Comparison of Conversation, Freeplay, and Story Generation as Methods of Language Sample Elicitation

Frenette Southwood
Ann F. Russell
Stellenbosch University,
Stellenbosch, South Africa

The spontaneous language sample forms an important part of the language evaluation protocol (M. Dunn, J. Flax, M. Sliwinski, & D. Aram, 1996; J. L. Evans & H. K. Craig, 1992; L. E. Evans & J. Miller, 1999) because of the limitations of standardized language tests and their unavailability in certain languages, such as Afrikaans. This study examined 3 methods of language elicitation, namely conversation (CV), freeplay (FP), and story generation (SG), on the following 5 measures to determine which method is best for clinical practice: number of utterances, variety of syntactic structures, mean length of the utterance (MLU), number of syntactic errors, and proportion of complex syntactic utterances as elicited from ten 5-year-old, Afrikaans-speaking boys. FP elicited significantly more utterances than did SG but elicited a smaller proportion of complex syntactic structures than did CV and SG. Furthermore, SG elicited longer utterances than did CV or FP. It is recommended that SG be used in clinical practice with 5-year-olds if the clinician wishes to observe maximum behavior. Where typical behavior is to be evaluated, the clinician can select a language elicitation method that best suits the client's personality and communication style, bearing in mind that FP does elicit a larger language sample.

KEY WORDS: language sample analysis, language assessment, normal language development, 5-year-olds

The spontaneous language sample forms an important part of the language evaluation protocol (Dunn, Flax, Sliwinski, & Aram, 1996; Evans & Craig, 1992; Evans & Miller, 1999). Because of the limitations of standardized language tests (cf., among others, Hawkins & Spencer, 1985), the results of these tests must be supplemented with a spontaneous language sample, especially if appropriate therapy goals are to be set (Blau, Lahey, & Oleksiuk-Velez, 1984).

Several methods of eliciting language samples are discussed in the literature. Conversation (CV), freeplay (FP), and story generation (SG) are three of the prominent ones.

CV, as proposed by Crystal, Fletcher, and Garman (1976, 1989), among others, consists of dialogue on some aspect of the client's experience that is unrelated to the immediate situation. Methods used to elicit a conversation include (a) the researcher asking the child questions about a variety of topics such as family, school activities, and TV programs (cf. Longhurst & Grubb, 1974), (b) the researcher giving the child verbal

imperatives such as "Tell me about something you like to play outside" (cf. Atkins & Cartwright, 1982), (c) the child describing action pictures or a silent movie (cf. Wren, 1985), (d) the child explaining to the researcher how to play a game (cf. Fujiki & Willbrand, 1982), and (e) the child describing one picture at a time from a set of pictures and the researcher guessing which picture is being described (cf. Wren, 1985). As a language elicitation method, CV is problematic (Dollaghan, Campbell, & Tomlin, 1990; Fujiki & Willbrand, 1982). One problem is the lack of spontaneity as far as children are concerned, a deficit that is more common than admitted (Wren, 1985). Wren even stated that dialogue might be unnecessary if the focus of the evaluation is on the elicitation of as many syntactic structures as possible, because an evaluation of syntax, unlike that of pragmatics, does not require interaction between conversational partners.

Crystal et al. (1976) have also proposed FP as being a good method of language elicitation. Language samples are commonly elicited during researcher–child interaction, when participants play alongside one another with age-appropriate toys (cf. Craig & Washington, 2000; Dunn et al., 1996; Gavin & Giles, 1996; Washington, Craig, & Kusmaul, 1998). FP samples have also been elicited during researcher–child interaction while constructing a puppet show (cf. Wren, 1985) and during primary caregiver–child interaction at the child's home or in the clinic (cf. Scott & Taylor, 1978). Several authors (e.g., Evans & Craig, 1992; Fujiki & Willbrand, 1982) have criticized FP as being a time-consuming method for the elicitation of an adequate sample (Dollaghan et al., 1990). This is probably because the child is under little pressure to speak. In addition, because little control can be exercised over the children's utterances, there is not necessarily an opportunity for them to use all their acquired syntactic structures.

SG (or narration) is defined as telling original stories (Roth & Speckman, 1989); retelling movies, fairy tales, or folk tales; or reporting a summary of the experiences of people or animals (Scott, 1988) in two or more consecutive phrases (Labov, 1972). Stories are a natural form of discourse (Feagans & Short, 1984; Johnston, 1982) that occur regularly in everyday conversations (Sleight & Prinz, 1985). The two main SG methods are (a) telling a story about a particular topic (Roth & Spekman, 1989) and (b) telling a story after receiving a prompt, such as a picture series, a request to tell a familiar story ([e.g., Goldilocks] cf. Liles, Coelho, Duffy, & Zalagens, 1989; Merrit & Liles, 1987, 1989; Sleight, 1987; Wren, 1985), or a story told by the researcher (cf. Peterson & McCabe, 1983).

