances produced by the same speaker. The total number of pause time in seconds during the sample was also calculated to further analyze the child's discourse skills. Pauses within utterances and between utterances of greater than two seconds were recorded in the transcript and measured to the nearest second (Leadholm & Miller, 1992).

The final class of language sample measures documented children's errors and dialectal differences. The total number of errors and omissions were calculated to document all semantic and syntactic errors as well as omission of words and bound morphemes that occur in obligatory contexts. To calculate total errors and omissions, the following SALT measures were summed together: omissions of words, omissions of bound morphemes, errors at the word level, and errors at the utterance level. Errors at the word level are overgeneralizations of syntactical rules, pronoun errors, and word choice errors (e.g., "goed"). Word order errors and coordination errors of tense or number are classified as utterance level errors. Omissions included children leaving out obligatory words and bound morphemes. (Leadholm & Miller, 1992). The percentage of intelligible utterances was calculated from the entire sample to determine what percentage of the total utterances were fully intelligible. Some children exhibited African American dialect use resulting in a language measure documenting the amount of dialect used during the sample which was calculated using the dialectal features outlined by Washington and Craig (1998). A percentage was used for this measure as well. For a complete description of the language sample measures, see Leadholm and Miller (1992) or Miller and Iglesias (2008).

CHAPTER 3: RESULTS

All 11 language sample measures were first calculated for the two complete 10-minute samples produced by the 20 participants. Table 1 displays the mean, standard deviation and range for each of the measures for session 1 and session 2. The group's average performance on all measures increased in session 2, except for the percentage of mazes and number of pauses To test the significance of the changes in measures, as series of analysis of variance (ANOVA) tests were completed with each measure as the dependent variable and session (time 1 versus time 2) as the between groups variable. Non-significant differences were observed for each contrast ($p > .15$) with the exception of percentage of intelligible utterances, which was significant at the ($p = .03$). While there were nearly no significant differences between time 1 and time 2, it cannot be assumed that there were no meaningful differences across sessions given the limited power of this study and the logical flaw in accepting the null hypothesis. Therefore, eta squared values were calculated for each measure to document the amount of variability explained by session for each measure (see table 1). Reviewing the effect size analyses confirmed that there was indeed a notable increase in measures for the children as a group, with session accounting for more than 10% of the variance in numerous measures.

Test-Retest Reliability

Separate test-retest reliability coefficients were calculated for each of the eleven language sample measures. Measures from the first 10-minute sample were compared to the measures from the second 10-minute sample using Pearson correlation coefficients. The strength of the correlations was determined using the scale created by Cohen (1988). A small or weak correlation was established at $r = \pm .10$ to + - .29, a moderate correlation at $r = \pm .30$ to + - .49, and a strong correlation at $r = \pm .50$ to + - 1.0. A significant correlation was found on all

Table 1.

Descriptive Statistics of the 11 Language Measures for Two 10-Minute Samples

		Productivity			Vocabulary/Grammar		Discourse			Errors & Differences		
		NTU	NTW	WPM	NDW	MLU	Turn Length	Mazes	Pauses	Intell	Errors & Omis.	Dialect
Sample 1	Mean (SD)	75.2 (35.96)	399.8 (261.69)	50.2 (33.22)	125.8 (62.81)	5.3 (1.67)	22.3 (29.22)	11.5% (6.86%)	2.4 (1.85)	94.0% (5.70%)	8.8 (6.65)	7.8 (10.12)
	Range	16 - 131	46-915	5.3-118	32-229	2.5-8.7	2.1-132.5	0%-21%	0- 6.4	82%-100%	0- 27	0-31
Sample 2	Mean (SD)	82.6 (40.5)	440.20 (275.1)	53.34 (33.3)	134.65 (60.4)	5.44 (1.7)	22.95 (28.1)	10.96% (5.7%)	2.16 (1.7)	95.85% (5.3%)	10.25 (8.7)	9.95 (10.8)
	Range	22-283	49-1027	4.9-120.5	31-224	2.5-9.3	2.3-128.8	0%-23%	.17-6.6	78%-100%	0-33	0-31
	eta ^{2}	.12*	.11*	.04*	.09*	.02*	.01*	.02*	.10*	.22**	.06*	.06*

