UNIVERSITY of LIMERICK

OLLSCOIL LUIMNIGH

Final Year Research Project SL6036; Student 13012037

Are Whole-Word Measurements of Speech Accuracy More Sensitive to Change than Segmental Measurements?

A research project submitted to the Department of Speech & Language Therapy, University of Limerick, as part of a requirement for the MSc in Speech and Language Therapy (professional qualification)

May 2015

Intended Journal for submission

International Journal of Language and Communication Disorders

Supervisor: Aileen Wright

Word count: 6,959

Abstract

Background: Measurement of change occurring during and after intervention is a critical part of evidence based practice. Therefore, it is imperative to find measurements that are sensitive enough to show real change. In intervention for speech sound disorders (SSDs), therapists have traditionally used a range of measurements. Currently there seems to be no consensus on the most valid method of measuring changes in speech accuracy (Flipsen et al. 2005). This has negative consequences for evidence-based practice.

Objectives: This project investigates whether a whole word measure of speech accuracy, Proportion of Whole Word Proximity (PWP), is more sensitive to change than the segmental measure, Percentage Consonants Correct (PCC), most commonly used in the literature.

Methods: 12 monolingual children with SSDs aged between 3;0 and 4;11 took part in the study. Participants were assessed in single word production using the *Diagnostic Evaluation of Articulation and Phonology* (DEAP) (Dodd et al. 2002), and in continuous speech using the *Renfrew Action Picture Test* (RAPT) (Renfrew 1997) at initial assessment and again 8 weeks later.

Results: Analysis showed a high correlation between PCC and PWP. For the whole group, there was no statistically significant difference between PCC and PWP in measuring change. However, this may have been due to the fact that no significant change occurred at group level. Analysis of individual cases where significant change did occur showed that the different methods were differentially sensitive to different types of change.

Conclusions: In this sample, PCC appeared to be a more sensitive measure of change in features, while PWP appeared to be more sensitive to changes in structure. Clinical implications are discussed and recommendations are made.

Keywords: Speech accuracy, measurement of change, percentage consonants correct (PCC), proportion of whole word proximity (PWP).

Acknowledgements

I would like to thank my parents for their constant support and encouragement throughout the course.

I would also like to thank my classmates, who have been great colleagues and friends over the past two years. In particular, I would like to thank my classmates who worked with me on this project for their hard work and unfailing support, particularly Bernadette and Jessica, who worked closely on this research question with me and were always helpful and generous with their time.

I would like to extend my gratitude to the parents/carers and children who participated in this project, without whom this research would not have been been possible.

Finally, I would like to express my appreciation towards my supervisor, Aileen Wright, who was a great source of guidance and support throughout.

Table of Contents

List of Figures

List of Tables

Introduction

The majority of children referred to speech and language services with a communication difficulty have a speech disorder (Dodd 2005). In fact, Broomfield and Dodd (2004a) suggest that one third of all speech and language therapy referrals are related to speech impairment. Children with speech sound disorders are at risk of literacy/ mathematical difficulties, school failure, and socialization problems (Williams et al 2010), and, as a result, writers such as Broomfield and Dodd (2005) highlight the critical importance of effective intervention. It is clear that intervention for speech sound disorders in children is vital. It is also clear that effective measurement of change in speech accuracy is of the utmost importance in order to show efficacy of treatment, to compare different methods of therapy, and to inform intervention policy.

From this, the problem arises of finding a measurement of speech accuracy that is sensitive enough to show real change. The literature has traditionally used a range of measurements, both subjective and objective, in order to measure speech accuracy. This leads to a lack of standardisation of results and at this point in time there does not seem to be a consensus on the most valid method of measuring speech accuracy (Flipsen et al 2005). This has negative implications for evidence-based practice, which is needed to improve intervention, increase resources, and enhance perception of the profession (Dodd 2007).

This project will first examine the main measures of speech accuracy from the literature on speech sound disorders. It will be argued that subjective measures of speech accuracy have issues of reliability, standardisation, and validity. It will also be suggested that objective measures of speech accuracy offer more potential in terms of reliability, standardisation, and validity. From these objective measures, *the question of whether segmental or whole word measures of speech accuracy are more sensitive to change in speech accuracy* *will be raised.* Segmental measures of speech accuracy are those focused at the level of the individual speech sound, while whole word measures are those focused at the level of the word. Few studies to date have explicitly compared sensitivity to change in segmental and whole word measures of speech accuracy. *Specifically, the sensitivity of the segmental measure of speech accuracy, Percentage consonants correct (PCC) will be compared to the whole word measure Proportion of Whole Word Proximity (PWP).*

Subjective measures of speech accuracy

One of the most common subjective approaches to determining severity of speech disorder is impressionistic judgment based on ordinal rating scales (Flipsen et al 2005). The Intelligibility in Context Scale (ICS) (McLeod et al 2012) is an example. Severity is usually categorized as mild, moderate, or severe. The subjective nature of these measurements raises issues of reliability, standardisation, and construct validity (Flipsen et al 2005). For instance, Shriberg & Kwiatkowski (1982) claim that differences in judges' interpretation of severity are to be expected. Rafaat et al (1995) reported only 61% exact agreement between 14 clinicians rating samples from 45 children on a 5- point scale. Flipsen et al (2005) found poor/fair listener agreement between ten very experienced Speech and Language Therapists (SLTs) (with at least 10 years' experience) using a mild/moderate/severe rating scale to evaluate seventeen speech samples.

These findings suggest that reliable, standardised, and valid measurements of intelligibility (and by extension, changes in intelligibility over time) must come from objective measures of speech accuracy.

Objective measures of speech accuracy

The use of objective measures of speech accuracy fall broadly into segmental and whole-word measures of accuracy (Flipsen et al 2005).

Segmental measures of speech accuracy

Segmental measures of speech accuracy are those that are focused at the level of the individual speech sound. Many have been put forward by Shriberg & Kwiatkowski (1982), and Shriberg et al. (1997a).

The most well-known and widely used segmental measure of speech accuracy is *percentage of consonants correct* (PCC) (Shriberg & Kwiatkowski 1982). PCC has been frequently cited in the literature, in textbooks (e.g. Bauman-Waengler 2012), and in standardized assessments such as the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al. 2002). PCC is calculated by dividing the number of correct consonants produced by the total number of target consonants. Shriberg et al. (1997a) argue that the PCC method, calculated from a 5-10 minute speech sample, is psychometrically robust. The robustness of PCC as an indicator of change has been supported in the literature by Newbold et al. (2013), amongst others.

As a development of PCC, Shriberg et al. (1997a) suggested *Percentage of Vowels Correct* (PVC), calculated by dividing the number of correct vowel productions by the number of target vowels, and *Percentage of Phonemes Correct* (PPC), which is calculated by dividing all correct consonants and vowels by the total number of phonemes. PVC and PPC, along with PCC, have also been used widely in the literature (e.g. Dodd et al., 2003).

PCC, PVC, and PPC have long been considered robust measures of speech accuracy, however inter-rater reliability for PCC and PPC tends to be much higher than for PVC. This may be due to the fact that transcribers vary more in their transcriptions of vowels (Powell, 2001). For example, Dodd et al. (2003) found correlations between scores of 0.886 for PCC ($p<0.001$), 0.315 for PVC ($p<0.01$), and 0.876 for PPC ($p<0.001$).

