Using Language Sample Databases

John J. Heilmann
East Carolina University, Greenville, NC
Jon F. Miller
Ann Nockerts
University of Wisconsin—Madison

Language sample analysis (LSA) has been used in research on language development for more than 50 years and has produced most of our knowledge about the development of language production in typical children (Brown, 1973; Slobin, 1967; Templin, 1957). The process has moved from handwritten transcription and analyses (Slobin, 1967) to computer software that is designed to improve the efficiency and accuracy of LSA. Current LSA software programs include the Systematic Analysis of Language Transcripts (SALT; Miller & Iglesias, 2008), the Child Language Data Exchange System (CHILDES; MacWhinney, 2000), and Computerized Profiling (CP; Long, Fey, & Channell, 2008). Because LSA documents children’s language use in naturalistic contexts, it has unparalleled validity for measuring language use and has been considered by many to be the gold standard for assessing language status in children (Tager-Flusberg & Cooper, 1999).

Early language sample studies revealed that measures from children’s spontaneous language use were remarkably robust (Miller & Chapman, 1981). Miller and Chapman documented strong age-related changes in children’s mean length of utterance in morphemes (MLUm) and identified LSA as a potentially useful clinical tool. This early work provided an impetus for the use of LSA to document language status and changes associated with intervention. However, interpreting the results of language samples was still difficult for clinicians due to the extensive knowledge of language development that is required without some other reference data available. This dilemma provided the motivation for creating the SALT databases.

Development of Language Sample Protocols

The first SALT database project began in collaboration with a group of speech-language pathologists (SLPs) from the Madison, WI Metropolitan School District. The collaboration began at the same time that the first computerized version of SALT was developed. Researchers from the University of Wisconsin–Madison and school-based clinicians met monthly to share case studies and explore the clinical utility of computer-assisted LSA. The first project that was identified as essential to making language sampling a viable clinical tool was the creation of databases of typical speakers. Over these transcripts were compared to measures from 244 transcripts in the SALT conversational database. A series of discriminant function analyses were completed to document the sensitivity and specificity of the language sample measures.

Results: The language sample measures were effective in classifying children based on their language status, with correct identification of 78% of the children with language impairment and 85% of the children who were typically developing.

Conclusion: The SALT databases provide a useful tool for the clinical management of children with language impairment.

KEY WORDS: language sample analysis, assessment, discourse, language sample databases
the course of months, language sample protocols were developed for conversation and narrative contexts to determine if these different speaking demands produced different language outcomes. Our case study experience suggested that we needed to follow specific guidelines in order for the samples to be uniform, controlling examiner behavior and content as much as possible to create comparable sets of samples. Conversation protocols included fixed topics such as school, family, and holiday events, which provided an opportunity for children to talk about things absent in time and space and make reference to past and future events. Examiners were instructed to ask open-ended questions, make requests for clarification, and avoid yes/no and other questions that commonly elicit one-word responses. The narrative protocol asked children to retell their favorite story, such as a television episode or movie. Examiners were to only prompt when necessary with open-ended statements (e.g., “tell me more” or “then what happened”). Requests for clarification were encouraged to let children know that the examiner did not understand but was very interested in what happened.

What Constitutes a Typical Distribution of Language Ability?

After finalizing the protocols and running a successful pilot program, we were next confronted with establishing the criteria for the children to be included in the database. We identified and attempted to control three variables that would affect the language sample measures: socioeconomic status (SES), identified special needs, and school ability. The Madison Schools administrative staff was supportive of the project and identified children to participate who represented the socioeconomic diversity of the state as measured by their eligibility for free and reduced lunch. The goal was to have 20% of the sample qualify for free and reduced lunch, a percentage that is consistent with state averages. Children receiving any special education services were excluded. We added a measure of school ability, which asked the teachers of each child to rank him or her as high, average, or low relative to all other children they had taught. This provided some control over the general abilities and diversity of the final sample. As an unfunded project, this model was used over direct measures of SES and cognitive ability out of necessity. As the work has progressed, this approach has proven sound in producing data sets that are comparable to each other (Miller, Heilmann, Nockerts, Andricachi, & Iglesias, 2006).

Language samples were first collected from 5-, 7-, and 9-year-old children to determine if there were age-related changes in language sample measures beyond the preschool years. Initial analyses of this database revealed significant age-related increases in key language measures and provided the motivation for the development of an extensive database for children 3–13 years of age. Upon comparing children’s language use across sampling contexts, we found that children produced more complex language in narratives than conversation but had equivalent vocabulary diversity in each sample type, as measured using number of different words (NDW). Children produced more reduplications and reformulations (i.e., mazes) in narration than conversation, confirming the increased formulation load of the narrative task (Leadholm & Miller, 1992).

Expanding the Databases

The first Wisconsin samples confirmed the value of these databases in providing detailed expectations for language production in school-age children. We continued to work on collaborative projects to extend the original Wisconsin conversation and narrative databases. One expansion to the original databases was extending the sampling contexts by eliciting story retells. The original narrative task used in the Wisconsin database allowed each child to select the story, movie, or television program to relate, which provided optimum familiarity with the story but prevented the comparison of specific vocabulary and narrative structure except in the most general terms. Other investigators had employed elicitation procedures that cued each child to tell the same story, which facilitated more accurate comparisons across children when analyzing vocabulary and story structure (Berman & Slobin, 1994). In addition, research on oral narratives using story retell tasks demonstrated its power in predicting literacy attainment (Snow, 1983).

From a clinical efficiency standpoint, narrative retells require fewer utterances while maintaining robust developmental data. Our initial databases used 100 complete and intelligible utterances as the standard sample size. The results of our research on these databases concluded that 100 utterances were sufficient to document age-related changes and changes in performance that were attributable to speaking condition. The narrative retell literature (e.g., Berman & Slobin, 1994; Snow, 1983) provided evidence that story length could be shorter than 100 utterances and still provide powerful predictive evidence for oral language and reading achievement. These results provided motivation for developing new story retell databases that could provide additional opportunities to document children’s oral language skills. Narratives were collected in both Wisconsin and California using the retell procedure, where students listened to the examiner tell a story, followed along with pictures in a wordless picture book, and then were cued to retell the same story to the examiner (adapted from Strong, 1998). The kindergarten and first-grade children were cued to retell Frog. Where are You? (Mayer, 1969), the second-grade children to retell Pookins Gets Her Way (Lester, 1987), the third-grade children to retell A Porcupine Named Fluffy (Lester, 1986), and the fourth-grade children to retell Dr. DeSoto (Steig, 1982).

