Journal of Communication Disorders 93 (2021) 106146

Contents lists available at ScienceDirect

Journal of Communication Disorders

journal homepage: www.elsevier.com/locate/jcomdis

# Validity of the SDDS: A 40-item vocabulary screening tool for 18- to 42-month olds in Czech

Filip Smolík [a,*], Ilona Bytešníková [b]

[a] Institute of Psychology of the Czech Academy of Sciences, Hybernská 8, CZ-110 00 Praha 1, Czech Republic
[b] Masaryk University, Faculty of Education, Department of Special and Inclusive Education, Poříčí 9, CZ-603 00 Brno, Czech Republic

ARTICLE INFO

*Keywords:*
Vocabulary screening
Validity of screening
Parent report

ABSTRACT

*Background:* Children with early language delays are at increased risk of persistent language impairment. Early identification and intervention are desirable. Parent-report inventories are useful screening tools, but the screening context places limits at their length. Validity of parent-report screening tools in languages other than English has been rarely reported in detail.
*Aims:* The aim was to establish the concurrent validity of an existing 40-item parent-report vocabulary screening tool in Czech, using a picture-based examiner-administered comprehension and production task as a concurrent measure of vocabulary.
*Methods & Procedures:* Parents of 200 children aged 1;3 to 3;6 were given the screening inventory, in which they were asked if their child says or comprehends each of 40 words. At the same time, children were administered a picture-based comprehension and production task. Concurrent validity of the inventory was examined using correlations, partial correlations, and regression analyses controlling for age.
*Outcomes & Results:* The partial Spearman correlation (controlling for age) between production scores from parent-report and production scores from the examiner-administered task was 0.53; for comprehension, the correlation between parent report and test scores was 0.36. These values are similar to those reported for short and full versions of MacArthur-Bates CDI for 2-year-olds.
*Conclusion & Implications:* A 40-item tool shows clear concurrent relations with an examiner-administered picture comprehension and production task. The study demonstrates that short parent-report tools may be useful in early identification of language impairments, and they may be a good option particularly in languages that have limited repertoire of assessment instruments.

## 1. Introduction

Language impairments in children present a serious risk for further development, including the development of social relationships, cognitive skills, educational achievement, and social adjustment (Hughes et al., 2017; McCabe, 2005; Redmond & Rice, 1998; St Clair, Pickles, Durkin & Conti-Ramsden, 2011; Whitehouse, Watt, Line & Bishop, 2009; Zambrana, Pons, Eadie, & Ystrom, 2014). It is thus important to identify children with language impairment early, so that appropriate intervention strategies may be deployed in the

Author note: The study was supported by the Czech Science Foundation (GA ČR) grant No. GA19–15576S awarded to the first author, and a Masaryk University grant MUNI/A/1429/2020 to the second author.
* Corresponding author.
*E-mail address:* smolik@praha.psu.cas.cz (F. Smolík).

https://doi.org/10.1016/j.jcomdis.2021.106146
Received 8 January 2019; Received in revised form 25 June 2021; Accepted 27 July 2021
Available online 4 August 2021
0021-9924/© 2021 Elsevier Inc. All rights reserved.

youngest possible age. To identify children early on, various types of assessment instruments are needed, from full-length multidimensional norm-referenced batteries to short tasks that can be used for screening and identification of children who are at risk. Screening tasks have a valuable role in practice because they can be used for time-efficient identification of children who need further attention. Full assessment is time-consuming and requires well-trained professionals, and the screening tasks can be used to select, with relatively low costs, children that should be referred for such detailed examination. In addition, the screening tools are easier to develop than full-fledged batteries, which is especially relevant in languages that have no formal assessment tools available. The present study reports the validity of SDDS (*Stručný dotazník dětského slovníku,* Short questionnaire of children's vocabulary; Smolík & Bytešníková, 2017), a short, 40-item parent-report checklist for screening of vocabulary development in children acquiring Czech, based on relations to another language development measure.

