The development of Phonological Mean Length of Utterance as a clinically relevant measure in phonologically impaired Dutch children

M.L. Meeuwes
3944921
Definitief onderzoeksverslag, 1 juli 2015
Universiteit Utrecht, masteropleiding Klinische Gezondheidswetenschappen, masterprogramma Logopediewetenschap, UMC-Utrecht
Begeleider: Dr. Mieke Beers
Cursusdocent: Dr. Rob Zwitserlood
Stage-instelling: Lectoraat Hogeschool Utrecht
Beoogde Tijdschrift: Clinical Linguistics and Phonetics, referentiestijl: American Psychological Association's guidelines, aantal woorden: 7000 inclusief referenties
Aantal woorden: 3798
Criteria voor transparante rapportage: STROBE
Aantal woorden Nederlandse samenvatting: 297
Aantal woorden Engelstalige samenvatting: 300

Introduction

Mapping the phonological development of children is often done on the basis of an assessment of speech. The assessment of segments (consonants and vocals) is important in order to give an insight into which segments are acquired productively and which phonological processes occur in a child's language development. With this information a phonological disorder can be diagnosed and a plan for intervention can be prepared. Phonological assessment is usually done by using a naming task, such as Metaphon[1] or Hodson & Paden (H&P)[2]. According to Taelman et al.[3], the use of spontaneous language has major advantages compared to a naming task: this will provide more relevant information about a child's speech. They[3-5] suggests that spontaneous language samples will lead to more representative insight on the phonological skills in daily speech.

Beers[6] designed the Phonological Analysis of Dutch (Fonologische Analyse Nederlands, FAN); which can be used to perform a detailed study of children's phonological abilities and is applied in diagnosing developmental speech disorders. A FAN[6] gives insight at the lexical level in terms of sound production rules and phonological contrasts. With this method can be determined whether there is a phonological impairment. In daily practice, speech-language therapists (SLTs) do not often use the FAN because of the high time investment that is needed. Crystal[7] suggested that a FAN must be based on a sample of 100 different realizations. Health insurers do not compensate the time invested in doing such an analysis for SLTs in private practices. There is need for a brief measurement procedure, which can determine a delay in the phonological development. SLTs can use the measurement procedure to determine whether it is necessary to start a time-consuming FAN.

Ingram[4] introduced PMLU as a diagnostic instrument to measure the phonological development of children. PMLU can be used to compare the complexity of the child's words to the words being attempted. In scientific literature PMLU is mentioned as a suitable tool for a brief measurement procedure for phonological development[3, 4, 8-13]. It measures the length of child's words and the number of correct consonants.

When children get older, their word complexity will increase and they acquire more segments. This will be reflected in their PMLU. The errors made by phonological impaired children, have an influence on the PMLU. Preliminary results from a pilot-study seem to suggest that SLTs can use PMLU as an instrument to differentiate between typically developing (TD) children and children with a phonological delay[14]. Rodenburg-Van Wee[9] compared the PMLU scores of phonological impaired children to those of TD children between 1;3-3;11 (years;months). The results showed that the PMLU scores of the

phonological impaired children between 4;0-6;0 were comparable to those of younger children. This indicated a delay in the phonological development in phonological impaired children.

In order to map the development of the whole-word Ingram[4] also suggested Proportion of the whole word Proximity (PWP) as a word measure. PWP is a measure that divides the child's PMLU score by the target word's PMLU score. An advantage of this measure is that the value is found to reflect a child's progression in phonological terms. For example, a child can choose for a simple word structure and pronounces it correctly. Then, the PMLU will score low, while the PWP will be quite high. On the other hand, this child can also choose many complex words and makes a lot of mistakes. The production PMLU can be similar to that of the child that uses only simple words. But it will result in a lower PWP. PWP can differentiate between these children, while PMLU cannot. Newbold et al.[15] suggested that PWP is related to intelligibility – although intelligibility obviously is most directly measured by having listeners identify the child's productions[4]. According to Newbold et al.[15] PWP captures change across time and tasks and is potentially an useful tool for evaluating speech outcomes in children with speech production problems.

According to the literature[9, 10, 12, 16, 17] PMLU and PWP can be applied to English, Finnish and Dutch TD children. PMLU and PWP can also be applied to English and Finnish speaking children with speech production problems[15-17]. But previous studies [3, 9, 18, 19] did not examine the PLMU and PWP used for Dutch phonologically impaired children. Because of specific language features of Dutch, it is important to study the use of PMLU and PWP in Dutch phonologically impaired children. First, the reliability of PMLU and PWP must be examined on the basis of the influence of language sample size on the calculation of PMLU and PWP. Language sample size is defined as the number of words in a language sample. If PMLU and PWP are proved clinically relevant for small sample sizes (for example 25 words), SLTs can use this measurement to determine whether it is necessary to start a time-consuming FAN.

Taelman et al.[3] proved that a PMLU score is more reliable when the size of the language sample increases. They found that a language sample of 25 and 50 words was too small to obtain reliable results of PMLU for Dutch children between 1;0-1;11. Ingram[4] suggested a minimum sample size of 25 words. This was based on the calculation of three different samples of 25 words from one child. According to Rodenburg-Van Wee[9] a sample size of 25 words seems to be sufficient to calculate the PMLU and PWP for Dutch TD children between 1;3-3;11 – but only if the sample contains random chosen words of different lengths and complexity. Preliminary results from a pilot-study showed no significant differences in sample

size for phonologically impaired Dutch children between 4;0-6;0[14]. This study had a study population of fourteen children, so the results must be interpreted with caution. Different language sample size of 25, 50, 75 or 100 words had probably no effect on the average PMLU score. Differences in PMLU scores were larger in individual scores than in groups. It seemed to be that most of the children needed a language sample size of only 25 or 50 words to calculate a reliable PMLU. More research with a larger population is desired.

