Intelligibility as a clinical outcome measure following intervention with children with phonologically-based speech sound disorders

Running head: Intelligibility as a clinical outcome measure

[1]Lousada, M., [2]Jesus, Luis M. T., [3] Hall, A. e [4]Joffe, V.

[1] School of Health Sciences (ESSUA) and Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal; e-mail: marisalousada@ua.pt

[2] School of Health Sciences (ESSUA) and Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal; e-mail: lmtj@ua.pt

[3] Department of Mathematics and Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal; e-mail: andreia.hall@ua.pt

[4] School of Health Sciences, City University London, Northampton Square, London, UK; e-mail: v.joffe@city.ac.uk

Corresponding author: Marisa Lobo Lousada, School of Health Sciences (ESSUA) and Institute of Electronics and Telematics Engineering of Aveiro (IEETA), University of Aveiro, 3810-193 Aveiro, Portugal. E-mail: marisalousada@ua.pt. Phone: 351234372470

Abstract

Background: The effectiveness of two treatment approaches (phonological therapy and articulation therapy) for treatment of 14 children, aged 4;0-6;7 years, with phonologically-based speech sound disorder (SSD) has been previously analysed with severity outcome measures (percentage of consonants correct score, percentage occurrence of phonological processes and phonetic inventory).

Considering that the ultimate goal of intervention for children with phonologically-based SSD is to improve intelligibility, it is curious that intervention studies focusing on children's phonology do not routinely use intelligibility as an outcome measure. It is important that the impact of interventions on speech intelligibility is explored.

Aims: This paper investigates the effectiveness of the two treatment approaches (phonological therapy and articulation therapy) using intelligibility measures, both in single words and in continuous speech, as the primary outcome.

Methods and Procedures: Fourteen children with phonologically-based SSD participated in the intervention. The children were randomly assigned to phonological therapy or articulation therapy (7 children in each group). Two assessment methods were used for measuring intelligibility: a word identification task (for single words) and a rating scale (for continuous speech). Twenty-one unfamiliar adults listened and judged the children's intelligibility. Reliability analyses showed overall high agreement between listeners across both methods.

Outcomes and Results: Significant improvements were noted in intelligibility in both single words (paired t (6) = 4.409, p = 0.005) and continuous speech (Asymptotic Z = 2.371, p = 0.018) for the group receiving phonology therapy pre- to post-treatment, but no differences in intelligibility were found for those receiving the articulation therapy pre- to post-treatment, either for single words (paired $t(6) = 1.763$, $p = 0.128$) or continuous speech (Asymptotic $Z = 1.442$, $p = 0.149$).

Conclusions and Implications: Intelligibility measures were sensitive enough to show changes in the phonological therapy group but not in the articulation therapy group. These findings emphasise the importance of using intelligibility as an outcome measure to complement the results obtained with other severity measures when exploring the effectiveness of speech interventions. This study presents new evidence for the effectiveness of phonological therapy in improving intelligibility with children with SSD

Keywords: intelligibility, outcome measure, intervention, children, speech sound disorders

What we know

Phonological and articulation therapy have been found to be effective in improving the speech of children with phonologically-based SSD with the phonological approach being the more effective of the two when severity measures are used.

What this paper adds

Intelligibility measures showed the phonological therapy to be effective in improving speech, both in single words and in continuous speech. However, intelligibility did not significantly improve for children receiving articulation therapy. These findings highlight the importance of using intelligibility as an outcome measure when working with children with phonologically-based SSD and places intelligibility as an essential measure to take into account when evaluating effectiveness.

Introduction

Intelligibility can be defined as how well a client's speech is understood by other individuals, i.e., understandability of speech (Pascoe et al., 2006). Whilst typically developing children of around 4 years of age are intelligible most of the time (Hodson and Paden, 1981), children with moderate to severe speech sound disorders (SSD) can show marked unintelligibility which can impact adversely on functional communication and social participation (Hustad, 2012). Intelligible speech is an essential component of effective verbal communication, and Miller (2013) places the measurement of intelligibility at the centre of clinical decision-making, monitoring and building a robust evidence base for the effectiveness of interventions across a range of impairments, both developmental and acquired (Miller, 2013, p.601).

Improving intelligibility is typically a key long term aim of speech and language therapists working with all client groups with speech impairments (Miller, 2013) including children with SSD (Flipsen, 1995, Dodd and Bradford, 2000). Despite its key role as a therapeutic goal, the intelligibility of children with SSD has been analysed in a very limited number of studies (for example, Kwiatkowski and Shriberg, 1992, Flipsen, 1995, Gordon-Brannan and Hodson, 2000, Klein and Flint, 2006, McLeod et al., 2012, Speake et al., 2012); and it appears to be rarely used, in a consistent, robust and reliable way, as a tool for assessment or of measurement of outcomes in routine clinical practice. Whilst a survey of clinical practice in the USA showed that three quarters of paediatric clinicians assessed intelligibility, the authors reported that this was accomplished through subjective means rather than through the use of any robust objective measurement (Skahen et al., 2007).

Intelligibility is usually variable across words and settings, making its measurement a challenging task (Miller, 2013). It can be influenced by a range of factors such as whether the listeners are familiar with talkers, the presence/ absence of speech cues, knowledge of the context of the speech sample and the number of times the speech samples are presented (Kwiatkowski and Shriberg, 1992, Flipsen, 1995, Pascoe et al., 2006, Ertmer, 2010).

Measurement of Speech Intelligibility

Two different methods of measuring intelligibility have been used with clients with speech impairments: word identification tasks and rating scales (Pascoe et al., 2006, Ertmer, 2010, Miller, 2013). *Word identification tasks* require listeners to write down the words that they have understood to be said (open-set) or involve the listener selecting words from a range of multiple-choice alternatives (closed-set). The speech sample can be single words, sentences or continuous speech, usually pre-recorded and randomised (Gordon-Brannan and Hodson, 2000). The listener's responses are typically scored for the number of words they match correctly so that a percent-intelligible score can be calculated (Gordon-Brannan and Hodson, 2000, Pascoe et al., 2006).

Rating scales, used typically with continuous speech samples, are of two types: interval scaling, and direct magnitude estimation. Interval scaling is more typically used, requiring the listeners to rate speech samples (e.g., sentences) along a continuum of intelligibility (e.g., on a 4 point-scale where 1 represents completely unintelligible and 4 represents completely intelligible). Direct magnitude estimation requires an estimate, usually a percentage, of parts of a sentence which are understood (e.g., an estimation of 50% would indicate that a listener has understood only about half of the message).

This method is implemented by playing samples (pre-recorded) to listeners and asking them to select a number/ descriptor to specify how well they understood the speech sample (Pascoe et al., 2006, Ertmer, 2010).

The use of rating scales is potentially quick and easy to implement, but has some disadvantages: listeners could use different criteria on which they base their judgments, and ratings for the same children across raters can be widely variable. Scaling may also not be sufficiently sensitive to small improvements. Some adaptations, however, can make this method more useful and robust. In order to minimise the problem with differences in listeners' criteria, and the subjectivity that this brings, a group of listeners could be asked to be raters for the same children. In this case it would also be important to have more than one speech sample for each child, , so the listener remains unfamiliar with the content of each sample. When different samples are obtained at pre- and post-intervention, Pascoe et al. (2006) suggest matching the speech samples in terms of length and content. Furthermore, asking the same listener to rate subsequent samples of the same talker (e.g., pre- and post-treatment samples) is more reliable than using different listeners for pre- and post-treatment (Ertmer, 2010).

There are advantages and disadvantages to both approaches, with knowledge of the stimuli being particularly advantageous when rating highly unintelligible speech (Flipsen, 1995, Pascoe et al., 2006, Baudonck et al., 2009). Word identification tasks have the advantage over rating scales of having high validity (Kwiatkowski and Shriberg, 1992), however, the main disadvantage of this method is that the transcription involved is time consuming and not always possible for severely unintelligible speech when target words are not known (Kwiatkowski and Shriberg, 1992, Konst et al., 2000, Klein and Flint, 2006, Pascoe et al., 2006, Ertmer, 2010).