### 1.1. Early screening for language impairment

The delayed or disordered development of language skills may be classified under various clinical labels. Children who do not use language at an age-appropriate level at 2 or 3 years of age, without any obvious medical cause, are described as *late talkers* (Rescorla, 2011), children with *late language emergence* (e. g. Rice, Taylor, & Zubrick, 2008) or a similar label. In children who are older, usually about 4 years old or above, the terms *language impairment, language disorder* (American Psychiatric Diagnostic and statistical manual of mental disorders, 2013), *specific language impairment* (e. g. Tomblin, Smith & Zhang, 1997), *developmental language disorder* (Bishop et al., 2017) or similar are used. Children are not diagnosed with language impairment before the age of 4 or 5 years because early language development is highly variable across children and some proportion of those who are late talkers by 2 or 3 years will catch up with their peers and their language skills will be in the normal range. The proportion of late talking children who develop long-term language impairment is a matter of debate, with quite variable estimates available. Dale, Price, Bishop and Plomin (2003) found that more than 40% of children who scored below 10th percentile on vocabulary at 2 years of age could be classified as language impaired at 4 years. While this means that early delay does not necessarily lead to later impairment, the likelihood of impairment is substantially increased if early delay is present. In contrast, Rescorla (2002) found that children classified as late talkers at the age of 2 had language scores mostly in the normal range when they were 5 years old and older. However, the children's performance was still below age average, and this persisted at least until 17 years of age (Rescorla, 2009). It is important to note that children recruited for this study had expressive-only delays, and this is apparently associated with better prognosis than combined delay in production and comprehension (Dale et al., 2003; Thal, Tobias, & Morrison, 1991; Zambrana, Pons, Eadie, & Ystrom, 2014). Overall, there is solid evidence that children with language delay at the age of 3 years or earlier are at increased risk of developing language impairment, and even those who later improve have lower language skills than their peers (McKean et al., 2017; Morgan, Farkas, Hillemeier, Hammer, & Maczuga, 2015; Rice, Taylor, & Zubrick, 2008; Zambrana, Pons, Eadie, & Ystrom, 2014). Screening for language delay at the age of 3 years and earlier is thus well-founded, and there is growing consensus that developmental intervention should be provided as early as possible, avoiding the *wait-and-see* approach (Capone Singleton, 2018).

### 1.2. Language screening and parent report

Various tools and procedures are available for early screening of language development. Some of these include testing by healthcare professionals such as speech-language pathologists, pediatricians or nurses, and they are often derived from larger full-scale assessment batteries (Seymour, Roeper & de Villiers, 2003; Uilenburg et al., 2018; Wiig, Secord & Semel, 2013). Other tools rely on parent report, capitalizing on the fact that parents know the child in a wide variety of situations. The advantage is that parent reports are based on regular and repeated behaviors of the child rather than short-time snapshots.

The parent report methods have long been used in various areas of child psychology (Humphreys & Ciminero, 1979). In language acquisition research and assessment, the first formal tools were developed in late 1980s and early 1990s, based on previous use in research. The two well-known instruments are the MacArthur-Bates Communicative Development Inventories (CDI; Fenson et al., 1993, 2007), and the Language Development Survey (LDS; Rescorla, 1989). Both contain extensive lists of words, and parents are asked to mark the words used or understood by their child. In addition to the word list, the tools contain questions about language and communicative development of children. The original CDI is available in two versions for different ages, 8 to 15 months and 16 to 30 months. In the former, Words and Gestures, additional questions target various nonverbal symbolic and communicative behaviors that are reliably related to the onset of language. The version for older children, Words and Sentences, contains additional questions about how children combine words and what kinds of sentences or word combinations they use. The Language Development Survey relies primarily on the word list but it asks parents to write down five longest word combinations they can remember from the child.

The CDI inventories were developed as general assessment tools for use in practice as well as research. In contrast, LDS was primarily intended as a screening tool for language impairment. Both tools are widely used for these purposes, but they may not be suited for routine screening because of the amount of time and attention needed to complete them. According to Rescorla (1989), most parents can fill LDS within 10 min, which seems efficient. CDI is about twice as long, various authors mention times between 30 and 45 min as the typical time needed to respond (Eriksson, Westerlund, & Berglund, 2002; Jackson-Maldonado, Marchman & Fernald, 2013). With young children or children who say only a few words, the time can be considerably shorter, but even reading the full list of words in CDI may take 10 to 15 min, and this time may not be available in the settings where screenings are typically administered, such as the pediatrician's office.

To some extent, shorter tools are available as components of broader developmental screening instruments such as Ages and Stages (Squires, Twombly, Bricker & Potter, 2009) or Vineland Adaptive Behavior Scales (Sparrow, Cicchetti & Balla, 2005). Parent-report

tools focused on language include the Developmental Vocabulary Assessment of Parents (Libertus, Odic, Feigenson, & Halberda, 2015), comprising over 200 items. While shorter than full-scale CDI or LDS, they are not suitable for first routine screening. The authors of CDI recognized the need for a shortened version of the inventory, and developed two inventories with 89 or 100 items for children of various ages (Fenson et al., 2000). For children aged 30 to 37 months, CDI III with 100 items was developed (Fenson et al., 2007). Short versions of CDI were also developed in other languages such as Spanish and Swedish (Eriksson, Westerlund, & Berglund, 2002; Jackson-Maldonado et al., 2013). The reports on the development of these short versions suggest that their measurement properties are comparable to the full versions and the decrease in accuracy compared to the full versions is small. Shortened parent-report vocabulary inventories are thus a viable option for developing screening tools.