Shriberg et al. (2005) also put forward two sets of segmental measures focusing on error types rather than distortions. The first set calculates the percentage of targets containing errors of omission, substitution, and distortion. They are the *absolute omission index* (AOI), *absolute substitution index* (ASI), and *absolute distortion index* (ADI). The second set represents the relative proportion of errors. They are the *relative omission index* (ROI), the *relative substitution index* (RSI), and *relative distortion index* (RDI). A significant negative correlation between PCC and ROI was reported (i.e. as PCC decreases, the proportion of omission errors increases). On the other hand, a significant positive correlation between PCC and RDI was observed (i.e. as PCC increases, the proportion of distortion errors increases). No significant correlation between PCC and RSI was found. This was interpreted by Shriberg et al. (2005) as suggesting that ROI may be an effective measurement of speech accuracy, however further research may be needed to show clinical effectiveness.

Elbert and Geirut (1986) put forward *Productive Phonological Knowledge* (PPK) of individual phonemes. This method assigns consonant phoneme production into one of six levels of knowledge (adultlike to nonadultlike). This measurement includes consideration of each phoneme by word position and morpheme. Calculating PPK requires elicitation of the 311 single word set developed by Geirut (1985). Data from the PPK task was developed by Dinnsen et al. (1987; cited in Williams 2000) to make a measurement called *percentage correct underlying representation* (PCUR). Forrest and Morrisette (1999) found a correlation of 0.82 between PCC and PCUR, which Flipsen et al. (2005) suggests may show that PKK and PCUR are valid measurements. Further research may be needed to show clinical effectiveness.

Whole word measures of speech accuracy

Measures can also be made at the level of the word. McCabe & Bradley (1973) proposed *whole word accuracy* (WWA). A development of this, called *proportion of whole word correctness* (PWC) was proposed by Ingram (2002). WWA/ PWC is a measure that determines what proportion of produced words are correct out of the entire vocabulary. Schmitt et al. (1983) studied WWA on conversational samples on children aged between three and seven. They found that WWA increased significantly between three and three-and-a-half, from 68% to 76%. There is no data on children below three, and further studies are required to gather normative data and show efficacy of WWA.

Phonological mean length of utterance (PMLU) was proposed by Ingram & Ingram (2001) and further developed in Ingram (2002). PMLU measures the number of segments in the word produced and the number of correct consonants. The number of segments accounts for the fact that words get longer as they become more complex. For example, [kat] has three segments and two consonants, so its target PMLU is 5, [tomato] has six segments and three consonants, so its target PMLU is 9, etc. (Ingram 2002). The PMLU can both inform the clinician of the complexity of the word that is being produced, and the relation between the child's form and the target (e.g. *tomato* produced as [mato] would yield a PMLU of 6 out of a potential 9). Vowels correct are not scored, since transcribers tend to vary more in their transcriptions of vowels (Powell 2001).

Directly related to PMLU is *proportion of whole word proximity* (PWP) (Ingram & Ingram 2001; Ingram 2002). PWP is calculated by dividing the PMLU of the target words into the child's PMLU. The level of difference between the target form and the child's production can then be scored and used as an indirect measure of intelligibility. For example, the productions of *zucchini,* [dzukini], [skini], and [zəzi] would have PWPs of 0.89, 0.78, and 0.56 respectively. [dzukini], with a score of 0.89, is clearly the production closest to the target.

A final whole word measure is *process density index* (PDI) (Wolk et al 1993). PDI refers to the average number of phonological process descriptors that are applied to each word in a sample. For example, if *ski* is produced as [gi], two processes (/s/ cluster reduction and initial voicing) are counted. PDI is then calculated by totalling all processes and dividing by the number of test words. PDI requires more research in order to show efficacy.

Whole word measures of speech accuracy vs. segmental measures

It is apparent that in practice and in the literature, measurement of speech accuracy prior to Ingram (2002) tended to focus on segments and not whole words. However, this has changed and nowadays it can be argued most researchers favour measurements based on the whole word rather than segments as a benchmark for diagnosis (Arias & Lleó 2014). This may be because whole word measurements claim to provide a way to further explore the phonological systems of children with speech sound disorders. Flipsen et al. (2001) compare PMLU with PCC on the phonological samples of seventeen children, aged from 2;11 to 5;3, diagnosed with a speech delay. The results of this study show that the PMLU may be a better predictor of severity than PCC. Burrows & Goldstein (2010), in their study of bilingual Spanish-English children, support the validity of *both* whole word measures (pMLU and PWP) and PCC for monolingual and bilingual children with speech sound disorders, as well as typically developing children. Newbold et al. (2013), in their study of four children with speech sound disorders, conclude that while PCC is a robust indicator of change. Furthermore, Newbold et al. (2013) claim that PWP also captures change, and is more sensitive than PCC to psycholinguistic variables.

This (arguable) favour towards whole word measures may be due to the fact that whole word measures may better reflect the child's acquisition of phonology. Ingram & Ingram (2001) support this by arguing that children acquire words, not individual vowels and consonants, and show little awareness of segments. Bunta et al. (2006, 2009) claim that maintaining a consistent level of phonological proximity to the target is a driving force that propels phonological acquisition.

Arias & Lleó (2014) consider this from the perspective of Optimality Theory. They write that the distance from production to target is largely consistent across languages because what separates the production from the target (i.e.

the lack of faithfulness) are the universal markedness constraints that stand in the way of the target productions. In Optimality Theory, the acquisition process is ended with the demotion of the originally higher ranked markedness constraints to form a hierarchy in which faithfulness constraints are undominated (i.e. faithfulness outranks markedness). It is possible that whole word measures are more sensitive to the development of this acquisition process.

An example of comparison may be made using the productions of *zucchini* from the previous example, and comparing PWP and PCC results. [dzukini] and [skini] score PWPs of 0.89 and 0.78, while the PCC for both words would be the same (66.7%). PCC in this example fails to account for the fact that [dzukini] is closer to the target than [skini]. Likewise, a child whose production of *frog* changes from [fɑg] to [fwɑg] would score PWPs of 0.71 and 0.86 respectively, while the PCC would remain the same at 66.7% for both productions.

Aims of investigation

It is clear that a sensitive measure of change in speech accuracy over time is needed to show efficacy and inform evidence-based practice (Dodd, 2007). This review of the literature suggests that subjective measures of speech accuracy have issues of reliability, standardisation, and validity (Flipsen et al 2005). It seems that objective measures are better suited to address these issues.

Since objective measures of speech accuracy have been divided in the literature between segmental and whole word measures, this raises the question of which are more sensitive to change in speech accuracy over time. *Few studies have explicitly examined the comparative sensitivity of segmental and whole word measures.* This project will carry out this comparison by focusing specifically on the segmental measure Percentage Consonants Correct (PCC) and the whole word measure Proportion of Whole Word Proximity (PWP).

Methodology

Ethical approval

Ethical approval for this project was received from the Research Ethics Committee, University Hospital Limerick. Parents of all the participants were provided with information outlining the aims and methods of the project, the use of their child's data and gave written consent.

Research design

Data for this study was drawn from a pre and post intervention study designed to compare changes in speech accuracy between the randomly assigned 'treatment now' group and 'treatment later' group, used as a control. It was expected that the intervention would produce the change in speech accuracy that could be used to compare the sensitivity of the two measures.

Participants

Parents of eligible children on local Speech and Language Therapy (SLT) waiting lists were invited to participate in the project. All participants had been referred because of speech sound difficulty and were awaiting assessment. SLT managers were given an email address, phone number, or postal address to contact for further information and consent forms if they wished to take part. Inclusionary and exclusionary criteria are outlined in Table 1.

