The Sensitivity of Measurements when Analysing Change in Children's Speech Accuracy

A Research Project submitted to the Clinical Therapies, University of Limerick, as part of the requirement for the Master of Science in Speech & Language Therapy

May 2015

Intended Journal for Submission:

International Journal of Language and Communication Disorders

Name: Jessica Lyons

Student Number: 12011053

Supervisor: Aileen Wright

Word Count: 8,209
Acknowledgements

I would firstly like to thank my mother for her constant support and encouragement throughout my studies. I could not have done any of this without you Mum. I would also like to thank my family and my boyfriend Craig who have been a constant support during the past two years.

I would like to extend my gratitude to my classmates. It has been a wonderful experience and an honour being part of the class of 2015. I would especially like to thank my research partner Bernadette, working together throughout this project has been a pleasure and I truly could not have asked for a better research partner to work with.

I would also like to thank the parents and children who participated in the study. Without their involvement, this study would not have been possible.

Finally, I would like to convey my sincerest appreciation to my supervisor Aileen Wright for her patience, guidance, time and valuable advice throughout the course of this project.
Abstract

Background: Objective measures of children’s speech accuracy are used to track changes in children’s speech over time. Although many existing measures are available, there is a lack of consensus regarding the reliability and sensitivity of these measures, and whether segmental or whole-word analysis is a more sensitive approach.

Objectives: To determine whether the whole-word measure of proportion of whole-word proximity (PWP) is more sensitive to change than the segmental measure of percentage of consonants correct (PCC) in the speech of children with speech sound difficulties (SSD’s).

Methods: Data was drawn from twelve monolingual English speaking children (7 males; 5 females) with SSD’s aged between 3;2-4;5. The phonology subtest of the Diagnostic Evaluation of Articulation and Phonology (DEAP), and the Renfrew Action Picture Test (RAPT) were administered to gain a single word and connected speech sample at initial assessment and again eight weeks later. PCC and PWP were calculated across all samples.

Results: Results indicated that there were no significant difference between the sensitivity of PCC and PWP when analysing change in children’s speech accuracy for the sample as a whole. Individual cases are discussed to illustrate how PCC captured more change in some cases, while PWP captured more change in others. These cases are exemplified.

Conclusions: This study was limited by the small amount of change that occurred in the group between the two time points. Where change was apparent, the two measures showed differential sensitivity to different types of change. Implications for practice are discussed.
Table of Contents

Acknowledgements i
Abstract ii
Table of Contents iii
List of Tables iv
List of Figures iv
Introduction 1
  What is a sensitive measure? 1
  Importance of Sensitive Measures 1
  Evaluation of Different Methods of Measurement 3
  Aims of Investigation and Hypothesis 7
Methodology 9
  Ethical Approval and Consent 9
  Experimental Design 9
  Participants 9
  Assessment Procedure 10
  Materials and Assessments Used 10
  Measures and Analysis 11
  Reliability 12
Results 13
  Test of Normality 14
  Test of Correlation 14
  Amount of Change by Group 14
  Amount of Change by Individual Participants 15
Discussion 26
  Limitations and Indications for Further Research 26
  Conclusion and Clinical Implications 27
References 29
Appendix A: Rules for the calculation of PCC 32
Appendix B: Rules for the calculation of PMLU & PWP 34
List of Tables

**Table 1:** Materials used at each assessment time point

**Table 2:** Participant data with PCC and PWP scores for Time 1 and Time 2 assessments

**Table 3:** Correlation between PCC and PWP across samples

**Table 4:** Mean scores awarded by PCC and PWP at Time 1 and Time 2 on all assessments

**Table 5:** Mean increase in scores between Time 1 and Time 2 on each measure for the whole group

**Table 6:** Individual analysis of change in PCC and PWP from Time 1 to Time 2

List of Figures

**Figure 1:** Bar chart showing the amount of change made by individual participants’ from Time 1 to Time 2 in single words

**Figure 2:** Bar chart showing the amount of change made by individual participants’ from Time 1 to Time 2 in connected speech

**Figure 3:** Bar chart showing the amount of change made by Child 1 from Time 1 to Time 2

**Figure 4:** Bar chart showing the amount of change made by Child 3 from Time 1 to Time 2

**Figure 5:** Bar chart showing the amount of change made by Child 5 from Time 1 to Time 2

**Figure 6:** Bar chart showing the amount of change made by Child 6 from Time 1 to Time 2

**Figure 7:** Bar chart showing the amount of change made by Child 8 from Time 1 to Time 2

**Figure 8:** Bar chart showing the amount of change made by Child 9 from Time 1 to Time 2

**Figure 9:** Bar chart showing the amount of change made by Child 12 from Time 1 to Time 2
Introduction

Objective measures of children’s speech accuracy are used to track changes in children’s speech over time. Measures sensitive enough to capture small changes in speech production are important in therapy and research as it allows the Speech and Language Therapist (SLT) to accurately evaluate a child’s progress and the efficacy of an intervention. In order to meaningfully quantify change, different types of objective measures are used. These include segmental measures and whole-word measures.

Although there are many existing measures available, there is a lack of consensus regarding the validity, reliability and sensitivity of measures and whether a segmental or whole-word approach is more sensitive to change in children’s speech accuracy (Preston, et al., 2011). According to Flipsen, et al. (2005) sensitive, accurate, and viable measures of speech production for clinical practice and research are one of the continuing needs in the fields of developmental phonology and speech and language therapy.

What is a sensitive measure?

In order for a measure to be considered sensitive, clinically relevant aspects need to be captured such as the degree of phonetic accuracy in a child’s productions and small changes of improvement in the nature of the error types that a child is producing (Preston, et al., 2011). These factors are critical for a measure to be viewed as sensitive as reliable data is necessary to accurately track changes in a child’s speech accuracy.

Importance of Sensitive Measures

If a method of measuring change is not sensitive enough to accurately track the smallest of changes in a child’s speech accuracy, improvements may not be noticed by the clinician. This may affect the clinician’s judgement of a child’s speech accuracy and have an impact on their decision making process regarding service delivery such as, the effectiveness of the current therapy for producing change, and whether or not to continue or change the chosen therapy method. Therefore, SLT’s need sensitive measures of change to be better informed about children’s true phonological abilities when planning individualised treatment (Flipsen, et al., 2005).
In terms of research, there are a large amount of studies surrounding the topic of intervention for speech sound difficulties (SSD’s). However, different studies use different methods of measurement to evaluate change and the efficacy of the intervention. For example, Gierut (1989) evaluated the efficacy of the maximal opposition approach where a standard generative phonological description of the child’s speech was used to analyse speech production. This is a qualitative descriptor of the structure and function of sounds in the child’s language. In contrast, Crosbie, et al. (2005) evaluated the effect of two phonological therapies for children with severe SSD’s where changes in speech accuracy were analysed using PCC. Since no universal method of measurement has been established, results are dependent on the type of measurement chosen in the study which makes it difficult to compare outcomes of speech accuracy after intervention. Therefore, if the chosen method of measurement is more sensitive to change, the more effective an intervention may appear to be in comparison to another.