### *1.3. SDDS: the Czech 40-item screening tool*

In Czech, there are only few tools available for early language assessment, and there are no psychometrically constructed instruments or procedures to screen young children for language development delays. This means that professionals have very limited range of options when examining language in young children. Recently, Dovyko (Dotazník vývoje komunikace: Communicative Development Questionnaire), a normed adaptation of the CDI: Words and Sentences was published (Smolík, Turková, Marušincová & Malechová, 2017) but this is not suitable for fast screening. At the same time, there is demand from practicing professionals for formal screening and assessment tools. To address this need, a screening inventory was developed: SDDS - *Stručný dotazník dětského slovníku* (Short questionnaire of children's vocabulary; Smolík & Bytešníková, 2017). This inventory used the approach taken in the short versions of the CDI, namely shortening the vocabulary inventory and using the scores to compare children, not to map the whole vocabulary. However, a 100-item inventory was still considered lengthy based on discussions with practitioners, and a shorter list of 40 words was developed, based on the age norms of the Czech CDI adaptation and informal feedback from professionals. The final product, SDDS, is a short questionnaire in which parents mark both whether the child uses or comprehends each word. The idea was to provide professionals in the field with an instrument that could be completed by parents in a very short period of time, e. g. in the waiting room of the pediatrician's office or even during the examination.

### *1.4. Establishing validity of the SDDS*

The SDDS was standardized on a sample of 1047 children from Brno and neighboring areas, and it has been published, along with the norms and psychometric information (Smolík & Bytešníková, 2017). The norming study revealed a good discrimination ability especially around the age of 2 years. Separate norms were created for comprehension and production. While the norming study used a large sample, it did not include any measures of concurrent validity, relying instead on arguments about the content validity. One reason for this was the absence of an appropriate measurement tool in Czech that could be used as the validating measure: there are no such norm-referenced tools in Czech for the ages around 2 or 3 years.

However, it is desirable to test the validity of the screening tool against a direct assessment of language or vocabulary skills. For this reason, the present study administered the SDDS screening tool along with the Czech version of the Crosslinguistic Lexical Task (CLT; Haman, Łuniewska & Pomiechowska, 2015; Łuniewska & Smolík, 2016), a tool that was developed in the European COST cooperation project on bilingual SLI. The method has not been normed in Czech (nor in any other language) but was developed with the intent to assess vocabulary. The development took into account naming agreement for stimulus pictures and the properties of the target words such as frequency, length, and age of acquisition ratings. Details of the procedure for developing and selecting the list of items were reported by Haman et al. (2017) and Haman et al. (2015). It would be ideal to use a standardized norm-referenced task for validating our SDDS inventory, but because no such tasks are available for this age range in Czech, CLT is a reasonable option to establish relations between the screening inventory and an examiner-administered measure. The advantage of CLT is that it has an international community of users and it is likely that more technical information becomes available with the increased usage of the task.

CLT consists of 128 items, of which 64 test comprehension and 64 production. The comprehension component is similar to the Peabody Picture Vocabulary Test (PPVT; Dunn & Dunn, 2007), and the production component is similar to the initial parts of Expressive Vocabulary Test (EVT; Williams, 2007). Even though the task does not have formal proof of concurrent validity yet, the fact that it uses similar type of tasks as well-known instruments supports its content validity. If there are adequate relations between SDDS and CLT, the validity of both tasks is supported to some extent. Ideally, this will be the first step in the process of describing the measurement properties of SDDS and CLT, and developing a set of norm-referenced instruments for Czech. We thus examined the relations between SDDS scores and the examiner-administered CLT vocabulary task using correlational analysis, regression analyses and sensitivity and specificity analyses, while controlling for the effects of age.

## 2. Method

### *2.1. Participants*

A total of 200 typically developing children aged 15 to 42 months were recruited for the study. The parents of potential participants were contacted via preschools, pediatricians, parent's clubs or social media, either by the staff of the cooperating institutions or directly by the examiners. Children were included only if parents reported no communication disorder or other developmental disorder, no frank neurological impairment or serious sensory impairment. Only three of the children had some exposure to languages other than Czech at home (Russian or Turkish) but their families reported that Czech was the primary means of communication and

one of the parents was native Czech speaker. The sample covered the whole age range for which SDDS is normed, i. e. 18 to 42 months, including a few younger children, but older children were represented somewhat more than younger children. Table 1 summarizes the distribution of children's ages as well as the demographic and perinatal information. The children were recruited in various regions of the Czech Republic, primarily in the South and East, and were from various types of communities, both large cities and the country. The study was approved by the Ethics Commission of the Institute of Psychology of the Czech Academy of Sciences.