* $p > .15$

** $p = .03$

Table 2

Test-retest Correlation Coefficients for the 11 Language Measures for the Two 10-Minute Samples

	Productivity			Vocabulary & Grammar		Discourse			Errors & Differences		
	NTU	NTW	WPM	NDW	MLU	Turn Length	Mazes	Pauses	Intell	Errors & Omis.	Dialect
r	.86*	.90*	.88*	.89*	.87*	.96*	.78*	.91*	.78*	.73*	.61**

* $p < .001$

** $p = .005$

measures (see table 2). Using Cohen's scale for strength of correlations all measures demonstrated strong correlations.

Generalizability Theory

After completing the reliability analyses, Generalizability Theory (G-Theory) analyses were completed to better understand the source of measurement error in the language samples. G-theory documents the sources of error obtained in a measure based on a given number of facets. The possible sources of error are considered simultaneously in order to observe the interactions among the facets of error (Scarsellone, 1998). G-theory analyses identify sources of error using the principles of Analysis of Variance (ANOVA), where the facets (or factors, in ANOVA) are analyzed to see how they affect the measurement error. While test-retest reliability analyses document the total amount of variability explained by measurement error, G-theory further documents facets that are responsible for the unexplained test error.

Two separate G-theory analyses were computed. In order to examine the effects of length each 10-minute language sample was broken into a three minute segment and a seven minute segment. Each sample was also separated into two different topics, school versus out-of school activities to assess the effects of topic on language measures. The sample arrangements can be viewed in Figure 1.

Because the facets were not fully crossed, separate G-Theory analyses were computed for the length and topic facets. Because the length analyses compared samples of varying length, the measures had to be converted to the form of a ratio except for WPM, MLU, intelligibility, and mazes. All other measures were divided by the sample length (i.e., 3 or 7) to generate comparable ratio measures (e.g., number of utterances per minute). For the length analysis, facets

included the child, session (1 versus 2), length (3 versus 7 minutes), and all associated interactions. For the topic analysis, facets included child, session (1 versus 2), topic (school versus out of school), and all associated interactions.

A G coefficient was also calculated for each measure in the two separate analyses resulting in 22 G coefficients which document how well the measures would generalize. For both sets of analyses (length and topic), strong G-coefficients were observed on all language measures except for errors and omissions in the topic analysis; these results were consistent with the results from the reliability analyses.

When interpreting the variance components, analyses of the individual facets (i.e., child, session, length, and topic) documented the amount of variability specific to the facet. The child facet documented the amount of variability explained by differences in children; if there is minimal error introduced from instrumentation variables (session, length, topic), a large proportion of the variance will be allocated to the children. The additional single facets (session, length, and topic) and the interaction between the instrumentation facets (session x length) documented variability due to instrumentation variables. When variability is allocated to the instrumentation facets and their interactions, it shows that differences in instrumentation affect all children similarly. For example, if substantial variance loads onto the session facet, then the children as a group performed differently across sessions. When there is an interaction between the child and one of the instrumentation facets, there is documentation that differences in instrumentation, such as sample length, change the rank ordering of the children. Thus, these interactions are some of the most important interactions to observe; if there is substantial variability loaded onto the child x instrumentation facets, the particular instrumentation facet is

problematic for facilitating reliable measurement. Because there is no error term in G-theory analyses, unexplained error loads onto the highest level interaction (i.e., child x session x length and child x session x topic) in addition to a true three-way interaction of the facets. Therefore, the highest level interactions can be difficult to interpret (see Brennan, 2001 for a review).

Length analyses. The percentage of variance due to the child, session (1 versus 2), length (3 versus 7 minutes), and all respective interactions are displayed in Table 3. For each language measure the greatest variance was loaded onto the child facet except for errors and omissions and dialect use, in which the child x session x length three-way interaction created the most variance. Because the variance attributed to the 3-way interaction also accounts for any error that was not due to the other facets in addition to the true 3-way interaction, the percentage of variability for errors and omissions and dialect use may not only represent error from the interaction of the 3 facets (child, session, and length). Having the majority of the variance loaded on the child for most measures is yet another confirmation that the measures from these samples, as a whole, are highly reliable and that the instrumentation facets contribute substantially less error than the error across children. While none of the instrumentation facets introduce copious levels of error, they will be reviewed to identify which facets are introducing the most error.