Although a high degree of sensitivity is a much sought after quality in a measure of speech accuracy, feasibility and ease of use have also to be discussed. This is a fundamental component in order for a measure to be used practically in the clinical context and for research purposes. For example, the Weighted Speech Sound Accuracy (WSSA) measure, developed by Preston et al. (2011), uses Logical International Phonetics Program (LIPP) software to differentially weight different types of speech sound errors according to various levels of phonetic accuracy (i.e. atypical errors are weighted more heavily than more common errors). Although Preston et al. (2011) have shown WSSA to be sensitive to the types of errors produced by children, the use of computerised software does not fulfil the practical aspect needed in a measure to allow ease of use from a clinical perspective. A computerised measure may not be accessible to an offsite clinician and furthermore, the expense of purchasing such software is an additional factor as the LIPP can cost up to US$1,000 where system requirements are also necessary. Therefore, when evaluating different methods of measurements, practicality and ease of use are necessary qualities in order for a measure to be widely used and accepted.
Evaluation of Different Methods of Measurement

Subjective Measures
Informal severity ratings are largely used in clinical practice due to their utility in the busy workload of clinicians. According to Flipsen, et al. (2005), the use of impressionistic judgements made by experienced clinicians take into account the number, type, and consistency of errors as well as intelligibility. However, the use of these subjective measures for determining severity in children is problematic. Flipsen, et al. (2005) examined the use of these subjective measures where ten SLT’s were asked to determine the severity of children’s speech production. They found that the overall variability of ratings analysed yielded an intraclass correlation value of .405 which reflected poor to fair agreement among the SLT’s. Flipsen, et al. (2005) concluded that impressionistic judgements were unreliable and among the lowest levels of evidence to be used in clinical decision making and in research. Therefore, the use of objective measures such as PCC, Proportion of Whole-word Proximity (PWP) or Whole-Word Accuracy (WWA), among others, were advised as more valid and reliable approaches. However, associating a meaningful numeric value to phonetic transcriptions of children’s speech has proven to be a tremendous challenge (Preston, et al., 2011).

Segmental Measures
Productive Phonological Knowledge (PPK): Elbert and Gierut (1986 cited in Flipsen, et al. 2005) described a procedure for calculating PPK of individual phonemes. PPK assigns a child’s production of each consonant into one of six types, ranging from adultlike to non-adultlike. For example, a child displaying type 1 knowledge of the [s] phoneme produces [s] correctly in all word positions at all times e.g. [sʌn], [mesɪ], [kɪs]; whereas type six knowledge, identified as non-adultlike, is where a child produces the [s] phoneme incorrectly in all word positions at all times e.g. [tʌn], [mɛtɪ], [kɪt]. Each type is determined according to the distribution of sounds in the child’s speech, the presence of phonological rules, and the nature of the child’s underlying representations as correct (adultlike) or incorrect (non-adultlike) (Williams, 2003). However, PPK does not fulfil the practicality component required by a measure to be used successfully in the clinical context as determining PPK requires a large sample of 311 single words. The same difficulties can
occur in research as Flipsen, et al. (2005) reported that the measure of PPK was not included in their study as the large single word sample required for PPK could not be elicited for the 17 children in the study.

**Percentage of Phonemes Correct (PPC) & Percentage of Vowels Correct (PVC):** Shriberg and colleagues have proposed a number of segmental measures such as PPC, PVC and PCC. PPC and PVC are not as commonly used in comparison to PCC. This is likely due to the measurement of vowel correctness in these measures. Accuracy of vowels has received much less attention than consonant accuracy as vowel errors are less frequent in children’s speech (Speake, et al., 2012). Furthermore, Ingram (2002) states that reliability of vowel transcriptions is typically poor and agreement of vowel correctness between transcribers is problematic. For the purpose of this study and due to its high acceptance in the literature and clinical practice, PCC is the chosen method of segmental measurement. A description of PCC is given with advantages and limitations outlined below.

**Percentage of Consonants Correct (PCC):** PCC is one of the most well-known and well established measures used both clinically and in research (Shriberg et al. 1997 cited in Saaristo-Helin, 2009). PCC compares the child’s production to the target on a phoneme by phoneme basis by dividing the total number of correct consonants by the total number of consonant targets and multiplying by 100 for an overall percentage of consonant correctness (Stoel-Gammon, 2010). For example, if a child produces /frog/ as [fwog], a PCC score of two would be awarded for the correct /f/ and /g/ consonants in the child’s production. Therefore, the child’s earned score of two is divided by the target PCC score of 3 and multiplied by 100, resulting in a PCC score of 66.6 i.e. 67% accuracy. PCC is frequently cited in research studies and referred to as the index of severity in many textbooks on SSD’s (e.g. Bauman-Waengler, 2008; Flipsen, et al., 2005). Due to its high acceptance and use in the literature, this gives the ability to compare effects of different studies and interventions when the same method of measurement has been used.

Newbold, et al. (2013) compared PCC to five commonly used measures of speech output which covered different levels of analysis (i.e. phonetic, phonological, segmental and word-based analysis). Data was drawn from four children with severe and persisting SSD’s at 4 years of age and again at 6 years of age. They concluded that in comparison to the
other methods of measurement, PCC was a more sensitive and robust indicator of change that was less affected by the choice of stimuli. Similar findings were made in a study by Almost and Rosenbaum (1998) where PCC was used as an outcome measure in a randomised controlled trial of 30 children with severe speech disorders. Following a four month period of intervention, results showed a significant difference in PCC between treated and untreated groups.