### 2.2. Measures

*SDDS.* The inventory consists of 40 items that are shown in the appendix, with approximate translations. The words are organized by thematic relations, grouping animals, household items, actions and other categories in a similar way as in CDI and Dovyko. The inventory includes 20 nominals, 9 verbs, 2 adjectives, and 9 words belonging to other categories (including but not limited to interjections, pronouns, and numerals). For each word, the parent marks whether the child uses the word, and whether s/he understands it. Separate scores were calculated for comprehension and production by tallying the number of words marked in each category.

*CLT.* Cross-linguistic Lexical Task consists of 128 items, 64 for comprehension and 64 for production. In each group, there are 32 items testing nouns and 32 for verbs. In the production items, the child is shown a single picture and asked to name it. In the comprehension items, children are shown a page with four pictures and asked to point to the one labeled by the experimenter. The task was scored separately for production and comprehension. Unlike the standardized measures such as PPVT or EVT (Dunn & Dunn, 2007; Williams, 2007), CLT does not present items with increasing difficulty, and thus does not use discontinuation rules based on repeated unsuccessful responses. All items must be presented. This was generally not a problem with the comprehension component, but with the production component, a total of 17 children could not cooperate: they would not respond to the prompts and they became frustrated and fidgety when confronted with additional items. The administration of the task was finished for these children. These cases were scored with 0 and included in our main analyses because this appeared to be the more conservative approach for estimating the validity of SDDS. However, we also report some analyses that excluded these children.

### 2.3. Procedure

If parents agreed with their children's participation, they worked with the examiners to schedule a testing session during which CLT was administered. The testing sessions could be at the child's preschool, parent club, at a pediatrician's office or in the children's home, always in a quiet room with the examiner, child and possibly the parent present. If parents were present, they were given SDDS task during the testing session. Otherwise, they received the form from preschool teachers or other collaborators with the instruction to return it within 7 days from the testing session. CLT was administered first with the comprehension component, then the production component. The picture stimuli were presented on a tablet screen.

**Table 1**
Demographic and perinatal data.

| Variable | N | % |
|---|---|---|
| Child age in months (years;months) | | |
| 15–17 (1;3–1;5) | 6 | 3.0 |
| 18–23 (1;6–1;11) | 28 | 14.0 |
| 24–29 (2;0–2;5) | 40 | 20.0 |
| 30–35 (2;6–2;11) | 40 | 20.0 |
| 36–42 (3;0–3;6) | 86 | 43.0 |
| Mother's education | | |
| elementary or vocational | 19 | 9.5 |
| completed secondary, some postsecondary | 82 | 41.0 |
| college or university degree | 99 | 50.5 |
| Father's education (if living in household) | | |
| elementary or vocational | 51 | 25.6 |
| completed secondary, some postsecondary | 73 | 36.7 |
| college or university degree | 76 | 38.7 |
| Number of siblings | | |
| 0 | 68 | 34.0 |
| 1 | 103 | 51.5 |
| 2 | 21 | 10.5 |
| 3 or more | 8 | 4.0 |
| Birthweight | | |
| 2500 g and above | 188 | 94.0 |
| under 2500 g | 12 | 6.0 |
| City/town of residence population | | |
| above 200,000 | 86 | 43.0 |
| 20,000 – 200,000 | 21 | 10.5 |
| 2000 – 20,000 | 28 | 13.5 |
| under 2000 | 66 | 33.0 |

### 2.4. Data analysis

The data were analyzed using correlational analysis and regression analysis examining the relations between the performance on SDDS and on CLT comprehension and production tasks. For correlational analysis, Spearman rank-order correlations were used because the relation between the two measures might not be linear. Both SDDS and CLT scores naturally increase with age and the raw correlations would simply reflect this effect; for this reason, partial correlations controlling for age were used.

Linear regression analysis was used to estimate the effect of using the 10th percentile cut-off criterion on SDDS. In this set of analyses, CLT production or comprehension score was used as the dependent variable. The status according to SDDS was used as a categorical predictor, with children at or below 10th percentile for the given age considered at risk, along with age as a continuous predictor. To control for the effects of age, it was included as an additional quantitative predictor. Age was initially included as orthogonal polynomial with linear and quadratic components to account for any changes in the growth rate of CLT scores. If the quadratic component had no significant effect, the models were refit and the model with linear age only is reported. Standardized regression coefficients ($\beta$) are reported here as measures of effect size. All analyses were performed in R (R development core team, 2020).

Sensitivity and specificity of SDDS for diagnosing impaired language development were calculated using different quantitative criteria on CLT to determine the language status and different screening criteria on SDDS. Because CLT is also a task in development, there are no established criteria for diagnosing language impairment using this task. We chose to use the bottom 15 percent criterion for classifying children as language-impaired according to CLT. Because our children were of different ages, the value of this criterion must be calculated by modeling quantile curves. The value of 15th percentile for CLT was determined, separately for the comprehension, production and total scores, by using the quantSheets function in the software package gamlss for R (Stasinopoulos et al., 2007).

