Temporal Reliability and the LARSP Profile: A Test – Retest Paradigm on Syntactic Production in Child Language

Jodi Tommerdahl
Department of Curriculum & Instruction
University of Texas at Arlington
SW Center for Mind, Brain and Education
701 Planetarium Place, Arlington, TX, 76019-0545, USA
E-mail: joditom@uta.edu

(Received: 8-10-13 / Accepted: 1-4-14)

Abstract
The LARSP (Crystal, Fletcher and Garman, 1976) is a language profile commonly used by Speech and Language Therapists in analysing the maturity of a speaker’s syntactic production based on a minimum of 50 utterances. This paper tests the temporal reliability of the LARSP profiles of child language by carrying out a test-retest procedure on repeated spontaneous language samples of three children aged 2;5 – 3;2. Profiles for each child were compared in terms of clause and phrase structure and at the word level. Results showed that despite variability between the type and number of individual morphosyntactic structures produced in a sample, the LARSP was reliable in determining overall development at both clause and phrase levels.

Keywords: Morphosyntax, Temporal Reliability, LARSP, Language Acquisition.

1. Introduction

A variety of tools exists to examine child language development and potentially identify language delay or impairment. The majority are standardized and often norm-referenced, allowing clinicians to compare the individual children’s performances to those of their peers. Highly respected tests, such as the Clinical Evaluation of Language Fundamentals (CELF-4) (Semel, Wiig and Secord, 2003) achieve this status partly because they have been tested for diagnostic accuracy, validity and reliability. While quantitative assessments are valuable for their ability to target specific skills with high degrees of reliability, qualitative assessments also exist, and play a complimentary role.

In the 1970’s and 80’s, qualitative assessments began to be developed by Crystal, along with his colleagues, Fletcher and Garman, (1976) covering diverse aspects of language including morphosyntax, phonology, prosody and semantics. These assessments, often referred to as profiles, were designed to be more comprehensive than formal tests and according to Crystal, profiles are meant to ‘provide a qualitative impression of linguistic behaviour, to supplement the quantitative scores arrived at through testing’ (1986). Profiles do not provide numerical...
scores, but instead create a picture of the individual’s linguistic strengths and weakness. Profiles are intended not only to describe an individual’s language capacity, but also to provide a basis for remediation planning which is generally not provided through standardized testing. (For an interesting discussion on the limitations of standardized tests, see Hasson and Joffe, 2007).

One of the most discussed criticisms of standardized assessments is the lack of ecological validity, or the quality of ‘real life’ language that is produced during a formal testing session. The underlying idea is that during standardized testing, language is elicited by an array of situations, many of which are unrepresentative of natural communicative experiences. A profile such as the LARSP however, allows the child to produce language freely in a more naturalistic context. Hewitt, Hammer, Yont and Tomblin (2005) find that analysing a spontaneous sample of language has the advantage of looking at ‘the natural behavior of children’ while more formalized testing looks at their behaviour in contexts that are ‘foreign to their experience’.

Language profiles using spontaneous samples overcome this barrier to ecological validity, at least partially. Analyses are based on language samples collected in natural conversation between a child and either a therapist or a caregiver, allowing the clinician to analyse production based on the child’s desire to communicate. This has clear advantages over the relatively unnatural situations of more formalized testing.

Linguistic profiles are not designed as a replacement for quantitative and formal language assessment tools, but are instead meant to be used alongside these tools to provide the most detailed insight into an individual’s linguistic ability. Each type has its own strengths and weaknesses with the strengths of each type often filling in the gaps left by the other. The more reputable standardized, quantitative assessments are subject to testing in areas such as validity and reliability in order to evaluate their appropriateness as clinical tools. Such measures of the assessment tools themselves lend credence to their use. The qualitative nature of the linguistic profile however, is not conducive to statistical measures in these areas. Given the nature of the spontaneous language sample as an ecologically valid, child-driven production of language, those who use language profiles in a clinical setting are unlikely to question their validity. The question of reliability though, is of concern. The aim of this article is therefore to carry out LARSP profiles on the same children in similar contexts in short succession to explore the LARSP’s reliability when applied to children’s spontaneous language samples.