Intelligibility as an outcome measure

The principal aim of intervention for children with SSD is to improve the ability to communicate effectively, i.e., to improve intelligibility (Flipsen, 1995, Dodd and Bradford, 2000). Therefore, evaluations of therapy for children with SSD should consistently include intelligibility measures to ensure that therapies are socially and functionally important (Pascoe et al., 2006). However, few studies use intelligibility as an outcome measure to monitor change following a period of intervention, and most of the ones that do, involve adults with dysarthria (Yorkston et al., 1990, Hunter et al., 1991, De Bodt et al., 2002, Kempler and Van Lancker, 2002, Kain et al., 2007).

Most intervention studies targeting children's phonology have not used intelligibility measures to evaluate effectiveness (for e.g. Gillon, 2000, Hesketh et al., 2000, Crosbie et al., 2005, Lancaster et al., 2010, Allen, 2013). Instead, studies typically provide alternative speech severity indices, such as percentage of occurrence of phonological processes and percentage of consonant correct (PCC) score (Pascoe et al., 2006). For example, Bowen and Cupples (1999) used PCC and percentage of occurrence of phonological processes to evaluate the effectiveness of the 'Parents and Children Together' (PACT) approach in the treatment of children with developmental phonological disorder and Hesketh *et al.* (2000) used measures of phonological output (PCC and a naming task) to investigate the effectiveness of two therapy methods (articulation therapy and metaphonological therapy) in children with phonological disorders.

Gillon (2000) assessed the efficacy of a phonological awareness intervention approach and compared it with two other approaches for children with spoken language impairment. The effects of the interventions on phonological awareness ability, reading performance and speech production (PCC and percentage of occurrence of phonological processes) were analysed with no measure of intelligibility. Whilst the measures used in all these studies were sensitive enough to detect change, it could be argued that they were unable to monitor change across all areas of speech production. For example, , a measure like PCC, which is commonly used, does not take into account vowel production, , phonotactics or suprasegmentals, all of which influence intelligibility (Bowen and Cupples, 2006). In a review of outcome measures for children with SSD undertaken by Baker and McLeod (2011), only 2 (Camarata, 1993, Yoder et al., 2005) of the 134 intervention studies they cited made reference to intelligibility (Baker and McLeod, 2011).

Pascoe et al. (2006) looked at intelligibility as an outcome measure for five school-aged children with phonological disorders with varied severity. The children were individually treated by the same speech and language therapist with the duration of treatment ranging from four to nine months depending on the children’s difficulties. A group of 33 unfamiliar listeners were recruited to judge the children’s intelligibility. Single words, sentences and spontaneous speech samples were obtained pre- and post-intervention. Severity speech indices (e.g., PCC) were also calculated for each child pre- and post-intervention so that severity and intelligibility measures could be compared. Two of the children showed improvement according to severity indices after treatment, with no corresponding improvement in intelligibility. Only one child showed improvement in intelligibility at post-intervention, whilst severity remained unchanged. No improvements were noted in the two remaining children, in either severity or intelligibility. Hence three out of the five children showed improvement in only one of the outcomes measures, and none of the children improved across both measures. Pascoe et al. (2006) highlighted the importance of using intelligibility as an additional outcome measure, but also noted that intelligibility should not be used as the only outcome measure for children with phonological disorders because clients may never attain high levels of intelligibility, and therefore changes in their speech might not be captured by looking at intelligibility alone. Intelligibility measures may also not be sufficiently sensitive to pick up the more subtle changes in the child's speech as a result of an intervention.

A recent study by Speake et al. (2012), evaluating the effectiveness of vowel-targeted intervention with two school-age children with persistent speech difficulties using both percentage of vowel correct (PVC) and intelligibility measures as outcomes, showed both measures to be effective in identifying change. Three types of stimuli (single word naming, imitated and spontaneous sentences) were taken from the pre- and post-intervention vowel assessments, and played to 19 peer listeners (aged 9 to 11 years) who were asked to write down what had been said. The results showed that the two children made significant improvement in vowel production according to PVC scores. In addition, the listeners perceived more productions accurately after the therapy. Vowel-targeted intervention therefore resulted in increased PVC scores as well as improved intelligibility. Typically, intelligibility studies have included adult listeners to rate recordings. In contrast, Speake et al. (2012) included a peer group of child listeners to rate intelligibility as a measure of change following intervention. Speake et al. (2012) found that intelligibility was influenced by the perception and own language experiences of the listeners. Thus, the authors recommended the use of a range of listeners with different characteristics (e.g., age, familiarity with the children's speech, language experience) when using intelligibility as an outcome measure for children with speech difficulties.

Improving or maintaining intelligibility is an essential aim for speech and language therapists working with any client with a difficulty in speech output (Miller, 2013), and, with regards to children with SSD, is often anecdotally, what parents and teachers report to be their main aspirations for intervention. It remains important, therefore, to incorporate intelligibility, together with other more frequently used severity outcome measures, when exploring the effectiveness of speech sound interventions in enhancing overall speech production.

Intervention approaches to SSD

Research has shown that traditional articulation therapy, a motor-based approach to remediating SSD and phonological therapy, a linguistic-based approach, can both be effective in improving speech for children with SSD (Kahmi, 2006). In routine clinical practice, it appears, however, that clinicians adopt a more eclectic approach taking components from different approaches (Joffe and Pring, 2008). The success of articulation therapy has been acknowledged in the literature as standing the test of time for many clients with SSD (Bernthal and Bankson, 2004). It has, however, been argued that the nature of the SSD should determine the type of intervention chosen, and that children with a phonologically-based SSD should receive a linguistic-based intervention (Crosbie et al., 2005). Whilst articulation therapy has been shown to be effective in remediating articulation difficulties (Powers, 1971), there is greater consensus that phonological therapy is more effective for children with phonologically-based SSD (Klein, 1996, Pamplona et al., 1999, Gillon, 2000, Teutsch and Fox, 2004).

In spite of this, however, articulation therapy is still used by many clinicians for a variety of clients (Kahmi, 2006), and this is the case in Portugal, where the routine clinical practice of SLTs in treating children with *all* subgroups of speech sound disorders continues to be articulation therapy. This blanket use of articulation therapy for all children with SSD in Portugal was one of the key motivations for this study.

Aims of the study

A previous paper by the authors (Lousada et al., 2013) provides details of the study investigating the effectiveness of two types of treatment approaches to remediate phonologically-based SSD in a group of 14 pre- or early school-age children (aged from 4.0 to 6.7 years) using a randomised control intervention study: articulation therapy (AT) (Van Riper and Emerick, 1984) and phonological therapy (PT), that combined phonological awareness therapy (Gillon and McNeill, 2007) and listening and discrimination activities (Lancaster, 2008). Severity measures (PCC, percentage occurrence of different phonological processes and phonetic inventory) were used to study the effectiveness of both treatment approaches. Both treatments were found to be effective in improving speech, with the PT being the more effective of the two (Lousada et al., 2013). In this paper we report on a separate series of analyses focusing on intelligibility, as an outcome measure, with the same group of children. This new investigation aims to compare the effectiveness of two types of treatment approaches (phonological versus articulation) using intelligibility as the primary outcome measure. Intelligibility measures will also be discussed in relation to a severity measure (PCC score) in order to investigate the relative sensitivity and utility of each as an outcome measure. It was hypothesized that children with phonologically-based SSD treated with phonological therapy would show greater improvement in intelligibility than those given articulation therapy,; as phonological therapy focuses on phonological contrasts and encourages the reorganization of the sound system which potentially allows for greater transfer and generalization across the sound system (Dodd and Bradford, 2000, Baker and McLeod, 2004). This comparison between phonology and articulation therapy for children with SSD was conducted in the context of a country (Portugal) where all children with SSD, up until a couple of years ago, were treated exclusively with articulation therapy.