Table 1: Inclusionary and Exclusionary Criteria

Inclusionary	*Exclusionary*
Aged between 3;0 and 4;11	Already in receipt of speech and language therapy
Monolingual English speaking	Speech deficits due to structural or organic causes
	Additional diagnosis of cognitive or sensory difficulties

Appointments were arranged for participants who were thought to satisfy inclusionary/exclusionary criteria. Of the 15 families that volunteered to participate, 13 fit the inclusionary criteria in Table 1, however Child 12 was retroactively excluded from the study due to issues surrounding intelligibility and subsequent difficulties in obtaining reliable measures of speech accuracy. Therefore, 12 children were included in the final results. This group was aged between 3;2 and 4;8, and was composed of 5 girls and 7 boys.

Materials and measures

Diagnostic Evaluation of Articulation and Phonology (DEAP), (Dodd et al. 2002). This assessment is standardised on children aged from 3;0 to 6;11. For the purposes of the project, three parts of this assessment were carried out:

- An articulation assessment, which samples all English speech sounds (consonants and vowels) in initial and final position. This determined stimulability for each speech sound.
- An inconsistency assessment, which involved production of 25 words on 3 separate trials. This evaluated the stability of each participant's phonological system.
- A phonological assessment, where 50 words are sampled via picture naming tasks. This evaluates the child's phonological system, and any phonological processes found were transcribed.

The accuracy of single word production was evaluated using PCC and PWP.

Renfrew Action Picture Test (RAPT), (Renfrew 1997).
This assessment is standardised on children from 3;0 to 8;0 years of age, and is designed to assess information content and grammatical usage in connected speech samples elicited using pictures with set questions. For the purposes of this project, we focused on the speech accuracy of the connected speech samples, measured using PCC and PWP.

Percentage Consonants Correct (PCC) (Shriberg & Kwiatkowski 1982) *and Proportion of Whole Word Proximity (PWP)* (Ingram & Ingram 2001; Ingram 2002).

PCC is a segmental measure of speech accuracy, while PWP is a whole word measure. PCC is calculated by dividing the number of correct consonants produced by the total number of target consonants. It is more commonly cited in the literature/ textbooks (e.g. Bauman-Waengler 2012), and in standardised assessments of speech accuracy for example the DEAP (Dodd et al. 2002). The robustness of PCC as an indicator of change has been supported in the literature, by Newbold et al. (2013), amongst others. For example, the word *zucchini* [zukini] has three consonants [z,k,n]. Productions of *zucchini,* [dzukini], [skini], and [zəzi] would have PCCs of 67%, 67%, and 34% respectively.

PWP is calculated by evaluating structure and correct production of consonants. For example, *zucchini* [zukini] would have a maximum score of 9 (6 for structure, 3 for correctly produced consonants). Feature accuracy is ignored when awarding points for structure. Therefore, as long as a consonant or vowel is produced in the right place, regardless of accuracy, 1 point is given. Then an additional point is given for each correctly produced consonant. The level of difference between the target form and the child's production can then be scored and used as an indirect measure of intelligibility. For example, the productions of *zucchini,* [dzukini], [skini], and [zəzi] would have PWPs of 89%, 78%, and 56% respectively. [dzukini], with a score of 89%, is clearly the production closest to the target. The validity of PWP has been supported in the literature by Burrows & Goldstein (2010), amongst others, while its sensitivity to psycholinguistic variables has been supported in Newbold et al (2013).

Procedure

Appointments were made for SLT Clinics in Ennis and Limerick. Time 1 (T1) assessments were carried out on the 12th and 13th of January 2015 and Time 2 (T2) assessments were carried out on the 9th of March 2015.

All participants were assessed each time on production of single words using the Diagnostic Evaluation of Articulation and Phonology (DEAP) (Dodd et al 2002). A connected speech sample was collected from responses to the Renfrew Action Picture Test (RAPT) (Renfrew 1997). This was a constrained sample, designed to elicit specific vocabulary so that comparisons could be made between words at T1 and T2. Three other assessments were carried out as part of another study, but these are not reported here. Throughout each assessment at T1 and T2, two student Speech and Language Therapists (sSLTs) were present.

One sSLT administered the assessments, while both transcribed and later reviewed the transcriptions. Each assessment session was recorded, after obtaining written parental consent to record. The recording was reviewed later and used to reach agreement on any discrepancies in transcription between the two transcribers. Participants were then randomly assigned to the 'treatment now' or 'treatment later' group, by their parent who blindly picked either a blue or red token from a bag.

All single word and connected speech samples were analysed using PCC and PWP at T1 and T2.

Intervention

The 'treatment now' group was given intervention immediately. The intervention was called the 'Mixed up Marty' programme, a phonological home programme designed to be carried out by parents/carers. Parents/carers were given a two hour training session by the sSLTs and then carried out the home programme over a period of 8 weeks. The 'Mixed up Marty' programme is designed to rectify deficiencies in the child's phonological system through a programme of stories and activities that the parent/carer carries out with the child and makes the necessary phonological contrasts (minimal pairs) explicit. It was delivered in a comic book format in four fortnightly editions. Parents/carers and the child then had the freedom to choose the stories/ activities/ nursery rhymes that they considered most enjoyable/ beneficial. Stories and activities then had to be carried out for a minimum of 15 minutes a day, 6 days a week for the 8 weeks. The programme was entirely input based, i.e. the child was not required to produce the phonological contrasts targeted in the programme, only to listen and understand the contrast in the minimal pairs (for example).

Reliability

Once a consensus was reached by both sSLT transcribers , all single word and connected speech samples were scored using PCC and PWP at T1 and T2. These results were then checked and rechecked for T1 and T2, using two sSLTs per participant, ensuring high- inter rater reliability.

Results

Data and results for all participants are shown in Tables 2 and 3. Table 2 shows the results of PCC and PWP in single word (SW) production for the 'treatment now' and 'treatment later' groups. Table 3 shows the results of PCC and PWP in connected speech (CS) for the 'treatment now' and 'treatment later' groups.

Table 2: Results for Single Word (SW) Production

Participant	*Gender*	*Age at T1*	*Treatment now/ later*	*SW PCC T1*	*SW PCC T2*	*SW PWP T1*	*SW PWP T2*
1	M	3;5	Now	75%	84%	88%	93%
2	M	3;11	Later	65%	65%	81%	83%
3	F	4;5	Later	71%	82%	87%	93%
4	F	3;9	Now	62%	62%	77%	77%
5	M	3;2	Now	29%	33%	54%	62%
6	F	3;4	Later	55%	50%	77%	73%
7	F	3;10	Later	50%	49%	67%	64%
8	F	3;6	Now	63%	61%	81%	82%
9	M	4;2	Now	63%	62%	79%	81%
10	M	4;1	Later	64%	71%	81%	83%
11	M	3;7	Now	80%	82%	88%	90%
13	M	4;4	Later	59%	60%	73%	76%

Table 3: Results for Connected Speech (CS)

Participant	*Gender*	*Age at T1*	*Treatment now/ later*	*CS PCC T1*	*CS PCC T2*	*CS PWP T1*	*CS PWP T2*
1	M	3;5	Now	86%	89%	91%	93%
2	M	3;11	Later	81%	78%	86%	82%
3	F	4;5	Later	88%	91%	93%	94%
4	F	3;9	Now	71%	71%	80%	82%
5	M	3;2	Now	43%	42%	65%	69%
6	F	3;4	Later	67%	77%	79%	85%
7	F	3;10	Later	65%	61%	73%	71%
8	F	3;6	Now	82%	73%	90%	82%
9	M	4;2	Now	72%	73%	80%	87%
10	M	4;1	Later	75%	83%	85%	89%
11	M	3;7	Now	81%	82%	90%	88%
13	M	4;4	Later	58%	58%	68%	76%

Non parametric tests were used due to the small sample size, and therefore no assumptions were made about normal distribution of outcomes.