A third set of narratives was compiled documenting the oral language skills of native Spanish-speaking children who were learning English as a second language in schools in the United States. These data were collected as part of a large multisite collaborative project examining the academic outcomes of English language learners (Francis et al., 2005). Oral language skills were assessed in both languages using narrative retells with wordless picture books.

An additional narrative database has recently been included in the SALT databases, which includes samples that were acquired during the norming of the Test of Narrative Language (Gillam & Pearson, 2004). Gillam and colleagues have provided access to 500 narrative language samples that were elicited using three separate tasks: a story retell with no picture cues, a narrative elicited using a sequence of five pictures, and a narrative elicited using a single complex picture.

The final database available in SALT includes samples that were collected from a large group of children from major and secondary urban areas throughout New Zealand. This database includes language samples from both children of European descent and children from the indigenous culture (Maori), all of whom were monolingual English speaking. Language samples were elicited in three contexts: conversation, child-initiated personal narrative, and a narrative retell task.

As the number of language sample databases increased, we were confronted with the task of organizing the databases based on the
sampling context, geographic location, and linguistic proficiency. One option was to keep the samples from each project separate. The second option was to combine databases that used the same sampling procedures and same population (e.g., typically developing [TD] English-fluent children who produced narrative retells). Using the second option, conversations and narrative retells that were collected in Wisconsin and California could potentially be combined. The relationship among these databases was examined, and only modest differences in language sample measures were observed when comparing across geographic locations (Miller et al., 2006). For example, when comparing two sets of narrative story retells, geographic location (Wisconsin vs. California) accounted for ≤8% of the variability in MLUm, NDW, mazes, and words per minute (WPM; see Table 1). Furthermore, measures from conversational databases were also consistent across samples from Wisconsin and California, with geography accounting for ≤5% of the variability in MLUm, NDW, mazes, and WPM (see Table 1). Because there were modest differences, these databases were combined. Surprisingly, conversation and story retell data sets from the United States and New Zealand were similar for most measures, with the exception of the 5-year-old children (Westerveld, Gillon, & Miller, 2004; Miller et al., 2006). Despite the similarities in measures, these databases remain separate to retain national integrity. Today, SALT includes six databases with samples from more than 4,000 children producing more than 6,500 language samples. These databases are unique in that consistent elicitation procedures were used for several different sampling contexts, with each context containing hundreds of samples from TD children. Table 2 summarizes the various databases and includes the number of samples in each.

Properties of the Databases

This large corpus of language samples allowed us to address a number of research questions about the LSA process, including the consistency of measures and the utility for school-age children (Leadholm & Miller, 1992; Miller & Klee, 1995; Miller et al., 2005). This work documented that measures of syntax, semantics, discourse, rate, and fluency that were collected from conversational and narrative samples for children 3–13 years of age are consistent and orderly within ages and contexts and are complementary to other data in the language development literature. Analyses of these databases have provided detailed descriptions of linguistic advances in TD children, tested models of language development, and critically evaluated measurement techniques for assessing children’s oral language skills (Leadholm & Miller, 1992; Miller, 2007; Miller et al., 2006; Miller & Klee, 1995). Analyses of the databases also documented age-related changes in general measures of language production for children 3–13 years of age, including MLUm, NDW, and WPM, which were correlated with age at \( r \geq .65 \). No changes with age were recorded for the other measures, including mazes, pauses, and errors. These initial databases documented the power of LSA by demonstrating that many measures continued to change with advancing age (e.g., MLUm and NDW), whereas others did not (e.g., mazes; Leadholm & Miller, 1992). Contrary to conventional belief (cf. Scarborough, Wyckoff, & Davidson, 1986), measures of language production, such as MLUm, continued to increase beyond age 5. In fact, most measures continued to increase until age 11. The results of these analyses confirmed that these general measures of language production documented continued change through childhood as well as provided the means to quantify mazes, pauses, and errors at the word and utterance level.

Accessing the Databases

Since its inception, the goal of the SALT Software Project has been to make LSA user friendly by developing a transparent transcription system, transcription training, error checking tools, measures at all language levels, and access to databases of more than 6,000 language samples from typical speakers. Users new to SALT can receive training across a variety of modalities to facilitate different learning styles: reading the documentation that comes with the software, viewing training videos on the SALT Web site, and participating in training workshops. Of the three commonly cited computer-based approaches to LSA (SALT, CHILDES, and CP), SALT is the only program that requires a fee for access to the program. This fee allows the program to continue to meet the needs of both clinicians and researchers by providing funding for the continued development of the software, readily available technical support, and maintenance and expansion of the databases. As existing databases are expanded and new databases added, they are made available on the SALT Web site. Although the SALT programs

| Table 1. Language sample measures from two different locations (Wisconsin and California) for two sampling contexts (conversation and narrative retell). |
|---|---|---|---|---|---|
| | Conversation | Narrative retell |
| | Wisconsin | California | \( \eta^2 \) | Wisconsin | California | \( \eta^2 \) |
| N | 236 | 159 | | 46 | 55 | |
| Age range (years;months) | 5.2–9.5 | 5.2–9.5 | | 7.10–8.10 | 7.0–8.10 | |
| MLUm | 6.1 (1.3) | 6.9 (1.5) | .08 | 9.2 (1.0) | 9.4 (1.0) | < .01 |
| NDW | 116 (11) | 113 (13) | .02 | 135 (29) | 122 (23) | .05 |
| % Maze words | 11% (5%) | 11% (5%) | <.01 | 10% (5%) | 12% (4%) | .03 |
| WPM | 80 (22) | 88 (21) | .03 | 94 (24) | 89 (20) | .01 |

Note. MLUm = mean length of utterance in morphemes; NDW = number of different words; % maze words = percentage of maze words to total words; WPM = words per minute; standard deviations are presented within parentheses.

\( ^{\text{a}} \)Eta squared (\( \eta^2 \)) values represent the amount of explained variance in language sample measures as a function of geographic location.
are needed to access these databases, users can freely update their existing software to include the new and expanded databases.

