

Metodické studie

VALIDITY OF LANGUAGE SAMPLE MEASURES TAKEN FROM STRUCTURED ELICITATION PROCEDURES IN CZECH

FILIP SMOLÍK

Psychologický ústav AV ČR, Praha

GABRIELA SEIDLOVÁ MÁLKOVÁ

Katedra psychologie PedF UK, Praha

ABSTRACT

Validity of language sample measures taken from structured elicitation procedures in Czech

F. Smolik, G. Seidlová Málková

Objectives. Test the use of language samples elicited during the presentation of standardized psychometric methods for the diagnostics of language development using methods for language sample analysis. Validate the use of Mean Length of Utterance (MLU) against standardized measures of grammatical development in a sample of children of the same age.

Subjects and settings. Total of 135 children aged on average 72.2 months (SD=3.7) participated. Linguistic productions of the children during the administration of the WPPSI vocabulary subtest were analyzed. This language sample was used for calculating the MLU and other indices. Additionally, children were administered Czech adaptations of standardized tests of grammatical development: TROG-2 for sentence comprehension and a test of morphological production.

Hypotheses. MLU and other sample-derived indices should show significant relationships between the test measures of grammatical comprehension and production.

Statistical analyses. Correlations, linear regression models.

Results. There were significant relationships between MLU and the tests of grammar comprehension and morphological production. The number of be-forms and prepositions per utterance were also related to the grammar tests. MLU was independently predicted by both grammar comprehension and the production of grammatical morphology. The results show validity of MLU in a sample of same-age children, which is rare in the available literature.

Study limitations. The language samples provided by children's definitions differed widely in size and style, which may underestimate the relationships between MLU and grammatical test measures.

key words:

mean length of utterance,
language development,
sentence complexity,
grammar acquisition

klíčová slova:

průměrná délka výpovědi,
vývoj jazyka,
komplexnost vět,
osvojování gramatiky

Transcripts of spontaneous language interaction have become a standard assessment tool in research on child language. In many countries, spontaneous language samples also form a standard component of examination by speech-language pathologists or

Došlo: 5. 10. 2010; F. S., Psychologický ústav AV ČR, Politických vězňů 7, 110 00 Praha 1; e-mail: smolik@praha.psu.cas.cz; G. S. M., katedra psychologie PedF UK, Praha

logopedists. Spontaneous language transcripts are based on recordings of an interaction between a child and an adult. Language used in the interaction is then transcribed according to a set of formatting rules such as SALT (Leadholm & Miller, 1992) or CHILDES (MacWhinney, 2000). The transcript is then analyzed using various indices of language development that are sensitive to the level of language proficiency shown by the child. Hux, Morris-Friehe, and Sanger in 1993 reported that 80 % of speech-language professionals in nine US states used some form of language sample analysis when evaluating children's language, and this number has been increasing since then (see Eisenberg, Fersko, & Lundgren, 2001).

The analysis of language samples differs from other diagnostic procedures in several respects. While standardized tests of language present children with artificial tasks eliciting speech or testing comprehension, speech samples are collected in a natural situation. The method thus has high level of ecological validity. At the same time, the method suffers from the same problems as all methods relying on elicited production, namely that children may choose to produce material that is not useful for their evaluation.

Probably the best known usage of language transcripts for research purposes is the longitudinal study of three American children reported in detail by Brown (1973). This study also popularized the most widely used index based on spontaneous language transcripts: the Mean Length of Utterance (MLU). Brown calculated the MLU from a sample of 100 utterances as the mean number of morphemes per utterance, and showed that MLU was a better index of overall language development than chronological age. MLU showed steady increase from the second year to the fourth year of age. There were substantial differences between the onset of growth in MLU across the three children studied by Brown, but once the MLU crossed a certain threshold, the growth rate was quite similar. Brown's analyses have contributed to the wide acceptance of MLU as the benchmark measure of language development.