The session facet accounted for less than 5% of variance on all language measures. The single facet of length accounted for some variance on the measures of NDW (6%) and MTL (13%) and a negligible amount of variance for all other language measures. These variances demonstrate that the children as a group had different lexical diversity and turn length values in the 3 and 7 minute samples; however, length accounted for less variance than the single facet of child for all measures. The session x length interaction accounted for little variation and all

Table 3

Percentage of Variance of Measures Due to Child, Session (1st vs. 2nd), Sample Length (3 vs. 7 minutes), and Interactions

	Productivity			Lexical Diversity/ Utterance Length		Discourse			Errors & Differences		
	NTU	NTW	WPM	NDW	MLU	MTL	Mazes	Pauses	Intell	Errors/ Omis.	Dialect Use
C	75.7%	80.5%	78.0%	69.3%	76.3%	53.7%	55.8%	83.3%	36.2%	40.6%	42.3%
S	0.06%	0%	0%	0.04%	0.03%	0%	0%	0.84%	3.9%	0.9%	0%
L	0.3%	0%	0%	7.5%	0.2%	12.6%	0%	0.03%	0%	0%	0.10%
C x S	1.9%	2.7%	7.1%	3.9%	6.2%	0%	0%	7.3%	0%	0%	6.3%
C x L	0%	0%	0%	0%	2.0%	13.1%	0%	0%	30.6%	0%	0%
S x L	0%	0.3%	0.1%	0%	0%	0%	0.2%	0%	0.7%	6.0%	1.6%
CxSxL (+ error)	22.1%	16.6%	14.7%	19.3%	15.3%	20.6%	44.0%	8.5%	28.7%	52.5%	49.6%

Note: C = Child, S = Session, and L = Sample Length

values were negligible except for errors and omissions. Six percent of variance in this measure was due to the interaction between the interaction of session and length.

As mentioned above the child and instrumentation factors are the most important facets to evaluate because they document how the rank ordering of the children changes as a function of the session and sample length. The child x session interaction accounted for a fair amount of variance for the measures of WPM, MLU, pauses, and dialect use. The percentage of variance for these interactions ranged from 6%-7%. Two of the child x length variables accounted for a fair amount of variance; 13% of the variance in the measure of MTL and 31% of the variance for intelligibility was due to this interaction. For the measures as a whole, there was a greater amount of variance observed in the child x session interaction than the child x length interaction. Of these two instrumentation facets, session had a greater impact than length and introduced more variability in the measures.

Topic Analyses. The percentages of variance for the facet of topic are displayed in Table 4. Again, the variance due to child, session, and topic individually are displayed and the interactions between child x session, child x topic, session x topic, and child x session x topic were calculated as well. The majority of the variance was due to the child for all language measures except intelligibility, errors and omissions and dialect use. The greatest amount of variance in the measures of intelligibility and dialect use was from the 3-way interaction between the child, session, and topic. This again may not be a true representation of the error associated with the interactions between the three facets because it includes error that is not accounted for in the facets examined. The interaction between the child and topic accounted for the greatest variance in errors and omissions which will be discussed further later.

Table 4

Percentage of Variance of Measures Due to Child, Session (1st vs. 2nd), Sample Topic, and Interactions

	Productivity			Lexical Diversity/ Utterance Length		Discourse			Errors & Differences		
	NTU	NTW	WPM	NDW	MLU	MTL	Mazes	Pauses	Intell	Errors/ Omis.	Dialect Use
C	72.5%	77.9%	75.0%	77.5%	74.9%	73.8%	47.4%	82.0%	38.6%	12.0%	29.5%
S	0.1%	0%	0%	0%	0%	0.3%	0%	0.7%	0%	0%	0.1%
T	0.6%	1.2%	1.4%	1.3%	0%	2.9%	0%	0.4%	3.5%	0%	0.1%
C x S	0.1%	3.3%	11.0%	2.7%	0%	0%	19.7%	2.1%	0%	14.4%	0%
C x T	7.8%	3.1%	3.4%	1.1%	0%	4.4%	10.4%	0%	0%	44.8%	0%
S x T	0%	0%	0%	0%	0%	0%	0%	0%	0%	6.0%	0%
CxSxT (+ error)	18.9%	14.5%	9.2%	17.4%	25.1%	18.6%	22.5%	14.8%	57.9%	22.7%	70.3%