Clinicians and researchers can use the SALT software to automatically generate a data file documenting their target child’s performance, average performance of age-matched children from the relevant database, and the statistical significance of differences in performance. When comparing a transcript to the database, the first task in generating these analyses is to choose the appropriate database to be used for comparison. As can be seen in Table 2, the SALT databases cover a variety of sampling contexts, age ranges, and data sources, all of which need to be considered when making this selection. After choosing the appropriate database, the user then matches the target transcript to appropriate samples selected from the database. The program automatically recognizes characteristics of the target speaker, such as age, gender, language, and speaking condition from information lines that are inserted at the beginning of the transcript. The program also considers the length of the sample to allow the user to match the target sample to the database based on the amount of language used (total number of utterances or words) or elapsed time. After comparing the characteristics of the target sample to the characteristics of all of the samples in the chosen database, the program preselects all database children within 6 months of age of the target child whose samples contain a sufficient number of utterances, words, or elapsed time. The age range and transcript cut of the comparison group can be changed by the user as necessary to create the ideal comparison set. The user then compares the target child to this comparison group, and the program automatically calculates the group values. This approach provides the user with total control over the comparison group within the limitations of the databases themselves.

### SALT Database Users

Clinical use of the SALT databases has been endorsed by a number of experts in the field (Gillam & Gorman, 2004; Johnston, 2006; Paul, 2007). Paul suggests using the database to compare measures that were generated from a child’s language sample, such as NDW and number of total words. She stated that identification of differences in performance between the target child and the database may assist in making a valid diagnosis of language impairment (LI) and can assist with the development of treatment goals. As an example, Paul stated that a child who produced fewer different words than children in the SALT database may have a vocabulary deficit and that an appropriate treatment would include focus on increasing expressive vocabulary. Johnston discussed using LSA throughout her book, focusing on bringing research to practice. She suggested that clinical evaluations use the databases along with more detailed analyses of syntax, semantics, and discourse to describe child performance. She incorporated the best of current research outcomes to describe language production to reveal child strengths and weaknesses leading to a focused intervention plan, if necessary. She devoted an appendix to SALT, discussing the utility and ease of use of the program. In their discussion of dynamic systems and LI, Gillam and Gorman acknowledged the multiple levels of linguistic proficiency needed for academic success and identified the need to evaluate a wide range of linguistic skills beyond simple measures of vocabulary and morphosyntax. The authors specifically recommended that clinicians not only collect language samples but also compare their target children’s performance to that of children in the SALT databases.

The databases from the SALT program have also been widely used by researchers in the field. The databases themselves have been used to provide control subjects in a number of studies examining children with speech and language deficits (Friel-Patti, DesBarres, & Thibodeau, 2001; Gillon & Young, 2002; Hall, 1996; Hall & Burgess, 2000; Pollack, Price, & Fulmer, 2003; Rvachew & Grawburg, 2006; Stockman, 2006). For example, Stockman documented that the MLUm’s of children using African American English were comparable to the MLUm’s of children from the reference database, who primarily used Standard American English. In addition, Rvachew and Grawburg used the SALT database to identify children with speech sound disorders who also had LI based on low MLUm scores relative to performance of children from the database.

### Use of LSA and SALT Databases for Identifying and Describing Children With LI

Although studies of children with LI have been extensively reported in the literature, there is still no consensus on the gold standard for identifying children with LI. In Tager-Flusberg and Cooper’s (1999) summary from a meeting to discuss a phenotype of specific language impairment (SLI), they identified that the criteria for LI is inconsistent throughout the research literature. Miller and Fletcher (2005) confirmed this variability by summarizing the criteria used for SLI in contemporary studies. Among studies, the severity of children’s LI varied widely, as did the tasks that were used to identify impaired performance.

Although standardized assessments are commonly used to identify children with LI in both research and practice, these measures do not consistently do a good job of identifying children who have a clinically identified LI. Plante and Vance (1994) identified four language tests that had strong published psychometric properties and administered these tests to two groups of children: those with LI and those without LI. They concluded that performance varied widely across the tests for both groups and that most of the tests were not effective in distinguishing between the TD children and the children.
with LI. Dollaghan and Campbell (1998) compared the effectiveness of the Test of Language Development 2: Intermediate (TOLD; Hammill & Newcomer, 1988) to a processing-based measure (the Nonword Repetition task) in identifying children who had been identified as having LI by an SLP. Upon completing a likelihood ratio analysis, they found that children who scored more than 1.5 SDs below the mean on the TOLD were only 3.73 times more likely to have been diagnosed with LI, whereas children who repeated fewer than 70% of the phonemes correctly on the Nonword Repetition task were 25.15 times more likely to have been diagnosed with LI. In sum, the TOLD was substantially less effective than the Nonword Repetition task in identifying children with LI.

Plante and Vance (1994) and Dollaghan and Campbell (1998) are just two studies that have demonstrated that standardized tests may not be effective in distinguishing children with LI. Rather, naturalistic and alternative assessment procedures that document the dynamic aspects of language use may be more effective in the clinical management of children with LI. LSA potentially overcomes many of the limitations of standardized language assessment measures by providing a means of collecting data on language use in naturalistic contexts. Aram, Morris, and Hall (1993) documented that MLU from a conversational language sample was the most sensitive measure for the identification of preschool children with LI and was substantially better than a battery of standardized tests. In a subsequent analysis, Dunn, Flax, Sliwinski, and Aram (1996) found that a subgroup of children who were clinically identified as having LI but did not meet the criteria from psychometric testing had substantial difficulty with their conversational language samples. Dunn et al. concluded that “objective measures derived from spontaneous language samples appear to be more closely related to the clinician’s perception of impaired language development for some children than standardized test discrepancy criteria” (p. 651).

Additional studies throughout the literature have documented that measures from conversational language samples are effective in identifying preschool-age children with LI (Bedore & Leonard, 1998; Fletcher & Peters, 1984; Gavin, Klee, & Membrino, 1993). These studies have found LSA to be effective in grouping children when using MLUm (Bedore & Leonard, 1998), measures of morphosyntax (Bedore & Leonard, 1998; Fletcher & Peters, 1984), and performance on the Language Assessment, Remediation, and Screening Procedure (Gavin et al., 1993), which is a method for describing children’s syntactic progress through 5 years of age (Crystal, 1982). In sum, the variety of measures and the naturalistic sampling conditions that can be effectively measured using LSA appear to be useful in documenting a range of linguistic skills that are present in the linguistic profiles of children with LI.