Even though MLU has been widely accepted as a measure of syntactic development, it is not the only measure that can be derived from spontaneous language samples. In a way, MLU is quite a rough measure, and there have been multiple attempts to develop more sensitive instruments. Alternative metrics have been proposed that evaluate spontaneous language samples with respect to the linguistic complexity of the utterances. Among the best known alternatives are the Index of productive syntax (IPSYN, Scarborough, 1990) and Developmental sentence scoring (Lee, 1974). Another problem is that all indices taken from spontaneous language samples are sensitive to the way the sample is collected. For instance, it is perfectly natural even for adults to use short sentences when responding to questions. To achieve optimal results, it is thus important to use communication style that invites full-sentence responses („Tell me more about it...“).

Validity and reliability of MLU and other indices

In order to provide a measure of language development, it needs to be established that the proposed measure is valid, i. e. reflects the characteristics of interest. Children with advanced language skills should score high on measures of language development, and vice versa. Instrument validity is typically established by comparing the measure with the results from different measurement tools that measure similar aspects of behavior or cognition. Validity of spontaneous language samples may thus be established by calculating these measures and correlating them with standardized tests of language development. In the case of a developmental test, there should also be a clear correlation with age.

One of the oldest attempts to establish validity of MLU was reported by Rondal (1978), who examined different aspects of speech samples from 42 children. Rather than comparing MLU with standardized tests, the study examined its relationship with other scores that reflected linguistic complexity of the observed language samples, such as the proportion of yes/no questions, number of modifiers per utterance etc. The study showed that MLU and other indices of linguistic complexity were significantly intercorrelated, which indicates that MLU reflects linguistic complexity. Blake, Quartaro, and Onorati (1993) examined the validity of MLU by comparing the MLU values with scores derived from a detailed qualitative analysis of language samples (LARSP) (Crystal, 1979), and found good validity.

Miller and Chapman (1981) published an influential study that established the relationship between MLU and age. This study has long been used as one of two main sources of normative data on MLU because it provides mean values and variability estimates for three-month intervals between 18 and 60 months of age. Another influential benchmarking study of MLU has been reported by Leadholm and Miller (Leadholm & Miller, 1992). Until recently, this was the most extensive set of normative data on MLU, establishing the relationship between MLU and age, and providing directions for the sample collection procedure, transcription, and transcript evaluation. The most recent normative study by Rice et al. (2010) gives normative data for American English-speaking children, both typically developing and language impaired.

The examination of MLU validity against other measures of language development provided mixed results. Klee and Fitzgerald (1985) found no relationship between MLU, age, and lexical scores from a standardized test. However, the study only examined 18 children, and the lack of statistically significant relationships could be due to the small power of the design. Dethorne, Johnson, and Loeb (2005) also failed to find a relationship between the scores from the PPVT test of lexical comprehension (Dunn, Dunn, William, & Wang, 1997), and MLU. In their study, MLU was significantly related to the number of different words in the language sample, and with the number of tense-marking morphemes in the sample. The study thus found significant relationships between different measures taken from the language sample, but these were not related to the scores on a lexical comprehension test. Contrary to the above reports, Rondal, Ghiotto, Bredart, and Bachelet (1987) found a good relationship between age and MLU, and MLU predicted the level of grammatical development on several different measures. A strong relationship between MLU and a standardized language test was also reported by Tomblin, Shonrock, and Hardy (1989), who used MLU as a criterion when evaluating the properties of the Minnesota Child Development Inventory. The expressive language scale of this inventory was predicted by MLU in children aged 23 to 28 months. Finally, Rice, Redmond, and Hoffman (2006) examined the relationship between MLU and other measures in children with specific language development and typically developing children. They found a strong relationship between MLU, IPSYn, DSS, and age. MLU was also significantly related to receptive lexical scores from PPVT. This supports the validity of MLU as a language development measure.

Overall, the validity of MLU and other spontaneous language sample measures warrants further research. It is likely that the reported validity of MLU depends on the exact method of language sample collection, the experience and skill of the examiner, and the size of the sample. It is also possible that the validity differs as a function of age or linguistic development level. It is generally agreed that MLU grows continually up to the values of 4.5 (Blake et al., 1993) or 6 (Rice et al., 2010) but the growth levels off after children reach this value. The language of children who reach this

value is so advanced that utterance length is determined mostly by situational factors and not by children's language skills. In any case, researchers agree that MLU and other language sample indices have good ecological validity for assessing productive language in children. Unlike standardized tests, they rely on natural communication and evaluate spontaneous linguistic productions in a conversational situation. This allows the examiner to obtain a realistic picture of the child's communication skills and style.