Method

Participants

The participants in the study include 14 children with phonologically-based SSD who participated in the intervention, and the 21 unfamiliar listeners who judged the children's intelligibility.

Children with phonologically-based SSD

Fourteen Portuguese children (10 boys and 4 girls) with phonologically-based SSD (mean age = 62.21 months; standard deviation = 11.00 months) were recruited through local speech and language therapists (SLTs). Children were diagnosed as having phonologically-based SSD after extensive assessment by a Speech and Language Therapist (SLT), an audiologist and a psychologist. Children with social or emotional problems, obvious neurological damage or childhood apraxia of speech were excluded. Speech errors were not due to oro-motor difficulties or structural problems, as assessed through an oral examination, and all speech sounds were stimulable when tested.

Participant inclusion criteria were: greater than 1.5 standard deviation (SD) below the mean on a standardised language test, the XX (TALC)[1] (Kay and Tavares, 2008), a level commonly in use by other research in diagnosing a language impairment (e.g., Swanson et al., 2005, Ebbels et al., 2007); audition of 20dB or lower in the frequencies 500Hz, 1000Hz and 2000Hz; and a discrepancy of, at least, 1 SD between language skills and non-verbal ability (NVIQ) with language being lower. The TALC contains two subtests: 1) Receptive subtest: semantic (vocabulary and semantic relations); syntax (sentence comprehension); and 2) Expressive subtest: semantic (vocabulary and semantic awareness); morphology (word structure); pragmatic (communicative intentions).

NVIQ was assessed with the Performance Scale of the Wechsler Preschool and Primary Scale of Intelligence – Revised (WPPSI-R) (Wechsler, 2003)[2].

---

[1] The standardisation of the TALC included 580 European Portuguese children. This test was used since it is the only test available to assess receptive and expressive language in Portuguese pre-school age children.

[2] The results were obtained with the Portuguese standardisation of this test (N=1352). The WPPSI-R is considered a reliable and valid assessment for Portuguese children (Seabra-Santos et al., 2003).

Some children (n = 6) had non-verbal abilities within the average range (above 85) and whose profile, therefore, was consistent with specific language impairment (SLI) (Leonard, 1998), while the remaining 8 had NVIQ ranging between 85 and 62 and can be considered to have more general language learning difficulties (see Table 1 for characteristics of the participants). The study was conducted with the approval of the local and national ethics committees and informed consent was collected from the parents of the children participating in the intervention study and the listeners judging intelligibility.

PUT TABLE 1 ABOUT HERE

The children were randomly assigned following simple randomisation procedures to one of two treatment groups (7 children in each group). A group of seven children was treated individually with Articulation Therapy (AT), and their progress compared with a group of seven children treated with Phonological Therapy (PT) (see Table 1).

An analysis of variance (ANOVA) was used to compare the PCC, receptive language, expressive language, NVIQ and age of the groups before the therapy and showed that, at pre-treatment there were no significant differences between groups in PCC ($F[1,12] = 0.304$, $p = 0.592$), receptive language ($F[1,12] = 2.346$, $p = 0.152$), expressive language ($F[1,12] = 2.120$, $p = 0.171$), NVIQ ($F[1,12] = 0.316$, $p = 0.584$) and age ($F[1,12] = 0.795$, $p = 0.390$).

Intervention

The intervention for both groups consisted of 25 individual weekly sessions of 45 minutes in duration, divided into 3 blocks (9 weeks + 8 weeks + 8 weeks) without any breaks. Each block had a different target. Both groups were treated by the same SLT, who was blind to the aim of the study, in order to reduce bias and minimise the influence of some confounding variables, for example, differences in practices across therapists. The study took place at the University of Aveiro in Portugal.

Phonological therapy included a combination of phonological awareness activities from Gillon and McNeill's (2007) programme and auditory discrimination and listening activities from Lancaster (2008). The phonological awareness activities from Gillon and McNeill's (2007) program included letter-sound knowledge, phoneme identity and phoneme matching, blending, segmentation, and phoneme manipulation. Focus for the first two sessions of each block was on listening and discrimination activities (Lancaster, 2008) whilst the remainder of the sessions focused on phonological awareness (Gillon and McNeill, 2007). During the phonological awareness activities, productions of the target sounds were elicited, as suggested by Gillon and McNeill (2007).

Articulation therapy consisted of a traditional articulation therapy approach following the 'Van Riper Method' (Van Riper and Emerick, 1984). In this approach the principal aim is to develop the child's ability to discriminate and articulate the target sound correctly in isolation, syllables, words, phrases and sentences. The first 2 sessions of each block focused on sensory perceptual training, with the remaining sessions focusing on production. Therapy was undertaken on one target sound at a time. Different activities were used during the sensory-perceptual training (e.g., detect sound errors in the clinician's speech). Direct instruction in the mechanism of sound production was used through different techniques (e.g., progressive approximation and phonetic placement). A more detailed description of each intervention approach can be found in Lousada et al. (2013).

Listeners

Twenty-one people (11 men and 10 women) between the ages of 22 and 44, unfamiliar to the children in the study, were recruited to judge the children's speech. Unfamiliar listeners were selected since the long term aim of therapy for children with phonologically-based SSD is to improve the ability to communicate with all people.

These listeners were professionals from different subject areas and worked or studied at the University of Aveiro. They all had normal hearing and were not familiar with speech-disordered children or with the aim of this study. The study was publicised at three departments at the University of Aveiro and the volunteers were recruited through an interview conducted by a SLT.

Speech samples

The study included the elicitation of two types of speech samples at pre- and post-intervention points: single words and continuous speech. Recordings of all speech samples were made in a sound-treated room at the speech, language and hearing (SLH) research laboratory at University of Aveiro, using a Cirrus Research MK224 microphone located 1 m in front of the child's mouth. The signal was preamplified (Cirrus Research MV 181 A), and then amplified and filtered by a Cirrus Research ZE 901B Preamplifier Power Supply. The acoustic signal was recorded using a Marantz PMD671 solid state recorder, with 16 bits and a sampling frequency of 48 kHz.

Single words (Sample A)

The children's phonological abilities were assessed with 67 single words from a standardised phonetic-phonological test (TFF-ALPE) devised for Portuguese Children (Lousada et al., 2012, Mendes et al., 2013). To assess word intelligibility, four groups of words were selected from the 67 words of the TFF-ALPE. Groups of words were matched according to number of syllables and syllabic structure (equal or similar). A total of 56 words (14 per group) were included in this sample (see Table 2). This procedure ensured that each rater listened only once to each word and reduced the confounding variable of familiarity for one set of words over another (Ertmer, 2010).

PUT TABLE 2 ABOUT HERE

Twenty-one participants acting as listeners took part in this experiment and were divided into seven clusters. Each cluster of three listeners rated one different set of words as illustrated in Table 3. For example listeners 1a, 1b and 1c rated words from group 1 of child CA (PT group) pre-treatment; words from group 2 of child CA (PT group) post-treatment; words from group 3 of child JC (AT group) post-treatment; and words from group 4 of child JC (AT group) pre-treatment. The groups of words were randomly selected for each phase (pre- or post-intervention) and for each type of intervention (articulation therapy or phonological therapy) ensuring equal distribution.

PUT TABLE 3 ABOUT HERE

Continuous speech (Sample B)

A continuous speech sample was also obtained through a picture description task (see Figures 1, 2 and 3 in Appendix A). Three pictures were specifically designed by the researchers and drawn by an illustrator to ensure that they depicted a wide range of everyday and familiar activities that would encourage speech and to maximise their attractiveness and subsequent engagement by the children. They depict different scenarios and actions in order to elicit a broad-ranging speech sample: Figure 1 – ‘Living Outdoors’, Figure 2 – ‘House and Garden’, Figure 3 – ‘The Zoo’.