In 1974, Longhurst and Grubb cautioned that different methods of language elicitation could lead to measurable differences in language use. Since this statement was made, several studies have been conducted to compare methods of language elicitation in children. When combining and comparing the results of studies of 3- to 5-year-olds by Atkins and Cartwright (1982), 8- to 9-year olds by Evans and Craig (1992), 4- to 5-year-olds by Stalnaker and Creaghead (1982), and 6-year-olds by Wren (1985), the elicitation methods can be ordered as follows, from the one eliciting the most language to the one eliciting the least: CV using picture interpretation during which open-ended questions were asked, CV during which questions were asked in an interview format, story-retelling after the story was read to the child, and FP consisting of researcher–child interaction with toys of the child's choice. Furthermore, FP with puppets was found to elicit more language than a story generation (SG) task did in which a picture series was used as a prompt.

The patterns are less clear when comparing methods across studies in terms of complexity of language elicited. Atkins and Cartwright (1982) found that CV in which the researcher asked open-ended questions during picture interpretation elicited more complex language than did CV during which the child responded to verbal imperatives from the researcher. Evans and Craig (1992) found that another method of CV, namely question-asking during an interview, elicited qualitatively better language than did FP consisting of researcher–child interaction using toys selected by the child. Wren's (1985) results are similar to those of Evans and Craig, with SG eliciting less complex language than CV but more complex language than FP. The results of Wagner, Nettelbladt, Sahlén, and Hilholm (2000), whose participants were children aged 4 years 11 months to 5 years 9 months, do not agree with those of Wren: narration (a combination of SG and story retelling) elicited more complex utterances than did CV, which consisted of a combination of discussion on topics of interest to the child's life and interview-style question-asking.

Apart from Wren's (1985) study, no one study has compared CV, FP, and SG, despite the fact that these methods are popular and are used by many clinicians. Wren also used a very specific method of FP (namely, playing with puppets and props) in his study, which may limit the generalizability of the results. The primary aim of our study was to supplement the available information regarding the influence of language elicitation methods on expressive syntax. This was done by an examination and comparison of the results of three language elicitation methods in order to decide which method was the most effective for 5-year-old kindergarten children.

CV and FP were selected as elicitation methods because, according to Dollaghan et al. (1990), these are two methods frequently used in clinical practice. SG was

also selected as an elicitation method because (a) it is a method of language elicitation used increasingly by clinicians during language evaluation (Merrit & Liles, 1989), (b) it offers the clinician the opportunity to study complex language use, and (c) it may challenge the language production skills of the client more than CV does (Wagner et al., 2000).

The primary aim of the present study was to compare CV, FP, and SG on the following five aspects of the language elicited: (a) number of utterances, (b) variety of syntactic structures, (c) number of morphemes per utterance, (d) number of errors on the Language Assessment Remediation and Screening Procedure (LARSP) profile of Crystal et al. (1976, 1989), and (e) proportion of complex syntactic structures, where *complex syntactic structures* are defined, as suggested by Wren (1985), as structures at or above Stage V of the LARSP profile.

In South Africa, no generally accepted protocol for the identification of children with language disorders exists. Of the existing tests of expressive language abilities, very few have been standardized on Afrikaans-speaking children. Yet South African clinicians are routinely expected to make judgments on the normalcy of the expressive language abilities of these children as part of school-readiness testing. For this reason, language samples are very frequently used by South African clinicians, not merely for monitoring the progress of language intervention, but also for diagnostic purposes. In this respect, South African clinicians do not differ from clinicians in the United States: The results of Kemp and Klee's (1997) survey revealed that 85% of clinicians use language samples to assess the language of children with language impairment. South African clinicians, however, have no normative data with which to compare the expressive language abilities of their clients. Therefore, a secondary aim of our study was to collect data on the expressive syntactic abilities of normal-developing Afrikaans-speaking kindergarteners.

Method

Participants

The first author visited a local kindergarten, with which she was familiar, during recess in order to meet all the staff and kindergarteners. The principal was asked to identify all the Afrikaans-speaking boys, between 5;0 (years; months) and 5;11, from monolingual Afrikaans-speaking homes. Afrikaans-speaking children were chosen because the latest census shows that this is the language spoken by the majority of the population in the area. Five-year-old boys were chosen because statistics show that more boys than girls of this age are sent for school-readiness testing. The fact that only boys and only 5-year-olds were used could be viewed as a limitation of this study, as it may severely reduce the generalizability of the obtained results.

All participants were to be typically developing, have normal hearing, language abilities, and intelligence according to the judgments of their experienced classroom teachers, and have had no previous referral to or treatment by a speech-language therapist. From the pool of potential participants, 10 were randomly chosen. The parents of these 10 boys were contacted by phone to obtain verbal consent for the inclusion of their child in the study. When consent was not given, another participant was selected at random from the pool. Again, the generalizability of the results of this study may be impacted negatively by the small sample size and the fact that all participants were typically developing.

Ten boys between the ages of 5 years 5 months and 5 years 11 months were eventually included in the study. Every participant was a member of a family with either two or three children and was most often the second-born child. Only one participant was a first-born child.

The mothers of all the participants had been their primary caregivers before the participants started attending kindergarten. One participant had a residential worker as primary caregiver during the mornings prior to his enrolment in kindergarten, but his mother was his primary caregiver during the remainder of the day. All mothers had completed at least 12 years of formal schooling, and six had received tertiary education. Most participants were read a story on a daily basis by one of the parents. Stories were more often read than told. The mother of every participant indicated that the participant enjoyed the story reading and story telling.