1.1 Reliability

1.1.1 Temporal Reliability

The specific type of reliability examined here is temporal reliability, sometimes called test-retest reliability. This asks to what degree the LARSP profile will give a similar result when it is used on different language samples from the same child, given that no linguistic development has had the chance to take place. For example, if a therapist collects a language sample from a child on a Tuesday morning and another one the following afternoon, will the LARSPs provide the therapist with the same information? Given the holistic and qualitative nature of the LARSP, numerical scores are not provided. Instead, an age range of typical language use at the clause and phrase levels is achieved, along with an ordered list of morphemes. This makes it possible for the LARSP to be tested in terms of general reliability to the extent that if repeated LARSPs provide the same information about the child’s developmental stage, the LARSP can be said to be more reliable than if it does not. Full statistical tests of reliability for the LARSP are not possible due to its nature. For a more in depth discussion of reliability and its basis in Classical Test Theory, see Tommerdahl and Kilpatrick (2014).
1.1.2 Reliability in Spontaneous Language Samples

As the LARSP is a measure of morphological and syntactic production, this discussion will be limited to what is known about the reliability of the frequency of morphological and syntactic production in spontaneous language samples.

MLU, or mean length of utterance, although not a direct example of morphosyntax, has long been recognized as a proxy for syntactic development (Klima and Bellugi, 1966; de Villiers and de Villiers 1973; Rondal, Ghiotto, Bredant and Bachelet, 1987; Hickey 1991) and has been shown to be highly reliable (Rondal et al., 1987; Casby 2011). However, the reliable nature of MLU appears to come from its global nature. Cole, Mills and Dale (1989), when comparing reliability of MLU and specific morphemes in children with developmental delays, found MLU to have a reliability value of .92 while tested morphemes (past tense, plural and present progressive) were all found to have values of less than .5. MLU was also found to be highly reliable by Heilmann, Nockerts and Miller (2010) even in very short samples of 1, 3 and 7 minutes.

Few studies have gone beyond global measures such as MLU to examine the reliability of more specific morphological and syntactic items or structures. This missing and much needed information has been discussed by researchers as long as 50 years ago when Minifie, Darley and Sherman (1963) pointed out that much of the research over the preceding 30 years was based on researchers’ assumptions that ‘the numbers they obtained with their scales and measures were isomorphically related to “real” language development.’ Unfortunately, more recently, Marinellie (2004: 519) was still calling the topic of reliability of syntax in spontaneous language samples ‘a relatively uncharted area’. However, a small number of studies examining the reliability of production do exist and will now be discussed.

In an examination of the reliability of the norm-referenced Developmental Sentence Scoring (Lee and Canter, 1971), a measure of syntax for clinical use based on the analysis of language samples, Johnson and Tomblin (1975) looked at frequency counts of specific morphosyntactic items such as indefinite and personal pronouns, negatives, conjunctions and verbs. Generally low reliability was found in smaller samples except for the case of personal pronouns.

The study mentioned above by Cole et al. (1989) also fulfils the criteria of examining the reliability of morpheme production in spontaneous language samples. These 100 utterance minimum samples were gathered from 10 children with developmental delays. Correlations between repeated samples were not based of frequency counts, but rather on the presence versus the absence of each morpheme in the samples. 25 specific morphemes were sought after in each sample. Multiplying the number of participants with the number of morphemes sought, the authors conclude that out of 250 chances for each item to appear, that 35 forms were used in the first sample only while 33 forms were exclusive to the second sample, with the authors declaring the need for further study.

Tommerdahl and Kilpatrick’s study (2013) uses a test-retest method to gather language samples from 2 and 3 year old children to analyse the reliability of production of a number of morphemes put forward as potential indicators of Specific Language Impairment including the copular, past tense, be as an auxiliary, and others. They then determined the reliability for each structure at 50, 100, 150 and 200 utterances. Results show that at 50 utterances, reliability is low for all items. By 200 utterances, the most reliable item tested was the contracted auxiliary at .78. Overall, reliability of most morphemes, even at larger samples sizes, was found to be low.