Besides validity, measurement tools are characterized by their reliability. This property refers to the extent to which a measurement tool provides the same results when applied multiple times. Reliable tests have low susceptibility to random error variation, and thus provide precise measurements. There are different ways of establishing reliability. Internal consistency is, roughly, the extent to which different parts and items of a test provide the same result. Test-retest reliability is the correlation between scores obtained from repeated administrations of the same measure. The reliability of MLU and other spontaneous language measures is a matter of discussion. Gavin and Giles (1996) reported quite high levels of test-retest reliability ($r > 0.9$) for language samples with more than 175 utterances. In samples between 50 and 100 utterances, observed reliabilities ranged from 0.64 to 0.90. Two other studies suggested reasonable levels of test-retest reliability (Cole, Mills, & Dale, 1989; Minifie, Darley, & Sherman, 1963). Internal consistency has been established by calculating MLU from different parts of the same samples, mostly with good results (Darley & Moll, 1960; Cole et al., 1989; Casby, 1984). The reliability estimates, however, are strongly influenced by the age range of children included in the analysis. If there is large variability in age, and thus in MLU, the correlations between measures from the same child will be higher. The most conservative estimate of reliability would thus be one calculated from a group of children of the same age. Only Darley and Moll (1960) reported such estimates, which suggested that MLU from samples around and under 50 utterances may not be sufficiently reliable. To summarize, there is good evidence for MLU reliability but given the sensitivity of the measure to a number of factors, the values have to be interpreted with care.

Evaluating language elicited in structured dialogue

A number of assessment instruments used in developmental diagnostics relies on elicited verbal responses. In some cases, the responses require the use of one or more sentences in order to be complete. In such situations, the children provide linguistic material that can be evaluated using the measures used for the analysis of spontaneous language. The present paper tests this approach to the use of children's diagnostic data. The goal is to establish whether language samples obtained by recording the responses in a standardized test provide a valid measure of grammatical development.

The language sample data analyzed in this report were recorded during children's examination using the Vocabulary subtest of the Czech working version of Wechsler Primary and Preschool Scale of Intelligence (WPPSI Wechsler, 2002). In this subtest, children are asked to define the meanings of up to 30 common Czech words, beginning with *bota* "shoe", ending with *mikroskop* "microscope". The scoring of the subtest does not require that children necessarily speak in full sentences and provide perfect definitions, but the task has the potential to elicit sentential and multi-sentence responses. Therefore, it might be used as a source of spontaneous language data.

Besides examining the use of WPPSI definitions as a source of spontaneous language, the present paper is the first to examine the properties of MLU and similar indices in Czech. The goal is to test the validity of MLU against two tests of gram-

matial development, and the WPPSI vocabulary estimate. Because vocabulary was estimated from the same data that was used to calculate the MLU, a high correlation is expected. The true test of MLU validity is the relationship between MLU and the grammatical tests. One of these tests assessed the comprehension of syntax, the other tested the production of grammatical morphology and the ability to inflect words appropriately.

Besides MLU, the study examined several other indices from the language samples. These indices captured the use of different word categories in children. Certain classes of words and morphemes are likely to reflect the sentence complexity better than others. This has been especially well documented in children with specific language impairment (Leonard, 1989). The categories examined in the present study were the forms of the verb “be”, which serve as copula or auxiliary forms in Czech, prepositions, personal and demonstrative pronouns, and connectives. The index for each category was calculated as the mean number of elements from each category per utterance. The goal was to evaluate whether the indices based on specific categories could provide richer information than MLU per se.

The questions posed by this study can thus be summarized as follows. 1) Does the transcription of WPPSI vocabulary definitions provide a useful sample of language for calculating MLU and other sample-based indices? 2) Does MLU provide a valid measure of grammatical development in Czech? 3) If yes, is it more related to the receptive or expressive knowledge of grammar? 4) Do specific indices reflecting the use of certain linguistic categories show stronger relationship with grammatical development than MLU? The study should thus introduce the MLU measure to the analysis of Czech. At the same time, it should contribute to the discussions about the validity of MLU. As the participants of the study were of the same age, the study thus provides a conservative and stringent test of MLU validity as a measure of grammatical development.