Experimental Protocol

Framework of Procedures

After verbal consent was given to participate in the study, a letter was sent to the parents of the participants to explain the study and to arrange a 1-hr visit to the Hearing and Speech Clinic of a university training hospital, during which the rest of the procedure was performed by the first author. During the visit, we obtained written consent from the mothers and their sons after we explained the study's aim, duration, procedure, risks, and advantages and the confidentiality of the study's results.

The auditory sensitivity and middle-ear functioning of the participants were then screened in a sound-treated room according to the guidelines of the American Speech-Language-Hearing Association ([ASHA] 1975, 1979). All the participants' auditory sensitivity and middle-ear functioning were essentially within normal limits bilaterally. Next, three language samples were elicited from each participant.

While the researcher elicited the language samples from each participant, the participant's mother completed a short case history form to provide the researcher with background information. We asked several questions regarding the participant's language development and current language abilities in order to obtain a parent report that could corroborate the teacher's report on the normalcy of the participant's language abilities. To date, only three standardized tests have been designed for evaluating the receptive and expressive language of Afrikaans-speaking 5-year-olds. For various reasons (e.g., the norms derived from a population differing from our participants, poor reliability, and inadequate sampling), these three tests would not have rendered an accurate profile of the participants' language abilities. Therefore, no formal language evaluation was performed.

Language Elicitation Methods

The language elicitation took the form of researcher–participant interaction for increased generalizability to clinical evaluations (Evans & Craig, 1992). Every participant was seen alone in a quiet room, while his mother, with his knowledge and assent, was seated in an adjacent room where she could observe his interaction with the researcher through a one-way mirror while completing the case history form. We made audiovisual and audiocassette recordings of every language elicitation session, using observable video and cassette recorders. Every child participated in each of the three language elicitation procedures for 15 min. These procedures were given in counterbalanced order, with 5-min rest periods between them.

CV. During CV, the researcher and participant sat next to one another at a low-seated table. The first author modified questions from the MWM Program for Developing Language Abilities (Minskoff, Wiseman, & Minskoff, 1972) and began asking them in the same order for each participant until 15 min had passed (see Appendix A for topic probe and sample questions). The most questions asked during the 15 min of any 1 participant was 126, and the least was 60, with 9 of the 10 participants answering more than 80 questions. Apart from the questions, the other verbal prompts offered by the researcher were "oh," "is that so?" "really?" "uh-huh," "I see," "yes," "and then?" and "that's right," and the nonverbal prompts were head nodding and smiling. As suggested by Minskoff et al., the conversation took the form of a telephone conversation, using two toy telephones, to make the activity more interesting for the participants.

FP. During FP, the only loose objects visible to the participant (excluding the video and cassette recorders) were a low-seated table, two low-seated chairs, and several experiment selected toys. The toys were chosen for their proven popularity with child clients of both genders who had previously visited the clinic and consisted of (a) a two-story doll house with a removable roof and a front that opened like a door, (b) doll house furniture in a heap next to the doll house, (c) building blocks in a cardboard container, (d) six puppets (representing a family), (e) a toy stove, and (f) toy kitchenware (e.g., crockery and a pan) in a plastic dish next to the toy stove. The layout of the room and the arrangements of the toys were the same for each participant.

The FP session was initiated by the researcher inviting the participant to join in a game of placing the furniture into the doll house. The rest of the session proceeded according to the guidelines provided by Crystal et al. (1989). The researcher played alone with the participant in a manner that she viewed to be appropriate and natural. If the participant was quiet for extended periods, the researcher prompted him with questions such as "What will happen next?" or "What are you doing?"

SG. We used a method similar to the one suggested by Peterson and McCabe (1983). They based their method on that of Labov (1972; cf. Rollins, McCabe, & Bliss, 2000) who recommended that a topic for discussion be mentioned. If the participant gives an indication of identification with the topic, he is encouraged to discuss the topic. An advantage of Labov's method is that the participant's interest and field of experience are taken into consideration when deciding on a topic, so that the discussion is under no circumstances centered on a topic about which the participant is not knowledgeable or which the participant considers boring.

The researcher repeatedly introduced a topic by telling a personal story and then asking the participant whether anything similar had ever happened to him. An example of a story prompt used by the researcher was "Oops! If this little boy isn't careful, he will definitely fall [referring to the boy in a picture]. I fell the other day. I ran so fast down a little hill that I couldn't stop when I reached the bottom. I fell on my knees. My knees were hurt and they bled. Has anything like that ever happened to you?"

If the participant answered in the negative, then a next story was told. If the participant answered positively, then he was asked to tell what had happened in that situation. In order to avoid the participant truncating the story due to interpreting the researcher's silence as disinterest (Rollins et al., 2000), the researcher filled pauses with prompts. The same verbal and nonverbal prompts were used as during CV, with the exception of "that's right."

According to Preece (1987), personal stories are the most popular type of story among 5- and 6-year-olds. Rollins et al. (2000) have recommended that the narrative assessment of preschool children be based upon a

child's personal story about a real past event. For this reason, the researcher use true personal stories as prompts during SG. Topics included the researcher falling off her bicycle, breaking her arm, being admitted to hospital, having chicken pox, going on holiday, making a mess in the kitchen, attending birthday parties, preparing food, being mischievous, and experiencing adventures with her pets. The researcher told stories to the participant and listened to the participant's stories until 15 min had passed. The most stories told to any one participant was 19 and the least was 12.