In addition to using frequency counts, this study also showed, similarly to Cole et al. (1989) how often the samples would differ as to whether an item was present versus absent. For example, when comparing the first 50 utterances of the children’s first and second samples to count the use of the past tense, the researchers found that 14 out of 23 children did not produce it in the first sample but then went on to produce it in the second. When using samples of 100 utterances, this number dropped to 10 out of 23. This highlights the danger of
a therapist using single samples in establishing whether a child uses a specific grammatical form in spontaneous language production.

A second study by Tommerdahl and Kilpatrick (2014) used a similar protocol but this time look at non-clinical items of morphosyntax as well. Results showed degrees of correlation to be relatively low for all items tested with the copula being the most reliable, reaching an r value of .64 at 100 utterances. This study also corroborated their earlier study's findings, showing that the absence of a specific morphosyntactic item in spontaneous samples up to 200 utterances was not sufficient to make the assumption that the child did not have the item in question in their productive repertoire. This was especially true of low frequency items such as the genitive. It was also shown that more established items of vocabulary were not more reliable than those still being acquired. Some items, such as the copula, were also shown not to be distributed evenly throughout the samples, but instead to appear in clusters. Overall, these studies, although few in number, show that specific morphosyntactic item production in spontaneous samples have relatively low reliability. Given what is known about this lack of reliability, it is good practice to question the reliability of whether a procedure such as the LARSP that is based on frequency counts of morphological and syntactic items in spontaneous samples may be subject to the same lack of reliability as individual morphemes or whether its nature of aggregating results of several structures within an age range will make its profile results relatively global and therefore more reliable.

### 1.1.3 The LARSP Profile

The LARSP (Language Assessment, Remediation and Screening Procedure) (Crystal, et al., 1976) is a linguistic profile, designed specifically to assess the performance of an individual in terms of productive syntax and morphology at a point in time. It is based on a spontaneous language sample which is recorded in a naturalistic play setting to be later transcribed and analysed.

The LARSP analysis sheet is a single page and is divided into sections. Sections A-D at the top contains boxes to be filled in regarding interaction between the therapist and the patient. This is followed by 7 stages of language shown on the chart which represent ever advancing stages of language development. Stage I is reserved for single word utterances and stages II-V are divided into clause and phrase level items. When a child’s sentence is analyzed, it is done so in terms of the items listed in the LARSP, with each item receiving a tick on the chart each time it’s produced in the sample. In this way, a frequency count of each item is carried out. A column outside of the clause and phrase headings is called word and is not divided into stages, but instead contains a list of morphemes such as the plural and past tense forms. Stage VI is a place to mark new constructions and to note errors. Stage VII focuses on discourse and style but will receive little attention here because the LARSP authors describe it as having ‘little real assessment value’ (Crystal, 1982: 41).

A single utterance can lead to one or to several marks being placed on the chart. Uttering the word ‘spoon’ would earn one mark at stage I alongside the N for noun. The utterance ‘I liked spoons’, however, would lead to marks at clause level III for SVO, at phrase level III for the use of a personal pronoun, and in the word column by the –ed indicating the past tense. Each of the seven developmental stages is associated with an age. Stage I represents productions made by a typically developing child from age 0:9 to 1:6. Each further stage represents another 6 month range with the seventh and final stage labelled as 4:6+. There are also dedicated transitional stages which run between stages II and III and III and IV. These are used when a child produces a sentence such as ‘big dog run’ which is labelled SV at clause level II but also gets a mark in the transitional stage between II and III to give credit for one of the clause elements containing more than one phrasal element (adjective and noun in this case).