METHOD

Participants

One hundred and thirty five Czech children participated in the study. The children were selected from an initial sample of 166 recruited for the cross-linguistic longitudinal study of early literacy development ELDEL. The excluded children did not have complete recordings of the vocabulary test, mostly because of their noncompliance in the task. All participants were typically developing children with Czech language background recruited from 21 kindergartens in Prague and two smaller cities. The children had no documented visual, speech, language or motor function impairments or behavior problems. All children included in the study were entering school on schedule, i. e. during the year after their sixth birthday. The mean age of the children when recordings of WPPSI testing were made was 72.2 months (SD 3.7).

Measures and procedures

All children were tested individually in their kindergarten, in a quiet room outside their classrooms. The data collection was performed by a team of 10 supervised research assistants. In line with the cross-linguistic purpose of the original large study, all measures were chosen to allow cross-linguistic comparisons of children’s performance within 4 European languages. In most cases, the measures were based on British assessment batteries. It is worth noting that measures of skills we were aiming for are not very well represented in the available Czech assessment tools. All data were collected using struc-

tured assessment tasks. The general procedure in each task involved one or more training items in the beginning, after which children received corrective feedback. No feedback was provided after the actual trials that were used for scoring.

Vocabulary For assessment of children's lexicon we adopted the English version of WPPSI-III (Wechsler, 2002) vocabulary scale. Data collected using this measure were used for obtaining the lexical scores, as well as the source of children's spontaneous production. The task requires children to give an oral definition for each of 25 words (20 nouns, 2 verbs, 3 adjectives). In the beginning, the practice item *bota* "shoe" was administered, after which children received corrective feedback if needed. Responses were scored with 0, 1, or 2 points. To be scored with 2 points, the child had to produce an appropriate synonym, a primary and defining feature of the word's referent, or several descriptive features. One-point responses lacked some important content, contained a less pertinent synonym, or attributes with correct but not defining features. The administration of the measure was recorded and later transcribed to obtain samples of children's speech production.

Receptive syntactic skills As a measure of receptive syntactic skills we adopted 4 blocks of a British receptive grammar test, TROGG-2 (Bishop, 2003). In this test, children hear a sentence and see four pictures. Their task is to point to the picture that corresponds to the sentence. Each block comprised 4 stimulus sentences, the whole task thus contained 16 items scored by 1 point each. Each of the four blocks tested comprehension of a different sentence type: subject relative clauses (*Kniha, která je červená, leží na tužce.* "The book that is red is on the pencil), sentences with inverted word order (*Krávu honí dívka.* "The cow (patient) chases the girl (agent)"), object relative clauses (*Šálek je v krabici, která je červená.* "The cup is in the box that is red"), and center-embedded sentences (*Ovce, na kterou se dívá dívka, běží.* "The sheep the girl looks at is running").

Morphosyntactic skills We created a set of tasks aimed at 6 morphosyntactic structures: noun plurals, third person singular verb forms, past tense forms, adjective agreement, masculine-feminine noun derivation, and subject-verb agreement. Some parts of the task were modeled after the Czech test *Zkouška jazykového citu* "Test of language sense" (Žlab, 1992). The examiner in these tasks produced the initial part of a sentence, and the children were asked to complete the sentences in line with the modeled example. An example of a plural item would be *Honza má banán, ale chtěl by čtyři...* "Johnny has a banana but he would like to have four ...". Two training items preceded each section of the morphological test.

Transcripts and the measures from spontaneous samples

A trained assistant transcribed the recordings of WPPSI vocabulary subtest administration. Children's definitions served as the language sample materials. Samples were transcribed orthographically and coded for several types of speech phenomena, including omissions, incomplete utterances, and unintelligible material. Only complete and intelligible utterances served as material for the analysis. Utterances were segmented according to a set of criteria. The primary criterion was prosodic. Utterances separated by a clear prosodic break and pause were considered separate utterances. If children spoke without apparent prosodic breaks, syntactic criteria were utilized: an utterance was defined as a main clause with all its dependent clauses. If multiple main clauses were coordinated using the connective *a* "and" or simple juxtaposition, the compound clause was segmented into multiple utterances. Coordinated clauses connected using other means (using connectives such as *but* or *or*) were not segmented into separate utterances.