However, concerns have been raised about the sensitivity of PCC. PCC is based on the transcribers’ judgments of speech sounds. Distortions such as dentalisation and lateralisation are scored the same as other error types (e.g. consonant deletions) even though they are unlikely to impact on intelligibility to the same extent (Flipsen, et al., 2005). Furthermore, PCC does not allow for different scoring of different error types. For example, a child may omit a consonant at first (e.g. [fɒɡ] for /frɒɡ/) and then progress to substituting the previously deleted phoneme (e.g. [fʰwɒɡ] for /frɒɡ/), yet the child still achieves the same score for accuracy even though the child has achieved a closer approximation to the target which therefore increases intelligibility of the given word. This type of progression is commonly seen in children with SSD’s but the lack of sensitivity of this measurement does not allow for these progressions to be quantified. Furthermore, Hall, et al. (1998 cited in Newbold, et al. 2013) discourage the use of PCC as an outcome measure of intervention due to PCC not being sensitive enough to articulatory changes.

**Whole-Word Measures**

Although segmental measures provide important information regarding children’s speech accuracy at the level of the phoneme, whole-word measures focus on children’s ability to approximate whole words (Burrows & Goldstein, 2010). It may be possible that whole-word measures show a difference in production that is not recognised by segmental measures. Velleman and Vihman (2002 cited in Watson & Terrell, 2012) suggested that children’s early word productions are lexically driven with little attention to individual speech segments. These early word productions are seen as an attempt to approximate the adult target. As children are exposed to more and more language, they start to notice patterns and regularities of sounds which they begin to incorporate into their own word productions. However, the need to approximate whole words continues to have an
impact on children’s phonological development (Bunta, et al., 2009). A number of potential whole-word accuracy measures are outlined below where advantages and limitations are discussed with reference to the literature.

**Whole-Word Accuracy (WWA) & Proportion of Whole-Word Correctness (PWC):** The WWA measurement proposed by McCabe and Bradley (1973 cited in Flipsen, et al., 2005) evaluates the ratio of correct words that contain no errors over the total number of words produced by a child in a sample. WWA has received little attention in research but is the equivalent to the more well-known measurement of PWC (Ingram, 2002). In the study conducted by Flipsen, et al. (2005), data was collected from 17 children aged between 2;11 and 5;3 with speech delays. They found significant correlations between PWC and consensus ratings of severity made by experienced clinicians. However, while WWA and PWC may be useful as measures of severity, they are not valuable measures of change since small changes, that are nonetheless significant, will not be apparent using these measures. Newbold, et al. (2013) also referred to PWC as a ‘blunt instrument’ as it only allows a child’s production to be either correct or incorrect.

**Phonological Mean Length of Utterance (PMLU):** The measurement of PMLU is highly sensitive to the structure of phonological words by measuring the number of correct consonants produced and the ability to produce words that are complex in length (Preston, et al., 2011). To calculate PMLU, the child receives one point for every consonant and vowel produced (regardless of accuracy), and one point for each consonant in the correct position. For example, if a child produces /frog/ as [fwog], a total of six points are scored out of a possible seven. Four points are received for each vowel and consonant produced, plus an additional two points for /f/ and /g/ as they are both in the correct position (Ingram & Ingram, 2001). According to Ingram (2002, p. 715), points for correct consonants “distinguish between children who might have similarly long words but who differ in the correctness of their productions”. Ingram (2002) reported results from the Flipsen et al. (2001) study which found that PMLU is a better predictor of the severity of a child’s speech impairment than PCC.

As previously mentioned with PCC, PMLU is also influenced by transcribers’ judgments of speech sounds, but this is to a lesser extent. As illustrated earlier, PCC does not award a
point for an incorrect consonant. Likewise with the scoring of PMLU, a point is also not scored here for accuracy, but one point is awarded for the position of that incorrect consonant which affects the child’s score to a lesser extent as positioning of the consonant (regardless of accuracy) has also been taken into account. For example, if a child produces /frɒg/ as [fɒg], a total of five points are scored out of a possible seven. Three points are received for each vowel and consonant produced, plus an additional two points for /f/ and /g/ as they are both in the correct position. If the child then progresses to producing /frɒg/ as [fɒg], this is awarded a higher score of six points. Therefore, PMLU has shown to be sensitive to the improvement in structure made by the child. However, in the example of a child’s production of the word /fɪʃɪŋ/ as [fɪtɪŋ] at T1 and [fɪsɪŋ] at T2, PMLU does not detect the closer approximation to the /ʃ/ phoneme at T2 which was also seen by PCC. Although PMLU provides important information about a child’s phonological skills since the complexity of target words is considered, the value of deriving a PMLU analysis may be strengthened by also including the measure of PWP (Watson & Terrell, 2012).

**Proportion of Whole-Word Proximity (PWP):** PWP is a whole-word measure directly related to PMLU that captures the child’s ability to approximate the adult target i.e. the degree of accuracy in producing words (Burrows & Goldstein, 2010). It is calculated by dividing the child’s earned PMLU score by the PMLU of the corresponding target word (Arias & Lleó, 2014). Using the previous example illustrated above where the child gained a PMLU of six for the word [fɪwɒg], proximity value is calculated by dividing the child’s earned score of six by the target PMLU (seven in this example), resulting in a PWP score of 0.86, i.e. 86% accuracy. In the study conducted by Newbold, et al. (2013) which evaluated five different measures of speech output, PCC and PWP were evaluated to be the most accurate indicators for capturing change over time.

**Aims of Investigation and Hypothesis**
In summary, less information is available surrounding the topic of whole-word patterns in comparison to the segmental aspects of speech acquisition during children’s development. Therefore less research has been conducted to examine children’s whole-word correctness as a measure of change in speech accuracy. Previous research has
provided preliminary findings on the appropriateness of the measure of PWP as a potentially valuable and sensitive measurement. The current study aims to compare the sensitivity of the widely used and accepted segmental measure of PCC with the well-known and supported whole-word measure of PWP when evaluating change in children’s speech accuracy over a period of time. Results will determine whether a whole-word measure is more sensitive to change than a segmental measure in children’s speech accuracy. Therefore the following research question has been proposed:

➤ Is the use of the whole-word measure of PWP more sensitive to change than the segmental measure of PCC as a measurement of speech accuracy?

The following is hypothesised:

➤ The use of PWP will be a more sensitive measurement than PCC when analysing change in children’s speech accuracy over a period of time as it reflects both segmental accuracy and phonological whole-word approximation to give an overall value of a child’s speech accuracy.
Methodology

Ethical Approval and Consent
Ethical approval for this study was granted by the Research Ethics Committee, University Hospital Limerick. Informed written consent was obtained from the parents prior to their child’s participation in the study.