Correlation between PCC and PWP:

A test of correlation (Spearman's rho) showed that PCC and PWP were highly correlated. For single word production at T1 and T2, correlation between PCC and PWP was 0.973489 and 0.97 respectively. For connected speech production at T1 and T2, correlation between PCC and PWP was 0.992972 and 0.934757 respectively.

This suggested that the relative difference between PCC and PWP remains highly constant for each sample. Therefore, as PCC rises or falls between T1 and T2, PWP will rise or fall by virtually the same amount (see Figures 1, 2). It should be noted here that since PWP measure both structure and correct consonants, it will always be a higher percentage than PCC.

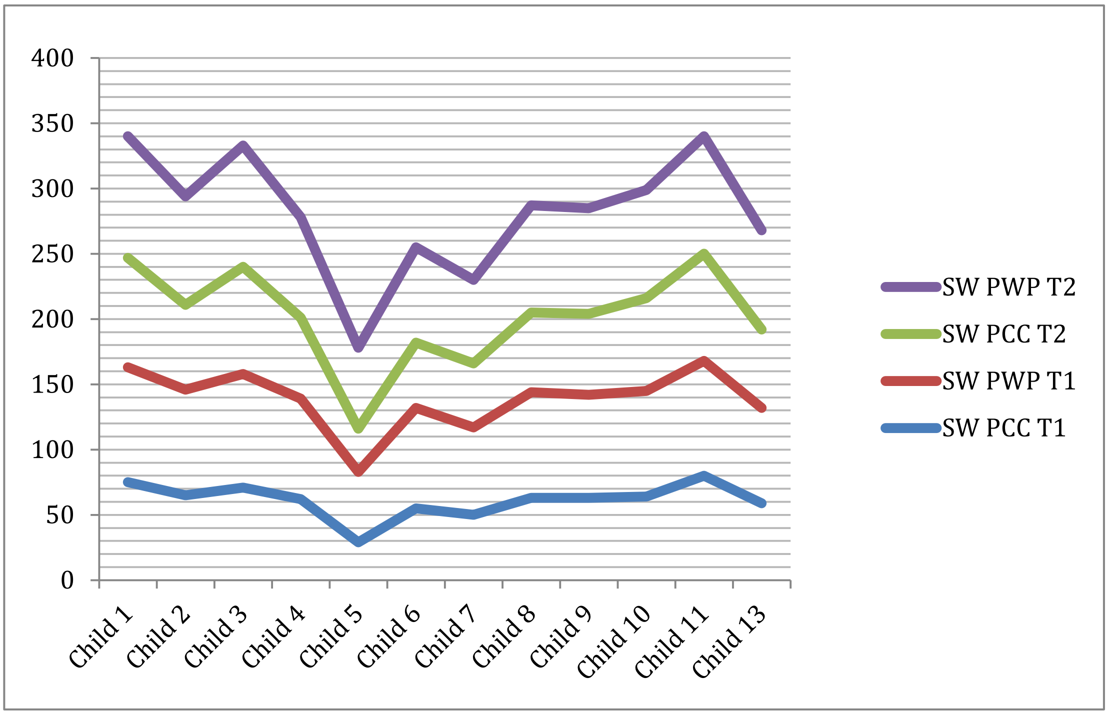

Figure 1: Correlation of PCC and PWP at T1 and T2 (Single word- SW)

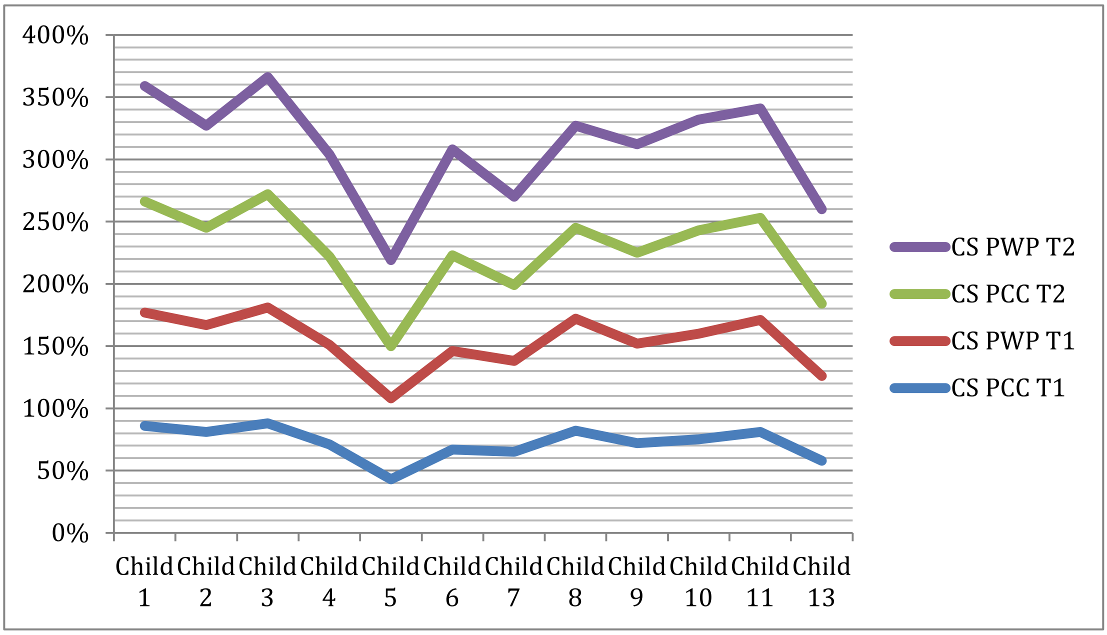

Figure 2: Correlation of PCC and PWP at T1 and T2 (Connected Speech- CS)

Analysis at group level

Table 4: Treatment Now Group Results

	Mean PCC T1	*Mean PCC T2*	*Mean PWP T1*	*Mean PWP T2*
Single word production (SW)	62 (SD: 17.8)	64 (SD: 18.45)	77.8 (SD: 12.5)	80.8 (SD: 10.98)
Connected speech production (CS)	72.5 (SD: 15.6)	71.67 (SD: 16.07)	82.67 (SD: 10.03)	83.5 (SD: 8.22

Table 5: Treatment Later Group Results

	Mean PCC T1	*Mean PCC T2*	*Mean PWP T1*	*Mean PWP T2*
Single word production (SW)	60.67 (SD 7.56)	62.8 (SD: 12.67)	77.67 (SD: 7)	78.67 (SD: 9.97)
Connected speech production (CS)	72.33 (SD: 11.09)	74.67 (SD: 12.79)	79 (SD: 12.3)	82.83 (SD: 8.42)

Table 6: Entire Group Results

	Mean PCC T1	*Mean PCC T2*	*Mean PWP T1*	*Mean PWP T2*
Single word production (SW)	61.33 (SD: 13.05)	63.42 (SD: 15.1)	77. 75 (SD: 9.69)	79.75 (SD: 10.06)
Connected speech production (CS)	72.42 (SD: 12.9)	73. 17 (SD: 13.93)	81.67 (SD: 9.23)	83.17 (SD: 7.94)

A Mann-Whitney U-Value test showed that there was no statistically significant change observed in neither the 'treatment now' nor 'treatment later' group in both measurements, in single word production (SW) and connected speech (CS) (Tables 4, 5).

In single word production (SW), the Mann-Whitney U-value test showed that changes in PCC in the 'treatment now' group and the 'treatment later' group was not statistically significant (z= 0.0801, p= 0.936). Furthermore, change in PWP was not statistically significant for either group (z= 0.4, p= 0.689) (Table 4).