The variables calculated from the transcripts were the mean length of utterance in words (MLU), and the mean number of words from certain grammatical categories per utterance. All the measures were calculated from the whole sample. MLU was calculated in words because Czech is a richly inflected language in which many words are multimorphemic, but the degree to which children analyze the words is questionable. The category-specific indices were calculated for the forms of the verb *byť* “be”, prepositions, personal pronouns, demonstrative pronouns, and connectives. These categories were selected because other studies suggested that similar categories in other languages are especially sensitive to language development level in children (Bedore & Leonard, 2001; Bortolini, Caselli, Deevy, & Leonard, 2002), or because they seem intrinsically related to sentence complexity (connectives). Both MLU and the category-specific indices were calculated from the whole sample available for each child, i. e. the number of utterances was different for each child.

RESULTS

Table 1 summarizes the descriptive statistics for all the measures used in the present paper. The values of MLU ranged from 1.2 to 11.2, suggesting that there was a considerable variability in the linguistic performance of the children. The high values of MLU were considerably above the typical values reported for English language samples collected from spontaneous conversations (cf. Rice et al., 2010). This suggests that the elicitation method provides different kind of data than spontaneous conversations, and at least in the advanced children leads to quite elaborate expressions. The descriptive statistics also shows the downside of the present sample. Some samples contained just a few utterances and practically no multi-word utterances. In total, there were 135 samples with available transcript data.

Table 2 shows the correlations between MLU and other language sample measures and the results of language tests and age. As expected, there is a strong relationship between MLU, as well as other language sample measures, and the vocabulary score. This is clearly because both measures were taken from the same language sample. Children who are able to provide good definitions score better in the lexical test, and because good definitions require elaborate language, they also score high on language sample measures.

MLU shows a significant correlation with both grammatical subtests. The value of the correlation coefficient suggests a weak to moderate relationship between the two variables (using Cohen’s classification, Cohen, 1988), with about 10 % of the MLU variance explained by the standardized grammar test scores. This may seem a rather weak relationship. However, given that the language samples were not large, and the collection method rather structured, it is noteworthy that there is a clearly significant relationship.

Two other indices from spontaneous language samples showed relationships with the grammatical subtests comparable to MLU. One of these is the number of be-forms per utterance. This index shows the strongest correlation with MLU of all language sample measures. Number of prepositions per utterance has just slightly lower correlation coefficient with the grammar tests than MLU. Two other sample measures (connectives and demonstrative pronouns) showed weaker but significant correlations with the subtests, and one (personal pronouns) did not show any significant correlations.

In order to investigate the specific relationships between language sample indices and grammar tests, regression analyses were used to estimate the contribution of grammatical subtests to the prediction of MLU and BPU (be-forms per utterance). The goal

was to establish whether the grammatical subtests were independently related to the MLU and BPU. The scores from TROG and morphology subtests were entered in the linear regression analysis as predictors in one step, and their marginal effects on the sample indices were evaluated. For MLU, there was a significant effect of both productive morphology ($\beta = 0.213$, $p = 0.026$) and syntactic comprehension -TROG ($\beta_2 = 0.193$, $p = 0.044$). Together, the two predictors explained 11.2 % of variance ($\eta^2 = 0.112$). For BPU, both subtests showed a significant effect as well (morphology: $\beta = 0.236$, $p = 0.013$; TROG: $\beta = 0.219$, $p = 0.020$), and they jointly explained 14.4 % of variance ($\eta^2 = 0.144$). This suggests that both grammatical subtests tap into domains that are important for children's performance in spontaneous language, especially for its grammatical aspect.

DISCUSSION

The present study had several different goals. On the most abstract level, it provides a useful contribution to the discussion about the validity of MLU and other indices taken from language samples. The findings show that MLU is related to other measures of progress in the grammatical acquisition. The relationships found between the sample indices and grammar subtests were not strong, but clearly significant. Given the specific method of sample collection, and the fact that the grammatical tests and language samples do not measure exactly the same aspects of language development, the result is rather encouraging, and shows that language samples provide meaningful measures of grammatical development.