Experimental Design
This study used a pre- and post-test design. As part of a wider study, each participant was randomised into a ‘treatment now’ and ‘treatment later’ group following T1 assessment. This was in order to partake in an intervention programme which was carried out by parents for eight weeks. T2 assessments took place following the eight week period where all of the participants’ speech was re-evaluated. No significant change occurred after the completion of the intervention programme. Therefore, the ‘treatment later’ group was also included in this study due to similar gains in speech accuracy as a result of maturation. The tests of speech output were the Diagnostic Evaluation of Articulation and Phonology (DEAP; Dodd, et al. 2002) to elicit single words, and the Renfrew Action Picture Test (RAPT; Renfrew, 1997) to elicit a constrained connected speech sample.

Participants
The study was completed with the cooperation of two Speech and Language Therapy (SLT) clinics in the mid-west of Ireland. The principal investigator contacted the two SLT services who sourced children from their waiting lists. These children were all referred to their service due to SSD’s. The SLT’s contacted parents of the children and sent them information about the study, a case history form and a consent form. Parents who were interested in taking part in the study sent the case history form and consent form back to their local SLT which were subsequently sent to the principle investigator of the study. From the group of parents interested in the study, a total of thirteen children took part. Inclusion criteria were: aged between 3;0 and 4;11; referred to SLT service due to SSD’s. Exclusion criteria were: bilingualism; speech deficits due to structural or organic causes; additional diagnosis of cognitive or sensory difficulties. One of the participants had to be excluded from the study as the assessments could not be completed during the
assessment session due to attention and behavioural difficulties. Therefore, this participant had to be excluded as sufficient data was not collected to include in the study. This reduced the number of participants to twelve. These children ranged in age from 3;2 to 4;5 (M = 3;10) and included seven males and five females.

**Assessment Procedure**

Appointments were situated in two SLT clinics in the mid-west of Ireland for both the T1 and T2 assessments. Each participant attended the clinic in their local area and was assessed individually in a quiet clinic room where each session lasted approximately 45 minutes. Each child had one or both parents present throughout the assessment session. As part of a wider study, a total of eight student Speech and Language Therapists (sSLT’s) were part of the data collection and analysis stage of this study. Two sSLT’s were present at each assessment session. At the beginning of the session, the assessing sSLT gathered the case history form and the consent form that were sent out prior to the assessment session. The assessing sSLT reviewed the case history form with the parent(s) to ensure all relevant information was collected. Parents then signed an audio release form to allow the assessment to be recorded. An Olympus Digital Voice Recorder VN5500PC was used for the audio recording. This recording was used to check and complete the transcription after the assessment session. The assessing sSLT and child were seated side by side at a table appropriate for the child’s height during the administration of the assessments. The second sSLT observed from a distance and co-transcribed the child’s responses during the assessments.

**Materials and Assessments Used**

**DEAP Diagnostic Screener:** The diagnostic screener of the DEAP was administered at T1 in order to evaluate each child’s single word production, phoneme stimulability, and single word inconsistency to indicate what further subtests were necessary. The phonology subtest was indicated for all twelve children.

**DEAP Phonology subtest:** The phonology subtest was administered to gain a single word sample of each child’s speech. This assessment consists of picture naming which involves the elicitation of fifty single words (Dodd, 2005).
**RAPT:** The RAPT was administered to gain a constrained connected speech sample to allow for analysis of speech accuracy in connected speech. The RAPT consists of ten picture cards that contain different scenes with accompanying questions for the assessor to elicit a response from the child such as “What has happened to the girl?”. When analysing a connected speech sample, colleagues of Shriberg have expressed concerns of the lack of standardization relative to the responses elicited in connected speech (Shriberg, et al. 1997). Therefore, the use of the RAPT during T1 and T2 provided a standard set of stimulus materials which aimed to provide content stability for the connected speech sample. The picture cards also provided context to help the assessor interpret unintelligible utterances produced by some of the children.

**Additional Assessments:** As part of a separate research project, a test of real word discrimination and a test of non-word discrimination were carried out at T1. A FOCUS (Focus on the Outcomes of Communication Under Six) questionnaire and Likert scale for the evaluation of the intervention programme were also filled out by the parent(s). Results of these assessments are not reported in the current study (see Barrett 2015, Kelleher 2015, Kerrigan 2015, Martin 2015, Quinlan, 2015).

<table>
<thead>
<tr>
<th>Time 1</th>
<th>Time 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case history</td>
<td>DEAP Phonology subtest</td>
</tr>
<tr>
<td>Consent form</td>
<td>RAPT</td>
</tr>
<tr>
<td>Audio release form</td>
<td>FOCUS Questionnaire</td>
</tr>
<tr>
<td>DEAP Diagnostic Screener</td>
<td>Likert Scale</td>
</tr>
<tr>
<td>DEAP Phonology subtest</td>
<td></td>
</tr>
<tr>
<td>RAPT</td>
<td></td>
</tr>
<tr>
<td>Test of non-word discrimination</td>
<td></td>
</tr>
<tr>
<td>Test of real word discrimination</td>
<td></td>
</tr>
<tr>
<td>FOCUS Questionnaire</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: Materials used at each assessment time point

**Measures and Analysis**

To address the research question, the phonology subtest of the DEAP and connected speech sample (obtained from the RAPT) were used to calculate and compare the changes in PCC and PWP across all twelve samples from T1 and T2. PCC was calculated using the method outlined by Shriberg et al. (1982 cited in Smit, 2004; see appendix A:
rules of calculation for PCC). PMLU and PWP were calculated in accordance to Ingram’s (2002) rules for calculation (see appendix B: rules for the calculation of PMLU and PWP).

Reliability

**Phonemic Transcriptions**

During each assessment session, both the assessing and observing sSLT made online transcriptions of the children’s speech during the administration of the DEAP subtest. To ensure accuracy of transcriptions, all speech data was recorded using an Olympus Digital Voice Recorder VN5500PC. The speech data from the DEAP subtest was re-transcribed by both sSLT’s via audio review. The connected speech sample was then transcribed on review of the audio recordings by both sSLT’s. If a disagreement occurred between the two transcribers, the assessing sSLT made the final decision.