Note: C = Child, S = Session, and T = Sample Topic

Both session and topic individually accounted for less than 5% of the variance in all observed language measures. All children performed similarly on these measures despite the change in session or topic. The session x topic interaction did not produce much variability in any measure except for errors and omissions (6%).

The child x instrumentation interactions are again important because of their impact on the rank ordering of children and which facets have a greater influence in the variability in measures. The child x session interaction demonstrated notable amounts of variation for WPM, mazes, and errors and omissions. The variations ranged from 11% to 20%. Unlike the previous analysis of the interaction of child x length, which did not have striking levels of variance attributed to it, the child x topic interaction had notable amounts of variation in certain measures. NTU, mazes, and errors and omissions had 10% - 45% variation due to the child x topic interaction. For this analysis, topic demonstrated more impact on some measures than did session and created substantially larger amounts of variation in those measures.

CHAPTER 4: DISCUSSION

The goal of this study was to document the reliability of measures from kindergarten children's conversational language samples to determine if clinicians can reliably use them as a diagnostic and progress monitoring tool. Because collecting and analyzing language samples from children can be time consuming, this study examined various facets that could have an effect on the analyzed language measures. If shorter language samples are still reliable and the length does not introduce substantial measurement error, then collection and transcription of language samples can become more feasible for practicing clinicians. Topic was also examined as a facet that may have impact on the reliability of language samples. For language samples to be used as a progress monitoring tool, many samples will need to be collected therefore many topics may be utilized across different sessions.

Two 10-minute language samples were collected from 20 kindergarten children who were at-risk for speech and language impairments. A structured elicitation protocol was used to collect the language samples in order to generate stable data. Two topics were used, either school activities or home activities, and each 10-minute sample was split into a 3 and 7-minute segment. The samples were then transcribed and analyzed using the SALT program. Eleven language measures were generated for each sample. Pearson correlation coefficients were calculated comparing measures from session one to session two. The second analysis involved G-theory to examine sources of measurement error. Child, session, length, and topic and their interactions were the facets in the present study examined for their impact on measurement error.

Changes in Measures from Session 1 to Session 2

Before initiating the reliability analyses, the descriptive statistics were reviewed and a positive trend was identified in the measures from time 1 to time 2. A series of ANOVAs with

effect sizes were calculated and revealed that there were notable increases in several measures from time 1 to time 2. For example, changes in NTU, NTW, NDW, pauses, and intelligibility from session 1 to session 2 accounted for 9 – 22% of the variance in the measures. In the G-theory analyses, notable amounts of variance loaded onto the single session facet for several measures. These analyses revealed that the children, as a group, demonstrated greater values in measures during the second session. This increase may have been due to familiarity with the examiner and/or familiarity with the task. It is important to note that these analyses only examined group differences and the changes do not affect the reliability of the measures. While measures from the children, as a whole, increased from time 1 to time 2, the reliability coefficients remained strong and there was minimal variance attributed to the child x instrumentation facets interactions. Thus, the rank ordering of the children remained the same across sessions. However, these changes in measures are important and have clinical implications.

Clinicians should recognize that these measures are sensitive to familiarity factors. When comparing a single sample to a normative sample, which can be found in the SALT database, clinicians need to use the same procedures that were used to generate data from the normative sample. Therefore the clinician should collect the sample during an early interaction with the child in order to compare it with the samples in the SALT database. Clinicians should also be aware of a warm-up effect that may be seen across the sessions when language samples are used as a progress monitoring tool. However in typical clinical practice a child becomes more familiar with the clinician and this warm-up may be seen from the first evaluation session to the treatment session. To further reduce the warm-up effect, a model of the activity or conversation from the examiner can help the child to feel more comfortable with a new conversation partner. The

results of this study also motivate further research to a) develop a normative sample with multiple data points and b) further modify the elicitation protocol to facilitate more group-level consistency across sessions.

