

VALIDITY OF SEVEN SYNTACTIC ANALYSES PERFORMED
BY THE COMPUTERIZED PROFILING SOFTWARE

by

Stacy Lynn Minch

A thesis submitted to the faculty of

Brigham Young University

in partial fulfillment of the requirements for the degree of

Master of Science

Department of Communication Disorders

Brigham Young University

August 2009

BRIGHAM YOUNG UNIVERSITY

GRADUATE COMMITTEE APPROVAL

of a thesis submitted by

Stacy Lynn Minch

This thesis has been read by each member of the following graduate committee and by majority vote has been found to be satisfactory.

Date

Ron W. Channell, Chair

Date

Martin Fujiki

Date

Shawn Nissen

BRIGHAM YOUNG UNIVERSITY

As chair of the candidate's graduate committee, I have read the thesis of Stacy Lynn Minch in its final form and have found that (1) its format, citations, and bibliographical style are consistent and acceptable and fulfill university and department style requirements; (2) its illustrative materials including figures, tables, and charts are in place; and (3) the final manuscript is satisfactory to the graduate committee and is ready for submission to the university library.

Date

Ron W. Channell
Chair, Graduate Committee

Accepted for the Department

Date

Ron W. Channell
Graduate Coordinator

Accepted for the College

Date

K. Richard Young
Dean, David O. McKay School of Education

ABSTRACT

VALIDITY OF SEVEN SYNTACTIC ANALYSES PERFORMED BY THE COMPUTERIZED PROFILING SOFTWARE

Stacy Lynn Minch

Department of Communication Disorders

Master of Science

The Computerized Profiling (CP) software extracts several quantitative measures from a transcribed sample of a client's language. These analyses include the Mean Length of Utterance in Words (MLU-W) and in Morphemes (MLU-M), the Mean Syntactic Length (MSL), the Syntactic Complexity Score (SCS), Developmental Sentence Scoring (DSS), the Index of Productive Syntax (IPSyn), and the Picture-Elicited Screening Procedure for LARSP (PSL). The validity of these measures was examined by comparing them to the number of finite nominal, adverbial, and relative clauses contained in samples from 54 first-, 48 third-, and 48 fifth-grade students and 24 young adults. The DSS and SCS correlated highly with the frequency of complex constructions; MLU-W, MLU-M, and MSL correlated moderately; and IPSyn and PSL correlated minimally at best.

ACKNOWLEDGMENTS

Many people deserve my thanks for their roles in the completion of this thesis. First, I would like to thank my mother, for giving me the option of either taking a summer job in fast food, or enrolling in online university classes during my high school years. Without her commitment to education, I certainly would not be where (or who) I am today. Second, thank you to my husband Steven. He was there for me on the first day of graduate school, and he will be there for the last (not to mention everything in-between). Third, thank you to all my teachers, from elementary to high school, college, and graduate school, who believed in me, let me know it, and encouraged me to live up to my potential. Fourth, thank you to the Walker family, for guiding me to one of my life's passions, and teaching me so many lessons about life along the way. And finally, a big thank you to Dr. Ron Channell, for helping me to break this project into manageable steps; and for never being too busy to offer frequent guidance and advice throughout this entire process. Truly, I could not have completed this work without any one of you.

Table of Contents

	Page
List of Tables and Figures.....	ix
List of Appendixes.....	x
Introduction.....	1
Review of Literature.....	4
Language Sample Analysis.....	5
Syntactic Complexity.....	6
Syntactic Productivity.....	7
Currently Available Measures of Complexity: An Overview.....	8
Length Measures.....	9
Syntactic Inventory.....	11
Weighted Measures.....	14
Automated Language Sample Analysis.....	19
Success of Automated Language Sample Analyses.....	24
Summary.....	27
Method.....	27
Participants.....	27
Materials.....	28
Procedure.....	29
Results.....	30

Length Measure Correlations.....	34
Syntactic Inventory Correlations	34
Weighted Measure Correlations	35
ANOVA and Post-Hoc Comparisons	36
Discussion.....	36
Weighted Measures.....	38
Length Measures.....	39
Inventory Analyses	40
Productivity.....	40
Conclusions.....	40

List of Tables and Figures

Table	Page
1. Pearson's r correlations among complexity measures for the combined samples.....	31
2. Productivity correlations for the combined samples.....	33
3. ANOVA between-group data for the first, third, fifth and adult subject groups ...	37

Figure

1. Significant correlations among complexity measures for the combined samples..	32
--	----

List of Appendixes

Appendix	Page
A. Steps for computing measures with CP	49
B. Individual group correlation tables for frequency	51
C. Individual group correlation figures for frequency.....	55
D. Individual group correlation tables for productivity.....	59

Introduction

Along with a variety of other results, the software package Computerized Profiling (CP; Long, Fey, & Channell, 2006) generates seven quantified syntactic measures of a client's language sample. These measures are the Mean Length of Utterance (MLU; Brown, 1973), calculated in both words and morphemes (MLU-W; MLU-M), Mean Syntactic Length (MSL; Klee & Fitzgerald, 1985), the Index of Productive Syntax (IPSyn; Scarborough, 1990), Developmental Sentence Scoring (DSS; Lee, 1974), and two methods of quantifying the results of a Language Assessment, Remediation, and Screening Profile (LARSP; Crystal, Fletcher, & Garman, 1989): a procedure called the Picture-Elicited Screening Procedure for LARSP (PSL; Ward & Fisher, 1990) and the Syntactic Complexity Score for LARSP (SCS; Blake, Quartaro, & Onorati, 1993). Though all of these measures yield a quantitative description which might characterize syntactic development, no study has as yet compared these measures to the actual presence of complex syntactic constructions produced by a client in a language sample. Such a comparison would give useful insight into the construct validity of each of these measures. In addition, comparing the resulting scores and correlations with each other would give further insight into the concurrent validity of these measures—demonstrating whether analyses designed to assess syntactic development would yield similar results.

These seven measures represent three different approaches to quantifying the syntactic performance manifest in a language sample. The MLU measures and MSL focus on describing the length of utterances, relying on the known positive relationship between age and utterance length. MLU calculations are perhaps the best known quantitative indices for describing language samples, and have been studied extensively.

MSL is a variant of MLU, described by Klee and Fitzgerald (1985), in which non-syntactic (single word) utterances are discarded in computation of the score. The notion of excluding certain utterances from the calculation of the MLU was also investigated by Johnston (2001), who suggested that such selectivity might remove, to some extent, the confounding influences of the pragmatics of the sampling situation from the measurement of syntactic ability.

Rather than focusing on utterance length, IPSyn and PSL quantify syntactic development by tabulating the number of different syntactic construction types used by the client. In IPSyn, point values are assigned to 56 grammatical forms, and a score is tabulated based on the number of different forms the sample contains. In a similar fashion, PSL quantifies syntax by noting the number of forms present at each LARSP level, and awarding points for the number of different constructions present.

A third approach to the quantification of syntactic development is taken by DSS and SCS. In general, this approach gives different scores to utterances based on the occurrence of specific syntactic constructions or the number of grammatical sections of an utterance. In DSS, scoring is based on the occurrence of developmentally-weighted syntactic forms. Unlike IPSyn and PSL, each occurrence of a form is awarded the assigned number of points; so that a child can receive points multiple times for the same construction. Also in contrast to IPSyn and PSL, the SCS quantifies complexity by identifying the clausal units in a sample and awarding points for each clausal unit used in an utterance.

Though a variety of approaches to measurement and quantification of syntactic development are possible, the most important question regarding any language measure is

its validity. Studies have compared some of these procedures. Kemper, Rice, and Chen (1995) compared the validity of six syntactic complexity measures, including DSS, Developmental Level (DLevel), IPSyn, Propositional Density (PropD), MLU, and Mean Clauses per Utterance (MCU). Fifty-two child language samples, collected from children ranging in age from five to ten years, were analyzed with each of the preceding measures, and growth trajectories were established for each measure. Results of the study indicated that DSS, DLevel, and MCU appear to measure the same underlying aspect of development, based on their similar growth trajectories. These measures showed rapid growth curves from four to six years of age, and eventually leveled off around eight years. MLU did not follow the same growth curve as these measures, but did correlate with DSS, MCU, and DLevel. Based on their growth trajectories, IPSyn and PropD did not appear to be sensitive to language changes in the ages examined in this study.

Rice, Redmond, and Hoffman (2006) similarly examined the validity of MLU using growth trajectories. A total of 124 language samples were analyzed using MLU, MLU in words, IPSyn, and DSS; and growth trajectories were established for each measure. The authors found significant correlation between MLU and the other three measures, indicating strong concurrent validity. MLU also showed stable growth trajectories up to age 10. Both of the growth trajectory studies yielded evidence of validity by documenting improvement in scores with age; affirming the relationship between age and improved grammatical development.

Another approach to studying the validity of these measures would be through examination of how they relate to the production of complex syntactic constructions in language samples. For example, Nippold, Mansfield, Billow, and Tomblin (2008)

examined the frequency of finite noun, relative, and adverbial clauses in language samples collected under different conditions. If these automated measures were shown to be sensitive to the frequency or the productivity (Bloom & Lahey, 1978) of those complex constructions which are indicative of syntactic development, the clinical use of these measures in the assessment process would be supported. In contrast, a lack of such a relationship might raise questions regarding the clinical use of these measures.

Accordingly, the present study compares the seven measures extracted by CP software from a language sample to the frequency and productivity of complex syntactic constructions in the sample; and also compares the seven measures' scores and complexity correlations with each other. A comparison of the frequency and productivity of complex syntactic structures to the scores from the MLU-W, MLU-M, MSL, SCS, DSS, IPSyn, and PSL for school-age child and adult samples yields insight into the construct validity of these measures, especially with regard to the higher-level language used by school-age children and adults. In addition, comparing the CP scores and complexity correlations with each other demonstrates their concurrence with similar existing measures, offering further information regarding the concurrent validity of these automated analyses. This validity information, in turn, provides greater insight to the clinician in employing the use of automated language sample analyses.

Review of Literature

The use of automated language sample analysis depends upon the collection of client language samples. Several manual protocols are currently available for language sample analysis. Automated programs, in most cases, consist of computerized versions of manual language sample analysis protocols. A discussion of language sampling practices; manual syntactic analyses; and computerized analyses follows.