In this study, more children (n=26) will be included. The aim of the study is to calculate the minimum language sample size for Dutch phonologically impaired children, which will lead to a clinically relevant and reliable PMLU and PWP.

Aim

The aim of this study is to determine the minimum language sample size in number of words for Dutch phonologically impaired children between 4;0-6;0 years of age, yielding to a clinically relevant and reliable PMLU and PWP.

Research question

This study focuses on the following research question:

What is the minimum language sample size for Dutch children (4;0-6;0) with a phonological impairment that leads to a clinically relevant PMLU and PWP?

Method

Participants

This study included twenty-six Dutch children referred for speech production problems (nineteen boys, aged between 4;1-6;0 and seven girls, aged between 4;0-4;9). The children were described as unintelligible by their environment. According to the speech and language therapists' assessment (Metaphon and FAN-analyses) all the children made frequent errors with vowels and consonants. The participants represented four age groups in order to equally distribute the children in the study population. There were no further analyses performed on these age categories. The age categories were: category A: 4;0-4;5, category B: 4;6-4;11, category C: 5;0-5;5 and category D: 5;6-6;0. The children were recruited from the pre-school groups from Kentalis, from the Audiologisch Centrum Utrecht and from private practices near Utrecht. The selection of criteria was identical to that used in recent research on PMLU[6, 9]. This was because of practical reasons; it creates a possibility for a larger validity study in the future.

The children must meet all of the following inclusion criteria: speech production problems, Dutch as native language, monolingual Dutch-speaking, IQ scores above 85[23] and aged between 4;0-6;0. Children with detectable hearing problems and significant hearing loss of >20 dB at the best ear[20], preterm-born[21] and dysmature children[22] and children with oral malformations were excluded.

Procedures and analyses

Data collection was done by recorded spontaneous language of the children. Before recording the children's spontaneous language, parents needed to give their informed consent. The data was derived from a 30-minute play situation between child and parent. Sessions were transcribed after the first five minutes of recording time. This was done in order to let the children get used to the recording situation. A hundred child realizations – a unique articulation of a target word - were transcribed. The target word is the adult realization of a word and may be a lexical item or a grammatical item. Only spontaneously realized words were accepted for the analysis; imitations, repeated utterances and unintelligible were excluded.

For the PMLU analysis, the PMLU of the child's target word and the child's own PMLU scores were calculated. The child's target words were first assigned points for all segments (one point for each consonant and vowel) and an additional point for all consonants. The child's own PMLU scores were calculated in the same way: by assigning a point to all segments in the word, and one extra point for each correct consonant. For example, a Dutch

word such as *banaan* 'banana' received a score of eight (five phonemes plus three correct consonants). Tallying up, over an entire sample, all PMLU scores and then dividing by the number of words, calculates the average PMLU. PMLU values were calculated manually and the calculation was conducted according to the rules proposed by Ingram and Taelman et al.[3, 4] (Table 1). The rules reduced the variation of the scores and gave a more reliable result. Ingram[4] suggested the first six rules. Rule number seven and eight were added based on the study of Taelman et al.[3]. Beside, conjugated words were also being included in the sample, contrary to what Ingram[4] suggested in rule number 2. According to Taelman et al.[3] including conjugated words will lead to a higher PMLU value in Dutch.

PWP compares the child's PMLU to that of the targets' PMLU by dividing the latter with the former. The proximity value of the production *naan (banaan)* 'banana' was 5/8, in other words 0.63. One again, all the PWP scores were tallied up and divided by the number of words to obtain an average PWP score.

The amount of words from the language samples varied from 67 to 100 words. The first 25 and the consecutive 50 words were used for the analysis, because there were at most 100 words available and there may no repeated utterances between the sample sizes (25 and 50 words). Double utterances in both sample sizes could influence the results. When in sample sizes 25 en 50 twenty words were the same, then the similarity would be higher. That will influence the reliability. The mean PMLU and mean PWP are calculated for the sample with 25 and 50 non-random chosen words for each child.

To verify the reliability of the PMLU calculations, the inter-rater reliability (IRR) was determined. Two independent raters calculated PMLU scores for 10% of the data. Three recordings were selected at random. The IRR was calculated using Cohen's Kappa[24, 25]. The IRR had a good agreement for child 4 (.884), child 12 (.983) and child 21 (.905). Further consensus was not necessary; there was almost no disagreement.

Statistical analysis

To examine whether a clinically relevant and reliable PMLU and PWP can be determined, reliability testing was used. Reliability is the degree to which the measurement is free from measurement errors[26]. The null hypothesis for this study is formulated so that the statistical test is proof of similarity; it states that the groups differ by more than a tolerably small clinically relevant amount, Δ (difference). The alternative hypothesis is that the groups differ by less than Δ - which means that they are similar. Thus, if one rejects the null hypothesis, one may correctly state that the groups are similar[27].