The speech and language therapist (first author) used the following prompt to elicit responses to the picture stimuli: “what is happening in the picture?” She encouraged the child to talk about all aspects of the picture by using the prompt “what is happening here?” and pointing to the specific objects or actions that had been omitted.

The same twenty-one listeners assessed the speech of the children using the continuous speech stimuli. Listeners rated single words and continuous speech of a different set of children.

Procedures for the listeners

All listeners' sessions were co-ordinated by the first author. Judgements were made in the same sound-treated room used for the recordings. One individual session was used for each listener with a mean duration of 40 minutes.

Single word condition (Sample A)

Each cluster of three listeners heard all the words produced by two children in both the pre- and post-treatment conditions. One of these children was treated with the AT and the other with the PT to ensure that each listener rated children receiving both types of intervention. This eliminated the possibility of differences in listeners' judgements influencing the efficacy analysis (Ertmer, 2010). The 7 groups of single word samples (two children pre- and post- in each group) were randomly distributed to 7 groups of listeners (see Table 3).

Judges listened to each word only once and orthographically transcribed each word. Instructions given were as follows: "You will listen to single words produced by the children.; You can only listen to each word once; After listening to each word, write down on the Excel sheet what you think the child is saying, for example if the child says "tevisão" and you identify this as "televisão" (television), please write "televisão" near the number of this word". The listeners were therefore encouraged to write what they thought the child was trying to say, rather than write verbatim the child's speech production (Pascoe et al., 2006). This was to attempt to measure intelligibility rather than another way of scoring severity. The pictures used to elicit the words were not provided.

Continuous speech condition (Sample B)

Similarly to the single word condition, each cluster of three listeners heard speech samples of two children (one sample taken at pre-treatment assessment and another sample taken at post-treatment assessment). One of these children was treated with the AT and the other with the PT. The 7 groups of continuous speech samples (two children pre- and post- in each group) were randomly distributed to 7 groups of listeners (see Table 4).

In this condition, a rating scale was used since four children had severe speech difficulties and it was not always possible to know exactly what the children were saying.

The participants were asked to listen to the samples and immediately afterwards to classify them according to their degree of understanding using the rating scale supplied. The following instructions were provided: "Select the number that best corresponds to the child's speech using this rating scale ("1 – impossible to understand"; "2- very difficult to understand"; "3 – difficult to understand"; "4 – easy to understand"; "5 – very easy to understand")". Judges listened to the continuous speech samples once only and did this without seeing the pictures used to elicit them. The 5-point Likert scale was pre-tested on a small pilot study and listeners reported that the scale was useful in identifying their level/degree of understanding and that they understood the instructions. A 5-point Likert scale is also suggested by McLeod et al. (2012) in a recent study to measure intelligibility of children.

Sequence of the listeners' sessions

The listeners first heard two samples of continuous speech (Sample B - part 1), each sample produced by a different child (either at pre- or post-intervention for each child). Then they heard Sample A (with single words) followed by Sample B – part 2, with the remaining assessment points for the same children they had heard at the start of the session. For example, the cluster of listeners 1a, 1b and 1c heard the continuous samples of child AM at pre-treatment and child MS at post-treatment (Sample B – part 1), then they heard Sample A, with single words (of child CA at pre- and post-treatment and child JC at pre- and post-treatment) and finally they heard the continuous samples of child MS at pre-treatment and child AM at post-treatment (see Table 4). The insertion of Sample A between the two ratings of sample B helped reduce the possibility of the listeners becoming too familiar with their selected children during condition B.

PUT TABLE 4 ABOUT HERE

Data Analysis

Speech intelligibility was assessed by two methods: single word identification procedure and a rating procedure of continuous speech.

Single word (Sample A) - identification procedure

A single word identification procedure was used to assess single word intelligibility. The percentage of words correctly identified by each listener was calculated for each child and for each assessment (pre- and post-intervention). The mean of the three listeners' percentages was used as the measure of intelligibility for each child. A paired t-test was used to assess differences between pre- and post-treatment values, for each treatment group.

Continuous speech (Sample B) - rating procedure

The ratings of continuous speech (1-5) were obtained for each child and for each assessment point (pre- and post-intervention). The mean of the three listeners' classifications was used as the measure for each child. In order to assess differences between pre- and post-treatment values, within each treatment group, Wilcoxon Signed Rank Test was used. Since continuous speech intelligibility values were based on ratings (1-5) a non-parametric analysis was selected as opposed to the parametric tests used for single words intelligibility (based on an average of percentage values).

Additionally, intelligibility scores obtained for both methods will also be discussed relative to their PCC scores, as suggested by Pascoe et al. (2006). Thus PCC scores (obtained with the same 67 single words) calculated in our previous study will be also presented.

Interrater reliability

Single word (Sample A) - identification procedure

Interrater reliability was analysed between the judges who listened to the same recordings (see Table 3) through the *kappa free-marginal* statistic (Randolph, 2005). This index of agreement was selected as it is appropriate for nominal data.

The *kappa* values obtained for the single word identification procedure are presented in Table 5. Values of *kappa* can vary between -1.0 and 1.0, with -1.0 indicating complete disagreement between raters and 1.0 indicating perfect agreement between raters. Landis and Koch (1977) suggested the following interpretation of *Kappa* coefficient: poor agreement (< 0); slight agreement (0.0 - 0.20); fair agreement (0.21 - 0.40); moderate agreement (0.41 - 0.60); substantial agreement (0.61 - 0.80); almost perfect agreement (0.81 - 1).

PUT TABLE 5 ABOUT HERE

The values obtained in the present study indicate substantial agreement for all groups of listeners (*kappa* > 0.61) except for group 6a, 6b and 6c for which the value (*kappa* = 0.90) shows an almost perfect agreement. The values obtained suggest strong reliability between listeners.

Continuous speech (Sample B) - Intelligibility rating

Considering the characteristics of the current study, interrater reliability was analysed through the *Intraclass Correlation Coefficient (ICC)* (1,k) (Shrout and Fleiss, 1979). This ICC design is indicated when each subject is assessed by a different set of randomly selected raters and reliability is calculated by taking an average of the k raters' measurements. This analysis uses one-way ANOVA results: between subjects mean squares (BMS) and within subjects mean squares (WMS), as evident below.

$$ICC(1,k) = \frac{BMS - WMS}{BMS}$$

Table 6 shows the ICC values obtained for the rating scale (continuous speech task): at pre-treatment for AT group; at pre-treatment for PT; at post-treatment for AT group; at post-treatment for PT group.

PUT TABLE 6 ABOUT HERE

ICC values above 0.75 represent excellent reliability, values between 0.4 and 0.74 indicate moderate to good reliability and values below 0.4 represent poor reliability (Fleiss, 1986). All the observed ICC values suggest excellent agreement between listeners except for the classifications of speech obtained at pre-treatment for children in the AT group which is considered 'moderate to good'. The same 21 listeners were involved in the calculation of each of the 4 ICC values.

Results

Speech Intelligibility

Single word (Sample A) identification procedure

The results obtained for each child and for each intervention group (AT and PT), before and after the intervention, are presented in Table 7. A mean of the three raters who listened to the same child (e.g., listener 1a, 1b and 1c) was calculated since excellent reliability was obtained between raters.

The results for children in the PT group showed an increase in percentage of intelligible words after the intervention, except for child AM, where a small decrease was observed (from 2.38% to 0.00%).

For children in the AT group, an increase in percentage of intelligible words was also observed, except for 2 children, DG and AP, where a decrease was evident from 23.81% to 14.29% and 9.52% to 7.14%, respectively.

PUT TABLE 7 ABOUT HERE

Statistical analyses were used to compare the differences between the percentage of intelligible words at pre- and post-treatment assessment in each group. A significant difference was found in the PT group (paired t (6) = 4.409, p = 0.005) pre- to post-treatment, with no significant difference seen in the AT group (paired t (6) = 1.763, p = 0.128).