Language Sample Analysis

Language sample analysis is a common clinical assessment tool, used to evaluate semantic, pragmatic, and syntactic abilities in the conversational language of children and adults. Analysis of language samples, “provides something that traditional tests do not; namely, the opportunity to examine the child’s linguistic system as it is put to use in communicative interaction” (Kemp & Klee, 1997, p. 161). Depending upon sample conditions and length; language sample analysis can provide the clinician with a picture of a client’s language that is commensurate with his or her natural abilities for language in everyday contexts.

Studies have documented the prevalence of language sampling and analysis practices among clinicians in the United States. Hux, Morris-Friehe, and Sanger (1993) assessed the language sampling practices of 239 speech-language pathologists, working in nine Midwestern states. Participating clinicians filled out and returned a questionnaire regarding language sampling attitudes and practices. On the subject of analysis, 49% of clinicians surveyed reported using non-standardized measures for analyzing language samples; and 31% reported using manual DSS as their preferred method of sample analysis. No other formal measure was strongly represented in the survey. Only 3% of clinicians reported using computerized software to aid language sample analysis.

Kemp and Klee (1997) surveyed language sampling practices of 253 ASHA-certified clinicians practicing in preschool settings in the United States. Of the clinicians surveyed, 85% reported using language sample analysis as part of assessment and treatment procedures. Those who reported not using language sample analysis (15%) reported a *lack of time* as being the paramount reason for not collecting language

samples. Forty-eight percent of the clinicians who used language sample analysis reported using informal and non-standardized forms of analysis. Frequently, clinicians reported that MLU was their non-standardized analysis method of choice. DSS was the second-most popular method of analysis (next to informal analysis), with 35% of clinicians reporting this as their preferred method of language sample analysis. Only 8% of clinicians in this study reported using computer programs for language sample analysis.

Syntactic Complexity

Language sample analysis is commonly used to assess a client's syntactic development. Syntax is a characteristic of the language domain of form, which also includes morphology and phonology. Syntax refers to the manner in which words are used to convey the speaker's meaning, according to the grammatical rules of a language.

Acquisition of syntax begins with the most simple grammatical forms, and gradually increases to include more complex grammatical structures. As children develop, clauses and phrases of increasing complexity are mastered and incorporated into conversational discourse (Diessel & Tomasello, 2000). Generally, the term *complex syntax* refers to utterances that contain multiple embedded or conjoined clauses within one sentence. There are three kinds of complex clauses, which can occur as finite or non-finite forms, and they include: noun clauses, adverbial clauses, and relative clauses.

The form of verb (finite or infinite) contained in a clause determines whether the entire clause is classified as finite or infinite. Non-finite verbs are verbs which are not inflected for person, tense, or number (i.e.: Mary has *to go*; The boys have *to go*). Finite verbs are verbs which are altered according to person, tense, and number (i.e.: Mary will

go; He *goes* every day; I *went* yesterday). A brief explanation and example of each complex form follows.

An adverbial clause is a clause (containing a subject and a verb) that post-modifies a verb in a sentence. Examples:

Finite adverbial clause: I went to the store *after I played with Hanna*.

Non-finite adverbial clause: I'm saving *to buy a toy*.

A noun clause post-modifies a noun. It consists of a verb and a predicate; and functions as a noun in the overall structure of a sentence. Examples:

Finite noun clause: Amy said *she can't play today*.

Non-finite noun clause: Mom told me *to clean my room*.

Relative clauses also post-modify nouns. These clauses contain a relative pronoun (who, that, etc.), which acts as the noun component of the clause, and a verb. Examples:

Finite relative clause: I know a man *who can stand on his head*.

Non-finite relative clause: There's this kid *that likes to eat ants*.

These complex structures are generally found more abundantly in the language of older school-age children and adults.

Syntactic Productivity

Children produce complex syntactic forms with some inconsistency before they are able to use them consistently. When a child is able to consistently produce a certain form, he or she is considered to be *productive* for that form. Productivity does not necessarily refer to the mastery of a form, but rather, a person's emerging capacity to master the form. Productivity is important to language sample analysis because it offers

the clinician insight into whether the client is able to produce a form generatively, or whether the production of the form occurred as a byproduct of context.

Bloom and Lahey (1978) describe two different methods for measuring productivity: a frequency count criterion, and a proportional analysis criterion. In a *proportional analysis* criterion method, the occurrence of the language form in obligatory contexts is calculated. The number of actual occurrences of the form is divided by the number of possible contexts within the sample, and a percentage is obtained.

A *frequency count* criterion for productivity measurement tallies only the occurrences of a form. When a certain number of occurrences is noted, the examiner assumes that the individual is capable of constructing the form in question; and that the form did not surface as a fluke of context. Bloom and Lahey (1978) describe an arbitrary criterion of five occurrences for use in determining syntactic productivity. This criterion is based on the analysis of a language sample consisting of at least 500 utterances. When using language samples shorter than 500 utterances, Bloom and Lahey suggest the use of a four-occurrence criterion for determining productivity. However, both the four- and five-occurrence frequency criteria suggested by Bloom and Lahey are arbitrary guidelines; and frequency criteria may be adjusted at the examiner's discretion in accordance with the length of the sample, and the conditions under which the sample was collected.

Currently Available Measures of Complexity: An Overview

Currently, many syntactic analysis procedures and protocols are available to aid clinicians in the syntactic analysis process. Available measures of complexity generally describe syntax in one of three ways: (a) quantifying the *length* of utterances; (b) tallying,

and sometimes scoring, the number of different kinds of syntactic forms used in a sample (*syntactic inventory*); or (c) awarding points to different structures, with later-occurring forms receiving higher point scores (*weighted*).

A review of some of the most commonly-used analysis procedures follows. Of the complexity measures that will be covered in this review, the MLU variants and the MSL measure length; LARSP, IpSyn and PSL examine the range of syntactic repertoire; and DSS and SCS quantify developmental complexity based on weighted scores.

Length Measures

Several currently-available measures, including MLU (with its variants) and MSL, measure the length of individual utterances within a sample; and compute an average utterance length for the sample.

MLU. Perhaps the most widespread and popular (Hux et al., 1993; Kemp & Klee, 1997) informal method of language sample analysis is the calculation of MLU (Brown, 1973). MLU has been widely used by practicing clinicians since its introduction in the early 1970s. There are two general methods of calculating the MLU: MLU in words (MLU-W), and MLU in morphemes (MLU-M). MLU-M calculation attempts to quantify language complexity based on the average number of morphemes present in each utterance; while MLU-W consists of an average of the number of words present in each utterance. To calculate both MLU-W and MLU-M, clinicians must divide language samples into separate utterances. The clinician finds the total number of morphemes or words present in the sample, and then divides the total number of morphemes or words by the total number of utterances. This gives an approximation of the child's average utterance length in morphemes or words.

The validity and reliability of MLU as a measure of syntactic development has been a long-standing topic of debate in child language literature. Miller and Chapman (1981) studied the correlation between age and MLU in young children. Samples were collected from 123 pre-school and early school-age children in Madison, Wisconsin, ranging in age from 17 to 59 months. Samples were collected either at home or in a clinic room during an unstructured free-play session with the children and their mothers. For data analysis, children were grouped into three-month age intervals, and the correlation between age and MLU score was calculated. MLU was found to have a significant correlation with age, $r = .88$. In addition, regression analyses were completed to determine the predictive value of MLU and age. Miller and Chapman concluded that MLU correlates highly with age, and that age and MLU can be used to reliably predict each other.

Although MLU does correlate reliably with chronological age (Miller & Chapman, 1981; Rondal, 1987), some researchers have questioned the measure's ability to accurately quantify language complexity. Studies have documented a solid correlation between MLU and the order of grammatical morpheme acquisition (de Villiers & de Villiers, 1973; Klee & Fitzgerald, 1985). One study has documented a correlation between MLU and other formal and informal measures of complexity (Rice et al., 2006). However, most have failed to delineate a solid relationship between the overall complexity of a language sample and the MLU score.

Mean Length of Utterance-2 (MLU-2). Johnston (2001) proposed an alternate method of MLU calculation termed MLU-2. In MLU-2, all single-word *yes* and *no* responses, imitative utterances, and elliptical question responses are removed from the

MLU calculation. Johnston implemented this method of analysis with 47 language samples collected from typically-developing and language-impaired preschoolers. The children's original MLU scores ranged from 2.0 to 6.5. Following MLU-2 calculation, Johnston found that the new scores were an average of 18% higher, with a range of individual score increases from 3% to 49%. Johnston noted that this method of MLU calculation may be attractive for practicing clinicians because of its apparent increased sensitivity to syntactic complexity.

MSL. Klee and Fitzgerald (1985) introduced a variant of MLU termed MSL. Mean syntactic length is computed in the same manner as MLU, with the exception that in MSL, single-morpheme utterances are discarded from the analysis. Klee and Fitzgerald predicted that MSL may correlate more closely with chronological age than the traditional MLU measure. Klee and Fitzgerald collected language samples from 18 children, and each sample was scored using MLU, MSL, and LARSP. Pearson correlations performed for age and MLU indicated a chance correlation between chronological age and MLU ($r = .26$). Mean syntactic length was found to correlate significantly with age in this particular study ($r = .52$). However, because only language samples from two- and three-year old children were used in this study, it cannot be assumed that the same MSL-age correlation would exist at other ages until further research delineates the relationship.

Syntactic Inventory

Some syntactic analyses assess language production by taking an inventory of the different structures present in a language sample. Currently available syntactic analyses

include LARSP (Crystal, Fletcher, & Garman, 1976), PSL (Ward & Fisher, 1990), and IPSyn (Scarborough, 1990).

LARSP. Crystal et al. (1976) developed a widely-used, descriptive procedure for analyzing the language samples of children. The assessment's protocol consists of a graphic representation of grammatical structures, with boxes for different structures on the word, phrase, clause, and sentence level. The clinician analyzes a 30-minute language sample to determine frequency count of occurrence of various structures on these levels. Each time a specific structure is noted, it is tallied in the corresponding box on the graph. The framework of LARSP is based on seven developmental levels of syntax acquisition; and therefore, it gives a general description of a developmental syntactic level. However, LARSP does not attempt to give a quantified score to a language sample, and as such, is considered a descriptive or qualitative measure of syntactic development.

PSL. Ward and Fisher (1990) developed a syntactic screening procedure which attempts to quantify the LARSP protocol, using a language sample collected through the presentation of 10 language-elicitation picture cards. While the scoring protocol was originally developed for use with picture screening cards, the scoring protocol may also be applied to naturalistic language samples entered into the automated CP program. PSL scoring is conducted as follows:

1. Each utterance is scored using LARSP, and each structure is marked on the LARSP scoring sheet.
2. The number of marked LARSP structures is counted at each LARSP stage (one occurrence of a structure is enough to be counted).