The hypothesis of this study is that there is a tolerably small difference, within the clinically relevant range, between the average of PMLU 25 and 50; the same applies to PWP. For this analysis, reliability testing is demonstrated using a range for clinical relevance; Δ = 1,00 for the PMLU scores and Δ = 0,05 for the PWP scores. Outside of these established limits of agreement, the confidence intervals, there is a clinically difference, the scores are with 95% confidence not similar to each other. Within these limits of agreement there is a negligible difference.

These limits were selected based on the phonological stages of PMLU and PWP of Rodenburg-van Wee[9] for TD children (Table 2). Ingram[4] suggested also phonological stages of PMLU for TD children, these are in line with the stages of Rodenburg-van Wee[9]. The phonological stages are based on a study with nine groups of five children. Therefore, clinically relevant ranges of PMLU and PWP must be interpreted with some caution.

Results will be outlined on the basis of a Bland-Altman plot[28, 29]. In this graphical method the differences between the two PMLU and PWP scores for each child are plotted against the averages of the two scores. Horizontal lines are drawn at the mean difference, and at the limits of agreement, which are defined as the mean difference plus and minus 1.96 times the standard deviation of the differences.

Results

PMLU analyses

The PMLU values for twenty-six children with a suspicion of a phonological impairment are presented in Table 3. The differences between PMLU 25 and 50 ranged between -1,32 and 0,58. To test for reliability, the limits of agreement for the difference between the two groups must be defined. The limits of agreements are defined as mean d (mean of the difference) – 1,96*SD (standard deviation of the difference) and d + 1,96*SD. In Table 3 the mean of the difference and the SD of the difference are shown. In Figure 1 the broken lines show the limits of agreement (0,567 and -1,327).

With 95% confidence a new difference between PMLU 25 and 50 will be between the limits of agreement. Because the value of the lower limit of agreement (-1,327) outreached the stated difference Δ = -1,00, there is probably no equivalence. The stated difference range must be interpreted with some caution; the result of this analysis must also be interpreted with some caution.

Child 9 and 19 are the outliers of the data, as shown in Figure 1. They could wrench the results. Table 4 and 5 give an insight in the use of mono-, di- or multisyllabic words from these children. They differed in score from the rest of the sample. These children used in ratio more di- and multisyllabic words in the language sample with 50 words, than in the sample of 25 words. There can be suggested that the words within the samples were not equally distributed.

An extra analysis was done to examine why these two children were outliers. An explanation for this could be that the discrepancy between the PMLU scores of language sample size 25 & 50 occurred because words were not selected at random. The scores were broken in a blocks of 25 words and 50 words based on a FAN sample.

For a second analysis, these two children were excluded and the limits of agreements were calculated once more. The results show that the lower limit of agreements is -1,107, there is still no equivalence. Although, the limit is closer to Δ = -1,00 than with these two children in the sample.

In 21 of the 26 children (Table 3), the PMLU scores were higher in the second sample with 50 words, than the first sample. In a FAN sample, repeated utterances are excluded. This results in a sample with at the beginning short and simple words (example: *ja, nee*), and at the end more complex and di- or multisyllabic words. In the language sample of 50 words it

could be the case that there were more complex words with different syllabic structures and consonants. Therefore, the PMLU scores of the language samples vary too much from each other.

PWP analysis

The accuracy of the children's productions can be seen in the whole-word proximity scores (PWP), presented in Table 6. The differences between PWP 25 and 50 ranged between -0,043 and 0,110. To test for equivalence, the limits of agreement for the difference between the two groups must be defined. In Table 6, the mean of the difference and the SD of the difference are shown. In Figure 2, the broken lines show the limits of agreement (-0,067 and 0,087).

With 95% confidence a new difference between PWP 25 and 50 will be between these limits. Because both limits (lower: -0,067 and upper: 0,087) outreached the range for clinical relevance Δ = 0,05, the scores are with 95% confidence not similar to each other when 25 and 50 not random chosen words were taken for calculation. The scores are probably not reliable.

In 14 of the 26 children (Table 6), the PWP scores were higher in the second language sample with 50 words, than the first language sample with 25 words. Almost half of the sample (n=12) had a lower score in the sample with 50 words than the sample with 25 words towards the PMLU scores. An explanation could be that the second sample contains more complex words. The more complex words in a language sample, the more mistakes a child could make. So the PWP scores for the second language sample will be lower.

The null hypothesis cannot be rejected for the PMLU and the PWP. The language samples of 25 and 50 words differ more than the tolerably small amount. For PMLU the amount was Δ = 1,00 and for PWP the amount was Δ = 0,05. The limits of agreements for PMLU were: 0,567 and -1,327 and for PWP: -0,067 and 0,087. The scores are with 95% confidence not similar to each other when 25 and 50 not random chosen words were taken for calculation; there is probably no reliability.

Discussion

The aim of this study was to determine the minimum language sample size in number of words for Dutch PI children between 4;0-6;0, yielding to a clinically relevant and reliable PMLU and PWP. This study shows that language sample size has influence on the calculation of the PMLU and PWP. The differences between the sample sizes were not situated between the established ranges of clinical relevance. The language sample sizes were with 95% confidence not similar too each other when 25 and 50 not random chosen words were taken for calculation.

This study has both strengths and limitations. The primary strength of this study is the amount of participants (n=26) compared to previous studies[4, 9, 14]. Therefore, the results of this study are more reliable. The second strength of this study is the use of statistics. The null hypothesis is formulated so that the statistical test is proof of similarity. To examine whether a clinically relevant and reliable PMLU and PWP can be determined, reliability testing was used. Rodenburg-van Wee[9] and Stolk et al.[14] used other statistical methods. Rodenburg-van Wee[9] used an ANOVA repeated measures with *sample size* as the within-subject factor. Stolk et al.[14] used the Friedman test because of non-parametric data, while in this present study reliability was used.