Heterogeneity in Children With LI

Tager-Flusberg and Cooper (1999) proposed that a variety of language constructs must be analyzed to account for the heterogeneity in children with LI, including measures of morphosyntax, semantics, vocabulary, phonology/articulation, and pragmatics. Although variability within the linguistic profiles of children with LI is generally accepted among clinicians, this heterogeneity is not often acknowledged in the research literature. Some effort has been placed on subtyping children with LI. The most common method of subtyping classifies children based on their relative expressive and receptive language skills (e.g., Archibald & Gathercole, 2006; Casalini et al., 2007; Simkin & Conti-Ramsden, 2006). A smaller body of literature has proposed several taxonomies for subgrouping children that go beyond relative weaknesses in receptive and expressive language skills (Aram & Nation, 1975; Bishop & Edmundson, 1987; Rapin & Allen, 1983; Tomblin, Zhang, Weiss, Catts, & Ellis Weismer, 2004). Three of the taxonomies have contrasted subgroups of children who have primary deficits in semantics and syntax versus primary deficits in phonological processing (Aram & Nation, 1975; Bishop & Edmundson, 1987; Tomblin et al., 2004). Additional subgroups were characterized by deficits in repetition ability/working memory skills (Aram & Nation, 1975; Tomblin et al., 2004). Two of the four taxonomy studies used observational and naturalistic language data in their analyses, which facilitated a more extensive subgrouping system. Their analyses revealed a group of children with primary deficits in pragmatics and social skills (Rapin & Allen, 1983; Tomblin et al., 2004), as well as subgroupings of children with verbal auditory agnosia, verbal dyspraxia (i.e., impaired production despite normal cognition), speech programming deficits, and word finding deficits (Rapin & Allen, 1983).

Early in the history of the SALT Software Project, we acknowledged and embraced this heterogeneity and worked firsthand with school clinicians to identify methods of documenting performance for the range of skills and deficits that the students presented. In developing the SALT analysis system and affiliated databases, we developed a set of transcription rules and associated analyses that allow clinicians to document a range of linguistic abilities and account for the heterogeneity in profiles of children with LI, including measures of linguistic form (e.g., utterance length, grammatical errors and omissions), content (e.g., number of different words), and use (e.g., turn taking). In addition, SALT provides the user with the ability to document children’s productivity and formulation skills that have been identified by clinicians as being meaningful to identifying and describing children with LI, including reduplications and reformulations (e.g., mazes) as well as verbal fluency and rate (e.g., WPM; Leadholm & Miller, 1992).

Although each sampling condition allows for simultaneous examination of the range of linguistic skills, conversational samples are particularly useful for eliciting both structural and pragmatic language sample features. Conversations, by nature, are more dynamic and interactive than other sampling contexts, such as narratives. To be effective in a conversation, speakers must formulate their ideas, use the appropriate vocabulary and grammar to express the ideas, monitor listener comprehension, reformulate ideas that were not understood by the listener, and incorporate new ideas and topics that were initiated by the conversational partner. By challenging children to effectively integrate this range of skills, examiners have an opportunity to document the range of skills necessary for accurate assessment of language production. Numerous studies have documented that measures from conversational language samples are powerful for identifying preschool-age children with LI (Aram et al., 1993; Bedore & Leonard, 1998; Dunn et al., 1996; Fletcher & Peters, 1984; Gavin et al., 1993). However, these studies have only examined preschool-age children and have focused primarily on measures of linguistic form (i.e., MLUm, morphosyntax, and syntactic skills). The purpose of the present study was to document the ability of a variety of SALT measures (i.e., measures of form, content, use, and productivity) to classify children based on their clinical diagnosis. This study is unique in its use of both linguistic and pragmatic measures to document group differences and group prediction in both preschool and school-age children.
The study will demonstrate the utility of SALT databases of typical speakers for answering the following important clinical questions:

- Do measures of form, content, use, and productivity acquired from conversational language samples predict language status in children who are 3–13 years of age?
- Do measures of form, content, use, and productivity acquired from conversational language samples predict language status within three age groups (preschool/kindergarten age vs. elementary age vs. middle school age)?
- Do measures of form, content, use, and productivity differ significantly in children with LI versus TD children?

**METHOD**

Conversational language samples were collected from 244 children from the Madison Metropolitan School District who were experiencing substantial language learning difficulties and were considered to have an LI. We used a clinical criterion for identifying the children with LI that included children who either were receiving language intervention services (90% of the sample) or were referred for language intervention services (10% of the sample). The children who were receiving services were identified by clinicians following the Madison Metropolitan School District guidelines, which required test performance to be $-1 SD$ or lower on a variety of measures, including descriptive procedures and performance on classroom tasks. All of the children receiving language intervention services had an individualized education plan (IEP) that only addressed language services. Children who had IEPs that identified additional disabilities (e.g., autism or cognitive disability) were excluded from the study. We acknowledge that the procedures for classifying children as language impaired differed from much of the literature, which typically uses discrete scoring criteria based on standardized tests. However, we were interested in identifying a group of children who were clinically presenting with substantial language difficulties. Several studies have documented that traditional assessment techniques that are commonly used in research studies are not sensitive to identifying children with true clinical LI (e.g., Dollaghan & Campbell, 1998; Plante & Vance, 1994) and that clinical criteria for LI may be more accurate when testing the ability of an assessment procedure to identify children with LI (Dollaghan & Campbell, 1998).

The children with LI ranged in age from 3 to 13 years, and 61% of the participants were boys. The school SLP recorded a conversational language sample of at least 10 hro ft r a i n i n gi nc o m p l e t i n gL S A .T h e language samples were recorded and later transcribed by a trained research assistant with at least 10 hr of training in eliciting consistent language samples. Conversational topics were exactly the same as those used to elicit conversational samples in the SALT database (e.g., home and school activities and holidays). Every SLP received 6–8 hr of training in eliciting consistent language samples. The language samples were recorded and later transcribed by a trained research assistant with at least 10 hr of training in completing LSA. The transcribers adhered to the standard SALT coding conventions, with transcribers adhering to the standard SALT coding conventions, with accuracy and agreement $> 95\%$ for these samples (Heilmann, Miller et al., 2008; Miller & Iglesias, 2008).