3. The number of marked structures at each stage is multiplied by the stage number (i.e. two structures at Stage II would translate to $2 * 2 = 4$).
4. The PSL score is the total of all the scores obtained for each stage.

Ward and Fisher concluded that the PSL provides a quick and practical screening instrument of syntactic development for clinicians. However, the instrument's only test population consisted of two kindergarten-age classes. Accordingly, the authors cautioned that the PSL procedure may not be effective with other age groups; and suggested that six to seven years would likely be the instrument's upper age limit. Beyond Ward and Fisher's initial investigation, no other known studies on the procedure have been conducted.

IPSyn. IPSyn (Scarborough, 1990) is a measurement tool for evaluating the complexity of preschool language samples. IPSyn assigns point values to 56 morphological and syntactic forms. These forms fall into four subscales: (a) noun phrases, (b) verb phrases, (c) questions and negations, and (d) sentence structures. The first two occurrences of each morphological or syntactic form are awarded points, and a score is tallied by adding all the points awarded. IPSyn provides an overall score of syntactic complexity, as well as separate scores for each of the four subscales.

Scarborough (1990) assessed the concurrent validity, content validity, and reliability of IPSyn as part of the measure's development. Concurrent validity was assessed using a common complexity analysis measure, MLU-M. IPSyn correlated significantly with this measure, suggesting strong concurrent validity. Content validity was assessed using widely-accepted developmental scales to reference the 56 forms assessed in IPSyn. The forms assessed in IPSyn were found to be similar to existing

scales of normal syntactic development. Reliability between trained examiners was also found to be high for IPSyn. Scarborough concluded that IPSyn is a useful measure that is sensitive to complexity, but does not provide specific information about a child's mastery of (or productivity for) grammatical forms. However, the tool is useful in determining emergence of grammatical forms.

Weighted Measures

SCS. SCS, developed by Blake et al. (1993), is also designed to quantify LARSP-scored language samples. In SCS, the scoring procedure is as follows:

1. Single-word utterances are discarded.
2. Every occurrence of LARSP clausal units (including subject, verb, object, and complement units) is counted and given one point each.
 - a. Subject: a noun, a pronoun, or a noun phrase.
 - b. Object: a noun, a pronoun, or a noun phrase.
 - c. Verb: the main verb, as well as any auxiliary verbs, particles, and infinitives which share the main verb.
 - d. Complement: a prepositional phrase, a predicate adjective, a predicate noun or pronoun, or an adverb.
3. Subject, object, and verb clauses are collectively counted as one unit, while each complement is counted separately.
4. Total clausal unit points are counted to determine SCS score.

Blake et al. reported that scores obtained using the SCS scoring protocol for LARSP correlated positively with MLU and LARSP mean clausal stage; indicating concurrent validity for the measure. However, beyond the authors' initial investigation, no further

studies have been conducted to either confirm or refute the validity of SCS as a quantifiable measure of syntactic complexity in language samples.

DSS. DSS is a systematic, standardized assessment tool for quantifying language complexity in child language samples. DSS assesses language complexity based on typical developmental acquisition of grammatical structures. Grammatical forms are assigned numerical scores based on complexity; with earlier-occurring (less complex) structures receiving lower numerical scores than later-occurring forms (Lee & Canter, 1971).

A clinician must have considerable knowledge of English morphology and syntax in order to use DSS as a language sample analysis method. To complete DSS analysis, the clinician must first collect a language sample that is at least 50 utterances in length. Only utterances that are complete sentences are analyzed; incomplete, one-word, repeated, and echolalic utterances are omitted from the analysis. Run-on sentences connected by the conjunction *and* are parsed to form separate sentences.

Grammatical forms are grouped into eight categories by DSS: indefinite pronouns or noun modifiers, personal pronouns, main verbs, secondary verbs, negatives, conjunctions, interrogative reversals, and *wh*-questions. Each of the eight categories contains several form subtypes of the corresponding grammatical category. Subtypes are grouped according to the order of normal acquisition and are assigned a numerical point value, ranging from one to eight, which corresponds to developmental order. Each utterance is scored according to the forms present. One point is added to the utterance if it is complete and correct; incomplete and incorrect forms are not scored, but are given an *attempt mark*. To calculate the final DSS score, point values for each of the 50 utterances

are added together, and divided by the number of utterances. The final score may be compared to Lee's (1974) norms for standardized scoring.

Because DSS yields a standardized score (Lee, 1974), results of analysis may be reliably interpreted by clinicians across settings. The measure provides detailed explanations of how to code utterances, and training software has been developed to assist clinicians in learning how to score language samples according to DSS procedures (Hughes, Fey, Kertoy, & Nelson, 1994). In addition, the measure is low-cost due to its status as public domain information (Channell, 2003), making DSS an affordable, widespread, and user-friendly method of language sample analysis.

DSS has been found to be a valid and effective measure of syntactic complexity in child language samples. Following the release of the preliminary version of DSS (Lee & Canter, 1971), Leonard (1972) used DSS to distinguish deviated language from normal language development. Nine children with normal language and nine children with deviant language participated in the study. Fifty-utterance language samples were collected from each child during a story-retell task. The samples were transcribed and analyzed using DSS. Results of the study indicated that there were no significant differences between the two groups when compared qualitatively (number of children using a particular form). However, when compared quantitatively (number of times the form occurred), significant differences among groups were observed. Because DSS is sensitive to the frequency of occurrence of forms in language samples, Leonard concluded that DSS may be the most effective method of distinguishing even mild language deviance from normal language development.

As part of the finalized version of DSS published by Lee (1974), Koenigsknecht (1974) studied the relationship between chronological age and the complexity score provided by DSS analysis; as well as the difference in grammatical forms present at different ages. Language samples from 200 children, ages 2;6 (years; months) to 6;11 were analyzed with DSS, and the resulting scores were correlated with chronological age. Grammatical structures were correlated with DSS score. Grammatical structures present in the samples were discovered to be significantly different at different age levels, indicating that the measure is sensitive to changing grammatical structures at different levels of syntactic development. The study also reported significant differences in DSS scores among age groups, validating the measure's developmental sequence of increasing complexity acquisition.

Since the introduction of DSS in 1971, the reliability, validity, and practicality of the instrument have been thoroughly scrutinized in clinical research. Johnson and Tomblin (1975) studied the reliability of DSS with relation to language sample size. Fifty preschool (ages 4;8 to 5;8) child language samples were collected, and each sample parsed into five segments of five utterances each, resulting in a total of 25 utterances per sample. Each five-utterance segment was then coded and scored using the DSS analysis protocol. Applying the five-utterance language sample segment scores to estimated reliability statistics, Johnson and Tomblin calculated the expected reliability of DSS scoring for samples of differing length (5 to 250 utterances). The authors' analyses indicated that the recommended 50-utterance sample length (Lee & Canter, 1971) is not sufficient to produce a reliable language score. Rather, in order to overcome the protocol's standard error of measurement, a language sample coded with DSS must

consist of 175 or more utterances to achieve sufficient reliability. While a reliable score is attainable using DSS, Johnson and Tomblin contend that the collection, transcription, and analysis of 175-plus qualifying utterances may exceed the bounds of the practicing clinician's time restraints.

The validity of using nonstandardized language sample collection methods for use in a standardized instrument has been questioned by some critics of DSS. Kramer, James, and Saxman (1979) examined the differences in MLU and DSS scores for preschool language samples collected at home versus language samples collected in the clinic. The study reported that participating children scored significantly higher in measures of MLU for samples collected at home versus samples collected in the clinic. DSS age-equivalency scores tended to be slightly higher in home-collected versus clinic-collected samples. However, DSS scores for samples collected at home versus collected in the clinic did not differ significantly; indicating that, although nonstandardized collection methods could be considered a weakness for the instrument, DSS may not be as sensitive to changing collection conditions as other measures of complexity.

Another validity-related shortcoming of DSS relates to the measure's ability to accurately quantify the language behavior of a child. Many grammatical forms are disregarded by the DSS analysis protocol (Owens, 2004), making it impossible for the practicing clinician to use DSS as the only method of language sample analysis. Additionally, the quantification of an inherently qualitative behavior lends itself to misrepresentation of specific language ability.

Hughes, Fey, and Long (1992) reviewed literature pertaining to the validity of DSS. In addition to the established shortcomings regarding DSS' validity, such as

exclusion of many grammatical forms, questionable validity of developmental sequencing, nonhierarchical nature of analysis, and nongeneralizable norms for culturally and linguistically diverse children, the authors concluded that similar DSS scores can represent significantly different errors. While a language-impaired child may receive a score similar to a child two years younger than him, the language impaired child will likely exhibit errors in many more categories, and exhibit a narrower repertoire of grammatical forms than the typically developing child. The quantified DSS score would not reflect these differences.

Automated Language Sample Analysis

Several automated language sample analysis software programs have been developed to decrease the amount of time required of clinicians in obtaining detailed analyses of language. Some currently-available language sample analysis programs include *Systematic Analysis of Language Transcripts* (SALT; J. Miller & Chapman, 2000), *Computerized Language Analysis* (CLAN; MacWhinney, 2000), *Automated LARSP* (Bishop, 1984; Crystal, 1982), *DSS Computer Program* (Hixson, 1983), and *Computerized Profiling* (CP; Long et al., 2006).

SALT 2008. Miller and Chapman introduced a computerized language sample analysis program for analyzing child language samples, which, in the following years, has undergone several software revisions. The most current version of the program is titled *SALT 2008* (J. Miller, 2008) and is available in five different versions: English, Bilingual Spanish-English, Research, Instruction, and Student versions. The program may be purchased online for \$35 to \$495, depending upon the version. *SALT 2008* is formatted to run only on Windows-enabled computers.

SALT 2008 analyzes language samples based on eight different domains: transcript length, syntax/morphology, semantics, discourse, intelligibility, mazes and abandoned utterances, verbal facility and rate, and omissions and error codes. Language samples must be transcribed and specially coded to indicate presence of mazes, bound morphemes, errors, etc. (J. Miller, 2008). The sample may then be analyzed by the program, and results compared to standardized scores collected from a corpus of school-age language samples elicited and recorded in Madison, Wisconsin. *SALT 2008* uses MLU as its measure of syntactic development.

CLAN. MacWhinney (2000) developed an automated program for analyzing child language samples as part of the Child Language Data Exchange System (CHILDES) project, a government-funded project intended to make a large collection of language samples available for the purposes of linguistic research. The program operates on IBM-compatible personal computers, Macintosh computers, and UNIX machines. The program and its documentation may be downloaded at no charge.