To appreciate the findings of this study, some aspects require further consideration. The current study has some limitations.

FAN analysis

According to Rodenburg-van Wee[9] there is no problem calculating PMLU or PWP on the basis of 25 words when words with different complexity and consonants are represented in a language sample. Results of this study showed that there was no equal distribution of all sorts of words. Because the language samples were based on the FAN method of including, this could have an effect on language samples. There were probably more complex words in the second language samples of 50 words than in the first language samples of 25 words. The results are therefore likely to be influenced.

Random versus not random chosen words

Second, the findings in this study are not in line with the results of other studies regarding PMLU. Results of the studies of Ingram[4] and Rodenburg-van Wee[9] showed that a language sample size of 25 or 50 words would be enough to get reliable PMLU and PWP scores for TD Dutch and English children. The difference is that they[9, 14] used random chosen words in the samples. This current study selected not random chosen words. Rodenburg-van Wee[9]

found mean differences between 0,03 and 0,24 for PMLU 25 and 50 while in this current study differences between -1,32 and 0,58 were found.

Jansen et al.[30] found that when words were not-random chosen, the results of PMLU are not reliable for TD Dutch children between 2;3-4;0 and PI Dutch children between 4;0-6;0. Random chosen words were reliable for the calculation of PMLU for these children.

Range of clinical relevance

Third, the differences (∆) for the PMLU score and PWP score were probably too strict. The difference values were selected based on the phonological stages of PMLU and PWP of Rodenburg-van Wee[8] for TD children. But according to the results of this study, this maybe needs reconsideration.

Current results are in accordance with Taelman et al.[3]. They concluded that the higher the PMLU score of the child is, the bigger the language sample size needs to be. The results are based on a study with TD Dutch children between 1;0-1;11. They examined the standard deviation and the confidence intervals. These two increased when a higher PMLU score was achieved and the language sample decreased.

The results of this current study offer several important implications: production PMLU cannot distinguish clear between children who produce short words with few mistakes and children who produce complex words with more mistakes in the production. PWP denotes in these cases the difference between target PMLU and production PMLU. According to Saaristo-Helin[8] the combination of PWP and PMLU gives a better reproduction of the phonological development of children.

Further research towards children with phonological impairments is needed. It turns out that phonological impairments could be detected and diagnosed earlier than the age of 4;0[31]. PMLU and PWP scores of children between 3;0-6;0 with phonological impairments must be obtained. Based on the mean PMLU scores and SD in further research, a statement can be made how a child with a suspicion on a PI varies from the mean and how considerable this deviation is.

A further point is that PMLU and PWP scores in Dutch must be calculated on the base of two FAN analyses. One analyses for the first and one for the second language sample. Complexity of words will be distributed equally and results will be more reliable. If PMLU and

PWP are proved clinically relevant for small sample sizes, SLTs can use this measurement to determine whether it is necessary to start a time-consuming FAN analysis.

Reference list

(1) Leijdekker-Brinkman W. De Metaphonbox. Behandeling van kinderen volgens de Metaphon-therapie. Lisse: Swets & Zeitlinger; 2002.

(2) Hodson, BW & Paden, EP. Targeting Intelligible Speech, San Diego, CA: College Hill Press; 1991.

(3) Taelman H, Durieux G, Gillis S. Notes on Ingram's whole-word measures for phonological development. Journal of Child Language 2005; 32(2): 391-405.

(4) Ingram D. The measurement of whole-word productions. Journal of Child Language 2002; 29: 713-733.

(5) Hase M, Ingram D & Bunta F. A comparison of two phonological assessment tools for monolingual Spanish-Speaking children. Clinical Linguistics & Phonetics 2009; 24(4-5): 346-356.

(6) Beers M. The phonology of normally developing and language-impaired children. Amsterdam: IFOTT; 1995.

(7) Crystal D. Profiling Linguistic Fisability. London: Edward Arnold; 1982.

(8) Saaristo-Helin K, Savinainen-Makkonen T, Kunnari S. The phonological mean length of utterance: Methodological challenges from a crosslinguistic perspective. Journal of child language 2006; 33(1): 179-190.

(9) Rodenburg-Van Wee M. Fonologische en lexicale ontwikkeling. De toepasbaarheid en betrouwbaarheid van fonologische woordmaten in de klinische praktijk: PMLU, PWP en PWC. Utrecht: Hogeschool Utrecht; 2013.

(10) MacLeod A, Laukys K, Rvachew, S. The impact of bilingual language learning on whole-word complexity and segmental accuracy among children aged 18 and 36 months. International journal of speech-language pathology 2011; 13(6): 490-499.

(11) Watson M & Terrell, P. Longitudinal changes in phonological whole-word measures in 2-year-olds. International journal of speech-language pathology 2012; 14(4): 351-362.

(12) Gerrits E & de Bree E. Early language development of children at familial risk of dyslexia: Speech perception and production. Journal of Communication Disorders 2009; 42(3): 180-194.

(13) Van Noort-Van Der Spek I, Franken M, Wieringa M & Weisglas-Kuperus, Phonological development in very-low-birthweight children: an exploratory study. Developmental Medicine & Child Neurology 2010; 52(6): 541-546.