In connected speech (CS), the Mann-Whitney U-value test also showed that there was no significant change in PCC for either group (z= -0.56, p= 0.575). Likewise, there was no statistically significant change for either group in PWP (z= -0.24, p= 0.81) (Table 5).

This lack of statistically significant change was also reflected at the level of the entire group (both 'treatment now' and 'treatment later' groups together- Table 6). A Wilcoxon Signed Rank Test showed no statistically significant improvement in single word production (SW) (PCC: z= -1.227, p= 0.22. PWP: z= -1.654, p= 0.098) or connected speech (CS) (PCC: z= -0.4587, p= 0.645. PWP: z= -1.059, p= 0.289).

Analysis of the group as a whole showed that, although some change in mean scores were observed on both PCC and PWP between T1 and T2 (Table 6), the difference was not statistically significant for either measure.

Analysis at individual level

The lack of statistically significant change at the group level motivated us to analyse the results individually to find examples of significant change. A Fisher's Exact Test found significant improvements at the single word level in three children (Child 1, 3, and 5), and in four children at the connected speech level (Child 6, 8, 9, and 13). Table 7 shows the p-values of change at the individual level (from Fisher's Exact Test). Significant p-values (<0.05) are shown in red.

Table 7: Statistical Significance of Change at Individual Level

	SW PCC	*SW PWP*	*CS PCC*	*CS PWP*
Child 1	0.078939	0.01093	0.347131	0.552257
Child 2	0.70999	0.503898	0.566164	0.284133
Child 3	0.03674	0.014291	0.267059	0.332451
Child 4	0.807394	0.792195	1	0.64902
Child 5	0.518342	0.035611	0.852432	0.552637
Child 6	0.480581	0.934207	0.070509	0.032913
Child 7	0.905885	0.350086	0.753709	0.609941
Child 8	1	1	0.135552	0.006813
Child 9	0.901624	0.637894	0.868495	0.04721
Child 10	0.206723	0.560309	0.080291	0.080291
Child 11	0.760272	0.474717	1	0.572154
Child 13	0.903999	0.493502	1	0.021717

Single word production:

Analysis of the three children who made significant improvement at the single word level showed a discrepancy between gains in PCC and gains in PWP. This was contrary to expectations, since the Spearman's rho test showed a very high correlation between PCC and PWP at the single word level (Figure 1- correlation between PCC and PWP was 0.973489 and 0.97 respectively). Analysis of Child 1 and 3 showed greater change in PCC than in PWP (Figure 3). It was found that, for Child 1, PCC improved by 9%, while PWP improved by 5%. Likewise, Child 3's PCC improved by 11%, while PWP improved by 6% (Figure 3).

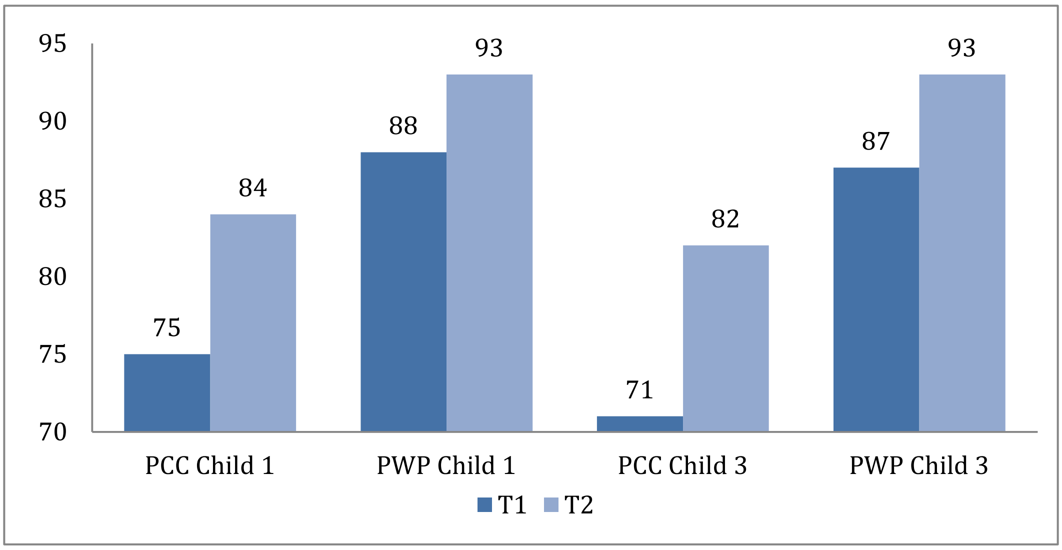

Figure 3: Greater Sensitivity of PCC to Change in Feature Accuracy at Single Word (SW) Level

In contrast, Child 5 showed greater improvement in PWP than in PCC. In fact, it was observed that PCC improved by 4%, while PWP improved by 8% (Figure 4).

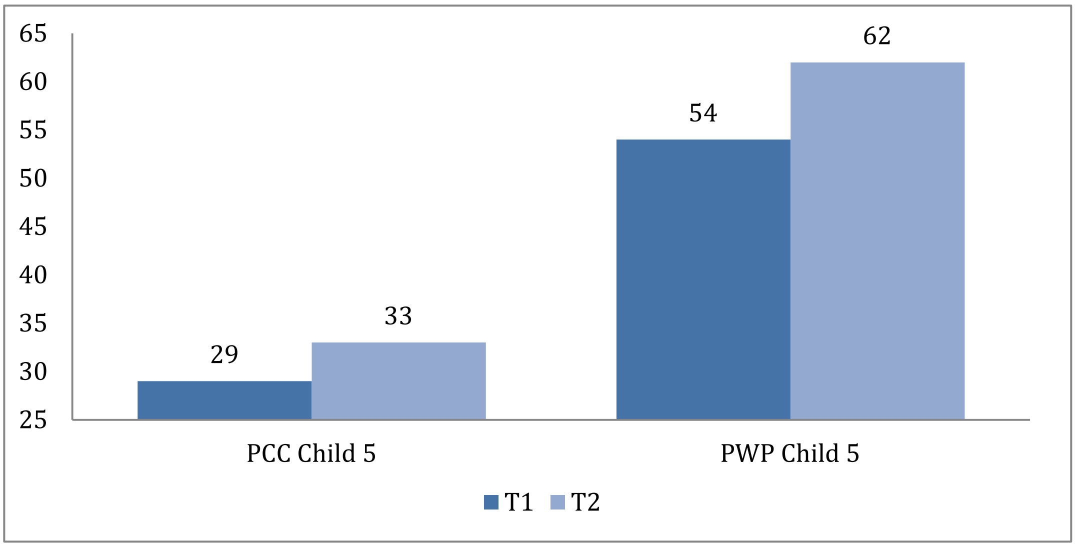

Figure 4: Greater Sensitivity of PWP to Change in Structural Accuracy at Single Word (SW) Level

Connected speech production

Analysis (using Fisher's Exact Test) at the connected speech level showed that four children (Child 6, 8, 9, and 13) showed significant change (Table 7). Since Child 8's results were negative (a decrease in PCC of 9% and in PWP of 8% between T1 and T2- Table 3) they were disregarded. However, Child 6, 9, and 13 made improvements in PCC and PWP (Table 3). Again, and contrary to expectations, a discrepancy was observed between PCC and PWP in all three children. This was unexpected, since a Spearman's rho test showed high correlation between PCC and PWP in the connected speech samples (Figure 2- correlation between PCC and PWP was 0.992972 and 0.934757 respectively).

Child 6 made greater improvement in PCC, which improved by 10%, while PWP increased by 6% (Figure 5).

In contrast, Child 9 and 13 made greater gains in PWP than PCC (Figure 6). Figure 6 shows that PWP increased by 7% and 8% respectively, while PCC increased by 1% for Child 9 and didn't change for Child 13.