Conti-Ramsden (1996) provided a user's review of CLAN, documenting the program's requirements and uses. Before entering a sample into a CLAN program, it must be transcribed according to Codes of Human Analysis of Transcripts (CHAT) conventions. CHAT uses keyboard symbols to code utterances according to the speaker, action, etc. Once the transcription is correctly coded, it can be run through one of over 30 CLAN programs. Each program conducts a different analysis of the sample, such as MLU, frequency count, word searches, co-occurrence analyses, interactional analyses, and many more.

Automated LARSP. Crystal et al. (1982; 1989) describe an automated language sample analysis program designed to assist the clinician in differentiating normal from abnormal language in children. The program is based on the LARSP language sample analysis protocol developed by Crystal et al. (1976). Bishop (1984) reviewed the automated program and offered practical guidelines for implementing its use in clinical settings. Manual LARSP analysis requires the clinician to have a thorough knowledge of grammar and syntax on the phrasal and clausal levels. The analysis method also requires that the clinician be able to frequently switch between clausal and phrasal elements during analysis. This process can be tedious and time-consuming, even for the clinician who is well-versed in the procedure.

The *Automated LARSP* computer program speeds LARSP analysis by automatically scoring phrasal and clausal elements according to LARSP scoring procedures. The program is able to score a sample by implementing the following procedure:

1. Utterances are broken up into words.
2. Words are labeled according to part of speech.
3. Each word is checked for points scored on the phrasal level, and a score is tallied.
4. Each sentence is parsed into phrases and the clinician is prompted to confirm the accuracy of computer-selected phrases.
5. Each phrase is checked for categories under which it might score points, and the points are tallied.

6. Each noun phrase is labeled either as a subject, object, complement, or indirect object.
7. Relationships between clauses are established (subordination, etc) and clinician is asked to check for accuracy of clausal relationships.
8. Each clause is checked for scored points at the transitional phrase-clause level, and points are tallied.
9. Each clause is checked for points scored at the clause level, and points are tallied.
10. A summary sheet is printed, which resembles the manual LARSP analysis form.

Automated LARSP closely follows the scoring guidelines of manual LARSP. However, the program's accuracy breaks down at scoring higher-level phrasal and clausal elements. In general, according to Bishop, the program reliably and accurately scores structures that are below LARSP's level IV. Accuracy of scoring decreases with increased complexity of samples. Therefore, the program may be deemed useful for analyzing the language samples of children who are young or have language disorders; but analysis of language-typical older children and adult language samples may prove difficult, given the amount of correction and revision required from the clinician.

DSS Computer Program. Developed by Hixon (1983), the *DSS Computer Program* analyzes language samples and provides a score based on Lee's (1974) DSS scoring system. The program is formatted to run on Macintosh computers, and (in 1986) was available for purchase for \$40. Current availability and pricing is unknown.

According to Klee and Sahlie (1986), the program analyzes utterances by parsing each utterance into units, and comparing those units to internal algorithms in order to determine point scores. Scores for each utterance are averaged over the 50-utterance sample, and a DSS score is computed. In addition, the program also computes an attempt score, which is the score the child would have received if every utterance had been produced correctly; and an error score, or the difference between the attempt score and the child's actual score. In order to obtain accurate results, the clinician must have a thorough knowledge of the original DSS scoring, as the program will invariably make errors in its analysis.

Computerized Profiling. The CP program (Long et al., 2006) is a freeware program available for download on the internet (www.computerizedprofiling.org). The program runs on all IBM-compatible personal computers, but not on Macintosh computers. CP analyzes each utterance in two phases, a grammatical tagging phase and a partial parsing phase. Grammatical tagging uses matrices of probability to predict the occurrence of grammatical structures. Probability of occurrence is based on the frequency of occurrence of the grammatical structures in the original sample population (Channell & Johnson, 1999). In the second phase, utterances are filtered using *shallow* or *partial* parsing. Shallow parsing is a method of quickly quantifying complexity, without performing an in-depth analysis. Utterances are compared with previously constructed templates to give a quantified estimation of syntactic complexity (Voss, 2005). One drawback of the program may be that CP runs under DOS, an unfamiliar operating system for younger clinicians.

Klee and Sahlie (1987) provided a user review of CP. The program performs several auto-analyses, covering the language domains of semantics, syntax, phonology, and pragmatics. The program integrates a diverse set of clinical linguistic analyses and combines them into one program. Measures of syntactic complexity are generated using DSS, and LARSP scoring guidelines and protocols. The authors noted that the program is easy to learn for clinicians who are already familiar with LARSP and DSS protocols. However, the program is not intended for use by clinicians who are unfamiliar with these measures, because clinicians will be required to correct the computer's analyses. For the LARSP analysis, the authors concluded that most of the utterances are coded incorrectly by the program; and often the clinician must spend more time correcting computer errors than it would have taken the clinician to code the utterance manually. No mention was made concerning the accuracy of DSS analysis. Since 1987, several other analyses have been added to CP, including IpSyn.

Success of Automated Language Sample Analyses

In 1983, researchers in Madison, Wisconsin began designing SALT and implementing its use in one school district of the city (Miller, 1992). The purpose of the program was to provide clinicians with an alternative assessment method for diagnosing language disorders, to the limited available standardized assessment batteries. The developers of the program hoped that the use of an automated computer program would increase the consistency of language sample interpretation among clinicians in the district. One difficulty the program faced, however, was that speech pathologists in the school district required a significant amount of training to master the program. In addition, transcribing and coding lengthy samples to enter into the program required

additional clinician time. The authors suggested solutions to these problems, such as thoroughly training a select group of clinicians, who could serve as mentors to other clinicians in the district; and contracting the work of language sample transcription to trained typists or college students. This study indicated that teaching clinicians to use language sample analysis software effectively might be a lengthy process.

Once a clinician is trained in the mechanics of a software program, however, the most obvious positive benefit provided by automated language sample analysis remains the speed and ease of obtaining a detailed syntactic analysis of language. Long (2001) conducted a study to determine the amount of time clinicians saved by using a computerized form of language sample or phonological analysis, versus a manual method. A total of 256 students and practicing clinicians participated in the study. Clinicians were asked to analyze language samples for phonological processes or syntactic development. Clinicians were allowed to choose the type of analysis they would perform, so long as they felt competent in performing the chosen analysis method. Results of the study found that, even for clinicians with limited training in automated analysis programs, automated language sample analysis was significantly faster than manual analysis. Long concluded that computerized analysis can make detailed language sample analysis efficient and practical, even for busy clinicians.

Even if automated analyses are efficient and practical, the most important aspect of their success relates to their ability to accurately identify language elements in a language sample. Quick, easy analysis methods are worthless if they are unable to produce valid results. Two studies have addressed this issue in regard to CP, revealing

that the analyses available on the program have met with varying levels of success in terms of their ability to accurately measure what they were designed to measure.

Long and Channell (2001) evaluated the accuracy of four automated language sample analyses performed automatically as part of the CP software program. A total of 69 child language samples (2;6 to 7;10) were used; including samples from typically-developing, language impaired, and speech-impaired children. Each of the samples was first coded manually by a clinician who was well-versed in the scoring procedures; which included MLU, IpSyn, LARSP, and DSS. Following manual analysis, the samples were scored using CP, and the correlation between manual and automated scores was tabulated. The authors found that agreement between manual and automated scores fell within the acceptable ranges of inter-rater reliability guidelines. This finding indicated that using automated analyses to analyze language samples may prove as accurate and reliable as using analyses coded manually by different clinicians.

Channell (2003) evaluated the accuracy of automated DSS scoring as computed by CP. Forty-eight school-age child language samples (including 28 samples from language-impaired children) were manually coded and scored for DSS; after which the samples were again coded and scored for DSS using CP software. Agreement between manual and automated codes and scores was calculated using correlational statistics. Results of the study indicated that the overall score agreement between manual and CP scoring was 78 percent. Scoring agreement for individual categories varied significantly among categories, and ranged from zero to 98 percent. Channell concluded that although the program does come close to reaching the 80 percent accuracy criterion suggested by

Hughes et al. (1994), further research to improve the accuracy of automated DSS is warranted.

Summary

A variety of measures have been developed to quantify the syntactic development manifest in children's spontaneous language samples; among these are the MLU-M, MLU-W, MSL, IPSyn, DSS, PSL and SCS. The scores can all be extracted from a language sample by the CP software. Though all of these measures yield a quantitative description which might characterize syntactic development, no study has as yet compared these measures to the actual presence of complex syntactic constructions produced by a client in a language sample. Such a comparison would give useful insight into the validity of these measures.

Method

This study analyzed a body of previously collected language samples. The pool of samples included samples collected from school-age children with typical language as well as samples collected from young adults with typical language.

Participants

The samples came from a corpus of language samples collected by researchers in the Los Angeles, California area during the early 1960s (Carterette & Jones, 1974; Jones & Carterette, 1963), and archived on the CHILDES database. Child language samples were collected from 54 first-, 48 third-, and 48 fifth-grade students at two different elementary schools. Child participants were primarily members of the middle socioeconomic class, spoke English as a first language, and did not have any documented speech or language disorders. Samples were also collected from 24 young adults who were participating in an introductory psychology class at a junior college in the area.

For both children and young adults, the language sample collection procedure ensued as follows: three participants (from the same age group) were brought to a room where there was a young, friendly clinician who encouraged the start of a conversation among the participants. The resulting conversations were recorded and transcribed.

The transcribed conversations were not divided according to speaker during transcription. This resulted in four large sample files, with each block representing the language of the representative age group in the sample, and not an individual speaker. It should be noted that while these sampling conditions would not be considered ideal for examining language on an individual basis; for the purposes of the current study (assessing the accuracy of syntactic analyses in identifying complex structure), the sampling procedures were deemed acceptable.

For the present study, each of the four age-group samples was divided into approximately 100-utterance segments to ease analysis by the computer software. Repeated utterances were not counted toward the 100-utterance tally, resulting in some variation in length among the samples, depending upon the number of repeated utterances present. The total number of samples extracted from the four age groups after segmentation was 107 samples.

Materials

The present study assessed the validity of seven quantified syntactic analyses performed by CP. All of the analyses were computed using CP version 9.7.0 (Long et al., 2006), installed and run on an IBM-compatible personal computer. The following CP analyses were conducted: MLU-W, MLU-M, MSL, DSS, SCS, IPSyn, and PSL.