Continuous speech (Sample B) - Intelligibility rating

The results obtained for intelligibility in continuous speech for each child before and after treatment are summarised in Table 8. A mean of the three raters who listened to the same child (e.g., listener 1a, 1b and 1c) was calculated as excellent inter-rater reliability was obtained.

All children in the PT group showed an increase in intelligibility in continuous speech after the intervention. For children in the AT group, an increase in intelligibility in continuous speech was observed in 5 of the 7 children.

PUT TABLE 8 ABOUT HERE

Table 8 shows the mean results obtained for condition B (continuous speech) at pre- and post-treatment assessments for both intervention groups (AT and PT). A significant difference was found in the PT group (Asymptotic Z = 2.371, p = 0.018) pre- to post-treatment. In contrast, results showed no significant difference in the AT group (Asymptotic Z = 1.442, p = 0.149). Although both approaches aimed to improve speech intelligibility, a significant effect in intelligibility of continuous speech was only observed for children in the PT group.

PCC scores and intelligibility measures

The PCC scores for the children in the PT and AT group at pre- and post-treatment assessment, reported in our previous study (Lousada et al., 2013), are summarised in Table 9. All children in the PT and AT group showed an increase in PCC after the intervention, however, the magnitude of the increase was very different from child to child (Lousada et al., 2013). Children in the PT group showed greater improvement in PCC than those in the AT group. However, unlike the intelligibility measures, the PCC scores of both intervention groups increased at post-intervention. Improvements in severity (as measured by PCC) were noted in some of the children (AM, DG, AP), but these improvements did not translate into improvements in intelligibility. We look more closely at these three case studies in the Discussion.

PUT TABLE 9 ABOUT HERE

Discussion

This study evaluated the effectiveness of two intervention approaches (PT and AT) for the treatment of 14 Portuguese children, aged 4;0-6;7 years, with phonologically-based SSD, by looking at intelligibility measures as the primary clinical outcome.

The percentage of intelligible words and intelligibility in continuous speech from pre- to post-treatment showed a significant difference only in the PT group suggesting that only children receiving PT improved significantly in intelligibility after treatment.

This finding supports Dodd and Bradford's (2000) conclusions, that articulation therapy is limited in its impact on the speech production of children with phonologically-based SSD, since this approach focuses on individual speech sound production at a motoric level, and not on the elimination of error patterns. In contrast, phonological therapy which aims to reorganise the child's linguistic system, has the potential to have a greater impact on intelligibility. So, whilst both groups improved in severity (according to PCC), only the PT group made significant changes in intelligibility.

It is also important to note that participants were children with concomitant language disorder and this may have contributed to the results as it is possible that children with more specific speech sound disorders would have shown greater improvement in intelligibility..

The results obtained for each child for the single word condition show that for the PT group, all children improved on intelligibility, except child AM, where a small decrease in the percentage of intelligibility was observed (from 2.38% to 0.00%). The percentage of intelligible words obtained at pre-treatment assessment (2.38%) was based on one word that was intelligible for only one listener. It can be concluded that for this one child, no improvement in single word intelligibility was observed. In spite of some improvement in PCC score (from 16.04% to 35.29%), this child continues to present with many phonological processes after intervention, and consequently his speech continues to be unintelligible. For example, this child said [pa] for target word “prato” (dish) at pre-treatment assessment and said [ˈpat] for the same target word at post-treatment assessment. In these examples it is difficult to identify the target word at either pre- or post-treatment assessments because [pa] and [ˈpat] could be many words: [ˈpatu]; [ˈpɾatu]; [saˈpatu] (these productions could only be correctly identified if they were produced in a sentence). This increase in PCC from pre- to post-treatment assessment, indicating an improvement in the child’s productions, as is evident from the above example, does not have any significant impact on word intelligibility for this child.

Children in the AT group obtained higher scores post-treatment on the intelligibility of single words, except for child DG and child AP, for whom the percentage of intelligible words decreased after treatment (from 23.81% to 14.29%, and 9.52% to 7.14% respectively).

For DG, in spite of a small improvement in one severity measure (PCC increase from 54.01% to 57.75%), his productions continue to be unintelligible after treatment. For example, this child produced [ˈkid] for target word “tigre” (tiger) [ˈtigɾɨ] at pre-treatment assessment and produced [ˈtiʀ] for the same target word at post-treatment assessment, showing an improvement in his production as the phonological process ‘backing’ (/k/ produced as [t]) was no longer used at post-treatment assessment. However, both productions are difficult to understand, at least when they are produced without any context, as can be seen from the intelligibility measure.

Concerning AP, an increase in PCC was also observed (increased from 21.39% at pre-treatment assessment to 28.88% at post-treatment assessment), although his productions also continue to be difficult to understand after treatment. For example, AP produced

[ˈtatu] for target word “carro” (car) [ˈkaʀu] at pre-treatment assessment and [ˈkatu] for the same target word at post-treatment assessment. This example shows again that in spite of an increase in the PCC (as a consequence of the absence of the phonological process “fronting” at post-treatment assessment), this child’s speech continues to be difficult to understand after the treatment, with words produced without any context. This pattern of results shares some similarities with those of Pascoe et al. (2006), where one of their participants showed some improvement in severity with no concomitant increase in intelligibility. All three of the children in the examples above (AM, DG, AP) make improvements in their speech production (as evident from PCC scores and absence of a phonological process), yet these improvements are insufficient to show any marked changes in overall intelligibility.

The results obtained for each child for the continuous speech condition suggest that all children in the PT group improved their speech intelligibility. In the AT group, an improvement was observed in 5 of the 7 children. No improvement was found in intelligibility ratings for DG and TM (both of whom showed a small decrease in overall intelligibility, as measured by a rating scale). The lack of improvement in intelligibility for these two children reflects to some extent, the severity measures, where PCC scores showed only a small improvement after intervention, smaller than all the other children in the AT group (see Table 9).

To improve speech intelligibility is the long term aim of intervention for children with speech problems (Flipsen, 1995, Dodd and Bradford, 2000). However, intervention studies with children with phonological delay or disorder have not routinely used intelligibility as an outcome measure (Baker and McLeod, 2011, Miller, 2013), and more typically use severity measures such as PCC and/ or percentage of occurrence of phonological processes (Pascoe et al., 2006). In our previous study using severity measures (Lousada et al., 2013), it was concluded that both treatments were effective at improving speech but that children receiving PT showed a more significant improvement in PCC scores than children treated with AT. The results obtained in the present study, using intelligibility as the outcome measure, provide some interesting additional insights into the children's performance and help to further differentiate the two interventions according to levels of effectiveness. Similarly to the severity outcome measures, the intelligibility measures showed the PT to be effective in improving speech, both in single words and in continuous speech. In contrast to the severity findings, however, intelligibility did not significantly improve for children receiving the AT. Therefore, whilst this approach was found to be successful when using severity outcomes in both single words and continuous speech, this was not the case when using intelligibility as the outcome measure; with no significant increase found in the AT group from pre- to post-treatment. Thus, it could be argued that intelligibility is a more stringent measure of change compared with other severity measures that are commonly in use. These findings emphasise the importance of using a range of outcome measures when working with children with phonologically-based SSD and places intelligibility as an important factor to take into account when evaluating clinical outcomes.

Conclusions and implications

This study emphasises the importance of using intelligibility as a clinical outcome measure following intervention with children with phonological problems. Findings showed that only the PT was effective in significantly improving children's speech as assessed by both severity and intelligibility measures. In spite of the difficulties in routinely assessing intelligibility in practice (e.g., finding a group of appropriate listeners, time constraints etc.), it is evident that intelligibility measures could add relevant information when assessing the efficacy of intervention for children with phonologically-based SSD; information that could be lost when using severity measures alone. It is important to use a combination of the two measures (PCC and intelligibility): intelligibility could be used as an end-point outcome measure to assess the long term goal of functional communication, whilst PCC can be used along the intervention trajectory to chart change and build towards the final outcome of improved intelligibility.