The researcher and participant jointly colored in a picture while generating stories, as suggested by Peterson and McCabe (1983). The purpose of the coloring-in activity was to put the participant at ease and to minimize self-consciousness. During the coloring-in activity, the researcher and the participant were positioned as they were during the CV session.

Data Transcription and Analysis

The three 15-min-long language samples of every participant were transcribed orthographically. All utterances occurring in these 15 min were considered as part of the sample. One-word utterances, utterances containing errors, responses to questions, elliptical responses, and utterances containing unintelligible segments were included in the transcription. The total number of utterances were then counted, and the mean thereof was determined for each language sample. After transcription, the utterances were analyzed and recorded on a LARSP profile, according to the specifications of Crystal (1979) and Crystal et al. (1989).

There are seven stages of syntactic development on the LARSP profile. For each stage, an indication is given of the typical age at which utterances at this level are expected to be the most complex utterances in the child's repertoire. Stage I (0 years 9 months to 1 year 6 months) contains 5 different single word utterances, such as *go!* Stage II (1 year 6 months to 2 years 0 months) contains 8 clauses and 8 phrases consisting of two elements, as in *dolly sleep*, *the dog naughty*, or *give a cookie!* Stage III (2 years 0 months to 2 years 6 months) consists of 10 three-element clauses and 8 phrases of up to three words, such as *I see you*, *the big ball mine*, or *don't mess here*. Stage IV (2 years 6 months to 3 years 0 months) contains 8 clauses consisting of four phrases and 8 phrases consisting of up to four words, such as *why does he stand there?* or *book on the table is mine*. Structures occurring at Stage V (3 years 0 months to 3 years 6 months) include the coordinate and subordinate connectives (e.g., *and* and *whenever*), coordinate and subordinate sentences, other sentences with more than one clause (e.g., *she is all that she's made out to be*), comparative sentences (e.g., *she runs faster than her mom can*), and postmodifying phrases (e.g., *the dog with the brown spot*). Structures at Stage VI (3 years 6 months to 4 years 6 months) include noun phrase initiators (e.g., *about half of the children*), complex verb phrases (e.g., *could have been ready*), passive sentence constructions, and exclamations (e.g., *How good it is to be here*). Structures at Stage VII (4 years 6 months and older) include comment clauses (e.g., *that's my chair, you know*) and sentences with emphatic word order (e.g., *what he wanted we all could have guessed*). The LARSP profile makes provision for the recording of grammatical and morphological errors but only once the child has reached Stage VI. Before this age, one expects to find such errors in the child's language, but from 4 years 6 months onward one expects the vast majority of such errors to be eradicated (Crystal et al., 1989). Therefore, the presence of such errors are noteworthy once the child reaches 4 years 6 months. Appendix B contains the error taxonomy of the LARSP, as well as examples of errors.

Children are awarded 1 point for each LARSP structure present in the transcript. For example, if a child says *I breaking the glass*, 1 point is awarded to each of the following structures: on clause level, *SVO* (subject-verb-object); on phrase level, *DN* (determiner-noun) and $Pron^{P}$ (personal pronoun); on word level; *-ing*; and, in the error box, Aux^{O} *Ø* (deletion of a nonmodal auxiliary verb). The total score any one structure can obtain, therefore, depends on the number of times such a structure occurs in the transcript of the language sample. It follows, then, that some structures will have high scores, because they were used frequently in the language sample, whereas other structures may have a score of 0 because the child did not use them at all. Errors are scored in the same way, with 1 point being awarded to an error for every occurrence thereof in the transcript.

In the present study, each utterance was analyzed on clause, phrase, and word level. The relevant structures contained in each utterance, as well as all errors, were then assigned a raw score of 1 for each occurrence.

The variety of syntactic structures (on the clause and phrase levels combined) and the number of errors (i.e., entries in the error box on the LARSP profile) were calculated from the completed LARSP profile.

The mean length of the utterance (MLU) elicited by each procedure was calculated by dividing the number of morphemes used in the sample elicited by the total number of utterances elicited. An utterance was considered to be a single word, a single phrase, or a clause with its own prosodic identification (cf. Garman, 1989).

The proportion of complex syntactic structures elicited by each of the three methods was calculated by dividing the number of structures occurring at Stages V, VI, or VII of the LARSP profile by the total number of structures.

Reliability

The reliability of the transcription and analysis of each language sample was checked by having a final-year student in speech therapy independently transcribe and analyze 10% of every language sample. The interjudge reliability for transcription and for analysis was calculated with the following formulas, respectively: total number of words agreed on divided by total number of words transcribed by researcher multiplied by 100, and total number of structures agreed on divided by total number of structures identified by researcher multiplied by 100. The interjudge reliability for the 10% sample was 99% for transcription and 98% for analysis.