Once the utterances have been individually analysed and placed on the LARSP chart, the profile can be used to determine whether the child is producing language at an age-appropriate level. This is done by checking whether the preponderance of markings on the
LARSP chart continue to the stage which is typical of the child’s chronological age. For example, Crystal, et al. (1976) provides a completed LARSP chart of a typically developing 3½ year old. Several markings at both clause and phrase level fill stages II to IV and both coordination and subordination are marked at the stage V clause level. After this stage, nothing else is marked. Therefore, the child has clearly progressed to a stage typical of children aged 3;0-3;6. It is not important that the greatest number of marks exist at stage IV or V in this case. Instead, it is that the marks are at all levels of the chart. Knowledge of the profiles of typically developing children is vital to being able to ‘read’ a LARSP. An experienced user can identify both normal language patterns which do not come far enough down the chart and recognize this as a language delay while a different pattern on the chart, with ‘holes’ in uncommon places may represent deviant development. Furthermore, some items on the chart, such as those representing pronoun use, have a tendency of accumulating a large number of marks, making an overall quantitative analysis of each level less than useful. It should be noted that the above paragraphs describe only the English LARSP and not the several versions that have been adapted to other languages (Ball, Crystal and Fletcher, 2012) including widely spoken languages such as German (Clahsen and Hansen, 2012) and French (Maillart, Parisse and Tommerdahl, 2012) and rarer languages such as Welsh (Ball and Thomas, 2012) and Frisian (Dijkstra and Schlichting, 2012). Each LARSP designed for non-English languages have been adapted to the specific needs of the particular language.

Clearly, reliability cannot be measured numerically because the LARSP provides no scores, but it can be examined by doing LARSPs of two different language samples produced by the same child close together in time in similar circumstances and comparing their results. The aims of this experiment are to determine whether separate samples gathered from the same child 1) achieve the same overall stage estimates and 2) have similar or different distributions of productions throughout the divisions of the chart.

In order to provide new information regarding whether the LARSP is a reliable tool, this paper will examine the results of LARSP use on spontaneous language samples gathered from three typically developing children using a test-retest procedure. Further studies on the reliability of the LARSP when testing different populations (those with language delay, SLI, autism, etc.) will be necessary for clinical use with these groups, but reliability must first be examined in regard to typically developing children as knowledge of their language profile is necessary for comparison with all other groups.

2. Materials and Methods

2.1 Participants

Three children’s language samples were chosen from a larger study of typically developing children’s morpheme production (see Tommerdahl and Kilpatrick, 2013). These particular children were chosen due to their age and to ensure that both genders were represented. The ages chosen represent periods of adequately robust language production to enable reasonably rapid sample gathering and to avoid ceiling effects. Alice (pseudonyms are used for all children) is a female aged 2; 5, Beatrice is a female aged 2; 8 and Charles is a male aged 3; 2. Each child was monolingual and had normal hearing. Each parent underwent a detailed interview regarding the child’s language development, language use and the history of the child and family members that might affect language development including ear infections and any neurological difficulties. Participants were recruited from playgroups in the local community in a large city in central England.

2.2 Sample Gathering

Each of the three children was recorded twice within a week while interacting in a play situation with their mother. Recordings were carried out in a Flexible Learning Room which was transformed into a play area with a large mat with pictures of a local village on it and a box of toys for the child to choose from which included a tea set, vehicles, building blocks and toy animals. The same toys were provided for each visit. A picture of a circus was
projected onto a whiteboard. The room had four hidden cameras and five microphones which allowed a technician in the adjacent room to select the camera angle that best showed the child.

The mother and child were allowed to interact in private in the playroom. The mothers were instructed to carry on a normal conversation with the child, much as they would at home. They were also asked, according to LARSP guidelines (Crystal, 1982) to spend part of the time talking about a recent event which had occurred prior to their session. This instruction was given in order to potentially bring about a greater variety of language. The mother and child stayed in the room for an average of 35 minutes while recordings were made.