## 3. Results

Descriptive data for all methods are shown in Table 2, separately for children in four age bands. There is a clear increase with age in all measures. Fig. 1 shows scatterplots demonstrating the increase of SDDS and CLT scores with age, also showing a linear regression line and a nonparametrically smoothed growth curve. The smoothed lines do not assume linear relation between the variables along the whole range of values of the x-axis variable, but rather show a smoothed empirical relation calculated locally for each point along the x axis. Fig. 2 shows the mutual relations between SDDS and the criterion task, CLT; close alignment between the regression line and the nonparametric spline suggests that the relation is linear, but it is not the case for the relations of both tasks with age. Because of the potential non-linearity in relations between age and the SDDS and CLT scores, the correlational analyses were performed using the Spearman rank-order correlation coefficients, and the growth models tested for the significance of the quadratic component.

Table 3 (below diagonal) shows the raw Spearman correlations between SDDS scores, CLT scores, and age, as well as (above diagonal) partial correlations between SDDS and CLT scores with age partialled out. All these correlations are statistically significant at the level of $p<0.001$. Validity of SDDS is best estimated from the partial correlations, and these suggest moderate relations between the two measures. The correlation coefficient between production scores on SDDS and CLT is 0.53, between comprehension scores 0.36. The relations also suggest some degree of modality-specific validity, especially for the production. SDDS production scores show stronger relation to CLT production ($r = 0.53$) than comprehension ($r = 0.39$). For SDDS comprehension, the relations to CLT comprehension and production are very similar ($r = 0.36$ and 0.32, respectively).

The analyses included 17 children who scored 0 on CLT production because they did not cooperate and produced no verbal labels for the pictures. When these children were excluded, the estimated coefficients were 0.55 for SDDS production and 0.41 for SDDS comprehension, comparing to 0.53 and 0.32, respectively, when all children were included. Given these numbers, we were satisfied that the children who did not provide verbal responses did not skew the validity estimates.

The analyses show a clear correlation between SDDS and CLT scores, but SDDS is likely to be used for making categorical distinctions about whether to refer the child for further evaluation or not. To compare the performance of children who do and do not pass a categorical criterion, we examined the performance on CLT in children who were at or below the 10th percentile on SDDS with children who were above the 10th percentile. The effect of age was taken into account by using linear regression, with the low vs. normal group as a categorical predictor and age as a continuous predictor. For the analysis of comprehension, age was entered with a linear and quadratic term to account for the nonlinearity in the growth of CLT. For production, the quadratic component was not significant and the reported model only includes linear effect of age. Table 4 shows that there were clear differences between the

**Table 2**
Descriptive statistics in four age bands.

| Age (months) | | SDDS Comprehension | | SDDS Production | | CLT Comprehension | | CLT Production | |
|---|---|---|---|---|---|---|---|---|---|
| | N | M | SD | M | SD | M | SD | M | SD |
| 15–23 | 34 | 24.9 | 9.7 | 8.5 | 10.29 | 27.32 | 13.77 | 13.50 | 11.24 |
| 24–29 | 41 | 34.7 | 6.1 | 26.0 | 25.88 | 44.10 | 16.28 | 27.98 | 16.35 |
| 30–35 | 40 | 36.5 | 4.6 | 32.5 | 31.12 | 47.55 | 13.63 | 28.23 | 14.90 |
| 36–42 | 85 | 36.8 | 5.4 | 38.0 | 36.32 | 50.73 | 11.58 | 37.55 | 13.54 |

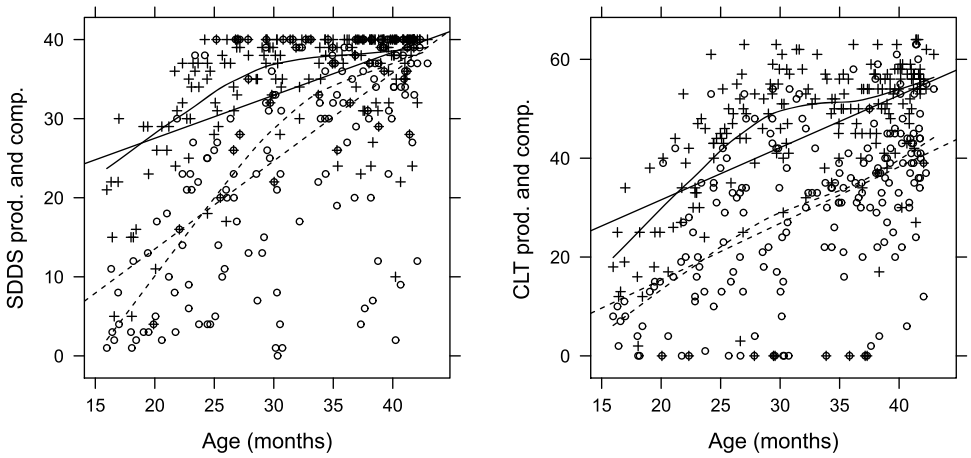


**Fig. 1.** The figure shows relations between both measures and age, with solid lines and crosses indicating comprehension scores, and dashed lines and circles production scores. The curved lines are nonparametric moving-average estimates.