Statistical Analysis

A repeated-measures one-way analysis of variance (ANOVA) was performed for each of the five measures, that is, for (a) the number of utterances, (b) the number of different structures, (c) the MLU, (d) the number of errors, and (e) the proportion of complex syntactic structures (cf. Craig & Washington, 2000). Where the overall F was significant, we made post hoc pairwise comparisons by using the Bonferroni multiple comparison procedure. An alpha level of .05 was used for all statistical tests.

Results

The means and standard deviations of the five measures of language elicited by CV, FP, and SG are found in Table 1.

Number of Utterances

For number of utterances elicited, the difference among methods was significant, $F(2, 18) = 4.69, p = .023$, and the coefficient of determination (R^2) was .640. Pairwise comparisons showed that there was no significant difference between the number of utterances elicited by CV and FP, $F(1, 9) = 0.005, p = .950$. The significant difference was between FP and SG, $F(1, 9) = 7.18$, $p = .025$, with FP eliciting significantly more utterances. It is interesting to note that no significant difference was found between the number of utterances elicited by CV and SG, $F(1, 9) = 14.21$, $p = .128$, even though the mean number of utterances elicited by CV and FP were comparable (Ms = 142.9 and 144.2, respectively).

There were very large differences in the standard deviations of the number of utterances elicited by CV, FP, and SG (SDs = 8.103, 65.267, and 36.040, respectively). For CV, the method with the smallest standard deviation, the range in number of utterances per participant was the smallest (133 to 153). This could be attributed to the fact that the CV elicitation method was very structured, with all participants being asked the same questions in the same order, and that all participants possibly assumed that they had to answer all the questions posed to them. SG showed the second highest standard deviation and elicited a range of 38 to 162 utterances. This method was less structured than that of CV, with the type of prompt used by the researcher (*Has something like that ever happened to you?*) eliciting responses ranging from one utterance (*yes* or *no*) to responses consisting of detailed accounts of the participants' experiences. FP elicited a range of 12 to 214 utterances. During this elicitation method, there was no pressure on the participant to verbalize. Where participants verbalized actively, no prompts were used. Quieter participants, who were engrossed in the FP activity and for whom prompts such as "*What are you doing now?*" were used, treated these prompts in the same way as they did the other verbalizations of the researcher, namely, as utterances that required either no response or a one-utterance response, such as *playing* or *putting this here*.

Table 1. Means and standard deviations of five measures of language elicited by conversation (CV), freeplay (FP), and story generation (SG).

	CV		FP		SG	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*
Utterances	142.9	8.103	144.2	65.267	101.7	36.040
Different	50.6	6.653	51.7	15.747	51.2	5.8
MLU	3.709	0.798	3.755	0.637	4.966	0.768
Errors	1.9	2.898	2.0	2.309	2.3	3.020
Complex	0.169	0.034	0.111	0.035	0.164	0.030

Note. Utterances = number of utterances elicited; Different = number of different syntactic structures elicited; MLU = mean length of utterance (in morphemes); Errors = number of syntactic errors elicited; Complex = proportion of syntactic structures on Stages V, VI, or VII of the Language Assessment Remediation and Screening Procedure (LARSP) profile.

Variety of Syntactic Structures

There was no significant difference in number of different syntactic structures elicited by each method, $F(2, 18) = 0.01$, $p = .989$, $R^2 = 0.338$. The three procedures, although differing in terms of their level of structure, still elicited a similar number of different syntactic structures, ranging from 50.6 to 51.7 on average.

MLU

For MLU, the difference among methods was significant, $F(2, 18) = 8.32$, $p = .003$, $R^2 = .558$. Pairwise comparisons revealed that the differences were not between CV and FP, $F(1, 9) = 0.02$, $p = .878$, but between CV and SG, $F(1, 9) = 10.67$, $p = .010$, and between SG and FP, $F(1, 9) = 10.92$, $p = .009$, with SG eliciting significantly longer utterances than did CV or FP. The longer MLU elicited by SG cannot be attributed to multiple conjoining using *and* or *and then*, as sentences conjoined by these connecting devices were counted as separate utterances, unless they were used as definite coordinating conjunctions. Rather, the longer MLU elicited by SG could be attributed to the type of responses required from the participants by this task and the lack of opportunity to use acceptable elliptical utterances during SG.

Number of Syntactic Errors

In terms of number of syntactic errors elicited, there was no significant difference among the methods, $F(2, 18) = 0.06$, $p = .941$, $R^2 = .382$. One reason for this lack of differentiation between elicitation methods could be that typically developing 5-year-olds, as those who participated in this study, have mastered most morphemes and have achieved "basic sentence competence" by age 5 (Owens, 2001) and, therefore, do not make many grammatical and morphological errors. Because typical language development was one of the criteria for participation in this study, we did not anticipate that the participants would exhibit many grammatical and morphological errors. Children with language impairment, however, usually make more errors and, therefore, their inclusion in a similar study might reveal a difference in the number of errors elicited by each method.