2.2 Preparation of the Sample

A trained speech and language therapist transcribed each recording orthographically, dividing samples into utterances according to provisions laid out in CHAT Transcription Format (MacWhinney and Snow, 1990). Each sample was chosen based on the following LARSP instructions:

‘Our sampling emphasis is therefore very similar to that recommended by Lee and Canter (1971), whose recipe it is to take 50 ‘complete, different, consecutive, intelligible, nonecholaic sentences elicited from a child in conversation with an adults, using stimulus materials…’ (317). But we do make a number of different decisions at various places.’ (Crystal, et al., 1976: 88).

Crystal et al. go on to discuss suggestions for sample choice regarding notions of being complete, different, consecutive, intelligible and nonecholaic. All of these suggestions were taken. Also, the first 100 utterances were not analysed, allowing the child to have a warm-up period. One additional rule made in choosing the 50 utterance sample was to not count more than the first of consecutive minor and identical responses. This means that if a child responded ‘yes’ several times in a row, only the first would be counted as forming part of the sample of 50 utterances. All of the minor responses were listed on the LARSP chart however. This decision was made to compromise between the advice given by Crystal et al. in the instructions and the need to have close to 50 major utterances when examining reliability. Also, for this reason, at the bottom of each chart, the total number of sentences will be higher than 50 as there will be 50 that correspond to this experiment’s criteria and others that are consecutive repetitions of minor responses.

The actual LARSP analysis was carried out both manually by the author and automatically with the Computerized Profiling (CP) System 9.7.0 (Long, Fey and Channell, 2008). While LARSPs produced by the CP system were similar to those carried out by hand, some examples of subordination were not included in the computerized results. In order to test the actual reliability of the LARSP on a small language sample, the crediting of every feature was important. Therefore, computerized results were used for sections B and C and the bottom section which lists total number of sentences, mean number of sentences per turn and mean sentence length in morphemes. The rest of the results showed on these charts came from human tabulation.

3. Results

LARSP charts showing the charts with their mark-ups follow from youngest to oldest. For each child, sample one is on the left and sample two is on the right.
Figures 1 and 2: LARSP charts of Alice, aged 2;5

Figures 3 and 4: LARSP charts of Beatrice, aged 2;8
Figures 5 and 6: LARSP charts of Charles, aged 3:2

As a LARSP chart is not designed to be analysed based on pure numerical characteristics, but rather as a holistic profile, results are presented in the form of a discussion of each set of LARSP charts.

Analysis of the LARSP Chart

The authors of the LARSP chart describe a list of factors regarding the linguistic analysis that is carried out after the profile is complete (Crystal, et al., 1976). First on this list is the insistence that a child’s profile, in a clinical case, be compared with the profile of other typically developing children of the same age. LARSPs of children without language disability should be the benchmark against which other children are compared. Furthermore, a child’s profile should be compared with his or her earlier profiles. These instructions underscore the need for information about LARSP reliability. In an example where a child’s language undergoes a LARSP at time A and time B, with intervening therapy, it is vital to know whether any changes in the results are due to a real change in language ability or whether it is more likely due to chance.

A major feature of LARSP analysis is to examine whether a significant amount of the language shown on the chart comes down to the stage that matches the child’s chronological age. For instance, if little language goes beyond Stage II, yet the child is 3 years of age, a delay or disorder may exist. Furthermore, both clause and phrase level markings should come to approximately the same stage level. As these are important characteristics for clinical information, they will be examined in the discussion around reliability.

Analysis of the word level is similar to that of clause and phrase level, with the top end of the word list representing inflections acquired earlier. However, they are approximations and not divided into stages. According to LARSP instructions, absences on the chart may indicate areas of therapy to be planned, however, the authors also make it clear that while ‘the presence of a score is a positive indication of ability, .the absence of a score may mean only that the sample is biased.’ (Crystal, et al., 1976:113). Scores shown on the error listing at stage VI may also indicate items that the child is currently acquiring, as they are to be marked only when the children produce errors in grammatical structures that they have shown that
they can produce correctly. These areas will each be addressed in a discussion of each child’s profiles.