The transcripts from the 244 children with LI were matched to 244 transcripts from the SALT database. Although SES was not measured for each individual child, the children in both groups attended similar schools that had a similar proportion of children eligible for free and reduced lunch (approximately 20% of the sample). The language samples from the children with LI were also age matched to the typical children in the SALT database (see the “All children” row in Table 3). A t test revealed that the two groups did not differ significantly in age ($t = 0.89, p = .38$). Cohen’s $d$ was calculated ($d = 0.08$) and revealed that the group differences produced a very small effect size (Cohen, 1988). To answer the second and third research questions, the sample was broken down into three separate age groups (see Table 3): preschool/kindergarten-age children (3:0 [years;months]–5:11), elementary school-age children (6:0–9:11), and middle school-age children (10:0–13:6). A series of $t$ tests revealed that there were no significant differences in age and there were small effect sizes between the children with LI and the TD children who were preschool/kindergarten age ($t = 1.7$, $p = .87$, $d = 0.03$), elementary age ($t = -2.8$, $p = .78$, $d = 0.03$), and middle school age ($t = -0.89$, $p = .37$, $d = 0.06$). There were more boys in the group with LI (61%) compared to children from the SALT database (48% boys), which was not surprising given the higher prevalence of LI in boys (Tomblin et al., 1997).

**Language Sample Measures**

Although more than 50 measures can be automatically generated by SALT, we opted to use a limited set of measures that reflect the range of linguistic skills that can be assessed using conversational samples. Most of these measures have also been extensively reported in the literature as being sensitive to both age and language status. MLU$_{in}$ was used as an index of the children’s grammatical skills, and NDW was used as an index of their vocabulary skills. To document children’s productivity skills, WPM was calculated by dividing the total number of completed words by the elapsed time. Children’s overall discourse and formulation skills were assessed using three separate measures: Mean turn length calculated the average number of main body words (excludes maze words) that the child produced in each of his or her conversational turns; percentage of maze words provided an estimate of the number of reduplications, re-formulations, filled pauses, and false starts by dividing the number of total words by the number of words within mazes; and between-utterance pauses documented the total amount of time in seconds when there were significant pauses (no speech) between utterances. All pauses of at least 2 s in length were included in this measure. The final class of measures documented the children’s omissions and errors. The omitted words and omitted bound morphemes measures calculated the total number of obligatory words and morphemes that each child omitted from his or her sample. The word errors measure calculated the number of times that a child produced a

<p>| Table 3. Summary of children with language impairment (LI) and a matched group of typically developing (TD) children from the Wisconsin conversational database. |
|---------------------|---------------------|---------------------|---------------------|</p>
<table>
<thead>
<tr>
<th><strong>Age range</strong></th>
<th><strong>Mean age in years (SD)</strong></th>
<th><strong>Number of children</strong></th>
<th><strong>Age range</strong></th>
<th><strong>Mean age in years (SD)</strong></th>
<th><strong>Number of children</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>3:0–5:11</td>
<td>5.0 (0.8)</td>
<td>5.0 (0.8)</td>
<td>3:0–5:11</td>
<td>5.0 (0.8)</td>
<td>5.0 (0.8)</td>
</tr>
<tr>
<td>6:0–9:11</td>
<td>7.7 (1.0)</td>
<td>7.7 (1.0)</td>
<td>6:0–9:11</td>
<td>7.7 (1.0)</td>
<td>7.7 (1.0)</td>
</tr>
<tr>
<td>10:0–13:6</td>
<td>11.9 (1.2)</td>
<td>11.9 (1.0)</td>
<td>10:0–13:6</td>
<td>11.9 (1.2)</td>
<td>11.9 (1.0)</td>
</tr>
<tr>
<td>All children</td>
<td>8.0 (2.6)</td>
<td>7.8 (2.6)</td>
<td>All children</td>
<td>8.0 (2.6)</td>
<td>7.8 (2.6)</td>
</tr>
</tbody>
</table>

Heilmann et al.: Using Language Sample Databases
Lexical item incorrectly. Examples of word errors include overgeneralizations (e.g., goed instead of went), pronoun errors (e.g., him’s instead of his), determiner use (e.g., a apple instead of an apple), and tense errors (had went instead of had gone). The 

terance errors measure calculated the number of utterances that had syntactic errors or utterances that simply did not make sense as well as utterances that contained three or more word-level omissions and/or errors. Although there is overlap between lexical and syntactic errors, word and utterance errors were differentiated to describe specific areas of difficulty for children with LI (i.e., predominantly word-level errors, utterance-level errors, or a combination of both). Each of these measures was automatically calculated by the SALT software and transferred into a data file. The data file was imported into SPSS (SPSS, Inc., 2008) for statistical analysis.

### RESULTS

Classifying Children With LI and TD Children

A series of discriminant function analyses were completed to determine if these 10 language sample measures could be used to predict group membership based on language status (LI vs. TD). This analysis identified the best model for predicting group membership based on the language sample measures and compared this classification to the gold standard for diagnostic category used in this study (i.e., whether or not the child is receiving language intervention or has been referred for services). The SALT variables were simultaneously entered into the analysis, which generated a summary of the sensitivity and specificity data. A test’s sensitivity reflects its ability to identify children who truly have LI, also known as the true positive rate. A test’s specificity reflects its ability to identify children who truly have no impairment, also known as the true negative rate.

The first discriminant function analysis analyzed all of the 488 participants simultaneously without accounting for any differences in age. The results are summarized in the top section of Table 4. The sensitivity estimate, or true positive rate, is in the top left quadrant of the table (69%); the specificity estimate, or true negative rate, is in the bottom right quadrant (84%). The false positive rate, which documents the percentage of children who were TD according to the gold standard but who were identified as LI by the language sample measures, is summarized in the upper right quadrant (29%). Conversely, the false negative rate, which documents the percentage of children who were identified as being LI according to the gold standard but who were identified as TD according to the language sample measures, is summarized in the lower left quadrant (15%).

Because there was a wide range of ages in this sample, subsequent discriminant function analyses were completed that accounted for the variance due to age. Separate linear regression analyses were completed for each of the language sample measures using chronological age to predict the language measure. The residuals from each of these regression equations were saved as a separate variable in SPSS. These residuals were entered as predictors into the discriminant function analysis, which accounted for the variability due to age yet maintained the variability specific to the language sample measures. The sensitivity and specificity data are summarized in the bottom of Table 4. After controlling for age, the sensitivity increased from 69% to 78%, and the specificity increased from 84% to 85%.

To test whether there were differences in sensitivity and specificity across age groups, a final set of discriminant function analyses was completed with the three age groups described earlier (preschool/Kindergarten age, elementary school age, and middle school age). Again, the residuals from the linear regressions were entered as the predictor variables. As seen in Table 5, the sensitivity and specificity values were the highest for the youngest group of children. There were slight decreases in sensitivity and specificity for the two older groups of children. However, sensitivity remained at 77% for the oldest group of children, demonstrating that a high percentage of the children with LI were being correctly identified by the language sample measures.