(14) Stolk E, Versteeg M & De Vries M. Phonological Mean Length of Utterance. Een beschrijvend onderzoek naar de Phonological Mean Length of Utterance (PMLU) bij Nederlandstalige kinderen met een fonologische stoornis. Utrecht: Hogeschool Utrecht; 2014.

(15) Newbold E, Stackhouse J & Wells B. Tracking change in children with severe and persisting speech difficulties. Clinical linguistics & phonetics 2013; 27(6-7): 521-539.

(16) Burrows L & Goldstein B. Whole word measures in bilingual children with speech sound disorders. Clinical linguistics & phonetics 2010; 24(4-5): 357-368.

(17) Kunnari S. Word length in syllables: evidence from early word production in Finnish. First language 2002; 22(2): 119-135.

(18) Saaristo-Helin K. Measuring phonological development: a follow-up study of five children acquiring Finnish. Language and Speech 2009; 53(1): 55-77.

(19) Saaristo-Helin K, Savinainen-Makkonen T & Kunnari S. Phonological mean length of utterance in specific language impairment; a multi-case study of children acquiring Finnish. Clinical Linguistis & Phonetics 2012; 26(5): 428-444.

(20) Yoshinaga-Itano C, Sedey AL, Coulter DK & Mehl AL. Language of early-and later-identified children with hearing loss. Pediatrics 1998; 102(5): 1161-1171.

(21) Largo RH, Molinari L, Pinto LC, Weber M & Due G. Language development of term and preterm children during the first five years of life. Developmental Medicine & Child Neurology 1986; 28(3): 333-350.

(22) Jansson-Verkasalo E, Valkama M, Vainionpää L, Pääkkö E, Ilkko E & Lehtihalmes M. Language development in very low birth weight preterm children: a follow-up study. Folia phoniatrica et logopaedica 2004; 56(2): 108-119.

(23) Gallinat E & Spaulding T. Differences in the performance of children with specific language impairment and their typically developing peers on nonverbal cognitive tests: A meta-analysis. Journal of Speech, Language, and Hearing Research 2014; 57(4): 1363-1382.

(24) Portney LG, Watkins MP. Foundations of Clinical Research: Pearson; 2007.

(25) Hallgren KA. Computing Inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol. 2012;8(1): 23-24.

(26) De Vet H, Terwee C, Mokkink L & Knol D. Measurement in medicine: a practical guide. Cambridge University Press: 2011.

(27) Barker L, Luman E, McCauley M & Chu S. Assessing equivalence: an alternative to the use of difference tests for measuring disparities in vaccination coverage. American Journal of Epidemiology 2002; 156(11): 1056-1061.

(28) Bland J & Altman D. Statistical methods for assessing agreement between two methods of clinical measurement. The Lancet 1986; 327(8476): 307-310.

(29) Bland J & Altman D. Measuring agreement in method comparison studies. Statistical methods in medical research 1999; 8(2): 135-160.

(30) Jansen H, Vegte N van de & Wouters M. Betrouwbare woordselectie voor de berekening van de PMLU - de klinische toepasbaarheid van de eerste 25 woorden als betrouwbare woordselectie voor het meten van de fonologische woordmaat Phonological Mean Length of Utterance (PMLU). Utrecht: Hogeschool Utrecht; 2015.

(31) Bron A, Groot de M, Scheper A & Verheugt J. Jessita tan oot de taa niet zeggen. Kinderen met fonologische stoornissen samen in behandeling. Logopedie & Foniatrie 2008; 10: 300-308.

Tables and figures

Table 1. Rules for the calculation of Phonological Mean Length of Utterance (PMLU)[6, 8]

1. *Sample size Rule*: Select at least 25 words, and preferably 50 words for analysis, depending on sample size. If the sample is larger than 50 words, select a selection of words that cover the entire sample, e.g. every other word in a sample of 100 words.
2. *Lexical-Class Rule:* Count words (e.g. common nouns, verbs, adjectives, prepositions and adverbs) that are used in normal conversation between adults. This excludes child words, e.g. mommy, daddy, tata etc. Counting child words can inflate the PMLU if a child is a reduplicator. Grammatical morphemes should be avoided, these can reduce the PMLU score.
3. *Compound Rule*: Do not count compounds as a single word unless they are spelled as a single words, e.g. 'cowboy' but not 'teddy bear', i.e. 'teddy bear' would be excluded from the count. This rule simplifies decisions about what constitutes a word in the child's sample.
4. *Variability Rule*: Only count a single production for each word. If more than one occurs, then count the most frequent one. If there is none, then count the last one produced. Counting variable productions may distort the count if there is a highly variable single word.
5. *Production Rule*: Count 1 point for each consonant and vowel that occurs in the child's production. Syllabic consonants receive one point. Do not count more segments than are in the adult words. For example, a child who says 'foot' as [hwut] had two consonants counted, not three. Otherwise, children who add segments will get higher scores despite making errors.
6. *Consonants Correct Rule:* Assign 1 additional point for each correct consonant. Correctness in vowels is not counted since vowel transcriptions are typically of low reliability. Syllabic consonants receive an additional point in the same way as nonsyllabic consonants. A child who applies liquid simplification, for example, will get 1 point for producing a vowel, but 2 points of the syllabic consonant is correct.
7. *Position Rule:* Only segments that are generalised in the right position, are counts as correct. This rule is valid for word level. When an initial consonant is getting a final consonant because of vocal deletion, this position is correct.
8. *Input Rule:* Children's words must be compared to the real words, and not to standard targets in spoken language or the written version of the word. Use of dialect and everyday speech has to be taken with.