**Analysis Procedure**

Reliability analysis for carrying out the calculation of PCC and PWP consisted of having the manual counts repeated on three occasions for all samples. After the two sSLT’s present for the assessment session completed and agreed on transcriptions and scoring of PCC and PWP values, two sSLT’s that were not present for the assessment session individually reanalysed all values awarded by both measures in single words and connected speech. This lead to an overall of three PCC and PWP values for each assessment i.e. original calculation, re-check 1 and re-check 2. If disagreement occurred, the scoring was reviewed until consistency was reached.
# Results

<table>
<thead>
<tr>
<th>Child</th>
<th>Gender</th>
<th>Age</th>
<th>Treatment Condition</th>
<th>Single Words</th>
<th></th>
<th>Connected Speech</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>PCC</td>
<td>PWP</td>
<td>PCC</td>
<td>PWP</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Time 1</td>
<td>Time 2</td>
<td>Time 1</td>
<td>Time 2</td>
</tr>
<tr>
<td>1</td>
<td>M</td>
<td>3;6</td>
<td>Now</td>
<td>75</td>
<td>84</td>
<td>88</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>M</td>
<td>3;9</td>
<td>Later</td>
<td>65</td>
<td>65</td>
<td>81</td>
<td>83</td>
</tr>
<tr>
<td>3</td>
<td>F</td>
<td>4;5</td>
<td>Later</td>
<td>71</td>
<td>82</td>
<td>87</td>
<td>93</td>
</tr>
<tr>
<td>4</td>
<td>F</td>
<td>3;11</td>
<td>Now</td>
<td>62</td>
<td>62</td>
<td>77</td>
<td>77</td>
</tr>
<tr>
<td>5</td>
<td>M</td>
<td>3;2</td>
<td>Now</td>
<td>29</td>
<td>33</td>
<td>54</td>
<td>62</td>
</tr>
<tr>
<td>6</td>
<td>F</td>
<td>3;4</td>
<td>Later</td>
<td>55</td>
<td>50</td>
<td>77</td>
<td>73</td>
</tr>
<tr>
<td>7</td>
<td>F</td>
<td>3;11</td>
<td>Later</td>
<td>50</td>
<td>49</td>
<td>67</td>
<td>64</td>
</tr>
<tr>
<td>8</td>
<td>F</td>
<td>3;6</td>
<td>Now</td>
<td>63</td>
<td>61</td>
<td>81</td>
<td>82</td>
</tr>
<tr>
<td>9</td>
<td>M</td>
<td>4;2</td>
<td>Now</td>
<td>63</td>
<td>62</td>
<td>79</td>
<td>81</td>
</tr>
<tr>
<td>10</td>
<td>M</td>
<td>4;1</td>
<td>Later</td>
<td>64</td>
<td>71</td>
<td>81</td>
<td>83</td>
</tr>
<tr>
<td>11</td>
<td>M</td>
<td>3;8</td>
<td>Now</td>
<td>80</td>
<td>82</td>
<td>88</td>
<td>90</td>
</tr>
<tr>
<td>12</td>
<td>M</td>
<td>4;5</td>
<td>Later</td>
<td>59</td>
<td>60</td>
<td>73</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 2: Participant data with PCC and PWP scores for Time 1 and Time 2 assessments (M = male; F = female)
Test of Normality

Although a Shapiro-Wilk test suggested the data was normally distributed, due to the small sample size, statistical analysis was carried out using non-parametric tests.

Test of Correlation

Spearman's rank-order correlation was run to determine the relationship between PCC and PWP. Results indicated that there was a strong, positive correlation (correlation coefficient) between PCC and PWP for all samples. This confirms that both measures are measuring the same factor i.e. speech accuracy (see table 3).

<table>
<thead>
<tr>
<th></th>
<th>Correlation Coefficient</th>
<th>Significance</th>
<th>Strength of relationship</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single words T1</td>
<td>.973</td>
<td>.000</td>
<td>Strong</td>
</tr>
<tr>
<td>Single words T2</td>
<td>.970</td>
<td>.000</td>
<td>Strong</td>
</tr>
<tr>
<td>Connected speech T1</td>
<td>.993</td>
<td>.000</td>
<td>Strong</td>
</tr>
<tr>
<td>Connected speech T2</td>
<td>.935</td>
<td>.000</td>
<td>Strong</td>
</tr>
</tbody>
</table>

Table 3: Correlation between PCC and PWP across samples

However, there was considerable difference between the scores awarded by PCC and PWP on both assessments at both time points (see table 4 for mean scores). The PWP scores were always greater than the PCC scores for all children. This shows that PWP attributes greater accuracy than PCC to productions.

<table>
<thead>
<tr>
<th></th>
<th>Single Words</th>
<th>Connected Speech</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PCC</td>
<td>PWP</td>
</tr>
<tr>
<td>Time 1</td>
<td>61</td>
<td>78</td>
</tr>
<tr>
<td>Time 2</td>
<td>63</td>
<td>80</td>
</tr>
</tbody>
</table>

Table 4: Mean scores awarded by PCC and PWP at Time 1 and Time 2 on all assessments

Amount of Change by Group

To explore whether PWP showed more difference than PCC between samples at T1 and T2, the mean difference between each score was calculated for each measure. Results showed that there were small amounts of change across the sample (see table 5).
Further investigation using a Wilcoxin Signed Rank Test was run to explore whether any significant difference occurred between the amount of change captured by PCC in comparison to PWP. Results showed no significant difference in the amount of change captured by PCC in single words compared to PWP in single words ($z = -0.205, p = .838$). Furthermore, no significant difference was apparent in the amount of change captured by PCC in connected speech compared to PWP in connected speech ($z = -0.472, p = .637$).

The investigation of the mean values showed that both PCC and PWP showed small amounts of change across the sample between T1 and T2 (see table 5 above). As neither measure showed significant change between the two time points, the following conclusions could be reached:

1. PWP is no more sensitive to change than PCC when capturing change in speech accuracy over a period of time.
2. The amount of change demonstrated by the group was not sufficient to show differences in sensitivity between PCC and PWP.

If no change was evident across the group, neither measure would be expected to be more sensitive than the other. Therefore, it can be concluded that the amount of change demonstrated by the group was not sufficient to show any differences in sensitivity between the two measures. As no sufficient change was shown at group level, examples of individuals who had made significant changes between T1 and T2 were sought to allow for comparison of PCC and PWP to be made.

**Amount of Change by Individual Participants**

Statistical analysis of the changes in each child’s speech accuracy at single word and connected speech level were analysed using a two-tailed Fisher’s exact test. Significant change occurred in Child 1, Child 3 and Child 5’s speech accuracy at single word level (see
Child 6, Child 8, Child 9 and Child 12 made significant change in speech accuracy at connected speech level (see figure 2).