One specific goal was to test if language sample measures could be applied to data collected in a structured manner during a standardized test administration. The results show that there is large variability among children in MLU and other language sample indices, which suggests that the procedure elicits potentially useful language samples. The samples differed in length but there were only few extremely short samples. The variability in length is somewhat problematic. Clearly, some children are quite talkative when asked for definitions, while other keep to the essentials. There was a moderate correlation between the number of utterances produced by the children, and the MLU ($r = 0.398$), suggesting that the observed MLU values may be affected by the talkativeness of the child. For unbiased MLU samples, it would be better if the sample lengths were more homogeneous across children. The measures are not directly comparable to MLU taken from spontaneous conversations, but they are useful for examining the children's language development.

Further goal of the paper was to test the use of language sample measures in Czech, and to examine indices that might be especially sensitive in Czech. MLU proved to be a usable index because it showed a reasonable variability and a significant relationship with other measures of grammatical development. Of the alternative measures, the mean number of be-forms per utterance proved to be the best one. This is in line with findings from English and other languages, in which the verb morphemes carrying tense have been found to be especially sensitive to grammatical development and language impairment (Rice, Tomblin, Hoffman, Richman, & Marquis, 2004; Rice, Wexler, & Hershberger, 1998; Steckol & Leonard, 1979; Bishop, Adams, & Norbury, 2006). The Czech verb *být* "to be" serves as an auxiliary or a copula, and both of these usages fulfill a predominantly grammatical function, including tense marking. The present findings thus provide cross-linguistic support to the view that tense-related measures are sensitive to the overall grammatical development.

Of the remaining indices, the use of prepositions also showed significant relation-

ships with the grammatical subtests. Prepositions mark syntactic complexity because they require that a phrase follows them. It is thus not surprising that the mean occurrence of prepositions predicts syntactic complexity. However, unlike with be-forms, this does not appear to add to the precision of MLU. The mean number of connectives per utterance also showed a significant relationship with the grammatical subtests, and so did the number of demonstrative pronouns, but these relationships were clearly weaker than in MLU or BPU. The use of personal pronouns did not show any significant relation to the grammatical measures.

Two tests of grammatical skills were used to validate language sample indices, a test of receptive syntactic knowledge, and a test of productive morphosyntactic skills. It was expected that the test of morphology could have closer relationship to the sample indices than the receptive grammatical test, because sample measures and the morphological test were based on productive language. To some extent, this was confirmed because the morphosyntactic subtest was a slightly stronger predictor of MLU and BPU than the TROG. However, both were significant independent predictors, suggesting that MLU and BPU reflect broader range of linguistic abilities.

Overall, the present data support the validity of MLU and other language sample measures. There is a clearly significant relationship between grammar tests and sample indices. Even though the relationship is not very strong in the present study, some previous studies failed to find any relationship between MLU and external measures (Klee & Fitzgerald, 1985). The present study was also limited by the data collection method. Unequal sample sizes in different children and the lack of conversational support from the adult probably mean that the MLU derived from the samples is measured with a large amount of error, and thus has limited reliability. Because validity cannot exceed reliability, the current validity estimates are likely deflated by the reliability limitations. It is likely that the correlations between grammatical subtests and MLU and other indices would be higher if the samples were longer and more homogeneous in size. In any case, the findings show that MLU and especially BPU are sensitive to grammatical development. Further research should validate MLU and BPU using data from conversational language samples collected by a trained interviewer. The present study indicates that this would be a worthwhile project.

Acknowledgements

This research paper is a part of the ELDEL research project (Enhancing Language Development in European Languages) funded by the EU, P7-PEOPLE-2007-1-1-ITN215961. Language samples were transcribed thanks to kind financial support from Faculty of Humanities at Charles University in Prague, project „Specifický vysokoškolský výzkum FHS UK v Praze pro rok 2009“.