Some materials in this research have been adapted from a standardised test for use in this study (e.g., matching a group of words to use at pre- and post-treatment assessment) and similar materials could be used for both severity and intelligibility measures in everyday clinical practice. These materials could also be useful for SLTs in the analysis of intelligibility of single words, to monitor change following intervention, in children with different disorders (e.g., other subgroups of SSD and dysarthria) as well as to compare the effectiveness of different types of interventions, as was done in this study. The pictures and rating scale are also potentially valuable clinical tools to use with severely unintelligible speech when target words are not known.

The role and importance of intelligibility as an assessment tool and as an outcome measure for children with SSD is well accepted, but more research is needed to determine how best to evaluate functional intelligibility in school-aged children, from the perspectives of the therapist, parent and teacher, and across a range of different contexts, using a procedure that is convenient, relatively quick and easily administered and with good validity and reliability.

Acknowledgments

The authors would like to thank to Professor Tim Pring, City University London for his collaboration in the study design.

This work was partially funded by FEDER through the Operational Program Competitiveness Factors - COMPETE and by National Funds through FCT - Foundation for Science and Technology in the context of the project FCOMP-01-0124-FEDER-022682 (FCT reference PEst-C/EEI/UI0127/2011) and Incentivo/EEI/UI0127/2013.

Appendix A - Figures

Figure 1. First picture used in the picture description task.

Figure 2. Second picture used in the picture description task.

Figure 3. Third picture used in the picture description task.

Tables

Table 1. Characteristics of participants: gender, age (months), Non-verbal intelligence (NVIQ) (Standard score), receptive language (raw score), expressive language (raw score), PCC and intervention group.

Child	Gender	Age (months)	NVIQ (Mean = 100; SD = 15)	Receptive language		Expressive language		PCC	Intervention group
CA	F	50	117	61	WNL	30	Below 2 SD	67.91%	PT
AM	M	64	66	63	WNL	28	Below 3 SD	16.04%	PT
MR	F	48	89	55	WNL	28	Below 3 SD	59.36%	PT
LA	F	62	83	64	WNL	22	Below 4 SD	73.80%	PT
DM	M	79	109	64	WNL	44	Below 1.5 SD	46.52%	PT
AD	F	50	82	58	WNL	3	Below 8 SD	19.79%	PT
RM	M	64	62	62	WNL	33	Below 3 SD	59.89%	PT
JC	M	77	63	65	WNL	26	Below 5 SD	31.02%	AT
MS	M	48	85	55	WNL	26	Below 3 SD	50.80%	AT
RF	M	57	84	53	Below 1.5 SD	24	Below 4 SD	71.66%	AT
DG	M	63	87	62	WNL	28	Below 3 SD	54.01%	AT
FP	M	75	66	57	Below 2 SD	22	Below 8 SD	23.53%	AT
AP	M	75	66	58	Below 2 SD	22	Below 8 SD	21.39%	AT
TM	M	59	116	55	WNL	21	Below 4 SD	48.13%	AT
Mean (Standard deviation)		62.21(11.00)	83.93 (18.96)						

Abbreviation: WNL - within normal limits.

Table 2. Words used in Sample A

Number of syllables	Group 1		Group 2		Group 3		Group 4	
	Word	Syllabic structure	Word	Syllabic structure	Word	Syllabic structure	Word	Syllabic structure
3	Sapato	CV-"CV-CV	Cabelo	CV-"CV-CV	Vassoura	CV-"CV-CV	Janela	CV-"CV-CV
2	Jipe	"CV-CV	Rato	"CV-CV	Pente	"CV-CV	Faca	"CV-CV
2	Bola	"CV-CV	Dedo	"CV-CV	Gato	"CV-CV	Chave	"CV-CV
2	Mesa	"CV-CV	Cama	"CV-CV	Carro	"CV-CV	Ponte	"CV-CV
2	Balde	"CVC-CV	Porco	"CVC-CV	Porta	"CVC-CV	Gordo	"CVC-CV
2	Carne	"CVC-CV	Força	"CVC-CV	Garfo	"CVC-CV	Polvo	"CVC-CV
2	Zebra	"CV-CCV	Cobra	"CV-CCV	Tigre	"CV-CCV	Vidro	"CV-CCV
2	Prato	"CCV-CV	Frango	"CCV-CV	Creme	"CCV-CV	Planta	"CCV-CV
4	Televisão	CV-CV-CV-"CVV	Telefone	CV-CV-CV-"CVV	Bicicleta	CV-CV-"CCV-CV	Almofada	VC-CV-"CV-CV
2	Chapéu	CV-"CVG	Caixa	CV-"CVG	Peixe	"CVG-CV	Queijo	"CVG-CV
2	Nariz	CV-"CVC	Comer	CV-"CVC	Pesca	"CVC-CV	Pasta	"CVC-CV
2	Brincar	CCV-"CVC	Quatro	CCV-"CVC	Dragão	CCV-"CVV	Livro	"CV-CCV
2	Água	"V-CGV	Unha	"V-CGV	Lua	"CV-V	Olho	"V-CV
3	Gravata	CCV-"CV-CV	Formiga	CCV-"CV-CV	Estrela	VC-"CCV-CV	Escrever	VC-CCV-"CVC

Table 3. Study design of Sample A – Single words

Listeners 1a, 1b and 1c*	Child CA (PT)	Pre-treatment	Post-treatment
	Words	Group 1	Group 2
	Child JC (AT)	Post-treatment	Pre-treatment
	Words	Group 3	Group 4
Listeners 2a, 2b and 2c	Child AM (PT)	Post-treatment	Pre-treatment
	Words	Group 4	Group 3
	Child MS (AT)	Pre-treatment	Post-treatment
	Words	Group 2	Group 1
Listeners 3a, 3b and 3c	Child RF (AT)	Pre-treatment	Post-treatment
	Words	Group 1	Group 2
	Child MR (PT)	Post-treatment	Pre-treatment
	Words	Group 3	Group 4
Listeners 4a, 4b and 4c	Child DG (AT)	Post-treatment	Pre-treatment
	Words	Group 4	Group 3
	Child LA (PT)	Pre-treatment	Post-treatment
	Words	Group 2	Group 1
Listeners 5a, 5b and 5c	Child DM (PT)	Pre-treatment	Post-treatment
	Words	Group 1	Group 2
	Child FP (AT)	Post-treatment	Pre-treatment
	Words	Group 3	Group 4
Listeners 6a, 6b and 6c	Child AP (AT)	Post-treatment	Pre-treatment
	Words	Group 4	Group 3
	Child AD (PT)	Pre-treatment	Post-treatment
	Words	Group 2	Group 1
Listeners 7a, 7b and 7c	Child RM (PT)	Pre-treatment	Post-treatment
	Words	Group 1	Group 2
	Child TM (AT)	Post-treatment	Pre-treatment
	Words	Group 3	Group 4

*1a – listener 1 in group 1; 1b – listener 2 in group 1, 1c – listener 3 in group 1.
Note: PT – phonological therapy; AT – articulation therapy

Table 4. Study design of Sample B – continuous speech

	Sample B – part 1		Sample A	Sample B – part 2	
Listeners 1a, 1b and 1c	Child AM (PT) Pre-treatment	Child MS (AT) Post-treatment		Child MS Pre-treatment	Child AM Post-treatment
Listeners 2a, 2b and 2c	Child RF (AT) Pre-treatment	Child MR (PT) Post-treatment		Child MR Pre-treatment	Child RF Post-treatment
Listeners 3a, 3b and 3c	Child LA (PT) Post-treatment	Child DG (AT) Pre-treatment		Child DG Post-treatment	Child LA Pre-treatment
Listeners 4a, 4b and 4c	Child FP (AT) Post-treatment	Child DM (PT) Pre-treatment		Child DM Post-treatment	Child FP Pre-treatment
Listeners 5a, 5b and 5c	Child AD (PT) Pre-treatment	Child AP (AT) Post-treatment		Child AP Pre-treatment	Child AD Post-treatment
Listeners 6a, 6b and 6c	Child TM (AT) Pre-treatment	Child RM (PT) Post-treatment		Child RM Pre-treatment	Child TM Post-treatment
Listeners 7a, 7b and 7c	Child CA (PT) Post-treatment	Child JC (AT) Pre-treatment		Child JC Post-treatment	Child CA Pre-treatment

*1a – listener 1 in group 1; 1b – listener 2 in group 1, 1c – listener 3 in group 1.
Note: PT – phonological therapy; AT – articulation therapy.