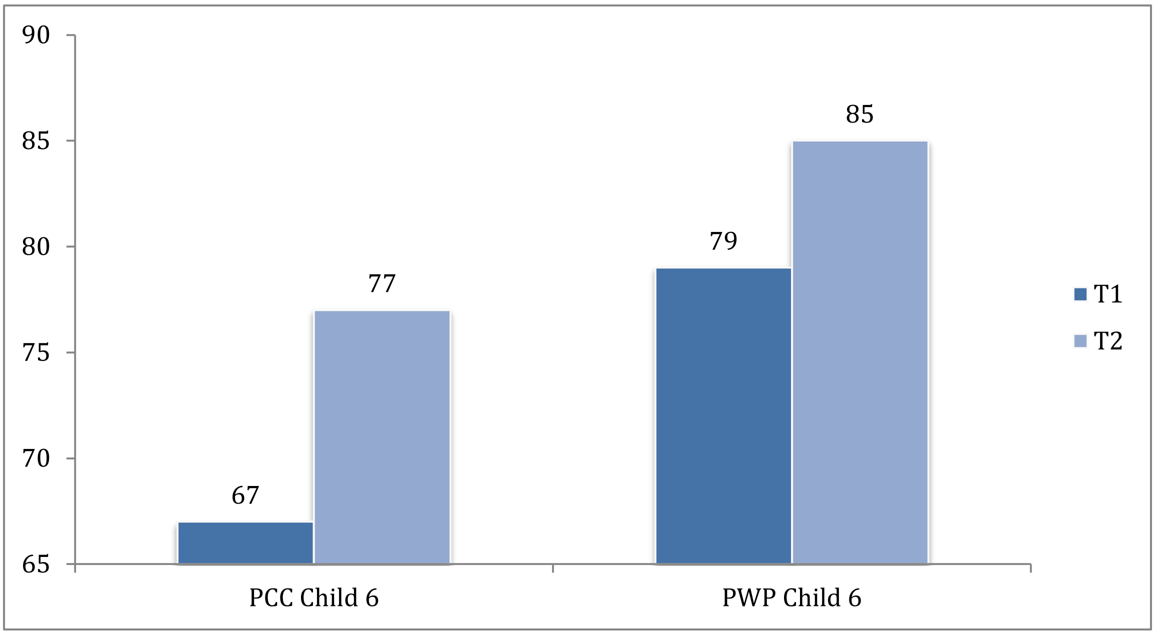

Figure 5: Greater Sensitivity of PCC to Change at Connected Speech (CS) Level

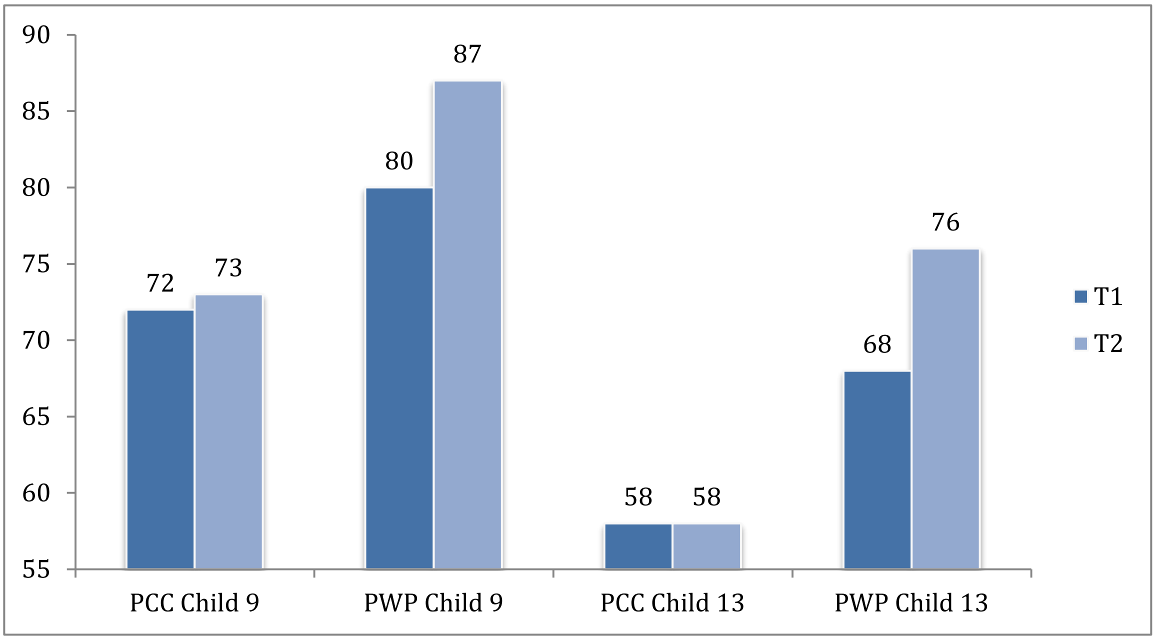

Figure 6: Greater Sensitivity of PWP to Change at Connected Speech (CS) Level

Further analysis of individual results

This apparent discrepancy between PCC and PWP in single word production and connected speech warranted further investigation. Research now had to focus on why PCC and PWP would show different levels of improvement for the same sample.

In single word production, types of errors were analysed from the sample elicited using the DEAP (Dodd et al 2002). It seemed that PCC and PWP were recording different types of improvement in Child 1 and 3 than in Child 5. The difference seemed to lie in improvement made in *feature accuracy* (i.e. phonological accuracy of produced sounds) as opposed to improvement made in *structural accuracy* (i.e. producing all syllables of words, and/or producing sounds in the right places of words, regardless of phonological accuracy).

Child 1 and Child 3, who made greater improvements in PCC than in PWP (Figure 3), appeared to make more improvements in feature accuracy than in structural accuracy. Child 1 made more improvement in feature accuracy in single word production (from 33 errors to 15, a decrease in errors of 55%) than in structural accuracy (from 6 errors to 4, a decrease in errors of 33%). Likewise, Child 3 improved in feature accuracy (from 27 errors to 16, a decrease in errors of 41%) more than in structural accuracy. In fact, Child 5 showed a decrease in structural accuracy (from 4 errors to 6, an increase in errors of 33%).

These improvements in feature accuracy, where PCC underwent greater change than PWP, can be exemplified by the following examples. Child 1 made changes such as the following between T1 and T2: **/bwɒg/ --> /fwɒg/** (target: /fɹɒg/), where PCC increased from 33.3% to 66.6%, (33.3% increase) while PWP increased from 71.4% to 85.7% (14.3% increase). **/ɛləbənt/ --> /ɛləfənt/** (target: /ɛləfənt/), where PCC increased from 75% to 100% (25% increase), while PWP increased from 91% to 100% (9% increase). **/z̪ɛvwə/ --> /zɛbɹə/** (target:/zɛbɹə/), where PCC increased from 33.3% to 100% (66.6% increase), while PWP increased from 75% to 100% (25% increase). Likewise, Child 3 made changes such as the following between T1 and T2: **/tʃɒtʃɪdʒ/ --> /sɒsɪd̪/** (target: /sɒsɪdʒ/), where PCC increased from 0% to 66.6% (66.6% increase). **/tʃɪdədz/ --> /sɪzəz/** (target: /sɪzəz/), where PCC increased from 33.3% to 100% (66.6% increase), while PWP increased from 62.5% to 100% (37.5% increase). **/z̪ɛfwa/ --> /zɛbwa/** (target: /zɛbɹa/), where PCC increased from 33.3% to 66.6% (33.3.% increase) while PWP increased from 75% to 87.5% (12.5 % increase). These examples explain the discrepancy between PCC and PWP for Child 1 and 3, while showing that PCC seems to have higher sensitivity to feature accuracy (consonant accuracy) than PWP.