REFERENCES

- Bedore, L. M., Leonard, L. B. (2001): Grammatical morphology deficits in Spanish-speaking children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 44, 905-924.
- Bishop, D. V. M. (2003): Test for reception of grammar, version 2 (TROG 2). London, Pearson Assessment.
- Bishop, D. V. M., Adams, C. V., Norbury, C. F. (2006): Distinct genetic influences on grammar and phonological short-term memory deficits: evidence from 6-year-old twins. *Genes, Brain and Behavior*, 5, 158-169.
- Blake, J., Quartaro, G., Onorati, S. (1993): Evaluating quantitative measures of grammatical complexity in spontaneous speech samples. *Journal of Child Language*, 20, 139-152.
- Bortolini, U., Caselli, C. M., Deevy,

- P., Leonard, L. B. (2002): Specific language impairment in Italian: The first steps in the search for a clinical marker. *International Journal of Language and Communication Disorders*, 37, 77-93.
- Brown, R. (1973): *A first language: The early stages*. Cambridge, Mass.: Harvard University Press.
- Casby, M. W. (1984): Application of intermittent time-sampling to calculations of mean length of utterance. *Perceptual and Motor Skills*, 58, 715-718.
- Cohen, J. (1988): *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ, Lawrence Erlbaum Associates.
- Cole, K. N., Mills, P. E., Dale, P. S. (1989): Examination of test-retest and split-half reliability for measures derived from language samples of young handicapped children. *Language, Speech, and Hearing Services in Schools*, 20, 259-268.
- Crystal, D. (1979): *Working with larsp*. London, Edward Arnold.
- Darley, F. L., Moll, K. L. (1960): Reliability of language measures and size of language sample. *Journal of Speech and Hearing Research*, 3, 166-173.
- Dethorne, L. S., Johnson, B. W., Loeb, J. W. (2005): A closer look at MLU: what does it really measure? *Clinical Linguistics & Phonetics*, 19, 635-648.
- Dunn, L. M., Dunn, L. M., William, K. T., Wang, J.-J. (1997): *Peabody picture vocabulary test-III*. Circle Pines, MN: American Guidance Service.
- Eisenberg, S. L., Fersko, T. M., Lundgren, C. (2001): The use of MLU for identifying language impairment in preschool children: A review. *American Journal of Speech-Language Pathology*, 10, 323-342.
- Gavin, W. J., Giles, L. (1996): Sample size effects on temporal reliability of language sample measures of preschool children. *Journal of Speech and Hearing Research*, 39, 1258-1262.
- Hux, K., Morris-Friehe, M., Sanger, D. D. (1993): *Language sampling practices: A survey of nine states*. *Language, Speech, and Hearing Services in Schools*, 24, 84-91.
- Klee, T., Fitzgerald, M. D. (1985): The relation between grammatical development and mean length of utterance in morphemes. *Journal of Child Language*, 12, 251-269.
- Žlab, Z. (1992): *Zkouška jazykového citu*. Ostrava, Mikrodata.
- Leadholm, B., Miller, J. (1992): *Language sample analysis: The Wisconsin guide*. Milwaukee, WI: Wisconsin Department of Public Instruction.
- Lee, L. L. (1974): Developmental sentence analysis: A grammatical assessment procedure for speech and language clinicians. Evanston, IL: Northwestern University Press.
- Leonard, L. B. (1989): Language learnability and specific language impairment in children. *Applied Psycholinguistics*, 10, 179-202.
- MacWhinney, B. (2000): *The CHILDES project: Tools for analyzing talk*. Mahwah, NJ, Lawrence Erlbaum.
- Miller, J. F., Chapman, R. S. (1981): The relation between age and mean length of utterance in morphemes. *Journal of Speech and Hearing Research*, 24, 154-161.
- Minife, F. D., Darley, F. L., Sherman, D. (1963): Temporal reliability of seven language measures. *Journal of Speech and Hearing Research*, 6, 139-148.
- Rice, M. L., Redmond, S. M., Hofman, L. (2006): Mean length of utterance in children with specific language impairment and in younger control children shows concurrent validity and stable and parallel growth trajectories. *Journal of Speech, Language and Hearing Research*, 49, 793-808.
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., Blossom, M. (2010): Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, 53, 333-349.
- Rice, M. L., Tomblin, J. B., Hoffman, L., Richman, W. A., Marquis, J. (2004): Grammatical tense deficits in children with SLI and nonspecific language impairment: Relationships with nonverbal IQ over time. *Journal of Speech, Language and Hearing Research*, 47, 816-834.
- Rice, M. L., Wexler, K., Hershberger, S. (1998): Tense over time: The longitudinal course of tense acquisition in children with specific language impairment. *Journal of Speech, Language and Hearing Research*, 41, 1412-1431.
- Rondal, J. A. (1978): Patterns of correlations for various language measures in mother-child interactions for normal and Down's syndrome children. *Language and Speech*, 21, 242-252.
- Rondal, J. A., Ghiotto, M., Bredart, S., & Bachelet, J.-F. (1987): Age-relation, reliability and grammatical validity of measures of utterance length. *Journal of Child Language*, 14, 433-446.
- Scarborough, H. S. (1990): *Index of productive syntax*. *Applied Psycholinguistics*, 11, 1-22.
- Steckol, K. F., Leonard, L. B. (1979): The use of grammatical morphemes by normal and language-impaired children. *Journal of Communication Disorders*, 12, 291-301.