Table 5. Kappa values obtained for 7 groups of listeners.

Listeners	Kappa
Listeners 1a, 1b and 1c	0.69
Listeners 2a, 2b and 2c	0.69
Listeners 3a, 3b and 3c	0.64
Listeners 4a, 4b and 4c	0.61
Listeners 5a, 5b and 5c	0.73
Listeners 6a, 6b and 6c	0.90
Listeners 7a, 7b and 7c	0.71

Table 6. Intraclass Correlation Coefficient (ICC) values obtained.

	ICC
Pre-treatment; AT	0.673
Pre-treatment; PT	0.924
Post-treatment; AT	0.949
Post-treatment; PT	0.877

Note: PT – phonological therapy; AT – articulation therapy.

Table 7. Intelligibility (percentage) in single words at pre- and post-treatment assessment for each child (mean of the three raters).

Child	Therapy	Intelligibility in single-words (%)			
		Pre-treatment		Post-treatment	
			Mean (SD)		Mean (SD)
CA	PT	50.00		71.43	
AM	PT	2.38		0.00	
MR	PT	16.67	25.85 (17.68)	64.29	58.84 (31.09)
LA	PT	42.86	Min – Max: 2.38-50.00	85.71	Min – Max: 0.00-85.71
DM	PT	23.81		78.57	
AD	PT	9.52		33.33	
RM	PT	35.71		78.57	
JC	AT	16.67		30.95	
MS	AT	40.48		47.62	
RF	AT	26.19	20.75 (11.80)	71.43	32.31 (22.37)
DG	AT	23.81	Min –Max: 4.76-40.48	14.29	Min – Max: 7.14-71.43
FP	AT	4.76		16.67	
AP	AT	9.52		7.14	
TM	AT	23.81		38.10	

Table 8. Intelligibility of continuous speech (rating scale) at pre- and post-treatment assessment for each child (mean of the three raters)..

Child	Intervention	Intelligibility in continuous speech			
		Pre-treatment		Post-treatment	
			Mean (SD)		Mean (SD)
CA	PT	3.67		4.00	
AM	PT	1.00		2.00	
MR	PT	3.00	2.48 (0.96)	4.33	3.62 (0.80)
LA	PT	3.00	Min – Max:	3.67	Min – Max:
DM	PT	2.67	1.00-3.67	4.33	2.00-4.33
AD	PT	1.33		3.33	
RM	PT	2.67		3.67	
JC	AT	1.67		2.00	
MS	AT	2.33	1.86 (0.50)	4.00	2.62 (1.11)
RF	AT	2.33	Min – Max:	4.00	Min – Max:
DG	AT	2.33	1.00-2.33	2.00	1.00-4.00
FP	AT	1.00		3.00	
AP	AT	1.67		2.33	
TM	AT	1.67		1.00	

Table 9. PCC score (percentage) obtained at pre- and post-treatment assessments for both intervention groups (PT and AT).

Child	Intervention group	PCC (%)			
		Pre-treatment		Post-treatment	
			Mean (SD)		Mean (SD)
CA	PT	67.91	49.04 (22.89) Min – Max: 16.04-73.80	75.40	67.23 (20.83) Min – Max: 35.29-89.30
AM	PT	16.04		35.29	
MR	PT	59.36		75.94	
LA	PT	73.80		89.30	
DM	PT	46.52		74.33	
AD	PT	19.79		40.11	
RM	PT	59.89		80.21	
JC	AT	31.02	42.93 (18.35) Min – Max: 21.39-71.66	39.57	50.42 (19.02) Min – Max: 28.88-77.01
MS	AT	50.80		70.59	
RF	AT	71.66		77.01	
DG	AT	54.01		57.75	
FP	AT	23.53		29.95	
AP	AT	21.39		28.88	
TM	AT	48.13		49.20	

References

ALLEN, M., 2013, Intervention efficacy and intensity for children with speech sound disorders. *Journal of Speech, Language and Hearing Research,* 56, 865-877.

BAKER, E. and MCLEOD, S., 2004, Evidence-based management of phonological impairment in children. *Child Language Teaching and Therapy,* 20, 261-285.

BAKER, E. and MCLEOD, S., 2011, Evidence-based practice for children with speech sound disorders. Part 1: Narrative review. *Language, Speech, and Hearing Services in Schools,* 42, 102-139.

BAUDONCK, N. L. H., BUEKERS, R., GILLEBERT, S. and LOIERDE, K. M. V., 2009, Speech intelligibility of Flemish children as judged by their parents. *Folia Phoniatrica et Logopaedica,* 61, 288-295.

BERNTHAL, J. E. and BANKSON, N. W., 2004, *Articulation and Phonological Disorders* (Boston: Allyn and Bacon).

BOWEN, C. and CUPPLES, L., 1999, A phonological therapy in depth: a reply to commetaries. *International Journal of Language & Communication Disorders,* 34, 65-83.

BOWEN, C. and CUPPLES, L., 2006, PACT: Parents and children together in phonological therapy. *Advances in Speech-Language Pathology,* 8, 282-292.

CAMARATA, S., 1993, The application of naturalistic conversation training to speech production in children with speech disabilities. *Journal of Applied Behavior Analysis,* 26, 173-182.

CROSBIE, S., HOLM, A. and DODD, B., 2005, Intervention for children with severe speech disorder: A comparison of two approaches. *International Journal of Language & Communication Disorders*, 40, 467–491.

DE BODT, M. S., MARIA E. HERNÁNDEZ-DíAZ HUICI, M. E. and VAN DE HEYNINGA, P. H., 2002, Intelligibility as a linear combination of dimensions in dysarthric speech. *Journal of Communication Disorders*, 35, 283-292.

DODD, B. and BRADFORD, A., 2000, A comparison of three therapy methods for children with different types of developmental phonological disorders. *International Journal of Language & Communication Disorders*, 35, 189-209.

EBBELS, S., VAN DER LELY, H. K. J. and DOCKRELL, J., 2007, Intervention for verb argument structure in children with persistent SLI: a randomized control trial. *Journal of Speech, Language and Hearing Research*, 50, 1330-1349.

ERTMER, D. J., 2010, Assessing Speech Intelligibility in Children with Hearing Loss: Toward Revitalizing a Valuable Clinical Tool. *Language Speech and Hearing Services in Schools*, 42, 52-58.

FLEISS, J., 1986, *Reliability of measurements* (New York: Wiley).

FLIPSEN, P., 1995, Speaker-listener familiarity: parents as judges of delayed speech intelligibility. *Journal of Communication Disorders*, 28, 3-19.

GILLON, G. T., 2000, The efficacy of phonological awareness intervention for children with spoken language impairment. *Language, Speech, and Hearing Services in Schools*, 31, 126-141.

GILLON, G. T. and MCNEILL, B. C., 2007, *Integrated phonological awareness: An intervention program for preschool children with speech-language impairment* (New Zealand: University of Canterbury).