Test-retest Reliability

This study demonstrated that measures from ten minute structured conversational language samples were reliable from time one to time two. Even though the group performed differently, the rank-ordering of the children was consistent from time one to time two, which can be observed through the strong correlations. Most of the measures generated very strong reliability coefficients. With the exception of mazes, intelligibility, errors and omissions, and dialect use, reliability coefficients were in the .86 - .96 range. By far, the strongest correlation was observed for turn length, which reflects the implementation of stringent rules for examiner input used in this study. It was noteworthy that the measures of productivity and vocabulary/grammar were, as a group, some of the most reliable measures. These data confirm that these measures of general language performance are reliable when a structured elicitation protocol is utilized. While previous studies have determined that language samples are reliable from time one to time two in relatively long conversational samples from young school-age children (Cole, Mills, & Dale, 1989; Gavin & Giles, 1994), this study demonstrated reliability for structured conversational language samples in school-aged children. Based on these findings, conversational language samples may be an effective tool in the assessment process of children with language impairment. They can determine a child's baseline language ability through various language measures. Conversational language samples can also be used to document change in a child's language throughout the therapy process.

Some of the language measures, while still demonstrating significant reliability, had correlations that were not as strong. Dialect use had a correlation of +.61. This may be because not all children were using dialectal features; 45% of the children used no dialectal features in their productions. Small changes in the use of dialect from time one to time two made a larger impact on the reliability of the measure whereas if all the children used dialectal features those small changes observed would not have had as significant of an impact on the measure. Relatively weak correlations were also observed for intelligibility and errors and omissions. These weaker correlations could be a result of a ceiling effect. Most of the children were very intelligible and there were few errors and omissions throughout the samples. Finally, children's use of mazes also generated weaker correlations, which was not surprising. Leadholm and Miller (1992) identified that children's use of mazes is not related to age and is more useful as a clinical descriptor.

G-theory analyses

The second analysis was completed to examine certain facets that contribute to the inevitable measurement error in language sampling. The facets of length and topic were analyzed separately and both included the facets of child and session along with the interactions between facets. In both analyses it was found that most of the variance in language measures was due to the child followed by the interaction between the child, session, and length/topic with some exceptions.

For the length G-theory analyses there were modest levels of variance accounted for by the session alone. All measures except for NDW and MTL demonstrated modest or no variance due to length alone, indicating that most measures are not impacted by length of the sample. The session and length interaction also had little effect on variance except for the measure of errors

and omissions. The child X session interaction produced little variance for all language measures with the exception of WPM, MLU, pauses, and dialect use. Child x length interactions had modest effects on variability for all but two of the eleven language measures (MTL and intelligibility). These data demonstrated that the more descriptive language sample measures (e.g., pauses, dialect use, turn length) are more susceptible to being affected by instrumentation facets, while measures of general productivity (e.g., NTW, MLU) tend to be less influenced by differences in instrumentation. These data also demonstrated that session had a greater impact on variability of measures than length, providing compelling evidence that shorter conversational language samples generate samples as reliable as longer samples. This implies that longer samples are not always necessary and shorter samples can be used in the assessment process without much risk to the reliability of measures.

In the second G-theory analysis, in which the facet of topic was analyzed, most of variation was attributed to the child facet, which was consistent with the length analyses. Little variation was accounted for by the session and topic individually for all measures. Session and topic interaction accounted for a fair amount of variance only in the measure of errors and omissions. WPM, mazes, and errors and omissions accounted for a fair amount of variance attributed to the interaction between the child and session. For these three measures there was a change for all children from session to session. The interaction of child and topic produced some variance in the measures of NTU, mazes, and errors and omissions. Therefore topic had some impact on the reliability of these measures and changed the rank ordering of the children.

Overall length and topic do not have a substantial impact on the reliability of most measures. The majority of variance in language measures was due to the child and the interaction between three facets indicating that reliable language measures can be obtained from shorter

language samples using different topics. However some measures did demonstrate a fair amount of variability due to the interaction of child and length/topic. This signifies a change in the rank order of children and influence on reliability of measures from the instrumentation facet.