3.1 Alice 2;5

Upon comparison of the two LARSPs based on Alice’s two language samples, it is visible that both charts show her level to be at stage 3 with movement beginning into stage 4. Furthermore, this is true of both clause and phrase levels which match in each instance. The expansion line linking the movement from stages III to IV each have 6 scores, albeit with different distributions. Interestingly, the individual structures marked at level III on each are highly similar, with VXY, SVC, SVO, SVA and both pronoun types being used and all others absent. The only exception to this is the auxiliary count as the first sample shows the use of a modal while the second shows two productions of a non-modal auxiliary.

Productions at level IV also bear a marked resemblance to each other in that QVS, SVOA, NegV and NegX are produced in each while the other structures are not produced. The only difference is the SVCA structure which is only produced in one sample.

At the word level, differences between the samples are more evident. The 3rd person singular is produced twice in sample 1 but not in sample 2. The contracted copula is produced in both samples but with great difference in frequency with a single production in the first sample and 6 productions in the second. Also, the contracted auxiliary is used once in the first sample only. When looking at inflections in the word column as being either present or absent, the two samples agree 10 out of 12 times.

Mean number of sentences per turn are listed as 1.2 and 1.4 in samples 1 and 2 respectively while a more striking difference exists in mean sentence length measured in morphemes by the PC programme with totals of 2.4 and 3.2.

3.2 Beatrice 2;8

Beatrice is three months older than Alice and both charts also show her to be at stage III with movement into stage IV, although this movement into stage IV is more established with Beatrice. Phrase and clause level appear to be equal in development. It is interesting to note that within any particular box, the variety of structures used is limited. Instead, the same structures are repeated. This is true in both charts. Both charts also show strong use of the expansion line between stages III and IV. The presence of 4-element clause items lies mainly within the Question category while each stage IV clause level statement box is limited to a single production.

An important area where the two charts differ is stage V which looks at the important developmental processes of acquiring coordination and subordination. Whereas the first language sample produced four instances of subordination, the second showed none.

As in Alice’s sample, more discrepancies exist at the word level. The present progressive, -ing, is used in one sample and not the other. The same is noted regarding –en, the genitive, and n’t although none of these are used more than twice in their matching sample. Overall, based on presence or absence of an inflection, the charts agree on 8 out of 12 items.

The mean number of sentences per turn is calculated to be 1.5 on the first sample and 1.1 on the second. MLU again varies greatly, being 3.4 in the first sample and 2.6 in the second.

3.3 Charles 3;2

Charles’ language is similar to the two girls’ in that stage III appears to be the most established. However, both of Charles’ LARSP charts show more advanced productions coming at both stages IV and V. Also, phrase, clause and word levels show equal development. From stage I through stage V, an imperfect, yet steady correlation exists
between the two charts in regard to which structures are used. This is true of both clause and phrase level. The expansion line between stages III and IV also show the same four out of five structures in use. Unlike the case of Beatrice, subordination usage is identical both in type and frequency in both samples.

At the word level, a greater spread of inflections used is seen although some discrepancies exist. –en is used in one sample twice and not at all in the other. 3s is not used at all in the first sample but is used five times in the second. Similarly, but in the opposite direction, the genitive is used three times in the first sample and not at all in the second. At the bottom of the word chart, -ly is used just once in one sample. Overall, based on presence or absence of a given inflection, the chart shows agreement on eight out of twelve items. This is surprising given that this child produced the largest amount, in both type and token if one can refer to morphemes in this way, of items in the word category.

Measures of mean number of sentences per turn for samples 1 and 2 were 1.2 and 1.1 while MLU was 3.7 and 3.8 respectively.

3. Discussion

Overall, the LARSP chart has demonstrated a strong degree of reliability in marking the stage at where the preponderance of a child’s productive syntax lies. In each of the three cases above, both test and retest were highly similar. Each of the cases also showed a strong correspondence between which clause and phrase level structures were used, despite differences in their frequency of usage. Expansion lines at the highest end of production also possessed a high degree of reliability. This is impressive given that a small sample of only 50 utterances was used, with some of these utterances containing single words such as ‘yes’ and ‘no’.