GORDON-BRANNAN, M. and HODSON, B., 2000, Intelligibility/severity measurements of prekindergarten children's speech. *American Journal of Speech-Language Pathology,* 9, 141-150.

HESKETH, A., ADAMS, C., NIGHTINGALE, C. and HALL, R., 2000, Phonological awareness therapy and articulatory training approaches for children with phonological disorders: a comparative outcome study. *International Journal of Language & Communication Disorders,* 35, 337-354.

HODSON, B. and PADEN, E., 1981, Phonological processes which characterize unintelligible and intelligible speech in early childhood. *Journal of Speech and Hearing Disorders,* 46, 369-373.

HUNTER, L., PRING, T. and MARTIN, S., 1991, The use of strategies to increase speech intelligibility in cerebral palsy: an experiment evaluation. *British Journal of Disorders of Communication*, 26, 163-74.

HUSTAD, K. C., 2012, Speech intelligibility in children with speech disorders *Perspectives on Language Learning and Education,* 19, 7-11.

JOFFE, V. and PRING, T., 2008, Children with phonological problems: a survey of clinical practice. *International Journal of Language & Communication Disorders,* 43, 154-164.

KAHMI, A., 2006, Treatment decisions for children with speech-sound disorders. *Language, Speech, and Hearing Services in Schools,* 37, 271-279.

KAIN, A. B., HOSOM, J. P., NIU, X., VAN SANTEN, J. P. H., FRIED-OKEN, M. and STAEHELY, J., 2007, Improving the intelligibility of dysarthric speech. *Speech Communication*, 49, 743-759.

KAY, E. and TAVARES, D., 2008, *Teste de avaliação da linguagem na criança* (Lisboa: Oficina Didáctica).

KEMPLER, D. and VAN LANCKER, D., 2002, Effect of Speech Task on Intelligibility in Dysarthria: A Case Study of Parkinson's Disease. *Brain and Language*, 80, 449–464.

KLEIN, E., 1996, Phonological/traditional approaches to articulation therapy: A retrospective group comparison. *Language, Speech, and Hearing Services in Schools*, 27, 314-323.

KLEIN, E. and FLINT, C., 2006, Measurement of intelligibility in disordered speech. *Language, Speech, and Hearing Services in Schools*, 37, 191-199.

KONST, E., H., W.-B., RIETVELD, T. and PETERS, H., 2000, An intelligibility assessment of toddlers with cleft lip and palate who received and did not receive presurgical infant orthopedic treatment *Journal of Communication Disorders*, 33, 483-499.

KWIATKOWSKI, J. and SHRIBERG, L. D., 1992, Intelligibility assessment in developmental phonological disorders: accuracy of caregiver gloss. *Journal of Speech and Hearing Research*, 35, 1095-1104.

LANCASTER, G., 2008, *Developing speech and language skills* (London: Routledge).

LANCASTER, G., KEUSCH, S., LEVIN, A., PRING, T. and MARTIN, S., 2010, Treating children with phonological problems: Does an eclectic approach to therapy work? *International Journal of Language & Communication Disorders,* 45, 174-181.

LANDIS, J. R. and KOCH, G. G., 1977, The measurement of observer agreement for categorical data. *Biometrics,* 33, 159-174.

LEONARD, L., 1998, *Children with specific language impairment* (Cambridge: MIT Press).

LOUSADA, M., JESUS, L. M. T., CAPELAS, S., MARGAÇA, C., SIMÕES, D., VALENTE, A. R., HALL, A. and JOFFE, V. L., 2013, Phonological and articulation treatment approaches in Portuguese children with speech and language impairments: a randomised controlled intervention study. *International Journal of Language & Communication Disorders,* 48, 172-187.

LOUSADA, M., MENDES, A. P., VALENTE, A. R. and HALL, A., 2012, Standardization of a phonetic-phonological test for European Portuguese children. *Folia Phoniatrica et Logopaedica* 64, 151-156.

MCLEOD, S., HARRISON, L. and MCCORMACK, J., 2012, The intelligibility in context scale: validity and reliability of a subjective rating measure. *Journal of Speech, Language and Hearing Research,* 55, 648-656.

MENDES, A., AFONSO, E., LOUSADA, M. and ANDRADE, F., 2013, *Teste Fonético-Fonológico ALPE* (Aveiro: Edubox).

MILLER, N., 2013, Measuring up to Speech Intelligibility. *International Journal of Language & Communication Disorders,* 6, 601-612.

PAMPLONA, M. C., YSUNZA, A. and ESPINOSA, J., 1999, A comparative trial of two modalities of speech intervention for compensatory articulation in cleft palate children, phonologic approach versus articulatory approach. *International Journal of Pediatric Otorhinolaryngology*, 49, 21-26.

PASCOE, M., STACKHOUSE, J. and WELLS, B., 2006, *Persisting speech difficulties in children - Children's speech and literacy difficulties: Book 3* (Chichester: Wiley).

POWERS, M. H., 1971, Clinical educational procedures in functional disorders of articulation. *In:* L. E. TRAVIS (ed.) *Handbook of speech pathology and audiology.* (Englewood Cliffs, N.J.: Prentice-Hall), pp. 875-910.

RANDOLPH, J. J., 2005, Free-marginal multirater kappa: An alternative to Fleiss' fixed-marginal multirater kappa. Paper presented at the Joensuu University Learning and Instruction Symposium. Joensuu, Finland.

SEABRA-SANTOS, M. J., SIMÕES, M. R., ALBUQUERQUE, C. P., PEREIRA, M. M., ALMEIDA, L. S., FERREIRA, C., LANÇA, C. and LOPES, A. F., 2003, Escala de Inteligência de Wechsler para a Idade Pré-Escolar e Primária - Forma Revista (W.P.P.S.I.-R.). *In:* M. M. GONÇALVES, M. R. SIMÕES, L. S. ALMEIDA and C. MACHADO (eds.) *Avaliação psicológica: Instrumentos validados para a população portuguesa.* (Coimbra: Quarteto), pp. 197-219.

SHROUT, P. E. and FLEISS, J. L., 1979, Intraclass correlations: Uses in assessing rater reliability *Psychological Bulletin*, 86, 420-428.

SKAHEN, S. M., WATSON, M. and LOF, G. L., 2007, Speech-Language Pathologists' Assessment Practices for Children With Suspected Speech Sound Disorders: Results of a National Survey. *American Journal of Speech - Language Pathology*, 16, 246-259.

SPEAKE, J., STACKHOUSE, J. and PASCOE, M., 2012, Vowel targeted intervention for children with persisting speech difficulties: impact on intelligibility. *Child Language Teaching and Therapy,* 28, 277-295.

SWANSON, L. A., FEY, M. E., MILLS, C. E. and HOOD, L. S., 2005, Use of Narrative-Based Language Intervention With Children Who Have Specific Language Impairment. *American Journal of Speech - Language Pathology,* 14, 131-143.

TEUTSCH, A. and FOX, A. V., 2004, Effectiveness of articulatory versus phonological therapy of phonological disorders in children: a pilot study. *Sprache-Stimme-Gehor*, 28, 178-185.

VAN RIPER, C. and EMERICK, L., 1984, *Speech correction: An introduction to speech pathology and audiology* (Englewood Cliffs, NJ: Prentice-Hall).

WECHSLER, D., 2003, *WPPSI-R. Escala de inteligência de wechsler para a idade pré-escolar e escolar. Edição revista* (Lisboa: CEGOC).

YODER, P. J., CAMARATA, S. and GARDNER, E., 2005, Treatment effects on speech intelligibility and length of utterance in children with specific language and intelligibility impairments. *Journal of Early Intervention,* 28, 34-49.

YORKSTON, K., HAMMEN, V., BEUKELMAN, D. and TRAYNOR, C., 1990, The effect of rate control on the intelligibility and naturalness of dysarthric speech. *Journal of Speech and Hearing Disorders,* 55, 550-60.