Proportion of Complex Syntactic Structures

For proportion of complex structures elicited, the difference among methods was significant, $F(2, 18) = 11.33$, $p = .0007$, $R^2 = .677$. Pairwise comparisons showed there was no significant difference between the number of utterances elicited by CV and SG, $F(1, 9) = 0.110$, $p = .745$. The significant differences were between CV and FP, $F(1, 9) = 18.30$, $p = .0005$, and between FP and SG, $F(1, 9) = 15.58$, $p = .0009$, with FP eliciting the smallest proportion of complex utterances. That CV and SG elicited proportionally more complex utterances than FP could be explained by the fact that the type of verbal prompts provided by the researcher during the CV and SG activities gave the participants opportunities to respond with a complex utterance.

Discussion

The effectiveness of three methods of language sample elicitation was compared with regard to their number of utterances, variety of syntactic structures, MLU, number of syntactic errors, and proportion of complex syntactic structures elicited. It was found that these three methods differ to a limited extent in their effectiveness.

FP elicited significantly more utterances than did SG but not more or less than CV. These results do not agree with those of Evans and Craig (1992), who found that CV elicited more language than did FP. This difference may be attributed to the age of the participants in Evans and Craig's study: They were older than our participants and were language-impaired, whereas ours exhibited normal language development.

In the literature, there is no consensus as to the preferred number of utterances in a language sample. Crystal et al. (1976) suggested that a 30-min interaction, usually yielding 100 to 200 utterances, would suffice. However, Kemp and Klee (1997) found that, in practice, only 24% of clinicians used samples containing more than 50 utterances. The majority (48%) of clinicians used 50-utterance samples, and 28% used fewer than 50 utterances. In our study, all three elicitation methods provided more than 100 utterances on average. Therefore, although FP resulted in significantly more utterances than did SG, all three methods provided sufficient utterances for clinical use.

SG elicited longer utterances than did CV or FP, supporting the results of Dollaghan et al. (1990) and Wagner et al. (2000). No significant difference was found between the number of syntactic errors elicited by the three methods.

There was a significant difference between the proportion of complex syntactic structures elicited by FP compared with the other two methods. These results do not agree with those of Atkins and Cartwright (1982) and Longhurst and Grubb (1974), who found that elicitation settings with less structure, such as FP, elicited more complex language than did more structured settings, such as CV. Longhurst and Grubb's results were, however, obtained from children who were mentally

challenged. This may be the reason for the difference between their results and those of the present study. Our findings regarding the number of complex syntactic structures do agree with those of Evans and Craig (1992), who found that CV elicited more complex language than FP, despite the fact that the participants in their study were language impaired, which may have influenced their results, as children who are language impaired tend to be less responsive in the presence of adults (Rosinski-McCledon & Newhoff, 1987), and this may have reduced the quality of their expressive language.

Cowan, Weber, Hoddinott, and Klein (1967) found that a change in researcher could result in a difference of up to four words in utterance length in homogeneous populations. It is possible that the researcher could also have influenced the number and complexity of the elicited utterances. This could explain the difference between our results and those of previous studies. Furthermore, Peterson and McCabe (1983) mentioned that the topic of discussion can influence the length of a story generated about it. A further possible explanation for the differences in results may, therefore, be that the topics of the conversations and stories were not kept constant across studies. In our study, the questions and topics of discussion were the same for each participant in order to control for the effect that these two variables may have had on the quality and quantity of his utterances.

In summary, in this study, we found that FP elicited more utterances than did SG, but that the utterances elicited by SG were longer than those elicited by CV and FP. Furthermore, the proportion of complex utterances elicited by FP was less than that elicited by CV and by SG.

Our results suggest an implication for clinical practice. Wiig and Semel (1984) stated that there is a discrepancy between that which the child with a language impairment can do (maximum behavior) and that which the child does do (typical behavior). It is recommended that the clinician selects an elicitation method that suits the goal of the evaluation. If maximum behavior is to be evaluated, an SG should be used, because this method renders longer utterances than do CV and FP, and also renders a larger proportion of complex utterances than does FP. If typical behavior is to be evaluated, any one of the three methods can be used, because CV, FP, and SG did not appear to differ in terms of number of different syntactic structures or number of syntactic errors that they elicit. Therefore, the clinician should select a method that best suits the client's personality and communication style, bearing in mind that FP does elicit a larger language sample size.

In future research, larger sample sizes should be used to increase the external validity or the extent to which the results can be generalized. Furthermore, participants who are language impaired should be included, because the typically developing children who took part in this study are expected to have higher language proficiency across all language elicitation contexts than that of children with language impairment. This high language proficiency could have led to the lack of differentiation observed when comparing the three language elicitation methods in terms of number of different syntactic structures and number of grammatical and morphological errors elicited.

A secondary aim of this study was to collect normative data on Afrikaans-speaking children. Because of the small sample size and the fact that children of only one gender and one age group were studied, this study can be viewed merely as an initial attempt to start the process of norm collection. What this limited study showed, was that the MLU of the Afrikaans-speaking 5-year-old boys is comparable with that of American 5-year-olds (cf. Owens, 2001), provided the clinician uses a elicitation method such as SG that limits the number of acceptable elliptical utterances. This comparable MLU could be explained by the fact that Afrikaans is a Germanic language, as is English, and is morphologically similar, though not identical, to English in terms of type of grammatical morphemes and sentence structure. Secondly, the results of this study suggest that Afrikaans-speaking children are at least as talkative during language sample collection as are their English-speaking American or British counterparts, when considering the size of the samples elicited in this study.