The LARSP was less reliable regarding complex productions reaching past the children’s established stages. This was demonstrated by the presence of subordinate clauses in one of a child’s samples, but with no productions at stage V in the other. It is perhaps telling to note that in the scoring of the language samples, certain structures appeared several times in close proximity to each other, and then ceased to appear further. Subordinate clauses were noticed to often fall into groups or clusters, as were some of these utterances containing single words such as ‘yes’ and ‘no’.

The word level was the area that most noticeably lacked reliability. This is unsurprising, considering these items are much more limited in scope than the items at clause and phrase level which could be filled by an enormous variety of utterances. These morphological items are much closer, and sometimes identical, to morphemes shown by Tommerdahl and Kilpatrick (2014; 2013) to possess low degrees of reliability both in frequency counts and in being present versus absent in a given language sample.

Also unreliable were the measures of mean number of sentences per utterance and MLU. This is of course odd, given that MLU has been shown repeatedly to be a highly reliable measure. It is likely that these results are due to the inclusion of the minor responses of ‘yes’ and ‘no’ into the overall count that was made by the CP system. Because some samples had long strings of repeated ‘no’ responses, and that the LARSP instructions did not clearly state that these should be excluded, the first of each string was analyzed in the LARSP analysis done by a human while all were included in the analyses done by the computer program. It is likely that these strings of single word responses greatly affected these measures, especially when given the short length of the sample. Given the ‘thin-ness’ of the samples of typically developing children when using 50 utterances, it seems wise to recommend the use of at least 50 utterances beyond the common responses of ‘yes’ and ‘no’. These utterances should not be wholly excluded from the analyses however, as they have value.

Overall, the LARSP, as it is meant to be used for measuring a child’s level of syntactic ability, appears to be reliable. However, caution is advised when using the LARSP output as a
tool for remediation of morphemes. As stated before, the LARSP authors have already cautioned their audience when explaining that the positive scores truly show a child’s ability while the absence of a score may simply represent a biased sample. On a deeper note, one could take the point of view that an analysis of the LARSP’s reliability is really an analysis of spontaneous language sample in terms of its morphosyntactic regularity. While several studies have found strong reliability for more global markers of language production such as MLU, number of different words and mean sentence length. (Minifie et al. 1963; Gavin and Giles 1996), studies of individual word and morpheme production show much lower reliability (Gavin and Giles 1996; Tommerdahl and Kilpatrick, 2013, 2014). It is therefore to be expected that the clause and phrase levels of the LARSP will be more reliable than the word level.

Regarding the use of the LARSP profile as grounds for developing therapy, this study shows that the LARSP is reliable in determining an individual’s degree of syntactic sophistication at the clause and phrase levels and. However, basing rehabilitation goals regarding morpheme production from a single language sample is not recommended. These three cases have shown that morphemes that have not been produced in a given sample may well be regularly produced in others. This coincides with Tommerdahl and Kilpatrick’s findings (2013) showed that even in samples of 100 utterances, 10 out of 23 samples showed that the past tense was not present when a full sample of 400 utterances showed that it was.

Again, Crystal et al. (1976) show their awareness of these qualities of language production and spontaneous sample by insisting that language profiles are best used in conjunction with standardized, formal testing and that therapy sessions use elicitation of structures absent in a spontaneous sample to better test the child’s ability. The findings of this study underscore the importance of these instructions.

Up to this point, relatively little research has been carried out regarding the reliability of morphological and syntactic production in spontaneous language production; hence, the importance of carrying out initial testing on a language profile dependent upon the regularity of language present in spontaneous samples. However, the LARSP goes beyond looking at the language of typically developing children and is meant to be used as a measure of morphosyntax production in a large number of populations, from children diagnosed with SLI to adults with aphasia post brain injury. While the body of work on the reliability of syntactic production in typically developing children is small, its counterpart in clinical populations is nearly non-existent. It is therefore crucial to test the reliability of the LARSP on different clinical populations, for if it is discovered that children with language difficulties coinciding with autism show very different profiles on different days, this will have strong implications for how the LARSP is used with this group. The importance of learning more about differences between linguistic groups in the acquisition of syntax is highlighted by recent findings showing that children with Specific Language Impairment show more difficulty in acquiring functional words that are infrequent than those that are frequent (Leroy, Parisse and Maillart, 2013). Much more work remains to be done in this area.