Procedure

Each language sample was analyzed both manually by the author, and automatically by CP to determine the frequency of complex forms, and the quantified analyses scores, respectively.

Manual analysis. Each language sample was analyzed manually to determine the frequency of occurrence of the following complex syntactic structures: adverbial clauses, noun (complement) clauses, and relative clauses; in their finite forms. Productivity was also determined for complex structures in each sample, with the criteria of two, three, and four productive occurrences of a form per sample.

To ensure inter-rater reliability, approximately one-third (34%) of the samples were analyzed by both the author and another individual trained in syntactic analysis. The level of agreement between examiners was determined to be 90%.

CP complexity analysis. Following manual analysis, each sample was entered into CP for automated quantification of complexity for the seven syntactic analyses available on CP. Prior to entrance into the program, each sample was formatted according to the program's requirements. Requirements include the following:

1. Enter only one utterance per line.
2. End each utterance with a period, comma, question mark, or exclamation mark.
3. Use lowercase, except for proper nouns and pronoun *I*.
4. Use parentheses to eliminate mazes or other unwanted items.
5. Use a non-alphanumeric character at the beginning of the line to eliminate an entire utterance.

The samples were entered into the CP program as SALT (.slt) files. To ensure consistency in analysis among the samples, when prompted by the program to make judgments about grammar coding, the author always accepted the program's default judgment. In addition, when prompted by the program to decide whether to "code repetitions as stereotypes," the author chose *yes*. Appendix A contains a detailed list of the procedures for analyzing a language sample in the CP program.

Data analysis. In the CP program, samples were tabulated for MLU-W, MLU-M, MSL, SCS, DSS, IPSyn, and PSL, and the resulting scores were recorded for each sample. Pearson's *r* correlations were performed on the seven data points for each grade-level sample, and also for a combined file which included all of samples. Productive occurrences of each form at the two-, three-, and four-occurrence level were also correlated with the seven CP analyses. In addition, a one-way analysis of variance (ANOVA) was conducted for the manual frequency count and CP analyses combined, post-hoc tests were conducted using the Student-Newman Keuls procedure, and a partial eta squared (η^2) analysis was used to examine the effect size of observed differences.

Results

The combined results of the first, third, fifth, and adult samples are reported in Table 1 and Figure 1 (See Appendixes B and C for individual group frequency correlation tables and figures). Productivity correlations were also conducted and are reported in Table 2 (See Appendix D for individual group productivity correlation tables). Unless otherwise noted, all correlations mentioned in the text were statistically

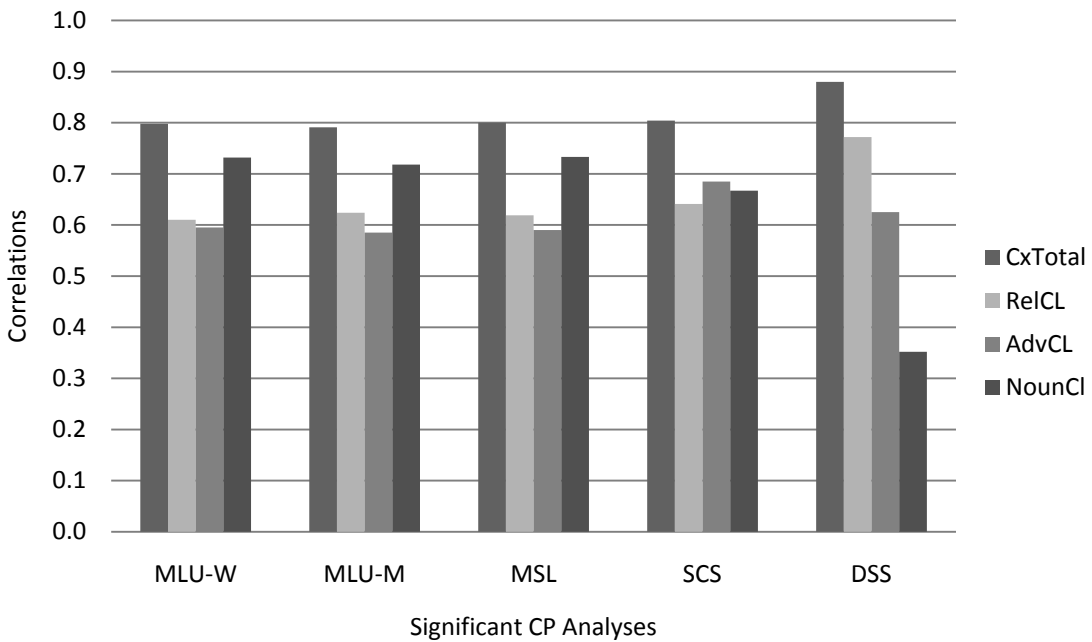
Table 1

Pearson's r Correlations Among Complexity Measures for the Combined Samples

	MLU-W	MLU-M	MSL	SCS	PSL	IPSYN	DSS
MLU-M	0.996**						
MSL	0.984**	0.991**					
SCS	0.945**	0.948**	0.953**				
PSL	0.429	0.435	0.461	0.430			
IPSYN	0.475*	0.486*	0.474*	0.417	0.589**		
DSS	0.879**	0.884**	0.892**	0.889**	0.540*	0.436	
RelCL	0.610**	0.624**	0.619**	0.641**	0.313	0.286	0.772**
AdvCL	0.595**	0.585**	0.590**	0.685**	0.173	0.260	0.625**
NounCL	0.732**	0.718**	0.733**	0.667**	0.352	0.286	0.776**
CxTotal	0.798**	0.791**	0.800**	0.804**	0.348	0.336	0.880**

* $p < .05$. ** $p < .01$.

Figure 1

Significant Correlations Among Complexity Measures for the Combined Samples

Note. Only significant correlations ($p < .05$) are included in the figure. Neither IPSyn nor PSL showed any significant correlations with the frequency count of complex structures.

Table 2

Productivity Correlations for the Combined Samples

	Two	Three	Four
MLU-W	0.460**	0.438**	0.484**
MLU-M	0.459**	0.447**	0.491**
MSL	0.468**	0.460**	0.511**
SCS	0.491**	0.507**	0.537**
PSL	0.121	0.129	0.212*
IPSYN	0.068	0.147	0.190
DSS	0.384**	0.436**	0.475**

*p < .05. **p < .01.

significant at $p < .01$, suggesting a very low probability that the observed correlations resulted from chance.

Length Measure Correlations

MLU-W, MLU-M, and MSL correlated highly with each other. They also showed significant correlations with SCS and DSS ($r = .710$ to $r = .875$). Length measures correlated with the total number of complex structures present in the sample ($r = .569$ to $r = .640$); and with each structure subtype, with a stronger correlation being observed overall between length measures and relative and adverbial clauses. In the adult subgroup, however, the strongest correlation existed between the length measures and noun clauses.

Measures of length also correlated with the number of complex clause types that the individual was productive for (either one, two, or three) at the two-occurrence, three-occurrence, or four-occurrence level ($r = .438$ to $r = .511$). Correlation was higher at the four-occurrence level than the two- and three-occurrence levels. In the individual age groups, correlation was higher for the first grade and adult samples than for the third and fifth grade samples (See Appendix D).

Syntactic Inventory Correlations

IPSyn and PSL revealed few significant correlations with other syntactic measures. PSL showed a correlation with MLU-W, MLU-M, and MSL at $r = .320$ to $r = .334$; and also correlated similarly with SCS. Both IPSyn and PSL showed minimal correlations with DSS. In the combined statistics, PSL showed a slight positive correlation with the total number of complex clauses ($r = .225$, $p < .05$) and a similar

correlation with the number of adverbial clauses. However, PSL also correlated with relative clauses in the third grade group; and with IPSyn in the adult group.

IPSyn demonstrated a correlation only with the number of noun clauses ($r = .316$). Individual group (first, third, fifth, and adult) statistics followed the same general pattern as the combined correlations, except that in the adult group, IPSyn also correlated with the length measures ($r = .47$ to $r = .49$; see Table B4 and Figure C4).

PSL correlated slightly with the number of four-occurrence productive clause types ($r = .212, p < .05$), but neither IPSyn nor PSL showed any other significant correlations for productivity overall. In the individual groups, PSL showed a correlation in the fifth grade samples ($r = .391, p < .05$) at the four-occurrence level; and IPSyn correlated at the four-occurrence level ($r = .388, p < .05$) in the third grade samples (See Tables D3 and D2, respectively).

Weighted Measure Correlations

DSS and SCS correlated with each other, and both showed significant correlations with measures of length ($r = .710$ to $.875$). Both measures also showed some correlation with PSL in the combined total, but did not correlate significantly with any measures for the individual subgroups. DSS correlated more highly with the total number of complex structures present ($r = .693$) than the individual structure categories of relative, adverbial, and noun clauses ($r = .447$ to $r = .483$). SCS followed the same pattern. Subgroup correlations for both SCS and DSS also exhibited a similar trend, correlating more highly with the total number of complex structures than the individual structure types.

The SCS measure correlated with the number of productive complex structure types at the two-, three-, and four-occurrence levels, with the strongest correlation

apparent at the four-occurrence level ($r = .511$). DSS also showed significant correlations with the number of productive complex structure types at the two-, three-, and four-occurrence levels, and, like SCS, showed the strongest correlation at the four-occurrence level ($r = .475$). In the individual subgroups, SCS showed the most correlation with productive occurrences in the first grade samples. DSS also correlated most strongly with the first grade samples,

ANOVA and Post-Hoc Comparisons

A one-way ANOVA was conducted to determine whether significant differences existed between the first, third, fifth, and adult subject groups for the seven syntactic complexity measures extracted from CP; results are reported in Table 3. Significant differences were found among the groups for SCS, IPSyn, DSS, and the number of noun clauses. Student-Neuman Keuls post-hoc tests revealed that for SCS, IPSyn, and the number of noun clauses, the adult group differed significantly from the three other groups. The first, third, and fifth grade groups did not significantly differ from each other for these measures. For DSS, significant differences were found between the adult group and the first and third grade groups. In addition, the first grade group differed significantly from both the fifth grade and adult groups.