### Differences in Language Sample Measures Across Groups

After documenting that the language sample measures as a group were effective in distinguishing between children with LI and TD children, we were interested in testing which variables were significantly different between groups across the age ranges. A series of

<table>
<thead>
<tr>
<th>Diagnosis based on treatment status</th>
<th>LI</th>
<th>TD</th>
</tr>
</thead>
<tbody>
<tr>
<td>All children: No control for age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI</td>
<td>69%</td>
<td>29%</td>
</tr>
<tr>
<td>TD</td>
<td>15%</td>
<td>84%</td>
</tr>
<tr>
<td>All children: Controlling for age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LI</td>
<td>78%</td>
<td>22%</td>
</tr>
<tr>
<td>TD</td>
<td>13%</td>
<td>85%</td>
</tr>
</tbody>
</table>

*Note. LSA = language sample analysis.*

<table>
<thead>
<tr>
<th>Diagnosis based on treatment status</th>
<th>Diagnosis based on LSA measures</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosis based on LSA measures</td>
<td>LI</td>
</tr>
<tr>
<td>3:0–5:11</td>
<td>87%</td>
</tr>
<tr>
<td>LI</td>
<td>13%</td>
</tr>
<tr>
<td>6:0–9:11</td>
<td>80%</td>
</tr>
<tr>
<td>LI</td>
<td>14%</td>
</tr>
<tr>
<td>10:0–13:6</td>
<td>77%</td>
</tr>
<tr>
<td>LI</td>
<td>19%</td>
</tr>
</tbody>
</table>

Table 5. Summary of sensitivity and specificity data broken down into three age groups.

90 LANGUAGE, SPEECH, AND HEARING SERVICES IN SCHOOLS • Vol. 41 • 84–95 • January 2010
univariate analysis of variance (ANOVA) tests were completed using each of the language sample measures as the dependent variable and group (LI vs. TD) as the between-subjects variable. These analyses were completed for the three age groups: preschool/kindergarten age, elementary school age, and middle school age. Because there were multiple comparisons completed, the significance value was reduced to \( p \leq .01 \). In addition, eta squared \((\eta^2)\) values were generated for each of the analyses as an estimate of the effect size, which documents the amount of explained variance in a dependent variable as a function of group. In the present analysis, large effect sizes demonstrated that children’s language status (LI vs. TD) explained a substantial proportion of the variance in the variable of interest (e.g., MLU\(_m\)). Because we were primarily interested in the degree of difference across the age groups, we presented the effect sizes and summary of the significance tests in Table 6.

Significant differences were observed between groups for both the grammar (MLU\(_m\)) and vocabulary (NDW) measures for each of the age groups. The largest differences were observed for the children in the youngest age group. Significant differences were also observed for the productivity measure (WPM). Effect sizes for WPM were again strongest for the preschool/kindergarten-age children, but were stronger for the middle school-age children than for the elementary-age children. For the discourse and formulation measures (i.e., mean turn length, percentage of mazes, and between-utterance pauses), mean turn length was the only measure that differed significantly for the youngest group. There were also significant group differences for mean turn length in the elementary school-age children, but this difference had a smaller effect size than that observed for the preschool/kindergarten-age children. Between-utterance pauses also differed significantly for the elementary-age group. No significant differences were observed for the discourse and formulation measures for the middle school-age children. All omissions were significantly different across each of the age groups. Omissions of obligatory words had larger effect sizes than omitted bound morphemes for each of the age groups. Effect sizes were the largest for the youngest group, but were relatively comparable across the two older groups of children. Significant differences were observed for the word and utterance errors, with the exception of word errors for the youngest group. Contrary to nearly every other measure, effect sizes for the word and utterance errors were highest for the oldest group and gradually decreased with decreasing age.

### Table 6. Effect sizes \((\eta^2)\) for group differences (LI vs. TD) for 10 language sample measures (nonsignificant differences are flagged)

<table>
<thead>
<tr>
<th>Language sample measure</th>
<th>3;0–5;11</th>
<th>6;0–9;11</th>
<th>10;0–13;6</th>
</tr>
</thead>
<tbody>
<tr>
<td>MLU(_m)</td>
<td>.27</td>
<td>.16</td>
<td>.09</td>
</tr>
<tr>
<td>NDW</td>
<td>.26</td>
<td>.22</td>
<td>.11</td>
</tr>
<tr>
<td>WPM</td>
<td>.18</td>
<td>.10</td>
<td>.15</td>
</tr>
<tr>
<td>Mean turn length</td>
<td>.12</td>
<td>.05</td>
<td>.01(^{NS})</td>
</tr>
<tr>
<td>% Maze words</td>
<td>(.01^{NS})</td>
<td>(.01^{NS})</td>
<td>(.05^{NS})</td>
</tr>
<tr>
<td>Between-utterances pauses</td>
<td>(.01^{NS})</td>
<td>(.05^{NS})</td>
<td>(.05^{NS})</td>
</tr>
<tr>
<td>Omitted words</td>
<td>.30</td>
<td>.16</td>
<td>.20</td>
</tr>
<tr>
<td>Omitted bound morphemes</td>
<td>.15</td>
<td>.10</td>
<td>.09</td>
</tr>
<tr>
<td>Word errors</td>
<td>(.04^{NS})</td>
<td>(.06)</td>
<td>(.14)</td>
</tr>
<tr>
<td>Utterance errors</td>
<td>.06</td>
<td>.09</td>
<td>.10</td>
</tr>
</tbody>
</table>

\(^{NS}\)Not significant at \( p \leq .01 \).

## DISCUSSION

### Clinical Utility of LSA Databases

The utility of the SALT databases in clinical decision making has been discussed in this article and endorsed by a number of clinicians/scholars (Gillam & Gorman 2004; Johnston, 2006; Paul, 2007) as effective in describing impaired language performance in children. The databases provide a broad range of measures of vocabulary, syntax, semantics, discourse, rate, and fluency that can identify strengths and weaknesses for clinical populations, which ultimately provides insight into areas requiring more exploration and possible intervention targets. The variety of elicitation contexts used in the SALT databases provides the performance expectations for typical children in a number of speaking conditions: conversation, narrative, and story retell. In the past, many clinicians have considered the language sample process to require too much time and to be too difficult to interpret without specific comparison data. The SALT Software Project has also worked to increase the usability of the language sample process by providing tutorials to aid in skill acquisition and through the development of many tools to reduce transcription time. Our research has documented that relatively short samples can provide robust and reliable data (Heilmann et al., 2008; Miller et al., 2006). For example, native Spanish-speaking children learning English as a second language provided story retells that were 35–65 utterances long, and measures from these samples significantly predicted reading ability in both English and Spanish (Miller et al., 2006). Transcribing these story retells took approximately 45 min, including coding for two additional measures (i.e., the subordination index and the narrative scoring scheme). The time commitment for completing analyses of children’s language use is comparable to most traditional assessment procedures yet has the added benefit of documenting multiple facets of children’s language use that cannot be captured using standardized assessments alone.