Figure 1: Bar chart showing the amount of change made by individual participants’ from Time 1 to Time 2 in single words

Figure 2: Bar chart showing the amount of change made by individual participants’ from Time 1 to Time 2 in connected speech
Children who showed change at single word level

*Child 1 (T1 data: male; CA: 3;6)*

A significant change in PWP at single word level was confirmed by Fisher’s exact test ($p = 0.01093$). Although change in PCC appears bigger (see figure 3), this did not prove to be significant. Child 1’s PCC increased by 9% (75-84%) and PWP increased by 5% (88-93%). Therefore, PCC was more sensitive than PWP in capturing change in this case.

![Figure 3: Bar chart showing the amount of change made by Child 1 from T1 to T2](image)

Overall, Child 1 corrected 11 substitution errors and 3 structural errors from T1 to T2. Therefore, PCC was more sensitive in capturing change in substitutions. The increased sensitivity of PCC can be seen in the following examples:

At T1 Child 1’s production of /elafant/ was produced as [elabant]. This was awarded a score of 75% accuracy in PCC. Three points were awarded for the correct [l], [n] and [t] consonants. The child’s earned score of three was then divided by the target PCC score of four and multiplied by 100, resulting in the PCC score of 75%. PWP awarded a score of 91% accuracy. A total of ten points were scored out of a possible eleven. Seven points were received for each vowel and consonant produced, plus an additional three points for [l], [n] and [t] as they are all in the correct position. The child’s earned score of ten was then divided by the target score of eleven, resulting in a PWP score of 91%. At T2 [elafant] was produced with 100% accuracy, increasing by 25% in PCC and only 9% in PWP (see table 6 for calculations). Therefore, greater change was seen in PCC as this child corrected
errors in substitutions from T1 to T2; whereas PWP, having already awarded partial credit for the position of the incorrect /b/ phoneme at T1, could only credit an increase of 9% accuracy at T2 to capture replacement of an error phoneme by a correct phoneme.

<table>
<thead>
<tr>
<th>Child’s Production</th>
<th>PCC Accuracy</th>
<th>PWP Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1: [ɛləbənt]</td>
<td>3/4 = 75%</td>
<td>10/11 = 91%</td>
</tr>
<tr>
<td>T2: [ɛləfant]</td>
<td>4/4 = 100%</td>
<td>11/11 = 100%</td>
</tr>
<tr>
<td>Change</td>
<td>25% increase</td>
<td>9% increase</td>
</tr>
</tbody>
</table>

Table 6: Individual analysis of change in PCC and PWP from Time 1 to Time 2

Additional examples can be seen in the child’s production of /glʌvz/ as [glʌbð] at T1 and [glʌbz] at T2, and in the production of /brɛd/ as [bwɛd] at T1 and [brɛd] at T2. These result in similar scores as shown above where PWP could only make minimal gains in comparison to PCC for the correction of substitutions at T2.

**Child 3 (T1 data: female; CA: 4;5)**

Fisher’s exact test indicated that Child 3 made significant change at single word level in both PCC ($p = 0.03674$) and PWP ($p = 0.014291$). PCC increased by 11% (71-82%) and PWP increased by 6% (87-93%). Although significant change occurred at single word level for both measures, PCC was more sensitive in capturing change for this child (see figure 4).

![Figure 4: Bar chart showing the amount of change made by Child 3 from Time 1 to Time 2](image-url)
Similarly to Child 1, Child 3 also made fewer phoneme substitution errors at T2 than at T1. Overall, Child 3 corrected 12 substitution errors and 1 structural error at single word level. Therefore, PCC was more sensitive in capturing change in substitutions for this child. The increased sensitivity of PCC can be seen in the following examples:

At T1 /kræb/ was produced as [kwæb] which gave a score of 67% accuracy in PCC. Two points were awarded for the correct [k] and [b] consonants. The child’s earned score of two was then divided by the target PCC score of three and multiplied by 100, resulting in the PCC score of 67%. PWP awarded a score of 86% accuracy. A total of six points were scored out of a possible seven. Four points were received for each vowel and consonant produced, plus an additional two points for [k] and [b] as they are both in the correct position. The child’s earned score of six was then divided by the target score of seven, resulting in a PWP score of 86%. At T2, Child 3 produced /kræb/ with 100% accuracy. This led to an increase of 33% in PCC but only 14% in PWP. Therefore, PCC showed greater sensitivity to change in this case.

This is also illustrated in the child’s production of /tuθbrʌʃ/ as [tuθbwʌs] at T1 and [tuθbrʌʃ] at T2; and in the production of /zɛbra/ as [zɛfrə] at T1 and [zɛbra] at T2. These examples would result in similar scores as outlined above.

Child 5 (T1 data: male; CA: 3;2)

Fisher’s exact test confirmed that Child 5 made a significant increase in PWP at single word level (p = 0.035611). PCC increased by 4% (29-33%) and PWP increased by 8% (54-62%). Therefore, PWP was more sensitive in capturing change for this child (see figure 5).

Analysis of this child’s single word productions found that Child 5 corrected 1 substitution error and 8 structural errors from T1 to T2. Therefore, PWP was more sensitive in capturing structural change for this child. The increased sensitivity of PWP can be seen in the following examples:

At T1 Child 5’s production of /faw/ was produced as [faɪ]. This was awarded a score of 50% accuracy in PCC. One point was awarded for the correct [f] consonant. The child’s earned score of one was then divided by the target PCC score of two and multiplied by 100, resulting in the PCC score of 50%. However, PWP awarded a score of 60% accuracy.
A total of three points were scored out of a possible five. Two points were awarded for the diphthong and consonant produced, plus an additional point for the [f] consonant as it is in the correct position. The child’s earned score of three was then divided by the target score of five, resulting in the PWP score of 60%. At T2, /faɪv/ was produced as [faɪz], increasing by 20% accuracy in PWP. However, no change was captured by PCC. Therefore, PWP showed a greater increase in accuracy, as PWP awarded credit for the gains in structure; however, PCC awarded no credit due to the addition of an incorrect phoneme at T2.

These changes can also be seen in the child’s production of /θri/ as [fi] at T1 which increased in accuracy to [fwi] at T2. Similarly, the child’s production of /treɪn/ as [ti] at T1 increased in accuracy to [tseɪn] at T2. Both PCC and PWP awarded credit for the addition of the final /n/ consonant in [tseɪn]; however, PWP also awarded further credit for the inclusion of the incorrect /s/ phoneme in the initial consonant cluster.