Discussion

Seven quantitative syntactic analyses performed by CP, including MLU-W, MLU-M, MSL, SCS, DSS, IPSyn, and PSL were extracted from first-, third-, fifth-grade, and adult language sample subgroups. These scores were correlated with manual frequency counts to determine the validity of the automated measures in reflecting complex syntax. Results indicated that syntactic analyses that weigh different forms

Table 3

ANOVA Between-Group Data for the First, Third, Fifth, and Adult Subject Groups

	Sum of Squares	Mean Square	F	η^2
MLU-W	0.437	0.146	0.151	0.000
MLU-M	1.158	0.386	0.323	0.001
MSL	0.790	0.263	0.274	0.001
SCS	1.356	0.452	3.851*	0.101
PSL	249.251	83.084	0.705	0.020
IPSYN	124.435	41.478	4.158**	0.108
DSS	30.242	10.081	4.218**	0.109
RelCL	35.686	11.895	2.412	0.007
AdvCL	86.639	28.880	1.476	0.004
NounCL	587.077	195.692	15.993**	0.318
CxTotal	584.650	194.883	3.904*	0.102

* $p < .05$. ** $p < .01$.

$df = 3$

according to developmental complexity, and award points for each occurrence of a form (SCS and DSS), exhibited the highest correlation with the actual frequency of complex syntax. Measures of length, including MLU-W, MLU-M, and MSL, showed moderate correlations with the frequency of complex structures; while measures that take an inventory of structures present, IPSyn and PSL, showed low or nonexistent correlations with the frequency of complex forms in the language of school-age children and adults.

Weighted Measures

Of the CP analyses conducted, SCS and DSS showed the highest degree of correlation with the frequency count of complex structures. Both SCS and DSS correlated most strongly with the total number of finite complex clauses (including relative, adverbial, and noun clauses) over one particular type of complex form. The strong correlations demonstrated by both SCS and DSS are not unexpected, and may be attributable to the manner in which these measures calculate the score for a sample. Both analyses award points to each occurrence of a clausal unit or complex structure; and thus may better reflect the actual presence of complexity than the other CP measures.

Interestingly, although neither SCS nor DSS was developed or normed for use with adolescent or adult syntactic analysis, both measures showed significant differences among several of the subject groups, indicating that these measures may be sensitive to the changing complexity present at some of the age groups sampled. Previous studies have indicated DSS' ability to differentiate normal from disordered language (Leonard, 1972; Rice et al., 2006) and verified the relationship between DSS and age (Koenigsknecht, 1974) in children. While no known, published studies have documented the ability of automated SCS to quantify syntax in adult language samples; the current

findings agree with previous studies that have effectively used DSS to describe and differentiate syntax in younger and older adult subjects (Kemper, Herman, & Lian, 2003; Kemper, Herman, & Liu, 2004; Small, Lyons, & Kemper, 1997).

Length Measures

MLU-W, MLU-M, and MSL showed moderate, significant correlations with the frequency count of complex structures. These measures, like the weighted measures, correlated most highly with the total number of complex clauses (as opposed to one syntactic form); and also showed similarly strong correlations with the number of noun clauses. The moderate correlations exhibited by MLU-W, MLU-M, and MSL demonstrate these measures' ability to detect increased utterance length as complex clauses are added to an utterance. However, the length measures are not sensitive to specific clausal elements, and therefore did not consistently produce the strong correlations that are observed with the weighted measures.

Length measure scores correlated strongly with DSS; but did not demonstrate significant differences among the groups in the one-way ANOVA. These findings suggest that although measures of length may be moderate, consistent correlates for the presence of syntactic complexity in a sample; they do not appear to be sensitive to the subtle changes in syntactic length and complexity presented by the sampled age groups. The measures' strong correlations with DSS coincide with previous studies examining the correlational relationship between MLU and DSS in child language samples (Kemper et al., 1995; Rice et al., 2006).

Inventory Analyses

IPSyn and PSL exhibited no significant correlations with the frequency counts of complexity for the combined sample data. The measures' scoring procedures of awarding points for only the first occurrence of each syntactic form likely contributed to their lack of correlation with the frequency of complex syntax. Typically-developing children should develop all syntactic forms surveyed by these inventories by kindergarten or first grade. Thus, it is not surprising that these analyses were unable to differentiate between different age groups in the ANOVA, or show significant correlations with the amount of complex syntax in a sample. However, IPSyn and PSL did show some minimal correlations with length measures; suggesting some agreement with previous correlational findings (Rice et al., 2006) which reported significant correlations between MLU and IPSyn scores in school age children up to age 10.

Productivity

Correlations among the CP analyses and the productive occurrences of a form followed the same pattern as the raw frequency correlations; with SCS, DSS, and the length measures demonstrating significant correlations with the productivity count; and IPSyn and PSL demonstrating no significant correlations with the productivity of complex forms. The determination of productivity correlations did not appear to add new information regarding the occurrence of complex syntax in a sample, in comparison with simply using the raw frequency count.

Conclusions

The findings presented in this study may prove particularly useful for practicing clinicians, who often face limitations on the amount of time they can spend conducting

language sample analysis. By using an automated program to analyze language samples, clinicians may significantly reduce the time required to obtain valid analyses, while obtaining similar results to the manual versions of the same analyses (Long & Channell, 2001). Results found in this study suggest that clinicians may be able to use the scores obtained from SCS and DSS with some level of confidence that the analyses are accurate correlates for the amount of complex syntax in a language sample without having to scour the sample manually to determine that information.

It should be noted that these findings are limited to the subjects represented in the current study, and that further research is needed before these findings can be applied to diverse sample populations. The current study used a relatively small corpus of language samples, collected in one geographical area, consisting of individuals from predominately white, middle-class backgrounds in the 1960s. Further research, involving a larger number of samples from diverse geographical areas, cultural, socioeconomic, disorder backgrounds, and a wider variety of ages is warranted to generalize these findings to other populations.

Despite its limitations, the current study demonstrates that syntactic analyses performed by CP can quickly produce quantified scores that reflect the actual presence of complex syntax in a language sample. SCS and DSS seem particularly promising, showing strong correlations with the frequency of complex structures for all of the age groups sampled. MLU-W, MLU-M, and MSL also show consistent correlations with all of the subject groups, but correlate less highly with frequency counts than DSS and SCS. IPSyn and PSL do not appear to be valid quantifiers of syntactic complexity in school-age or adult language samples. In practice, automated analyses can be conducted quickly

and efficiently and, when valid, greatly reduce the amount of analysis time required of busy clinicians. Reducing the time required of clinicians in syntactic analysis leaves time for other aspects of assessment and treatment, ultimately improving the quantity and quality of assessment and treatment available for clients.

References

- Bishop, D. V. (1984). Automated LARSP: Computer-assisted grammatical analysis. *British Journal of Disorders of Communication, 19*(1), 78-87.
- Blake, J., Quartaro, G., & Onorati, S. (1993). Evaluating qualitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language, 20*(1), 139-152.
- Bloom, L., & Lahey, M. (1978). *Language Development and Language Disorders*. New York, NY: John Wiley and Sons, Inc.
- Brown, R. (1973). *A first language: The early stages*. Cambridge: Harvard University Press.
- Carterette, E. C., & Jones, M. H. (1974). *Informal speech: Alphabetic and phonemic texts with statistical analyses and tables*. Berkeley, CA: University of California Press.
- Channell, R. W. (2003). Automated developmental sentence scoring using computerized profiling software. *American Journal of Speech Language Pathology, 12*(3), 369-375.
- Channell, R. W., & Johnson, B. W. (1999). Automated grammatical tagging of child language samples. *Journal of Speech, Language, and Hearing Research, 42*(3), 727-734.
- Conti-Ramsden, G. (1996). Computer applications: CLAN (computerized language analysis). *Child Language Teaching and Therapy, 12*(3), 345-349.
- Crystal, D. (1982). *Profiling linguistic disability*. London: Edward Arnold.
- Crystal, D., Fletcher, P., & Garman, M. (1976). *The grammatical analysis of language disability*. London: Edward Arnold.

- Crystal, D., Fletcher, P., & Garman, M. (1989). *The grammatical analysis of language disability: A procedure for assessment and remediation* (2nd ed.). London: Cole and Whurr.
- de Villiers, J. G., & de Villiers, P. A. (1973). A cross-sectional study of the acquisition of grammatical morphemes in child speech. *Journal of Psycholinguistic Research*, 2(3), 267-278.
- Diessel, H., & Tomasello, M. (2000). The development of relative clauses in spontaneous child speech. *Cognitive Linguistics*, 11(1/2), 131-151.
- Hixson, P. (1983). DSS Computer Program. Omaha, NE: Computer Language Analysis.
- Hughes, D. L., Fey, M. E., Kertoy, M. K., & Nelson, N. W. (1994). Computer-assisted instruction for learning developmental sentence scoring: An experimental comparison. *American Journal of Speech Language Pathology*, 3(3), 89-95.
- Hughes, D. L., Fey, M. E., & Long, S. H. (1992). Developmental sentence scoring: Still useful after all these years. *Topics in Language Disorders*, 12(2), 1-12.
- Hux, K., Morris-Friehe, M., & Sanger, D. D. (1993). Language sampling practices: A survey of nine states. *Language, Speech, and Hearing Services in Schools*, 24(2), 84-91.
- Johnson, M. R., & Tomblin, J. B. (1975). The reliability of developmental sentence scoring as a function of sample size. *Journal of Speech and Hearing Disorders*, 18(2), 372-380.
- Johnston, J. R. (2001). An alternate MLU calculation: Magnitude and variability of effects. *Journal of Speech, Language, and Hearing Research*, 44(1), 156-164.