Topic had a greater influence on variability in three measures. When topic was changed, NTU, mazes, and errors and omissions were affected. Length had an influence on the rank ordering of children for two measures, MTL and intelligibility. Despite these few exceptions, the findings revealed that the child facet contributed to most of the variance in measures. While there was some variance in measures noted because of instrumentation, the g-coefficients and reliability coefficients were still strong and the variations for the most part were considerably less than those for the child and 3-way interaction. Consequently shorter language samples can be utilized as well as different topics during the collection of samples.

Limitations

There were some limitations to this study. The primary limitation was the heterogeneity of the sample. While all children were involved in Tier 2 of the response to intervention program, they exhibited varying language skills. These varying language skills resulted in most of the variability of language measures being accounted for by the child.

Future studies can examine different facets than those studied here. The facet of topic may be looked at further to determine whether other topics affect the measurement error more significantly. Also the length of the same could be examined further to determine if an even shorter sample, such as one minute, still produces adequate reliability. To investigate the reliability of shorter conversational language samples in populations other than children at-risk for speech and/or language disorders these procedures can be replicated. Also various elicitation procedures including narratives should be examined for reliability of language measures.

Conclusion

This study demonstrated that conversational language samples as short as three minutes can be utilized in clinical practice to assess and document progress in children with language disorders. The language measures calculated in the present study were found to be reliable from one session to the next in the school-aged population. A structured elicitation protocol allows clinicians to collect more reliable language samples.

Another question of this study concerned the feasibility of language samples due to time constraints of practicing clinicians and the need to collect multiple language samples from the same child. The results of the G-theory analysis demonstrated that clinicians are able to reliably use conversational language samples of shorter lengths. Length was not found to have a significant influence on measurement error for most language measures. Topic was also not found to have a significant influence on measurement error for the language measures indicating various topics can be used to collect samples without impacting the language measures.

Clinically these findings demonstrate that structured conversational samples as short as three minutes using the topics of school or home activities can be reliably collected from school-aged children. Standardized procedures should be used in order to obtain the most reliable samples and assess the child's true language abilities

REFERENCES

Abbeduto, L., Benson, G., Short, K., & Dolish, J. (1995). Effects of sampling context on the expressive language of children and adolescents with mental retardation. *Mental Retardation,* 33(5), 279-288.

Aram, D., Morris, R., & Hall, N. (1993). Clinical and research congruence in identifying children with specific language impairment. *Journal of Speech and Hearing Research, 36,* 580 – 591.

Brennan, R. L. (2001). *Generalizability Theory*. New York: Springer-Verlag.

Cole, K.N., Mills, P.E., & Dale, P.S. (1989). Examination of test-retest and split-half reliability for measures derived from language samples of young handicapped children. *Language, Speech, and Hearing Services in Schools,* 20, 259-268.

Curenton, S.M., & Justice, L.M. (2004). African American and caucasion preschoolers' use of decontextualized language: Literate language features in oral narratives. *Language, Speech, and Hearing Services in Schools,* 35, 240-253.

Evans, J.L., & Craig, H.K. (1992). Language sample collection and analysis: Interview compared to freeplay assessment contexts. *Journal of Speech and Hearing Research,* 35, 343-353.

Gavin, W.J., & Giles, L. (1996). Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research,* 39, 1258-1262.

Heilmann, J., Miller, J.F., Iglesias, A., Fabiano-Smith, L., Nockerts, A., & Andriacchi, K.D. (2008). Narrative transcription accuracy and reliability in two languages. *Topics in Language Disorders,* 28(2), 178-188.

Heilmann, J., Miller, J., & Nockerts, A. (In Press) Using language sample databases. *Language, Speech, and Hearing Services in the Schools.*

Leadholm, B., & Miller, J.F. (1992). *Language sample analysis: The Wisconsin guide.* Madison, WI: Department of Public Instruction.

Loban, W. (1976). *Language development: Kindergarten through grade 12*. Urbana, IL: National Counsel of Teachers of English.

McCauley, R. (2000). *Assessment of Language Disorders in Children.* New York: Psychology Press.