In terms of collecting norms, this study could be considered a pilot study. To attain the secondary goal of this study, children of various ages and of both genders with typically developing language should be included to provide normative data on the five measures examined in this study.

Acknowledgments

We thank the Director of the Cape Education Department for permission to perform the study; the staff, parents, and students at a local kindergarten for their participation; and Ondene Van Dulm for her helpful comments on the manuscript.

References

American Speech and Hearing Association. (1975). Guidelines for identification audiometry. *ASHA, 17,* 94–99.

American Speech and Hearing Association. (1979). Committee on Audiometric Evaluation: Guidelines for acoustic immitance screening of middle ear function. *ASHA, 21,* 283–288.

Atkins, C. P., & Cartwright, L. R. (1982). An investigation of the effectiveness of three language elicitation

procedures on Head Start children. *Language, Speech, and Hearing Services in Schools, 13,* 33–36.

Blau, A. F., Lahey, M., & Oleksiuk-Velez, A. (1984). Planning goals for intervention: Language testing or language sampling? *Exceptional Children, 51,* 78–79.

Cowan, P. A., Weber, J., Hoddinott, B. A., & Klein, J. (1967). Mean length of spoken response as a function of stimulus, experimenter, and subject. *Child Development, 3,* 191–203.

Craig, H. K., & Washington, J. A. (2000). An assessment battery for identifying language impairments in African American children. *Journal of Speech, Language, and Hearing Research, 43,* 366–379.

Crystal, D. (1979). *Working with LARSP.* 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* (2nd ed.). London: Edward Arnold.

Dollaghan, C. A., Campbell, T. F., & Tomlin, R. (1990). Video narration as a language sampling context. *Journal of Speech and Hearing Disorders, 55,* 582–590.

Dunn, M., Flax, J., Sliwinski, M., & Aram, D. (1996). The use of spontaneous language measures as criteria for identifying children with specific language impairment: An attempt to reconcile clinical and research incongruence. *Journal of Speech and Hearing Research, 39,* 643–654.

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

Evans, L. E., & Miller, J. (1999). Language sample analysis in the 21st century. *Seminars in Speech and Language, 20,* 101–116.

Feagans, L., & Short, E. J. (1984). Developmental differences in the comprehension and production of narratives by reading disabled and normally achieving children. *Child Development, 55,* 1727–1736.

Fujiki, M., & Willbrand, M. L. (1982). A comparison of four informal methods of language evaluation. *Language, Speech, and Hearing Services in Schools, 13,* 42–52.

Garman, M. (1989). The role of linguistics in speech therapy: Assessment and interpretation. In P. Grumwell & A. James (Eds.), *The functional evaluation of language disorder* (pp. 29–57). London: Croom Helm.

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

Hawkins, P., & Spencer, H. (1985). Imitative versus spontaneous language assessment: A comparison of CELI and LARSP. *British Journal of Disorders of Communication, 20,* 191–200.

Johnston, J. R. (1982). Narratives: A new look at communication problems in older language disordered children. *Language, Speech, and Hearing Services in Schools, 13,* 144–155.

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,* 161–176.

Labov, W. (1972). *Language in the inner city.* Philadelphia: University of Pennsylvania Press.

Liles, B. Z., Coelho, C. A., Duffy, R. J., & Zalagens, M. R. (1989). Effects of elicitation procedures on the narratives of normal and closed head-injured adults. *Journal of Speech and Hearing Disorders, 54,* 356–366.

Longhurst, T. M., & Grubb, S. (1974). A comparison of language samples collected in four situations. *Language, Speech, and Hearing Services in Schools, 5,* 71–78.

Merrit, D. D., & Liles, B. Z. (1987). Story grammar ability in children with and without language disorder: Story generation, story retelling and story comprehension. *Journal of Speech and Hearing Research, 30,* 539–552.

Merrit, D. D., & Liles, B. Z. (1989). Narrative analysis: Clinical applications of story generation and story retelling. *Journal of Speech and Hearing Disorders, 54,* 429–438.

Minskoff, E. H., Wiseman, D. E., & Minskoff, J. G. (1972). *The M.W.M. program for developing language abilities: Expression manual.* Ridgefield, NJ: Educational Performance Associates.

Owens, R. E. (2001). *Language development: An introduction* (5th ed.). Boston: Allyn & Bacon.

Peterson, C., & McCabe, A. (1983). *Developmental psycholinguistics: Three ways of looking at child's narrative.* New York: Plenum Press.

Preece, A. (1987). The range of narrative forms conversationally produced by young children. *Journal of Child Language, 14,* 353–373.

Rollins, P. R., McCabe, A., & Bliss, L. (2000). Culturally sensitive assessment of narrative in children. *Seminars in Speech and Language, 21,* 223–234.

Rosinski-McCledon, M. K., & Newhoff, M. (1987). Conversational responsiveness and assertiveness in language impaired children. *Language, Speech, and Hearing Services in Schools, 18,* 53–62.

Roth, F. P., & Speckman, N. J. (1989). The oral syntactic proficiency of learning disabled students: A spontaneous story sampling analysis. *Journal of Speech and Hearing Research, 32,* 67–77.