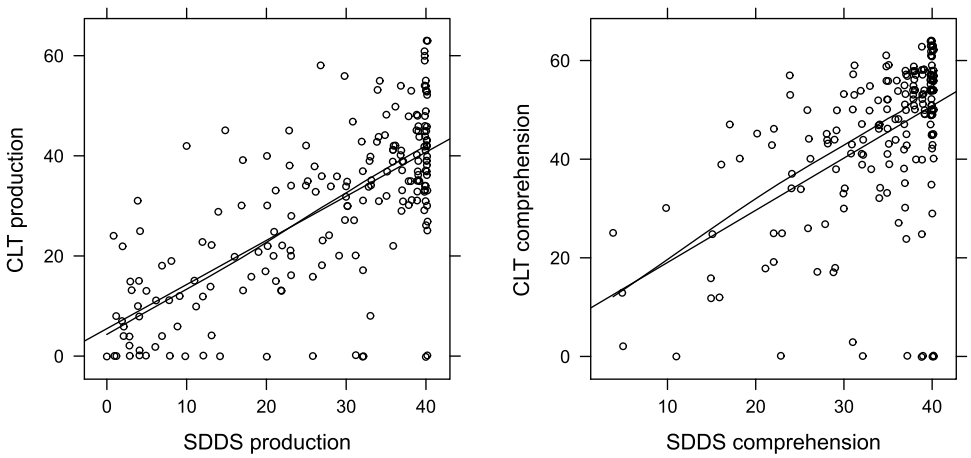


**Fig. 2.** Relations between SDDS and CLT, separately for each modality. The curved lines are nonparametric moving-average type estimates.

**Table 3**
Correlations between all the measures as well as age (below diagonal), and partial correlations between the measures, with age partialled out (above diagonal). All p's<0.001.

| | SDDS prod. | SDDS comp. | CLT prod. | CLT comp. |
|---|---|---|---|---|
| SDDS production | | 0.63 | 0.53 | 0.39 |
| SDDS comprehension | 0.69 | | 0.32 | 0.36 |
| CLT production | 0.71 | 0.49 | | 0.73 |
| CLT comprehension | 0.54 | 0.52 | 0.79 | |
| Age (days) | 0.64 | 0.54 | 0.53 | 0.51 |

groups. When the low-performing group was based on the SDDS production score, the standardized regression coefficient for SDDS status was 0.34 ($p<0.001$), and the average difference between CLT production scores between the low and normal group was over 15 points. Using the SDDS comprehension score to define the low-performing group resulted in mean difference of about 8 points on CLT, and the corresponding effect size in standardized regression coefficient metrics was 0.18 ($p = 0.002$). These results show a clear difference in vocabulary performance between children who are at the 10th percentile or below on SDDS and those who are above this cutoff. This difference is particularly strong for the production scores, confirming that the production scores on SDDS provide more precise measure of language status than comprehension scores.

The difference in CLT vocabulary scores between children above or below 10th percentile on SDDS confirms that low status on SDDS is associated with decreased scores on examiner-administered lexical task. However, diagnostic accuracy of screening tools is usually described using sensitivity and specificity, which are reported for different screening and reference criteria in Table 5. In

**Table 4**
Regression models showing the effect of being at or below 10th SDDS percentile on CLT scores.

| | Production<br>Standardized $\beta$ | $t$ | $p$ |
|---|---|---|---|
| Intercept | — | 5.78 | <0.001 |
| Age (days) – linear | 0.56 | 10.02 | <0.001 |
| SDDS production > 10th percentile | 0.34 | 6.12 | <0.001 |
| | Comprehension<br>Standardized $\beta$ | $t$ | $p$ |
| Intercept | — | 15.80 | < 0.001 |
| Age (days) – linear | 0.52 | 8.81 | < 0.001 |
| Age (days) – quadratic | −0.15 | −2.60 | 0.010 |
| SDDS comprehension > 10th percentile | 0.18 | 3.11 | 0.002 |

general, specificity is mostly above 0.8, with lowest values of 0.76. This means that most children who are identified in screening really have low vocabulary when measured in a test. On the other hand, sensitivity was low, ranging from 0.33 to 0.59, with best results for the combined criteria on SDDS comprehension and production. Low sensitivity means that a relatively large number of children with low vocabulary are not identified by SDDS. However, the number should be evaluated in comparison to other parent report tools used for language screening.