- Jones, M. H., & Carterette, E. C. (1963). Redundancy in children's free-reading choices. *Journal of Verbal Learning and Verbal Behavior*, 2, 489-493.
- Kemp, K., & Klee, T. (1997). Clinical language sampling practices: Results of a survey of speech-language pathologists in the United States. *Child Language Teaching and Therapy*, 13(2), 161-176.
- Kemper, S., Herman, R., & Lian, C. (2003). Age differences in sentence production. *The Journal of Gerontology: Psychological Sciences*, 58B(5), 260-268.
- Kemper, S., Herman, R., & Liu, C. (2004). Sentence production by young and older adults in controlled contexts. *The Journal of Gerontology: Psychological Sciences*, 59B(5), 220-224.
- Kemper, S., Rice, K., & Chen, Y. (1995). Complexity metrics and growth curves for measuring grammatical development from five to ten. *First Language*, 15, 151-166.
- Klee, T., & Fitzgerald, M. D. (1985). The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12(2), 251-269.
- Klee, T., & Sahlie, E. (1986). Computer applications: DSS Computer Program. *Child Language Teaching and Therapy*, 2(1), 97-100.
- Klee, T., & Sahlie, E. (1987). Computer applications: Computerized Profiling. *Child Language Teaching and Therapy*, 3(1), 87-93.
- Koenigsknecht, R. A. (1974). Statistical information on developmental sentence analysis. In *Developmental sentence analysis: A grammatical assessment procedure for*

- speech and language clinicians* (pp. 222-268). Evanston, IL: Northwestern University Press.
- Kramer, C. A., James, S. L., & Saxman, J. H. (1979). A comparison of language samples elicited at home and in the clinic. *Journal of Speech and Hearing Disorders*, 44(3), 321-330.
- Lee, L. L. (1974). *Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians*. Evanston, IL: Northwestern University Press.
- Lee, L. L., & Canter, S. M. (1971). Developmental Sentence Scoring: A clinical procedure for estimating syntactic development in children's spontaneous speech. *Journal of Speech and Hearing Disorders*, 36(2), 315-340.
- Leonard, L. B. (1972). What is deviant language? *Journal of Speech and Hearing Disorders*, 37(4), 427-446.
- Long, S. H. (2001). About time: A comparison of computerized and manual procedures for grammatical and phonological analysis. *Clinical Linguistics and Phonetics*, 15(5), 399-426.
- Long, S. H., & Channell, R. W. (2001). Accuracy of four language analysis procedures performed automatically. *American Journal of Speech Language Pathology*, 10(2), 180-188.
- Long, S. H., Fey, M. E., & Channell, R. W. (2006). Computerized Profiling (CP) (Version 9.7.0). Cleveland, OH: Department of Communication Sciences, Case Western Reserve University.

- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Miller, J. (1992). Implementing computerized language sample analysis in the public school. *Topics in Language Disorders, 12*(2), 69-82.
- Miller, J. (2008). SALT Software. Retrieved November 4, 2008, from <http://www.saltsoftware.com/>
- Miller, J., & Chapman, R. (1981). The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research, 24*(2), 154-161.
- Miller, J., & Chapman, R. (2000). Systematic Analysis of Language Transcripts (Version 6.1, Windows) [Computer software]. Madison, WI: Language Analysis Laboratory, Waisman Center on Mental Retardation and Human Development.
- Nippold, M. A., Mansfield, T. C., Billow, J. L., & Tomblin, J. B. (2008). Expository discourse in adolescents with language impairments: Examining syntactic development. *American Journal of Speech Language Pathology, 17*(4), 356-366.
- Owens, R. E., Jr. (2004). *Language disorders: A functional approach to assessment and intervention* (4th ed.). Boston, MA: Pearson Education.
- Rice, M., Redmond, S., & Hoffman, L. (2006). Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language, and Hearing Research, 49*(4), 793-808.
- Rondal, J. A. (1987). Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language, 14*(3), 433-446.

Scarborough, H. S. (1990). Index of Productive Syntax. *Applied Psycholinguistics*, 11(1), 1-22.

Small, J., Lyons, K., & Kemper, S. (1997). Grammatical abilities in Parkinson's disease: evidence from written sentences. *Neuropsychologia*, 35(12), 1571-1576.

Voss, M. (2005). *Determining syntactic complexity using very shallow parsing*.
University of Georgia, Athens.

Ward, E., & Fisher, J. (1990). A picture-elicited screening procedure. *Child Language Teaching and Therapy*, 6(2), 147-159.

Appendix A

Steps for Computing Measures with CP

1. Open CP. Type *Ret** (or *Enter*) three times to bring up the *Main Menu*.
2. Type *1, 1, 3*; then type *2* for *SALT format*. Select the desired file. Type *C* for *Continue*, then type *1* for *C as target*, and type *1* again for the child utterances to be analyzed. Type *Ret* to accept all default classifications of the 's element. This step may need to be repeated several times until all defaults are accepted. If an utterance is longer than 20 words, a number must be typed to split the utterance. Type a number to choose a point at which to split the utterance. Then type *Retto* accept the corpus file name. Type *Esc** to return to the *Main Menu*.
3. Type *5* for *LARSP*, type *1* to create the *LARSP file*, select the desired file, then type *Y* (for Yes) to code all repetitions as stereotypes. Type *Ret*, then type *Y* for *Yes to Analyze all single-word utterances as Stage 1*. When it finishes all the utterances, type *3* to *tabulate the LARSP file*. Type *Ret* three times to skip the top of the profile, then type *Ret* to start the tabulation. Type *P* for *LARSP Profile*, type *1*, then review profile to get the Number of Utterances, MLU-W, MLU-M, and MSL. The SCS measure can be found at the bottom of the page. Then type *Esc 2* times to return to the *Main Menu*.
4. Type *5* for *LARSP* again, then type *6* this time to choose *PESP Score*. Select the desired file, then type *V* for *View/Print*. The score is on the next to last line. Type *Esc 2* times to return to the *Main Menu*.
5. Type *7* for *DSS*, type *1* for *Create DSS*, then select the desired file. Type *C* for *Continue*, then type *V* for *View Profile*, then type *N* for *Norms*. Type in a dummy

age (66), then type *Ret* to get the DSS score. Type *Esc* 2 times to return to the *Main Menu*.

6. Type 6 for *IPSyn*, type 1 to *Create IPSyn*, then select the desired file. Type *Ret* to select the default 25 limit. Type *Ret* to begin on utterance 1. Then a pop-up window asks *Run Index Utterances to Identify Repetitions?* Type *Y* for *Yes*, then type *C* for *code repetitions*. Type *Y* for *Continue*, then type *Esc* to return to the *Main Menu*. Type 6 for *IPSyn* again. Type 1 to *Create File*, select the desired file, then type *Ret* to accept the limit of 25. Type *Ret* to begin on utterance 1. A pop-up window will ask *Run Utterance continue to find cutoff?* Type *Y*, then type *Esc*. Type *Ret* on *Limit 25*, then type *Ret* for *Begin on 1*, and type *Ret* for *End on Calculated End Utterance*. Type *Ret* to truncate, then type *E* for *Edit/Print Profile*. The *IPSyn* score is three-fourths of the way down the page.

*Note: *Ret* = Return, *Esc* = Escape

Appendix B

Individual Group Correlation Tables for Frequency

Table B1

Pearson's r Correlations Among Complexity Measures for the First Grade Samples

	MLU-W	MLU-M	MSL	SCS	PSL	IPSYN	DSS
MLU-M	.993**						
MSL	.978**	.971**					
SCS	.874**	.854**	.910**				
PSL	-.102	-.153	-.036	-.067			
IPSyn	-.079	-.077	-.139	-.126	-.156		
DSS	.796**	.811**	.796**	.766**	-.147	.273	
RelCL	.726**	.738**	.736**	.682**	-.105	.051	.652**
AdvCL	.534**	.501*	.592**	.706**	.281	-.207	.458*
NounCL	.161	.167	.196	.394	-.366	.040	.211
CxTotal	.654**	.637**	.711**	.854**	.014	-.111	.606**

*p < .05. **p < .01.

Table B2

Pearson's r Correlations Among Complexity Measures for the Third Grade Samples

	MLU-W	MLU-M	MSL	SCS	PSL	IPSYN	DSS
MLU-M	.992**						
MSL	.965**	.984**					
SCS	.844**	.859**	.880**				
PSL	.360	.343	.388*	.259			
IPSyn	-.365	-.339	-.257	-.169	.329		
DSS	.787**	.801**	.833**	.833**	.380	-.088	
RelCL	.442*	.418*	.400*	.392*	.399*	.055	.344
AdvCL	.325	.362	.446*	.527**	-.016	.170	.590**
NounCL	.108	.121	.118	.194	.204	.198	.022
CxTotal	.422*	.445*	.495**	.588**	.230	.239	.521*

*p < .05. **p < .01.

Table B3

Pearson's r Correlations Among Complexity Measures for the Fifth Grade Samples

	MLU-W	MLU-M	MSL	SCS	PSL	IPSYN	DSS
MLU-M	.995**						
MSL	.976**	.974**					
SCS	.848**	.849**	.883**				
PSL	.465**	.463**	.436**	.370*			
IPSyn	-.061	-.061	-.091	-.039	.035		
DSS	.650**	.647**	.709**	.769**	.093	.061	
RelCL	.434**	.419*	.417*	.294	.126	-.083	.142
AdvCL	.537**	.555**	.640**	.769**	.323	-.109	.676**
NounCL	-.031	-.015	-.022	.059	-.260	.205	.375*
CxTotal	.561**	.577**	.640**	.734**	.186	-.027	.743**

*p < .05. **p < .01.

Table B4

Pearson's r Correlations Among Complexity Measures for the Adult Samples

	MLU-W	MLU-M	MSL	SCS	PSL	IPSYN	DSS
MLU-M	.996**						
MSL	.984**	.991**					
SCS	.945**	.948**	.953**				
PSL	.429	.435	.461*	.430			
IPSyn	.475*	.486*	.474*	.417	.589**		
DSS	.879**	.884**	.892**	.889**	.540*	.436	
RelCL	.610**	.624**	.619**	.641**	.313	.286	.772**
AdvCL	.595**	.585**	.590**	.685**	.173	.260	.625**
NounCL	.732**	.718**	.733**	.667**	.352	.286	.776**
CxTotal	.798**	.791**	.800**	.804**	.348	.336	.880**

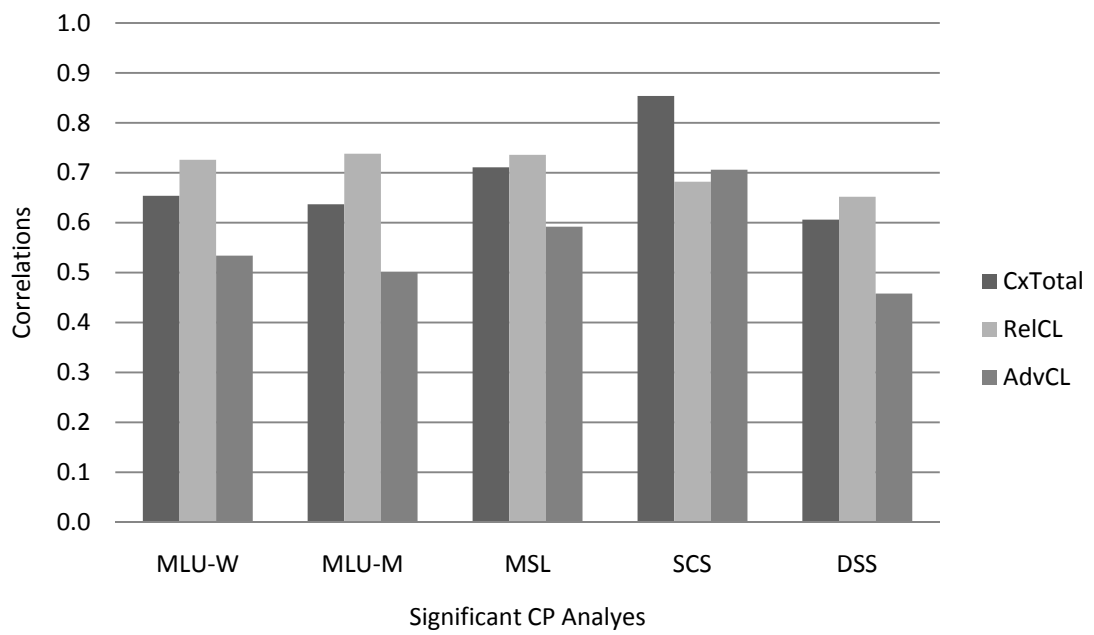
*p < .05. **p < .01.