Children who showed change at connected speech level

Child 6 (T1 data: female; CA: 3;4)

Fisher’s exact test indicated that Child 6 made significant change at connected speech level in PWP (p = 0.032913). Despite the fact change in PWP proved to be significant, the change in PCC appears to be bigger (see figure 6). Child 6’s PCC increased by 10% (67-
and PWP increased by 6%, (79-85%). Therefore, Child 6’s PCC score was more sensitive than PWP in capturing an increase in change for this child.

The increased sensitivity of PCC can be seen in the following examples:

At T1 Child 6’s production of /kæt/ was produced as [kæx]. This rendered a score of 50% accuracy in PCC. One point was awarded for the correct [k] consonant. The child’s earned score of one was then divided by the target PCC score of two and multiplied by 100, resulting in the PCC score of 50%. PWP awarded a score of 80% accuracy. A total of four points were scored out of a possible five. Three points were received for each vowel and consonant produced, plus an additional point for the [k] consonant as it is in the correct position. The child’s earned score of four was then divided by the target score of five, resulting in the PWP score of 80%. At T2, /kæt/ was produced with 100% accuracy, increasing by 50% in PCC but only 20% in PWP. Therefore, PCC showed a greater increase in change; whereas PWP, having already awarded partial credit for the position of the incorrect phoneme /x/ at T1, could only make a gain of 20% at T2.

This can also be seen in the child’s production of /glæsæz/ as [gwædæs] which increased in accuracy to [gwæsæz] at T2. Therefore, as more substitution errors were corrected than structural errors, PCC was more sensitive in capturing change for this child.
**Child 8 (T1 data: female; CA: 3;6)**

Fisher’s exact test confirmed a significant change in PWP at connected speech level (p = 0.006813). Child 8’s speech accuracy actually decreased from T1 to T2. Although the change in PWP was significant, the change measured by PCC and PWP only differed by 1%, with PCC showing a slightly greater amount of change. PCC decreased by 9% (82-73%) and PWP decreased by 8% (90-82%). Therefore, PCC and PWP were both sensitive in capturing change where PCC showed a 1% increase in sensitivity in this case (see figure 7).

![Amount of change made by Child 8 from T1 to T2](image)

Figure 7: Bar chart showing the amount of change made by Child 8 from Time 1 to Time 2

Examples to illustrate the difference in scoring of PCC and PWP were limited due to the fact that the words produced at T1 were considerably different to the words produced at T2 in the connected speech sample. However, the increased sensitivity of PCC can be seen in the following example:

Child 8’s production of /glæsəz/ was produced as [glæsəs] at T1 which gave a score of 75% accuracy in PCC. Three points were awarded for the correct [g], [l] and medial [s] consonant. The child’s earned score of three was then divided by the target PCC score of four and multiplied by 100, resulting in the PCC score of 75%. PWP awarded a score of 90% accuracy. A total of nine points were scored out of a possible ten. Six points were received for each vowel and consonant produced, plus an additional three points for [g], [l] and medial [s] as they are all in the correct position. The child’s earned score of nine was then divided by the target score of ten, resulting in a PWP score of 90%. At T2,
/glæsəz/ was produced as [glæθə] which led to a 25% decrease in PCC and a 20% decrease in PWP. Therefore, PCC was more sensitive, but by only 5% more than PWP. This is due to the fact that PWP still gave credit for the position of the [θ] phoneme at T2 even though it was incorrect. However, PCC deducted all credit for the substitution of the correct phoneme with the incorrect [θ] phoneme at T2. Therefore, PCC appeared to be more sensitive in capturing change in the decrease in accuracy for this child.

**Child 9 (T1 data: male; CA: 4;2)**

Fisher’s exact test indicated that Child 9 made significant change at connected speech level in PWP ($p = 0.04721$). Child 9’s PCC increased by only 1% (72-73%), while PWP increased by 7% (80-87%). Therefore, Child 9’s PWP score was more sensitive in capturing an increase in change from T1 to T2 (see figure 8).

![Figure 8: Bar chart showing the amount of change made by Child 9 from T1 to T2](image)

Similarly to Child 8, Child 9 produced considerably different vocabulary in the connected speech sample from T1 to T2, so straight comparisons of word productions are not possible. For example, at T1 the child used the word ‘teddy’ when describing a picture at T1, whereas ‘bear’ was used at T2. Likewise, the word ‘lady’ was used at T1, whereas ‘girl’ was used at T2. Furthermore, fewer words were produced at T2 during the recording of the connected speech sample. These factors made it impossible to compare word productions and provide an explanation for the increased sensitivity of PWP in this case.
Child 12 (T1 data: male; 4;5)

Fisher’s exact test confirmed that Child 12 made a significant increase in PWP in connected speech (p = 0.021717). At T1, this child’s PCC was 58% and PWP was 68%. However, at T2 PCC showed no increase, whereas PWP increased by 8%. Therefore, PWP was more sensitive in capturing change for this child (see figure 9).

The increased sensitivity of PWP can be seen in the following examples:

Child 12’s production of /dʒʌmp/ was produced as [dʌmt] at T1 which rendered a score of 50% accuracy in PCC. Two points were awarded for the correct [m] and [t] consonants. The child’s earned score of two was then divided by the target PCC score of four and multiplied by 100, resulting in the PCC score of 50%. PWP awarded a score of 67% accuracy. A total of six points were scored out of a possible nine. Four points were received for each vowel and consonant produced, plus an additional two points for [m] and [t] as they are both in the correct position. The child’s earned score of six was then divided by the target score of nine, resulting in a PWP score of 67%. At T2, /dʒʌmp/ was produced as [dʌmped]. PCC did not show any change in accuracy as the correct [p] consonant was added but the correct [t] consonant was lost. However, PWP showed an 11% increase in accuracy as credit was still given to the inclusion of a final consonant (regardless of accuracy). Therefore, PWP showed greater sensitivity to the increase in accuracy as PCC did not award any credit for the change in structure from T1 to T2.
Similar to the example above, Child 12’s production of /steɪr/ as [steə] at T1 increased in accuracy to [steɪrd] at T2 where PWP was again more sensitive to structural change at T2. These changes led to a higher PWP score due to the decrease in structural errors at T2. Therefore, PWP was more sensitive in capturing change for this child.

In summary, out of the seven children who made significant change from T1 to T2, four children showed larger change in PCC than in PWP, and three children showed greater change in PWP than in PCC.
Discussion

The purpose of this study was to compare the sensitivity of the segmental measure of PCC with the whole-word measure of PWP when evaluating change in children’s speech accuracy over time. It was hypothesized that the use of PWP would be a more sensitive measurement of change than PCC as it reflects both segmental accuracy and phonological whole-word approximation. The results did not support this hypothesis as PCC and PWP showed differential sensitivity to different types of change in speech accuracy.