The SALT databases provide valuable insight into the language production abilities of typical children. Initial analyses of the conversational database, for example, revealed that children produce fewer questions as they get older, suggesting that not all language features increase with age (Leadholm & Miller, 1992). We cannot interpret that children lose the ability to ask questions but, rather, that a conversational context will not elicit as many questions from older children. This finding reminds us that missing data do not equal absent knowledge. If we want to explore children’s question asking, we need to develop contexts that require children to ask them. Another example has to do with mazes (i.e., reduplications and reformulations): Typical children produce more mazes in narration than conversation and more mazes as they attempt longer utterances in both contexts. We had assumed that with increasing age, children would produce fewer mazes and become more fluent. This appears not to be the case (Leadholm & Miller, 1992). Although some measures are not sensitive to developmental changes in the normative databases, they may still have clinical implications for describing intrusive behaviors such as the frequency of repetitions and revisions, pauses, and errors.

### Using Databases to Identify Children With LI

The power of the databases for answering research questions was demonstrated with our example exploring impaired language
We examined the utility of the SALT databases to address three longstanding questions in the literature: (a) Can LSA measures distinguish typical from impaired language? (b) Do measures from conversational language samples identify preschool- and school-age children equally well? and (c) What measures distinguish typical from atypical performance? We used a broad age range and a large number of participants who had been identified as TD and LI. The initial analyses revealed that a collection of language sample measures identified 84% of the TD participants as typical and 69% of the children with LI as having true impairments. When the analyses were repeated controlling for age, 78% of the participants with LI were accurately identified, whereas the accuracy of identifying TD participants remained relatively constant at 85%. The series of ANOVAs that followed the sensitivity and specificity analyses revealed that group membership (LI vs. TD) accounted for a moderate amount of the variability in the majority of the LSA measures. Vocabulary, grammar, productivity, and discourse/formulation measures were significantly different within each of the age groups evaluated in the present study. When the analyses were broken down into three age groups, the sensitivity and specificity analyses revealed that the language sample measures were most effective in identifying the youngest group of children. However, the language sample measures were still effective in identifying children with LI throughout the school-age years.

In this study, the majority of the sensitivity and specificity estimates were in the “fair” range according to the commonly cited criteria established by Plante and Vance (1994), which stated that sensitivity and specificity values between 90% – 100% are considered “good,” values between 80% – 89% are “fair,” and estimates below 80% are less than adequate. In the present study, sensitivity for the middle school-age children was the only estimate that fell below 80%; all other sensitivity and specificity values were between 80% – 89%. Although these data may suggest that the measurement properties of conversational language samples are fair at best, the results from the present study are quite consistent with other published sensitivity and specificity analyses.

Spaulding, Plante, and Farinella (2006) reviewed the sensitivity and specificity data for standardized tests from nine published test manuals and from four studies that independently analyzed the properties of standardized tests (Gray, Plante, Vance, & Henrichsen, 1999; Merrell & Plante, 1997; Perona, Plante, & Vance, 2005; Plante & Vance, 1994). A total of 18 analyses were completed on 16 separate standardized tests. Spaulding et al. identified that 11 of the 18 analyses reported sensitivity and/or specificity estimates below 80%, four of the analyses had sensitivity and specificity values in the fair range (80%–89%), and only three tests had both sensitivity and specificity values in the good range (>90%). One may assume that the three tests that demonstrated good sensitivity and specificity should be used most often to identify children with LI. However, these tests look at a very limited skill set (i.e., expressive morphosyntax), and the analyses were generated using a very constrained and homogeneous group of children with LI (i.e., preschool-age children who performed poorly on an additional test of morphosyntax). Merrell and Plante documented good sensitivity and specificity values for the Test for Examining Expressive Morphology (TEEM; Shipley, Stone, & Sue, 1983) and the Patterned Elaboration of Syntax Test (PEST; Young & Perachio, 1993), two tests that focus entirely on expressive morphosyntax. The participants in this study were preschool-age children who scored at least 3 SDs below the mean on the Structured Photographic Expressive Language Test: Second Edition (SPELT–2; Werner & Kresheck, 1983), another test of expressive morphosyntax. Perona et al. documented good sensitivity and specificity levels for the SPELT–3 (Dawson, Stout, & Eyer, 2003) on a group of preschool-age children who had standard scores <75 on the TEEM. Because these tests were limited to the analysis of expressive morphosyntax in preschool-age children who performed poorly on other tests of morphosyntax, the results cannot be generalized to the heterogeneous group of children with LI. Furthermore, these tests of morphosyntax have modest utility in assisting with the developing treatment goals.

Sensitivity and specificity analyses are only two of many important measurement properties to consider when choosing the most appropriate assessment tools. Clinicians must also consider the validity of the task, potential for introducing test bias, usefulness in generating treatment goals, and ability to use the measure to monitor progress and document outcomes. Rather than focus on identifying the single assessment tool that provides the strongest sensitivity and specificity estimates, a better approach may be to identify the benefits and limitations of a range of assessment tasks and develop assessment protocols that facilitate accurate and unbiased assessment of children with LI. The present study has demonstrated that LSA and analyses using the SALT databases can be valuable steps in an effective assessment protocol. The results reported here may serve as preliminary data to justify examination of other groups of impaired speakers and alternative definitions of LI as well as further evaluation of the relative importance of the different LSA measures in predicting group membership.