## 4. Discussion

The results largely confirm that a 40-item parent-report inventory can be used as a screening measure of language development with acceptable validity. While the validity of parent report tools for child language has been repeatedly established, the shortest existing instruments consist of 89 items (Fenson et al., 2000; Jackson-Maldonado et al., 2013). It was thus unclear whether shortening the list of words even further would not render the instrument invalid – the limited length is likely to decrease reliability, which is a limiting factor for the validity of a test. The usual strategy to overcome the uncertainty of parent reports is to use large numbers of items, but the requirements of a screening tool are exactly the opposite, and 40 items might be too little to provide enough data. However, the results show that the 40-item scale can provide meaningful data on early vocabulary.

It could be argued that the relation between SDDS and CLT tasks is not very strong; the partial correlation coefficient is 0.53 between the production measures, and 0.36 between comprehension measures (see Table 3), which corresponds to about 25% or 13% of variance in CLT explained by SDDS. However, these numbers must be evaluated in the context of similar studies and similar tools. There are no generally accepted rules for how strong the correlation between a test and a criterion task should be for the test to be considered valid; rather the procedure of establishing validity requires integrating wide range of information about the performance of the instrument (Cicchetti, 1994; Newton & Shaw, 2013; Standards for Educational & Psychological Testing, 2014). The current study provides one step in this process, and the results should be interpreted with regard to other studies that examined the validity of screening and parent-report tools. In this respect, the current results are fully comparable with criterion validity figures available for the complete CDI and other screening tools. Feldman et al. (2005) reported correlation coefficient of $r = 0.32$ between CDI at 2 years of age and PPVT at 3 years, and of $r = 0.41$ between the 100-item version CDI-III at 3 years and PPVT at the same age. These numbers best correspond to our figure of $r = 0.39$ between SDDS in production and CLT in comprehension (Table 2). There are few validity estimates comparing production scores on CDI with formal production tasks, but Reese & Read, 2000 reported a correlation between CDI at 2;1 years and EVT (Williams, 2007) at 2;8 years, with the value of $r = 0.50$. This is very close to our value of 0.52 between the productive measures reported in Table 2. Libertus, Odic, Feigenson, & Halberda, 2015 tested the validity of their 204-word parent-report instrument. In a regression model with the parent-report scores as the dependent variable and the age and examiner-administered vocabulary task (PPVT) as predictors, the standardized regression coefficient for PPVT was 0.42, which is roughly comparable with the partial correlation of 0.39 in our results between SDDS production and CLT comprehension scores (although partial correlations are somewhat different from standardized regression coefficients, i. e. semipartial correlations). Heilmann, Weismer, Evans, & Hollar,

**Table 5**
Sensitivity and specificity of different SDDS criteria for language delay according to CLT (total $N = 200$).

| | CLT Score <=15th percentile<br>Comprehension<br>($N = 29$) | | Production<br>($N = 29$) | | Total<br>($N = 43$) | |
|---|---|---|---|---|---|---|
| SDDS screening criterion | Sensitivity | Specificity | Sensitivity | Specificity | Sensitivity | Specificity |
| <= 10th percentile in | | | | | | |
| production ($N = 25$) | 0.38 | 0.92 | 0.41 | 0.92 | 0.33 | 0.93 |
| comprehension ($N = 30$) | 0.38 | 0.89 | 0.41 | 0.89 | 0.33 | 0.90 |
| both ($N = 13$) | 0.48 | 0.84 | 0.55 | 0.85 | 0.44 | 0.85 |
| <= 15th percentile in | | | | | | |
| production ($N = 37$) | 0.45 | 0.86 | 0.48 | 0.87 | 0.40 | 0.87 |
| comprehension ($N = 35$) | 0.38 | 0.86 | 0.41 | 0.87 | 0.33 | 0.87 |
| both ($N = 16$) | 0.52 | 0.76 | 0.59 | 0.77 | 0.49 | 0.77 |

2005 examined a group of 24-month-old late talkers and found a correlation of $r = 0.63$ between the production scores on CDI and the expressive portion of the Preschool Language Scale-Third Edition (Zimmerman, Steiner, & Pond, 1992). This is higher than the correlation between productive scores on SDDS and CLT observed here ($r = 0.53$), but the full-length CDI is a more comprehensive measure so it is understandable. Comparisons with these studies also confirm that validity estimates for the parent-reported comprehension scores are lower than for production scores, both in SDDS and other studies.

Overall, comparing the correlations between CLT and SDDS in our study with the existing research suggests that the new 40-item parent-report instrument can be used as a measure of vocabulary skills. It does not have considerably lower validity estimates than full-length CDI, or the 100-item CDI-III. The reduction to 40 items thus does not limit the instrument's validity and the use of short parent-report instruments is supported by the current study.