Table 2. The phonological stages of PMLU and PWP of Rodenburg-van Wee[9] for TD children

Group	Age	Range PMLU production
A, B	1;3 – 1;11	3,5 – 4,5
C, D, E	2;0 – 2;8	4,5 – 5,0
F, G, H, I	2;9 – 3;11	5,0 – 6,0

Group	Age	Range PWP production
A	1;3 - 1;8	0,65 – 0,8
B, C	1;9 – 2;2	0,8 – 0,85
D, E, F	2;3 – 2;11	0,85 – 0,9
G, H, I	3;0 – 4;0	> 0,9

Table 3. Analysis 1. Summary of the mean PMLU scores for sample size 25 and 50, the difference between the sample sizes and the mean of the sample sizes of 26 Dutch phonologically impaired children

Child	PMLU25	PMLU50	difference	mean
1	4,56	5,16	-0,60	4,86
2	5,20	5,52	-0,32	5,36
3	4,68	5,38	-0,70	5,03
4	4,28	4,2	0,08	4,24
5	4,76	4,29	0,47	4,53
6	4,16	4,36	-0,20	4,26
7	5,04	5,18	-0,14	5,11
8	2,72	3,5	-0,78	3,11
9	4,56	5,88	-1,32	5,22
10	4,32	4,62	-0,30	4,47
11	4,48	4,8	-0,32	4,64
12	4,4	4,72	-0,32	4,56
13	5,48	5,88	-0,40	5,68
14	4,16	4,94	-0,78	4,55
15	4,48	4,98	-0,50	4,73
16	4,04	4,94	-0,90	4,49
17	5,52	4,94	0,58	5,23
18	4,8	5,18	-0,38	4,99
19	4,6	5,8	-1,20	5,20
20	5,04	4,6	0,44	4,82
21	5,4	5,48	-0,08	5,44
22	4,72	5,22	-0,50	4,97
23	5,24	5,14	0,10	5,19
24	5,2	5,38	-0,18	5,29
25	4,64	5,3	-0,66	4,97
26	4,96	5,8	-0,84	5,38
	Mean		-0,38	
	Standard deviation (SD)		0,47	
	1,96*SD		0,947	

Table 4. Summary of the mono-, bi- or multisyllabic words used in the language sample of 25 and 50 words for child number 9

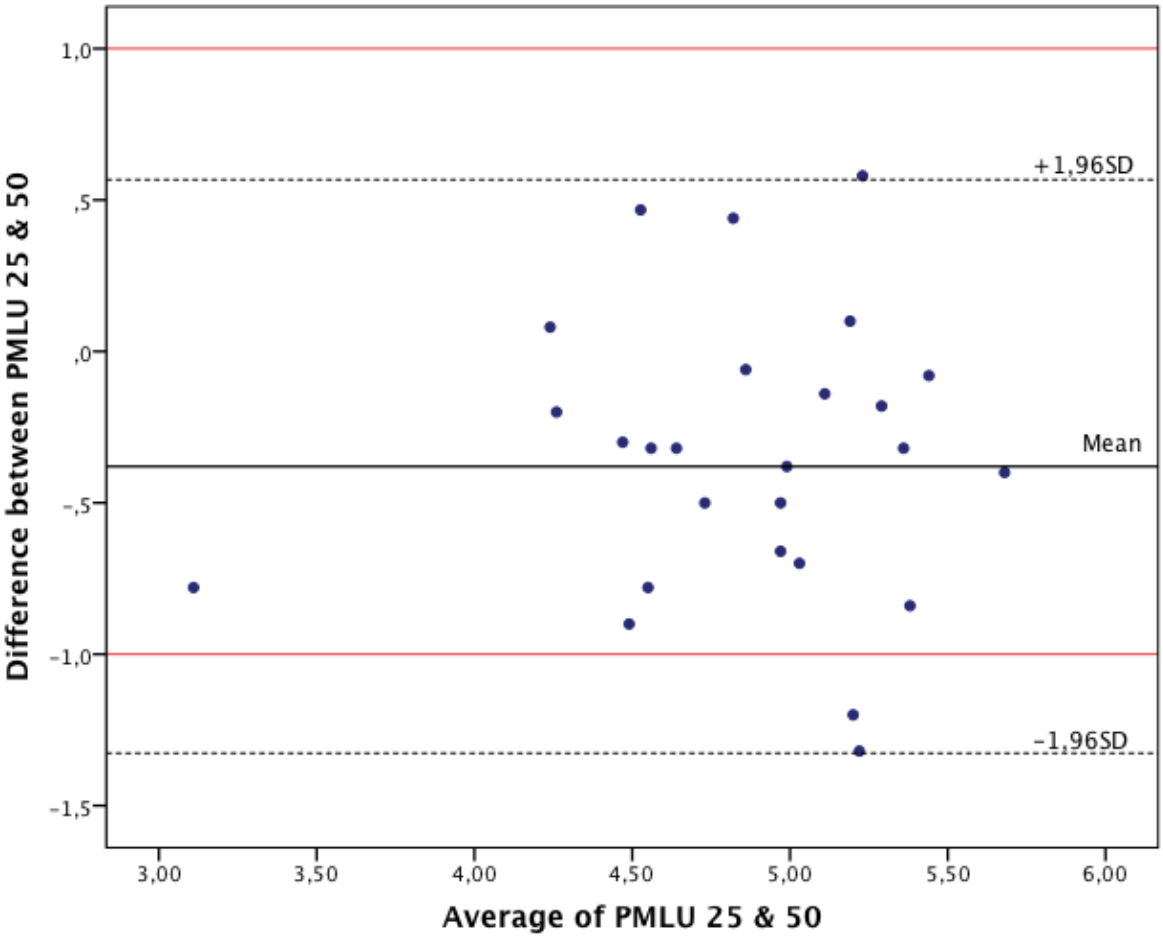
	PMLU 25	PMLU 50
monosyllabic words	64%	50%
disyllabic words	28%	40%
multisyllabic word	4%	12%