On the other hand, Child 5, who showed greater change in PWP than in PCC (Figure 4), appeared to make greater improvement in structural accuracy than in feature accuracy. In fact, more errors were made in feature accuracy at T2 (from 30 errors to 52, an increase in errors of 73%). In contrast, structural accuracy improved (from 37 errors to 33, a decrease in errors of 11%).

This can be exemplified by the follow examples from Child 5: **/hɛkɒjə/ -->** **/haijɪkəʔɪ/** (target: /hɛlikɒptə/), where PWP increased from 57% to 71.4% (14.4% increase) while PCC did not change from 40% (0% increase). **/paɪʃ/ -->** **/spaz/** (target: /splaʃ/), where PWP increased from 33.3% to 66.6% (33.3% increase) while PCC increased from 25% to 50% (25% increase). /**seɪ/ --> /dweː/** (target: /skweə/) where PWP increased from 42.8% to 57% (14.2%) while PCC did not increase from 33.3% (0% increase).

Further analysis of the connected speech samples was deemed unsuitable for this study, as it has been argued that whole word measurements that account for structure in connected speech are affected by morphosyntax, and therefore would not be a pure measure of phonological development (Taelman et al 2005, p404).

Discussion

Research question and results

This project set out to answer the question whether a whole word measure of speech accuracy (PWP) is more sensitive to change than a segmental measure (PCC). Analysis of results suggested that, while PCC and PWP are both sensitive to change and highly correlated, PCC appears to be a more sensitive measure of change in feature accuracy, while PWP appears to be more sensitive to change in structural accuracy.

Comparison of results with the literature

There have been few studies in the literature explicitly comparing the sensitivity of PCC and PWP to change in the speech accuracy of children with Speech Sound Disorders (SSDs). Burrows and Goldstein (2010) note that research into PWP in particular has tended to focus on typically developing children, but its use on children with SSDs has been limited. Nevertheless, Arias & Lleó (2014) claim that, nowadays, most researchers tend to favour measures based on the whole word, rather than of particular segments as a benchmark for diagnosis.

Ingram (2002) claims that whole word measures may have the potential to discover aspects of phonological acquisition that have been missed due to focus on segmental measures of speech accuracy. However, researchers such as Tyler & Lewis (2005) argue that PCC has been more often used and clinically validated.

The results of this project appear to correspond to some extent with those in the literature. Newbold et al. (2013) find that both PCC and PWP are valid measurements of change over time, and can differentiate between children who make different amounts of progress. This has been suggested by the results of this project.

Furthermore, Newbold et al. (2013) claim that PWP appears to be more sensitive to psycholinguistic variables. Arias and Lleó (2014) support this claim, and write that whole word measurements, as put forward by Ingram (2002), acknowledge the role that prosodic hierarchy has in the development of phonology and thus may better reflect psycholinguistic development. This is discussed in Arias and Lleó (2014) from the point of view of Optimality Theory. What separate the production from the target (the lack of faithfulness) are the universal markedness constraints that prevent complete production of the target. As discussed in the introduction, in Optimality Theory, the acquisition of phonology is ended with the demotion of the originally higher ranked markedness constraints to form a hierarchy in which faithfulness constraints are undominated. In other words, faithfulness outranks markedness. Likewise, the results of this project suggest that PWP, being more sensitive to improvement in structural accuracy, may indeed better reflect development of underlying phonological structure than PCC.

However, Newbold et al (2013) also conclude that PCC was found to show more maximal change in outcomes. This was not necessarily proved in this project, which found that outcomes in PCC and PWP appear to be dependent on the type of improvement made by each child. Nevertheless, the results of this project do suggest that PCC is also a sensitive measure of change over time. It therefore seems to be a useful measure for monitoring treatment gains in research and in clinical outcome measures, as claimed in Newbold et al. (2013).

The results of this project suggest that PCC is more sensitive to improvement in feature accuracy (i.e. accuracy of consonants). This was also found by Bunta et al (2009), who conducted a study on bilingual and monolingual Spanish and English-speaking peers. The study compared whole word measures (pMLU and PWP) and consonant accuracy (PCC) on elicited single words. It was found that, when comparing the Spanish and English of the bilingual group, the whole word measure (pMLU) and PCC displayed significant differences. It was concluded that the bilingual group differentiated between their language in terms of phonological whole word complexity (reflected by pMLU) and consonant accuracy (reflected by PCC). However, this study differs from this project in that PWP did not show a difference between the languages, suggesting a consistent level of proximity to the target. It was suggested that phonological proximity to the target is an important driving force in phonological acquisition (Bunta et al 2009, p174).

The results of this study may differ from those of the literature for a number of reasons. There was a lack of significant change at the group level, and so the results of three children who showed significant change at the single word level had to be investigated. This significantly reduced the sample size. The intervention received by the children between T1 and T2 also differed from the literature, in that it was an experimental, input-based, parent-led programme. This programme differed in method, intensity, delivery, and dosage to those in the literature. It was also noted that all but two of the group were poor non-word discriminators, which may have had an effect on both the amount of change made by the participants, and the type of change made by each participant. Finally, data was obtained through elicting single words through the DEAP (Dodd et al 2002), and not through other means such as repetition or spontaneous, connected speech such as in Newbold et al. (2013), amongst others.

Clinical implications

The differential sensitivity of PCC and PWP to different aspects of phonological production may well have clinical implications. For instance, Watson and Terrell (2012) write that whole-word production capabilities may be an important factor to consider when determining the need for intervention services. Targeting whole word accuracy may be beneficial. The addition of relatively complex words in intervention may positively impact results, i.e. the inclusion of words which are phonologically more complex (compared to typical productions) may serve to motivate improved production. This project suggests that PWP may be a more sensitive measure of change in intervention such as this, which targets whole word accuracy. Addressing complexity in phonological therapy has been shown to be beneficial to some children (Geirut 2007), although that paper was focused on difficult sounds and not whole word complexity.

On the other hand, intervention which targets specific phonological processes and is designed to target feature accuracy may be better served by the apparent higher sensitivity of PCC to feature accuracy. However, as pointed out in Newbold et al. (2013), PCC does not directly show whether individual consonantal targets have improved, and so further phonological analysis may be required to measure improvement on specific segmental targets. Also, it should be noted that change in PCC does not necessarily reflect change in word intelligibility, since PCC does not take into account the word as a unit, and so PWP may be a more useful tool for measuring word intelligibility (Newbold et al. 2013).

It seems that both measurements have strengths, which compliment each other, and this may justify the use of both as measurements of change in speech accuracy over time. Burrows and Goldstein (2010) found no significant difference between PCC and PWP in measuring change between bilingual and age-matched monolingual peers with SSDs. They concluded that whole word measures, in conjunction with segmental measures, appear to be valid for use with monolingual and bilingual children with SSDs. This follows the recommendations of Miccio et al. (1999), who advocates the use of multiple analyses to gauge phonological skill.

Limitations

The differences between the results of this project and those found in the literature may also be attributed to some of the limitations of this study.

Primarily, the lack of significant change at the group level (for either the 'treatment now' group, the 'treatment later' group, or the group as a whole) may well have had an affect on the results obtained. Furthermore, the sample size of those who made statistically significant change at the single word level was relatively small (n= 3), in comparison to those in the literature.