Appendix C

Individual Group Correlation Figures for Frequency

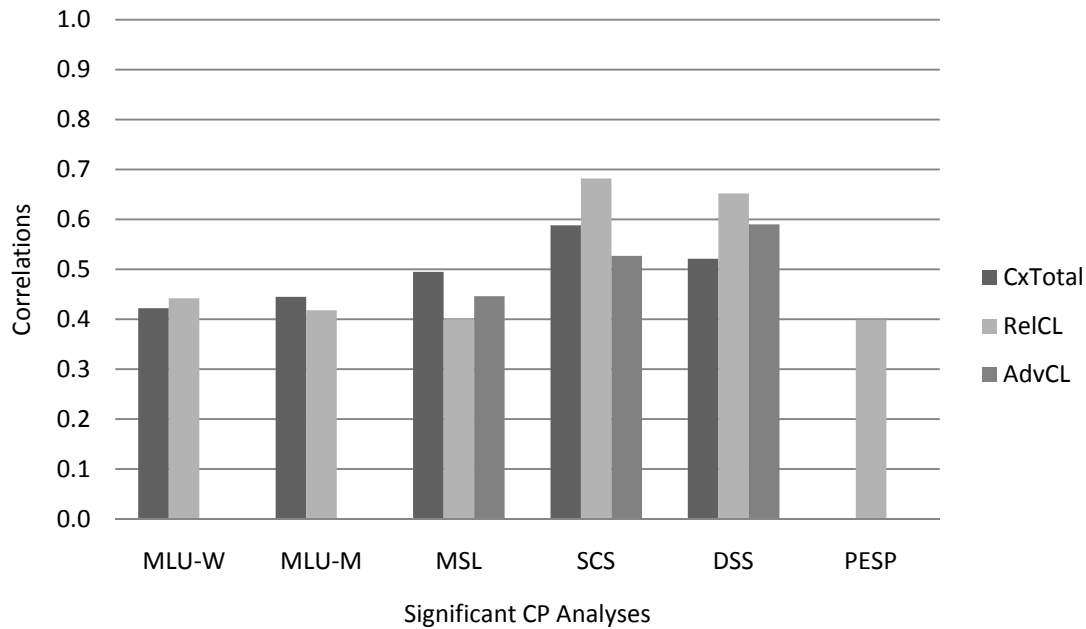
Figure C1

Significant Correlations Among Complexity Measures for the First Grade Samples



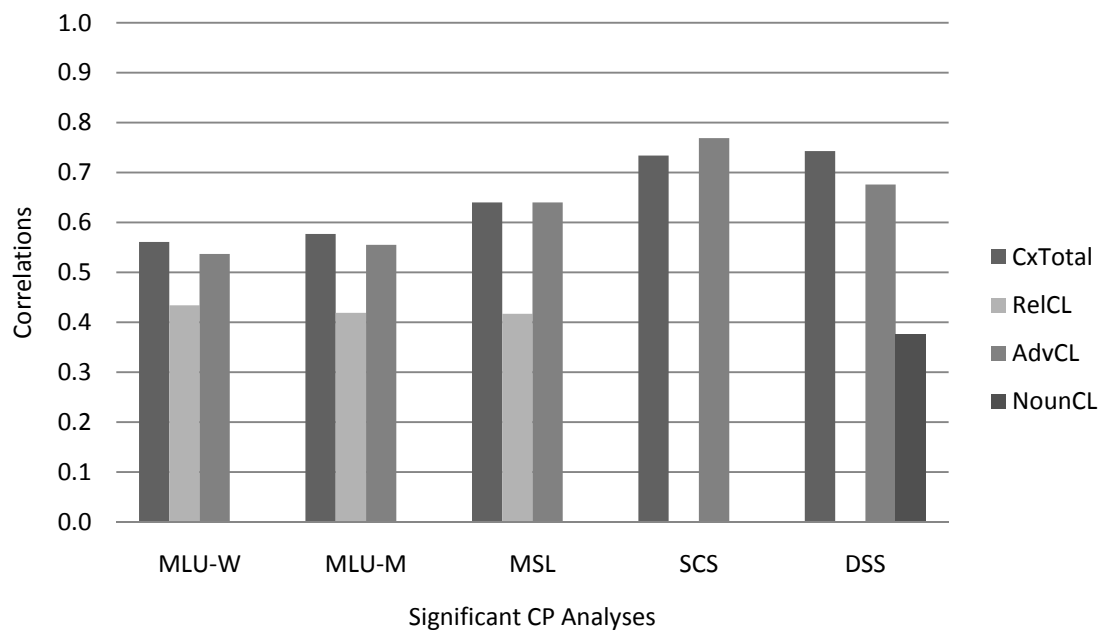
Note. Only significant correlations ($p < .05$) are included in the figure. PSL, IPSyn, and noun clauses demonstrated no significant correlations for the first grade samples.

Figure C2

Significant Correlations Among Complexity Measures for the Third Grade Samples

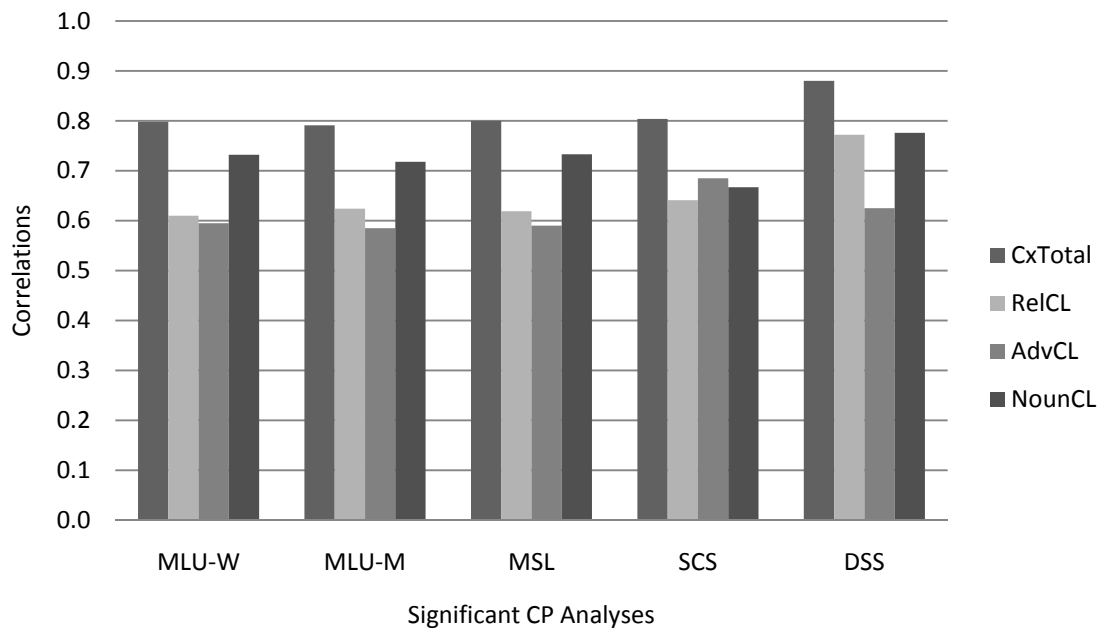
Note. Only significant correlations ($p < .05$) are included in the figure. IPsyn showed no significant correlations with the occurrence of complex structures. Adverbial clauses did not correlate significantly with MLU-W, MLU-M, or PSL. PSL only showed significant correlations with relative clauses.

Figure C3

Significant Correlations Among Complexity Measures for the Fifth Grade Samples

Note. Only significant correlations ($p < .05$) are included in the figure. IPSyn and PSL showed no significant correlations with the frequency of complex structures. Noun clauses showed significant correlations only with DSS; and relative clauses correlated significantly with all remaining measures but SCS and DSS.

Figure C4

Significant Correlations Among Complexity Measures for the Adult Samples

Note. Only significant correlations ($p < .05$) are included in the figure. PSL and IPSyn demonstrated no significant correlations with the frequency of complex structures.

Appendix D

Individual Group Correlation Tables for Productivity

Table D1

Productivity Correlations for the First Grade Samples

	Two	Three	Four
MLU-W	.760**	.697**	.635**
MLU-M	.755**	.708**	.639**
MSL	.795**	.670**	.639**
SCS	.747**	.685**	.665**
PSL	-.023	-.209	-.148
IPSYN	.030	.229	.225
DSS	.640**	.712**	.619**

Table D2

Productivity Correlations for the Third Grade Samples

	Two	Three	Four
MLU-W	.443*	.178	.224
MLU-M	.449*	.189	.241
MSL	.434*	.210	.276
SCS	.411*	.288	.326
PSL	.069	.223	.257
IPSYN	-.149	.195	.388*
DSS	.251	.090	.363

* $p < .05$. ** $p < .01$.

Table D3

Productivity Correlations for the Fifth Grade Samples

	Two	Three	Four
MLU-W	.368*	.475**	.565**
MLU-M	.356*	.461*	.556**
MSL	.371*	.470**	.579**
SCS	.324	.425*	.531**
PSL	.231	.260	.391*
IPSYN	.062	-.104	-.123
DSS	.286	.371*	.416*

*p < .05. **p < .01.

Table D4

Productivity Correlations for the Adult Samples

	Two	Three	Four
MLU-W	.322	.514*	.596**
MLU-M	.313	.518*	.595**
MSL	.282	.544*	.579**
SCS	.347	.592**	.624**
PSL	.126	.186	.252
IPSYN	-.071	.109	.230
DSS	.231	.561*	.572**

* $p < .05$. ** $p < .01$.