Previous Research

There are no studies, to the author’s knowledge, that directly compare the segmental measure of PCC with the whole-word measure of PWP. A study conducted by Newbold, et al. (2013) evaluated five different measures of speech output where data was drawn from four children with severe and persisting SSD’s at 4 years of age and again after a two year period. Results indicated that in comparison to the other methods of measurement, both PCC and PWP were shown to be sensitive to capture change over time and may be useful tools for evaluating speech outcomes. However, the present study has added the findings that PCC and PWP show differential sensitivity to different types of change. PCC was shown to be particularly sensitive to change in substitution errors. As PWP already gave credit for a placeholder in the structure, less credit was available to award when that placeholder changed from an incorrect to a correct consonant. Furthermore, PWP was shown to be more sensitive to change in structural errors as PCC gave no credit for the correct positon of an incorrect phoneme; however PWP awarded credit for the inclusion the incorrect phoneme regardless of accuracy.

Analysis of the data also found that neither measure showed sensitivity to change at feature level. For example, if the target word /sʌn/ improved in accuracy from [dʌn] to [zʌn], a child would have increased their accuracy by achieving a greater approximation to the correct consonant and the addition of another correct feature (i.e. fricative).

Limitations and Indications for Further Research

The present study was limited by the small sample size where it is believed that a larger sample would strengthen its findings. Furthermore, there was a limitation in the amount...
of change made by the group, where inclusion of children who made more significant amounts of change in speech accuracy may have strengthened the study.

Although an attempt was made to constrain the connected speech sample through the use of pictures obtained from the RAPT, the children varied their vocabulary in the T1 and T2 assessments. This was a limitation as the analysis of changes in speech accuracy was carried out by mapping the words used at T1 and T2 against one another to look at changes to compare the sensitivity of PCC and PWP. An alternative assessment such as the sentence imitation task proposed by Johnston, et al. (2004 cited in Flipsen, et al., 2005) could have overcome this limitation as this task was reported to show no significant difference in PCC obtained in this task compared to a conversational speech sample.

**Conclusion and Clinical Implications**

The use of objective measures is often mandatory in SLT services as it allows the SLT to evaluate the efficacy of intervention by measuring change in a child’s speech accuracy over time and serves as a speech outcome measure. This study highlights the importance of examining the scoring system used when evaluating the outcome of intervention. The goal of this study was to determine whether the whole-word measure of PWP is more sensitive to change than the segmental measure of PCC in the speech of children with SSD’s. The results of the present study provide evidence that PCC is a better indicator of change for children who make substitution errors, while PWP is a better indicator of change for children who make structural errors in their speech. Therefore, neither measure was shown to be better than the other as sensitivity of measures were dependent on the error type.

The clinical implications of these findings provide preliminary evidence for the use of PCC as an outcome measure when working with children with substitution errors, and the use of PWP when working with children with structural errors in their speech. This would lead to the use of measures that are more sensitive to capture the small changes in children's speech accuracy that are important for evaluation of progress in therapy and research. Practicality and ease of use were also discussed to be fundamental components of a measure in order for it to be widely used and accepted. The measures of PCC and PWP both proved to be practical measures for the use in clinical conditions and in research as
they allowed for relatively quick calculations of speech accuracy to be made, incur no expense and no equipment is necessary for calculations.
References


Appendix A

Rules for the calculation of PCC
Rules for the calculation of PCC (Shriberg & Kwiatkowski, 1982 cited in Smit 2004)

1. Determine the meaning of the words/utterances

2. Determine which consonants to use
   a) Include:
      Intended or target sounds in words
   b) Exclude:
      Vowels;
      Unintelligible or partially unintelligible utterances;
      Consonants in third or more repetition of same word;
      Repeated syllable (include only consonants in the first syllable).

3. Score correct consonants
   a) Correct scores also include:
      Dialectal variations or as they are said in conversation
      e.g. ‘and’ → [n]; ‘fishing’ → [fɪʃɪn]; ‘I called him’ → /ʌɪkɑldɪm/
   b) Incorrect errors include:
      Distortions of the target consonant;
      Omission of the target consonant;
      Substitution of the target consonant with another consonant;
      Addition of a sound to the correct/incorrect target consonant.

4. Calculate the PCC
   a) Count up total number of consonant targets and the number that are correct;
   b) Divide the number correct by the total consonant targets;
   c) Multiply by 100 to determine the PCC.
Appendix B

Rules for the calculation of PMLU & PWP
Rules for the calculation of PMLU & PWP (Ingram, 2002; Rules 7 and 8: Bónová, et al., 2005 cited in Saaristo-Helin, 2009)

1. **Sample-size Rule:** Select at least 25 words, and preferably 50 words for analysis, depending on sample size. If the sample is larger than 50 words, select a selection of words that cover the entire sample, e.g., every other word in a sample of 100 words.

2. **Lexical-class Rule:** Count words (e.g., common nouns, verbs, adjectives, prepositions and adverbs) that are used in normal conversation between adults. This excludes child words, e.g., mommy, daddy, tata, etc. Counting child words can inflate the PMLU if a child is a reduplicator.

3. **Compound Rule:** Do not count compound words as a single word unless they are spelled as a single word, e.g., ‘cowboy’ but not ‘teddy bear’. This rule simplifies the decisions about what constitutes a word in the child’s sample.

4. **Variability Rule:** Only count a single production of each word. If more than one occurs, then count the most frequent one. If there is none, then count the last one produced. Counting variable productions may distort the count.

5. **Production Rule:** Count 1 point for each consonant and vowel that occurs in the child’s production. Do not count more segments than are in the adult word. For example, a child who says ‘foot’ as [hwut] has two consonants counted, not three. Otherwise, children who add segments will get higher scores despite making errors.

6. **Consonants Correct Rule:** Assign 1 additional point for each correct consonant. Correctness in vowels is not counted since vowel transcriptions are typically of low reliability. Syllabic consonants receive an additional point in the same way as nonsyllabic consonants. A child who applies liquid simplification, for example, will get 1 point for producing a vowel.

7. **Positional Rule:** Only segments realized in the proper position in the word are counted as correct.

8. **Input Rule:** Children’s words must be compared with their real, often nonstandard targets in spoken language, not the written version of the word.

   **PWP:** Divide the earned PMLU score by the PMLU of the corresponding target words. Multiply by 100 to determine PWP.