Tomblin, J. B., Shonrock, C. M., Hardy, J. C. (1989): The concurrent validity of the Minnesota child development inventory as a measure of young children's language development. *Journal of Speech and Hearing Disorders*, 54, 101-105.

Wechsler, D. (2002): Wechsler Preschool and Primary Scale of Intelligence Third Edition (WPPSI III). London, Pearson Assessment.

SOUHRN

Validita řečových vzorků získaných strukturovanými elicitacími procedurami v češtině

Některé standardizované diagnostické metody pro vyšetření slovní zásoby se opírají o to, jaké děti definují zadaná slova. Tato studie použila definice, které děti (N=135) poskytly v subtestu slovník z WPPSI, k prozkoumání strukturních vlastností jejich jazykové produkce. Při analýze byly použity metody analýzy řečových vzorků, přičemž cílem bylo testovat validitu indexů řečových vzorků jako je průměrná délka věty (MLU). Výsledky dokládají, že řečové vzorky získané během standardizovaných testových in-

terakcí lze použít jako data pro analýzu, i když neposkytují tak bohatý zdroj informací jako vzorky získané ve spontánní interakci. MLU vykazovalo statisticky významnou korelaci se skóry porozumění syntaxi z testu TROG a z experimentálního testu ohýbání slov v produkci. Průměrný výskyt tvarů slovesa *být* v jedné větě vykazoval ještě silnější souvislost s testy gramatického vývoje než MLU. Korelace mezi indexy získanými z řečových vzorků a testy gramatického vývoje byly obecně poměrně nízké (kolem 0,3), ale vzhledem k počtu probandů vysoce statisticky významné. Potvrdilo se tak, že MLU a další indexy řečových vzorků jsou citlivými měřítky gramatického vývoje v češtině.

Table 1 Descriptive statistics (raw scores) for performance on core measures: Receptive syntactic skills, Morphosyntactic skills, Vocabulary, and language sample measures (MLU, total number of words in language sample, total number of sentences in language sample, number of different words in a sample, type token ratio, mean number of prepositions in one sentence, mean number of forms of verb "být" in one sentence.)

Measure	Mean (SD)	Range
Age (months)	72,23 (3,7)	65-80
Core measures		
Receptive syntactic skills	11,6 (2,6)	15-16
Morphosyntactic skills	17,2 (4,1)	0-23
Vocabulary	20,2 (7,2)	0-35
Language sample measures		
MLU	5,7 (2,0)	1,2-11,2
Total number of words	271,6 (179,1)	12-997
Total number of sentences	45,2 (19,5)	10-119
Different words	137,6 (76,5)	4-429
TTR	0,54 (0,08)	0,33-0,84
Prepositions per utterance	37,6 (15,1)	0,0-78,6
Be-forms per utterance	32,1 (19,9)	0,0-94

Table 2 Correlations (Pearson) between core and language sample measures

Measure	MLU	Prep p. u.	Be-forms p.u.	Different words	No. of sentences	No. of words
Receptive syntactic skills	0,303**	0,313**	0,340**	0,181**	0,087	0,157
Morphosyntactic skills	0,313**	0,244**	0,349**	0,160	0,060	0,154
Vocabulary	0,692**	0,570**	0,516**	0,675**	0,519**	0,635**

**Significant at the 0,01 level

Copyright of Ceskoslovenska Psychologie is the property of Institute of Psychology of the Academy of Sciences and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.