Miller, J. (1987). A grammatical characterization of language disorder. *Proceedings of the First International Symposium on Specific Speech and Language Disorders in Children.* London, England.

Miller, J. (1991). Quantifying productive language disorders. In J.F. Miller (Ed.) *Research on child language disorders: A decade of progress (pp. 211 – 220).* Austin, TX: Pro-Ed.

Miller, J. & Klee, T. (1995). Computational Approaches to the Analysis of Language Impairment. In P. Fletcher & B. MacWhinney (Eds.), *The Handbook of child language* (pp. 545-572). Oxford: Blackwell.

Paul, R. (2007). *Language Disorders from Infancy through Adolescence. Third Edition.* St. Louis, MO: Mosby.

Scarsellone, J.M. (1998). Analysis of observational data in speech and language research using generalizability theory. *Journal of Speech, Language, and Hearing Research,* 41, 1341-1347.

Schiavetti, N., & Metz, D.E. (1997). *Evaluating Research in Communicative Disorders. Third Edition.* Needham Heights, MA: Allyn & Bacon.

Sturm, J.M., & Nelson, N.W. (1997). Formal classroom lessons: New perspectives on a familiar discourse event. *Language, Speech, and Hearing Services in Schools,* 28, 255-273.

Tilstra, J., & McMaster, K. (2007). Productivity, fluency, and grammaticality measures from narratives: Potential indicators of language proficiency? *Communication Disorders Quarterly,* 29(1), 43-53.

Washington, J., & Craig, H. (1998). Socioeconomic status and gender influences on children's dialectal variations. *Journal of Speech, Language & Hearing Research*, *41*(3), 618.

Westerveld, M.F., Gillon, G.T., & Miller, J.F. (2004). Spoken language samples of New Zealand children in conversation and narration. *Advances in Speech-Language Pathology,* 6(4), 195-208.

APPENDIX A: ASSENT AND CONSENT FORMS

Child Verbal Assent Script

We want to learn how children learn to talk and use English so I'm going to have you talk to me for a few minutes about things you like to do in school and things you like to do at home. This will help us to hear the kinds of words you use and how many words you use. Some of your classmates are going to help me out too and talk with me. We will go out in the hall or into a quiet classroom and I will ask you some questions and you can tell me about your school, family, or what you do for fun. If you decide you don't want to participate anymore, you can tell me. It will be okay and you will not get in trouble and nobody will be mad at you. If you have any questions about the study you can ask me. Also if you think of any questions later, you can call me or my supervisor. Would you like to help me out and talk with me for a few minutes?

Parental Consent Form

Dear Parent/Guardian,

I'm presently working on my Masters of Communication and Science Disorders at East Carolina University. As part of my degree requirements, I am planning an educational research project to take place at your child's school that will help me to learn more about children's language skills by collecting a language sample and documenting their use of the English language. The fundamental goal of this research study is to determine better procedures for assessing children's language abilities.

As part of this research project, your child will participate in the collection of a language sample over two 10 minute sessions within one to two weeks that will allow me to record your child's speech sample. As this study is for educational research purposes only, the results of each recorded sample **will not** affect your child's grade. I am requesting permission from you to use your child's data (i.e. language sample) in my research study. Please understand that your permission is entirely voluntary.

If you have any questions or concerns, please feel free to contact me at school at

(440)570-9188 or by emailing me at lad0728@ecu.edu. If you have any questions about the rights of your child as a research participant, you may contact *The University and Medical Center Institutional Review Board* at 252-744-2914.

Please detach and return the form below by . Thank you for your interest in my educational research study.

Lindsay DeBrock

Researcher/Investigator

As the parent or guardian of _____,

(write your child's name)

☐ **I grant my permission for Lindsay DeBrock to use my child's data in her educational research project regarding language abilities. I voluntarily consent to Lindsay DeBrock using any of the data gathered about my student in her study. I fully understand that the data will not affect my child's grade, will be kept completely confidential, and will be used only for the purposes of her research study.**

☐ **I do NOT grant my permission for Lindsay DeBrock to use my child's data in her educational research project regarding language abilities.**