Conclusion and recommendations

The goal of this project was to investigate the relative sensitivity of a whole word measure of speech accuracy (PWP) as compared to a segmental measure (PCC), more commonly used in the literature. It was found that these measures appeared to be differentially sensitive to different kinds of change. PWP appears to be a more sensitive measure to changes in structural accuracy, while PCC appears to be more sensitive to changes in feature accuracy (consonant accuracy). Both measurements are highly correlated, and it seems that both are valid measures of change in speech accuracy over time. This has clinical implications, and it seems that a whole word measure such as PWP, in addition to the segmental measure PCC, may be valid for use with children with SSDs, as suggested by Burrows and Goldstein (2010). This is in line with recommendations made by Miccio et al. (1999), who advocate the use of multiple analyses to evaluate the phonological skills of children, including those with SSDs.

It is recommended that this study be replicated with a larger sample size, who display statistically significant change in speech accuracy, in order to further investigate the results of this study.

References:

Arias, J. & Lleó, C. (2014) 'Rethinking assessment measures of phonological development and their application in bilingual acquisition', *Clinical linguistics & phonetics,* 28(3), 153-175.

Bauman-Waengler, J. (2012) *Articulatory and phonological impairments: A clinical focus.* London: Pearson.

Broomfield, J. & Dodd, B. (2004a) 'Children with speech and language disability: caseload characteristics', *International Journal of Language and Communication Disorders. 39* (3), 303-324.

Broomfield, J. & Dodd, B. (2005) 'Clinical effectiveness' in Dodd, B., *Differential diagnosis and treatment of children with speech disorder.* London: Whurr, 211-229.

Bunta, F., Davidovich, I. & Ingram, D. (2006) 'The relationship between the phonological complexity of a bilingual child's words and those of the target languages', *International journal of bilingualism,* 10, 71-88.

Bunta, F., Fabiano-Smith, L., Goldstein, B. & Ingram, D. (2009) 'Phonological whole word measures in 3-year-old bilingual children and their age matched monolingual peers', *Clinical linguistics and phonetics,* 23, 156-175.

Burrows, L. & Goldstein, B.A. (2010) 'Whole word measures in bilingual children with speech sound disorders', *Clinical linguistics & phonetics,* 24, 357-368.

Dodd, B. (2005) 'Children with speech disorder: defining the problem' in Dodd, B., *Differential diagnosis and treatment of children with speech disorder.* London: Whurr, 3-23.

Dodd, B. (2007) ‘Evidence-based practice and speech-language pathology: Strengths, weaknesses, opportunities and threats’, *Folia Phoniatrica et Logopaedica, 59*, 118-129.

Dodd, B., Crosbie, S., Zhu, H., Holm, A. & Ozanne, A. (2002) *Diagnostic Evaluation of Articulation and Phonology (DEAP)*. London: Psychological Corporation.

Dodd, B., Holm, A., Hua, Z., & Crosbie, S. (2003) ‘Phonological development: a normative study of British English-speaking children’, *Clinical linguistics and phonetics,* 17(8), 617-643.

Elbert, M. & Geirut, J. (1986) *Handbook of clinical phonology.* Boston: College-Hill Press.

Flipsen, P., Hammer, J. & Yost, K. (2005) ‘Measuring severity of involvement in speech delay: segmental and whole word measures’, *American journal of speech language pathology,* 14, 298-312.

Flipsen, P., Yost, K., Valentine, D., Edick, L., Hume, S., King, D. & Vaughan, T. (2001) *PMLU vs. PCC as indices of severity: a preliminary evaluation.* Poster presented at the annual meeting of the American Speech-Language-Hearing Association, November, New Orleans.

Forrest, K. & Morrisette, M.L. (1999) ‘Feature analysis of segmental errors in children with phonological disorders’, *Journal of Speech, Language, and Hearing Research,* 42, 187-194.

Geirut, J. (1985) *On the relationship between phonological knowledge and generalization learning in misarticulating children.* Unpublished doctoral dissertation, University of Indiana, Bloomington.

Geirut, J. (2007) 'Phonological complexity and language learnability', *American Journal of Speech-Language Pathology, 16,* 6-17.

Ingram, D. (2002) 'The measurement of whole-word productions', *Journal of Child Language,* 29, 713-733.

Ingram, D. & Ingram, K.D. (2001) 'A whole-word approach to phonological analysis and intervention', *Language, Speech, and Hearing Services in Schools*, 32, 271-283.

McCabe, R.B. & Bradley, D.P. (1973) 'Pre and post articulation therapy assessment', *Language, Speech, and Hearing Services in Schools,* 4, 13-22.

Mcleod, S., Harrison, L.J. & McCormack, J. (2012) 'The intelligibility in context scale: validity and reliability of a subjective rating measure', *Journal of Speech, Language, and Hearing Research,* 55(2), 648-656.

Miccio, A.W., Elbert, M. & Forrest, K. (1999) 'The relationship between stimulability and phonological acquisition in children with normally developing and disordered phonologies', *American Journal of Speech-Language Pathology, 8,* 347-363.

Newbold, E.J., Stackhouse, J. & Wells, B. (2013) 'Tracking change in children with severe and persisting speech difficulties', *Clinical Linguistics & Phonetics,* 27(6-7), 521-539.

Powell, T.W. (2001) 'Phonetic transcription of disordered speech', *Topics in Language Disorders,* 21(4), 52-72.

Rafaat, S.K., Rvachew, S. & Russell, R.S.C. (1995) 'Reliability of clinician judgments of severity of phonological impairment'. *American Journal of Speech-Language Pathology,* 4(3), 39-46.

Renfrew, C. E. (1997). *Action Picture Test, (4th Edition).* Bicester: Winslow.

Schmitt, L.S., Howard, B.H. & Schmitt, J.F. (1983) 'Conversational speech sampling in the assessment of articulation proficiency', *Language, Speech, and Hearing Services in Schools,* 14, 210-214.

Shriberg, L.D., Austin, D., Lewis, B.A., McSweeny, J.L. & Wilson, D.L. (1997a) 'The percentage of consonants correct (PCC) metric: extensions and reliability data', *Journal of Speech, Language, and Hearing Research,* 40, 708-722.

Shriberg, L.D. & Kwiatkowski, J. (1982) 'Phonological disorders III: a procedure for assessing severity of involvement', *Journal of Speech and Hearing Disorders,* 47, 256-270.

Shriberg, L.D., Lewis, B.A., Tomblin, J.B., McSweeny, J.L., Karlsson, H.B. & Scheer, A.R. (2005) 'Toward diagnostic and phenotype markers for genetically transmitted speech delay', *Journal of Speech, Language, and Hearing Research,* 48, 834-852.

Taelman, H., Durieux, G. & Gillis, S. (2005) 'Notes on Ingram's whole words measures for phonological development', *Journal of Child Language, 32,* 391-405.

Tyler, A.A. & Lewis, K.E. (2005) 'Relationships among consistency/variability and other phonological measures over time', *Topics in Language Disorders, 25,* 243-253.

Watson, M.M. & Terrell, P. (2012) 'Longitudinal changes in phonological whole word measures in 2 year- olds', *International Journal of Speech-Language Pathology, 14(4);* 351-362.

Williams, A.L. (2000) ‘Multiple oppositions: Case studies of variables in phonological intervention’, *American Journal of Speech Language Pathology,* 9, 289-298.

Williams, A.L., McLeod, S. & McCauley, R.J. (2010) ‘Introduction’ in Williams, A.L., McLeod, S., & McCauley, R.J. *Interventions for speech sound disorders in children.* Baltimore, MD: Paul H Brookes, 1-26.

Wolk, L., Edwards, M.L. & Conture, E.G. (1993) ‘Coexistence of stuttering and disordered phonology in young children’, *Journal of Speech and Hearing Research,* 36, 906- 917.