Scott, C. M. (1988). A perspective on the evaluation of school children's narratives. *Language, Speech, and Hearing Services in Schools, 19,* 67–82.

Scott, C. M., & Taylor, A. E. (1978). A comparison of home and clinic gathered language samples. *Journal of Speech and Hearing Disorders, 43,* 482–495.

Sleight, C. C. (1987). Adolescents' narrative use of orientations and codas. *Language and Speech, 30,* 169–180.

Sleight, C. C., & Prinz, P. M. (1985). Use of abstracts, orientations, and codas in narration by language-disordered and nondisordered children. *Journal of Speech and Hearing Disorders, 50,* 361–371.

Stalnaker, L. D., & Creaghead, N. A. (1982). An examination of language samples obtained under three experimental conditions. *Language, Speech, and Hearing Services in Schools, 13,* 121–128.

Wagner, C. R., Nettelbladt, U., Sahlén, B., & Hilholm, C. (2000). Conversation versus narration in pre-school children with language impairment. *International Journal of Language and Communication Disorders, 35,* 83–93.

Washington, J. A., Craig, H. K., & Kusmaul, A. J. (1998). Variable use of African American English across two language sampling contexts. *Journal of Speech, Language, and Hearing Research, 41,* 1115–1124.

Wiig, E. H., & Semel, E. (1984). *Language assessment and intervention for the learning disabled* (2nd ed.). Columbus, OH: Charles E. Merrill.

Wren, C. T. (1985). Collecting language samples from children with syntax problems. *Language, Speech, and Hearing Services in Schools, 16,* 83–102.

Received November 8, 2002

Accepted March 17, 2003

DOI: 10.1044/1092-4388(2004/030)

Contact author: Frenette Southwood, Department of General Linguistics, Stellenbosch University, Private Bag X1, Matieland, 7602, Stellenbosch, South Africa. E-mail: fs@sun.ac.za

Appendix A. English translation of topic probe and sample questions asked during CV.

"Right, [participant's name], here are two telephones. This pretty one is yours. We are going to play a game where I phone you. OK? Here we go! [Researcher says participant's telephone number aloud while dialing.] Ring! Ring! [Participant picks up receiver.] Hello! Hello! Who am I speaking to now? Hello, [participant's name]. It's Frenette Southwood speaking. I want to ask you a few questions, and if you can answer everything correctly, then you win the competition. Right!"

Topic Probe	Sample Question
First I want to ask you about yourself.	What clothes are you wearing now?
Now I want to ask a little about your family.	Who is the youngest in your family?
Now I want to ask a little about your house.	Can you tell me about the rooms in your house?
Now I would like to know a little more about your school.	What do you like the most at school?
Let's talk a little about watching television.	What programme on TV do you not like at all?
Now I want to hear a little about your pets.	What type of animal is it?
Now I want to know about how you play.	Tell me, how does one play hide-and-seek?
Let's talk a little about holidays.	How did you get there [the place in which the participant vacationed]?
Now I want to ask some questions about the shops.	How do you get to the shop from your house?
[Participant's name], tell me a little bit about your neighbourhood.	Who lives next-door to you?
Now I want to know [about careers]:	And teachers? What do they do?
Let's talk a little about food.	Why is it good to have breakfast?
Let's talk about cars.	Can you drive yet? Why (not)?
Just a few more questions.	Why are you here?

Appendix B. Language Assessment Remediation and Screening Procedure (LARSP) error taxonomy.

Structure	Abbreviation on LARSP profile	Example of construction (where ambiguous, intended meaning or complete utterance in parentheses)
Errors of connectivity		
Incorrect use of *and*	and	he ate the sandwich and he made it (where *it* refers to the sandwich)
Incorrect use of coordinating conjunctions	c	his mommy scolded at him so he was naughty
Incorrect use of subordinating conjunctions	s	he got ill during the night because he's mom took him to the 24-hour clinic
Errors at clause level		
Omission of an element of the clause	Element Ø	he said to her (= he said something to her)
Incorrect placement of elements	Element ⇆	he came to see if just we are OK
Errors of concord	Concord	my cats eats
Errors at NP (noun phrase) level		
Incorrect use of a determiner	D	many coffee
Omission of a determiner	D Ø	he cuddled little dog
Incorrect placement of a determiner	D ⇆	dog a hungry (= a dog hungry)
Incorrect use of a preposition	Pr	he put the food in the table
Omission of a preposition	Pr Ø	he put the food the table
Incorrect placement of a preposition	Pr ⇆	he put the food the table on
Incorrect use of a personal pronoun	$Pron^{P}$	him saw the little girl standing there
Errors at VP (verb phrase) level		
Substitution of one modal auxiliary for another	Aux^{M}	he can read (= he must read)
Omission of a modal auxiliary	Aux^{M} Ø	he jump (= he can jump)
Substitution of one nonmodal auxiliary for another	Aux^{O}	he do eating (= he is eating)
Omission of a nonmodal auxiliary	Aux^{O} Ø	he eating (= he is eating)
Omission of a copula	Cop	the doggie happy (= the doggie is happy)
Errors at word level		
Errors in irregular pluralization	N irreg	fishes
Errors in regular pluralization	N reg	dogses (= dogs)
Errors in verb endings	V	singen