Signature of

Parent/Guardian:_____Date:_____

University and Medical Center Institutional Review Board
East Carolina University • Brody School of Medicine
600 Moye Boulevard • Old Health Sciences Library, Room 1L-09 • Greenville, NC 27834
Office **252-744-2914** • Fax **252-744-2284** • **www.ecu.edu/irb**
Chair and Director of Biomedical IRB: L. Wiley Nifong, MD
Chair and Director of Behavioral and Social Science IRB: Susan L. McCammon, PhD

TO: Lindsay Brock, Graduate Student, c/o John Heilmann, PhD, CSDI, ECU—3310 LAHN Building

FROM: UMCIRB KK

DATE: May 4, 2009

RE: Expedited Category Research Study

TITLE: "Stability of Conversational Language Samples from Children in Kindergarten: The Effects of Time, Sample Length, and Topic"

UMCIRB #09-0399

This research study has undergone review and approval using expedited review on 4.29.09. This research study is eligible for review under an expedited category because it is on collection of data from voice, video, digital, or image recordings made for research purposes. It is also a research on individual or group characteristics or behavior (including, but not limited to, research on perception, cognition, motivation, identity, language, communication, cultural beliefs or practices, and social behavior) or research employing survey, interview, oral history, focus group, program evaluation, human factors evaluation, or quality assurance methodologies. (NOTE: Some research in this category may be exempt from the HHS regulations for the protection of human subjects. 45 CFR 46.101(b)(2) and (b)(3). This listing refers only to research that is not exempt.)
The Chairperson (or designee) deemed this **unfunded** study **no more than minimal risk** requiring a continuing review in **12 months**. Changes to this approved research may not be initiated without UMCIRB review except when necessary to eliminate an apparent immediate hazard to the participant. All unanticipated problems involving risks to participants and others must be promptly reported to the UMCIRB. The investigator must submit a continuing review/closure application to the UMCIRB prior to the date of study expiration. The investigator must adhere to all reporting requirements for this study.
The above referenced research study has been given approval for the period of **4.29.09** to **4.28.10**. The approval includes the following items:

- Internal Processing Form
- Parental permission (received 4.17.09)
- Child Verbal Assent Script (received 4.17.09)
- Letter of Support: Principal (dated 4.17.09)
- Letter of Support: Superintendent of Pitt County Schools (dated 4.21.09)

The Chairperson (or designee) does not have a potential for conflict of interest on this study.

The UMCIRB applies 45 CFR 46, Subparts A-D, to all research reviewed by the UMCIRB regardless of the funding source. 21 CFR 50 and 21 CFR 56 are applied to all research studies under the Food and Drug Administration regulation. The UMCIRB follows applicable International Conference on Harmonisation Good Clinical Practice guidelines.

APPENDIX B: ELICITATION PROTOCOL

Hi *name of student*, my name is Lindsay. We're going to talk for a little bit today about activities you do at school and about activities you do outside of school. I will ask you a few questions and then you can tell me about it. I want to remember what you say so I'm going to tape record our conversation. Then I will listen to it later.

Questions for school activities

1. What kinds of things do you like to do in class?
2. What do you do at recess?
3. Can you tell me about your favorite part of the school day?
4. What kinds of things do you do during center time?
5. What do you do in gym class?
6. What do you do in music class?
7. What are you making in art class?
8. What kinds of things do you not like to do in class?
9. Can you tell me about your class pet?
10. Can you tell me about lunch time?
11. What are you learning about? (in math, reading)
12. What do you do when you get to school?
13. Tell me about your friends at school.
14. What books do you like to look at/read?
15. Tell me about your teacher.
16. Tell me about your school.

Questions for out-of school activities

1. What can you tell me about your family?
2. What do you like to do when you're not in school?
3. What toys do you like to play with at home?
4. What do you like to do on the weekend?
5. Do you like to play any sports or games?
6. What do you like to do outside?
7. What do you do like to do with your brothers and/or sisters?
8. Tell me about where you live.
9. What does your family do for dinner?
10. Tell me about your pets.
11. What do you do on vacations?
12. What do you do when you visit your grandparents?
13. What do you do like to do in the summer?
14. What do you not like to do at home?
15. Who do you like to play with outside of school?

Prompts

1. Can you tell me more about that?
2. That’s interesting, tell me more.