Table 5. Summary of the mono-, bi- or multisyllabic words used in the language sample of 25 and 50 words for child number 19

	PMLU 25	PMLU 50
monosyllabic words	72%	32%
disyllabic words	16%	48%
multisyllabic word	12%	20%

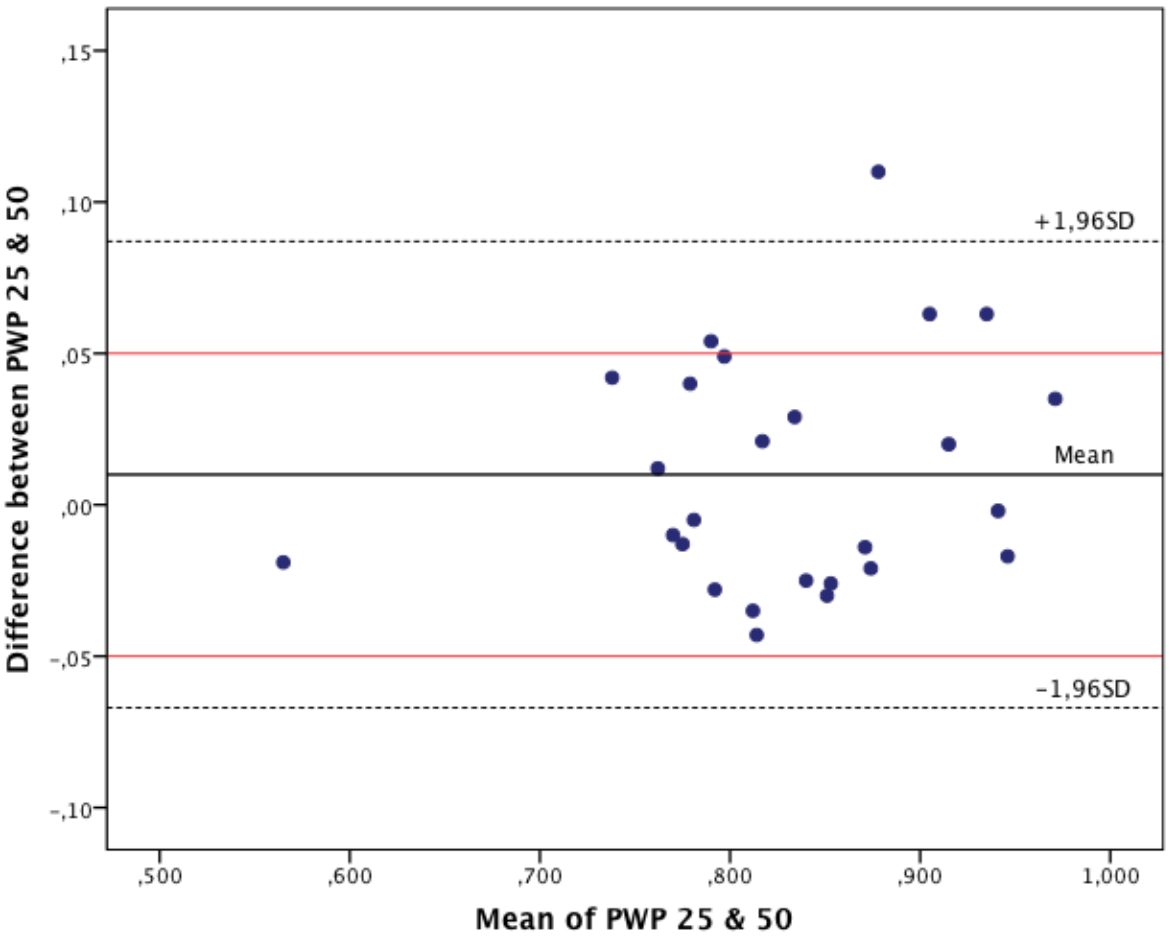
Table 6. Analysis 2. Summary of the mean PWP scores for sample size 25 and 50, the difference between the sample sizes and the mean of the sample sizes of 26 Dutch phonologically impaired children