Sensitivity and specificity analyses can also assist in identifying the properties of a language measure across different groups of children. This study revealed that measures from conversational language samples were most effective in distinguishing the children in the youngest group. Previous studies have documented that measures from conversational samples are robust during the preschool years (Leadholm & Miller, 1992) and are effective in classifying children based on their language status (Aram et al., 1993; Bedore & Leonard, 1998; Dunn et al., 1996; Fletcher & Peters, 1984; Gavin et al., 1993). There are very few studies in the literature that have compared the conversational language skills of older children with LI to their TD peers. Conversational samples were thought to be best suited to younger children, with older students requiring more demanding contexts to reveal their difficulties. This project documented that conversational samples are sensitive to differences in group status in older children. In addition, these results motivate additional research on children in more demanding speaking contexts that are consistent with school curriculum expectations, such as narration and exposition. Most state curriculum standards require high levels of oral language performance that parallel the written language curriculum. Failure to meet such standards may provide other criteria for classifying LI.

It is important to keep in mind that the children with LI in this study were receiving services or were referred for service. They are not necessarily the same children who have participated in studies of children with LI found in the literature. Differences between groups rest in part on the definition of LI and, more specifically, the gold standard for identification. A careful examination of the research literature over the past 10 years reveals that we do not employ a consistent definition of LI across studies (Miller & Fletcher, 2005). The study reported here used a clinical definition of impairment, and the resultant sample would not meet the formal definition of LI in every case. However, school personnel and parents agreed that special speech and language services were justified.
The results of this project motivate a reconsideration of what constitutes LI. The SALT databases provide expectations for naturalistic speaking tasks, conversation, and narration. These expectations based on typical speakers allow us to consider alternative models that go beyond the early school years and consider the impact of speaking condition on performance.

**Clinicians as Researchers**

The data used in this study were collected from practicing school-based clinicians. In fact, the majority of the SALT databases contain samples that were collected by practicing SLPs. This research model has assisted with the development of clinically feasible protocols that can be accurately and efficiently completed in a clinical setting to create large databases of typical performance. Recently, a new database was created for children 12–15 years of age using an expository sampling context (Malone et al., 2008). The project was initiated by an SLP from a school district outside of Milwaukee, WI who had provided services to students in middle and high schools. This investigation was motivated by the lack of good evaluation procedures that effectively met state guidelines for exposition at the eighth and twelfth grades, which require that students give a speech and write about how something works (e.g., how a bill becomes law). The SALT group from the Madison schools and the SALT software staff met with this SLP to develop a protocol for a collaborative project to develop a new database. The first task was to define the age range, estimate the required number of participants, and develop the elicitation protocol. After completing the protocol, approval was obtained from the institutional review boards from the University of Wisconsin and each of the participating school districts. Each district assisted in identifying potential participants from class rosters and mailed consent letters to parents. After the parents returned the signed letters to the SALT software group, a representative sample was selected. Twenty-eight SLPs from cooperating districts agreed to collect two to five samples each during noninstructional time. By the end of the spring term, 87 samples were collected by district SLPs and were transcribed by SALT staff. The samples were analyzed to evaluate consistency of participant performance and to compare the 13-year-olds to existing 13-year-olds in the conversation and narrative databases, which revealed that the expository context facilitated substantially more complex language use than both conversational and narrative elicitation procedures, a finding similar to Nippold, Hesketh, Duthie, and Mansfield (2005). Through development of a new database and use of the existing SALT databases, these school-based clinicians were instrumental in answering basic questions about the effects of elicitation contexts and assisted in the development of a new database that will help clinicians who are working with adolescent children.

**Use of SALT Databases for Research**

The data from the present study motivate researchers to rethink their criteria for LI. In studies of children with impaired language, researchers may consider performance on the language sample task relative to the SALT database as part of the inclusionary criteria and definition of LI. Use of language sample criteria for LI would be particularly useful for nonmainstream populations for whom standard-ized tests are often biased. In studies that choose to use traditional means of identifying LI, researchers have the ability to compare children’s performance on language sample tasks to the SALT database and thereby further describe and/or subtype their sample of children with LI. Each of these analyses can be completed with the current version of SALT by having the participants complete the same protocol described in the documentation for the databases. After the samples have been transcribed, researchers can compare each participant’s performance to that of children in the database and summarize his or her performance for each measure of interest. The researcher can establish criteria for grouping or subgrouping (e.g., –1 SD on MLUm, NDW, and/or WPM) or describe the performance of the participants as a group (e.g., 90% of the children with LI as determined by standardized testing were also at least 1 SD below the mean for MLUm compared with the SALT database). Such work would provide greater description of children with LI and extend our understanding of the underlying nature of the impairment.

The SALT development team is currently updating the software to provide users with greater flexibility in accessing the databases in the upcoming research version of SALT, due out this year. In this next version, users will have the ability to generate data files that contain actual measures for individual participants. Such access will allow researchers to create batch files containing all of the transcripts that meet the specified criteria (e.g., 50 complete and intelligible utterances), reducing the need to complete analyses on each individual participant. This procedure will also allow researchers to obtain measures for each of the individual children in the database, rather than a summary of the group means and standard deviations. Having individual scores from children in the database will provide access to the variance structure for the different measures and facilitate meeting the assumptions of statistical analyses (e.g., ANOVA). Researchers will be able to perform more efficient and accurate matching of experimental groups and complete group significance tests to document the similarities and differences between their samples and the SALT database. These data will also allow researchers to make direct comparisons of their new databases to these existing databases in order to answer questions about measurement properties of language samples (e.g., picture description vs. child-initiated narratives from the database). Furthermore, new studies may test how language sample measures from a target population compare to measures from children in the database (e.g., measures from narrative retells produced by school-age children with LI compared to measures from the SALT narrative retell database). This new access to the databases can also be used for independent evaluations of the measurement properties of the samples and could extend our analyses of similarities and differences across ages and sampling contexts.

**Conclusion**

In the age of accountability and least-biased assessment, it is the responsibility of clinicians and researchers alike to critically evaluate their assessment practices. The present study provides evidence that LSA and comparisons to the SALT databases can be highly useful tools for documenting the performance of children with LI. By incorporating natural language data into our assessment procedures, we will continue to better understand the nature of LI and better serve our clients with language learning difficulties.

**ACKNOWLEDGMENTS**

The SALT databases and project described in this article would not have been possible without the assistance of numerous collaborators and funding
REFERENCES


Received July 3, 2008
Revision received December 22, 2008
Accepted February 18, 2009
DOI: 10.1044/0161-1461(2009/08-0075)

Contact author: John J. Heilmann, Department of Communication Sciences and Disorders, Mail Stop #668, East Carolina University, Greenville, NC 27858-4353. E-mail: heilmannj@ecu.edu.

Heilmann et al.: Using Language Sample Databases 95