In closing, this examination of the LARSP, based on a small sample of three typically developing children, has produced the following results. Note that these results pertain only to typically developing children.

1) Using only 50 utterances, the LARSP is reliable at determining the level of a typically developing child’s overall morphosyntactic production.
2) At 50 utterances, the LARSP is fairly reliable in representing which syntactic structures are used at both clause and phrase levels.
3) At 50 utterances, the LARSP is less reliable regarding the production of specific morphological inflections.

Given the extensive resources that are needed to record, transcribe and analyse children’s language, these results are positive news for clinicians working in this area. This study of the reliability of typically developing children’s language clearly illuminates the need for future
work in this area on clinical populations. Such work is necessary for both using the LARSP tool appropriately, and for learning more about the characteristics of disordered language that may be of future diagnostic use.

<table>
<thead>
<tr>
<th>Name</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td></td>
</tr>
<tr>
<td>Unanalysed</td>
<td></td>
</tr>
<tr>
<td>Unint.</td>
<td>Symbolic noise</td>
</tr>
<tr>
<td>Problematic</td>
<td>Incomplete</td>
</tr>
<tr>
<td>B</td>
<td>Resp.</td>
</tr>
<tr>
<td>Stimulus type:</td>
<td></td>
</tr>
<tr>
<td>Questions</td>
<td></td>
</tr>
<tr>
<td>Others</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 1</th>
<th>Minor Responses</th>
<th>Vocatives</th>
<th>Other</th>
<th>Problems</th>
</tr>
</thead>
<tbody>
<tr>
<td>(0;9-1;6)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Majo r</td>
<td>Comm.</td>
<td>Quest.</td>
<td>Statement</td>
<td></td>
</tr>
<tr>
<td>‘V’</td>
<td>‘Q’</td>
<td>‘V’</td>
<td>‘N’</td>
<td>Other</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 2</th>
<th>Conn Clause</th>
<th>Phrase</th>
<th>Word</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1;6-2;0)</td>
<td>VX</td>
<td>QX</td>
<td>SV</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SO</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>SC</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NegX</td>
</tr>
<tr>
<td>X+S:NP</td>
<td>X+V:VP</td>
<td>X+C:NP</td>
<td>X+O:NP</td>
</tr>
<tr>
<td>X+V+A:AP</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 3</th>
<th>XXY let XY do XY</th>
<th>QXY VS(X)</th>
<th>SVC VCA</th>
<th>D Adj N</th>
<th>Cop</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2;0-2;6)</td>
<td>VXY</td>
<td>QVS</td>
<td>SVO VOA</td>
<td>AdjAdj N</td>
<td>Aux m</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SVA VOO</td>
<td></td>
<td>Pr DN</td>
<td>Aux o</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NegXY Other</td>
<td></td>
<td>Pron p</td>
<td>other</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Pron o</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XY+S:NP</td>
<td>XY+V:VP</td>
<td>XY+C:NP</td>
<td>XY+O:NP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>XY+V+A:AP</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stage 4</th>
<th>(+S VXY+ QVS)</th>
<th>QXY+ SVCA</th>
<th>AAXY NP Pr NP</th>
<th>Neg V</th>
</tr>
</thead>
<tbody>
<tr>
<td>(2;6-3;0)</td>
<td></td>
<td></td>
<td>Pr D Adj N</td>
<td>NegX</td>
</tr>
</tbody>
</table>
Acknowledgement

The author would like to thank Steven Long for his assistance in the use of his Computerized Profiling software. This project was carried out thanks to Nuffield Foundation Grant SGS/35006.

References