Child	PWP25	PWP50	difference	mean
1	0,799	0,758	0,040	0,779
2	0,840	0,866	-0,026	0,853
3	0,836	0,867	-0,030	0,851
4	0,817	0,763	0,054	0,790
5	0,759	0,717	0,042	0,738
6	0,765	0,775	-0,010	0,770
7	0,938	0,955	-0,017	0,946
8	0,556	0,575	-0,019	0,565
9	0,794	0,829	-0,035	0,812
10	0,864	0,878	-0,014	0,871
11	0,768	0,756	0,012	0,762
12	0,827	0,852	-0,025	0,840
13	0,966	0,903	0,063	0,935
14	0,778	0,806	-0,028	0,792
15	0,769	0,782	-0,013	0,775
16	0,827	0,806	0,021	0,817
17	0,863	0,884	-0,021	0,874
18	0,989	0,954	0,035	0,971
19	0,934	0,823	0,110	0,878
20	0,821	0,772	0,049	0,797
21	0,936	0,874	0,063	0,905
22	0,792	0,835	-0,043	0,814
23	0,940	0,942	-0,002	0,941
24	0,778	0,783	-0,005	0,781
25	0,849	0,819	0,029	0,834
26	0,925	0,905	0,020	0,915
	Mean		0,010	
	Standard deviation (SD)		0,039	
	1,96*SD		0,077	



Lines at +1 and -1: the tolerably small amount that PMLU 25 and PMLU 50 may differ from each other. Lines at +1,96SD and -1,96SD: limits of agreement. Line at -0,38: mean.

Figure 1. Bland-Altman plot PMLU



Lines at +0,05 and -0,05: the tolerably small amount that PWP 25 and PWP 50 may differ from each other. Lines at +1,96SD and -1,96SD: limits of agreement. Line at -0,010: mean.

Figure 2. Bland-Altman plot PWP

Abstract

Title: The development of Phonological Mean Length of Utterance as a clinically relevant measure in phonologically impaired Dutch children

Background: Phonological Mean Length of Utterance (PMLU) is introduced as a diagnostic instrument to measure the phonological development of children. PMLU can be used to compare the complexity of the child's words to the complexity of the words being attempted. PWP is a measure that divides the child's PMLU score by the PMLU score of the target word. The combination of PWP and PMLU gives a reproduction of the phonological development of children. These word measures are clinically relevant for SLTs when they are reliable for small language sample size.

Aim: The aim is to determine the minimum language sample size in number of words for Dutch phonologically impaired children between 4;0-6;0 years of age, yielding to a clinically relevant and reliable PMLU and PWP.

Methods: The study includes 26 Dutch children with a suspicion of a phonological impairment. Children's spontaneous language was recorded. PMLU and PWP scores were calculated for the language sample sizes 25 and 50. Reliability testing was used to examine whether a clinically relevant and reliable PMLU and PWP can be determined.

Results: The null hypothesis cannot be rejected for the PMLU and the PWP. The language samples of 25 and 50 words differ more than the tolerably small amount.

Conclusions: This study shows that language sample size has influence on the calculation of the PMLU and PWP scores. The differences between the sample sizes were not between the established ranges of clinical relevance. The language sample sizes were with 95% confidence not similar too each other when 25 and 50 not random chosen words were taken for calculation.

Recommendations: For further research PMLU and PWP must be calculated on the base of two FAN analyses. Complexity of words will be distributed equally and results will be more reliable.

Key-words: Phonological Mean Length of Utterance, Proportion of the whole word Proximity, sample size, phonological impairment, children

Samenvatting

Titel: De ontwikkeling van *Phonological Mean Length of Utterance* als klinisch relevante woordmaat voor Nederlandse kinderen met een fonologische stoornis
Inleiding: *Phonological Mean Length of Utterance* (PMLU) is geïntroduceerd als diagnostisch instrument om de fonologische ontwikkeling van kinderen te meten. PMLU kan worden gebruikt om de complexiteit van een kinderuiting met de doeluiting te vergelijken. PWP is een woordmaat die de PMLU score van de kinderuiting deelt door die van de doeluiting. De combinatie PMLU en PWP geeft een weergave van de fonologische ontwikkeling van kinderen. Deze woordmaten zijn klinisch relevant voor logopedisten wanneer zij betrouwbaar kunnen worden gebruikt voor kleinere taalsamples.
Doel: Het doel is om te bepalen wat de minimale grootte van een taalsample voor Nederlandse kinderen met een fonologische stoornis tussen de 4;0-6;0 jaar moet zijn, om een klinisch relevante en betrouwbare PMLU en PWP te verkrijgen.
Methode: De studie bevatte 26 Nederlandse kinderen met vermoedelijk een fonologische stoornis. Kinderen hun spontane taal werd opgenomen. PMLU en PWP scores zijn berekend voor de taalsamplegroottes van 25 en 50 woorden. Betrouwbaarheid werd getest om een klinisch relevante en betrouwbare PMLU en PWP te bepalen.
Resultaten: De hypothese kon niet worden verworpen voor PMLU en PWP. De taalsamples van 25 en 50 woorden verschilden meer dan een aanvaardbare hoeveelheid punten van elkaar.
Conclusie: De studie laat zien dat taalsamplegrootte invloed heeft op de berekening van PMLU en PWP. De verschillen in score tussen de taalsamples lagen niet binnen de vastgestelde ranges voor klinische relevantie. De taalsamples waren met 95% betrouwbaarheid niet gelijk aan elkaar wanneer 25 en 50 niet random gekozen woorden werden meegenomen in de berekening.
Aanbevelingen: Voor verder onderzoek moeten de PMLU en PWP scores worden berekend op basis van twee FAN analyses. De complexiteit van de woorden zal gelijk worden verdeeld over de samples en de resultaten zullen betrouwbaarder zijn.

Trefwoorden: fonologische stoornis, Phonological Mean Length of Utterance, Proportion of the whole word Proximity, taalsample, kinderen