For practical use of an instrument, sensitivity and specificity are particularly important. Our data show that SDDS has acceptable specificity but the sensitivity is low with most criteria. This means that children who are identified by the screening criterion usually do have low vocabulary, but a large number of children with low vocabulary are not identified by the screening tool. However, some previous studies reported even lower values of sensitivity for well-established screening tasks, e. g. nonword repetition (Archibald & Joanisse, 2009). A recent metaanalysis found the mean of sensitivity from 6 studies to be 0.66 (Sim, Thompson, Marryat, Ramparsad & Wilson, 2019), which is only slightly higher than the sensitivity of 0.59 provided by the best criterion reported here. Although our SDDS data do not reach the values considered as acceptable by Plante & Vance, 1994, they are nevertheless in line with tools used in practice. We thus recommend caution in the use of SDDS but consider it comparable with existing instruments. Based on the results, we recommend using combined comprehension and production criteria rather than modality-specific scores. Given the good specificity values, low result in SDDS should be viewed as an important warning sign that warrants more detailed examination by professionals. Values above the screening range, on the other hand, should not be used to exclude the possibility of language delay, especially if children score in lower percentiles.

Most existing research on the validity of parent-reports in language screening is from English-speaking environments using tools in different varieties of English. Some evidence for the validity of parent-report tools, especially the CDI but also LDS, is available from a number of languages (including Polish which is closely related to Czech: Rescorla, Constants, Bialecka-Pikul, Stepien-Nycz, & Ochal, 2017; see Dale and Penfold, 2011, for general overview), but test validity can be influenced by many factors and should thus be tested when tools are adapted in new environments, which was the goal of the current study. For example, parents in a different communities and cultures may approach the instrument with different beliefs, which might lead to differences in the quality of reporting. With the short inventory presented here, the validity could be compromised if items are not be representative of the whole vocabulary. The current study, however, indicates that this is not the case for the Czech SDDS measure.

The results reported here are especially encouraging for researchers and clinicians working in smaller languages with limited repertoire of assessment and screening tools. The present findings indicate that a short screening inventory can be comparable to much longer inventories. If a practical tool for efficient screening and early identification of language- impaired children is needed, a short inventory is a viable option. Such a short inventory is easier to develop than a full-length instrument such as the MCDI, and it is thus a good option as an initial step in developing a portfolio of diagnostic instruments.

We believe that SDDS should be introduced in clinical practice as a part of routine pediatric health checks. This could improve identification of children who are otherwise not identified. This is especially important because a relatively large proportion of Czech children do not go to daycare before the age of 4 or 5, and thus their limited skills may not become apparent. The introduction of SDDS into clinical practice should raise awareness of language and communication problems among pediatricians, and it is likely to be accepted by health professionals much better than long instruments such as CDI (Dovyko). Even LDS (Rescorla, 1989) would be probably considered as impractical for routine use due to its length, if it had a Czech version.

A strong aspect of the present study is that it examined the full age range of children for which the SDDS was normed, i. e. from 18 to 42 months, including a few younger children. Usually, validation of developmental measures is done in a narrow age band to avoid the confounding effect of age. However, this means that the instrument is not tested in the whole age range for which it is normed, and that the correlations with other measures may be lower in other age bands. The validity estimates presented here reflect the overall performance of the tool across the whole age range, with the age effects removed using partial correlations.

### *4.1. Limitations*

One limitation of the present study is the nature of the criterion task. Instead of being a well-documented extensive measure of language development, it was itself a new tool that is currently under development. However, this only reflects the situation that motivated the development of the 40-item SDDS inventory, namely the severely limited range of assessment and screening instruments for Czech. Further research will be important in establishing predictive and convergent validity, as well as other measurement properties.

For a screening tool, it would be most desirable to have an estimate of predictive validity, but with a new method, it is not surprising that this is not available. Another potential limitation is that the criterion test might not be suitable for the whole age range of children participating. CLT requires understanding and cooperation from the child, and children may be shy or unresponsive to an examiner they don't know. Even though children generally like the picture task formats and the pictures in CLT were well received, the test may not be performing so well in children below 2 years who have limited ability to cooperate. However, inspection of plots in Fig. 1 suggests that the floor effects on CLT tasks with young children are weak or modest, so this concern is to some extent alleviated by the data.

## 5. Conclusion

Parent-report inventories on child vocabulary have proven as valuable tools in research and in clinical practice. The usual strategy has been to use long word lists covering a large proportion of possible early words. This produces detailed snapshots of vocabulary development but may be impractical and unnecessary for estimating the developmental level. Shorter instruments of about 100 items have been developed for this purpose, and the present report shows that further shortening to 40 items is a viable option. Concurrent validity of a 40-item parent-report screening inventory is no worse than the validity of full CDI, its short versions, or some other parent report tools. This is true especially when the production and comprehension sections of the screening tool are used in a combined criterion. The results are particularly encouraging for work in languages that have no formal language screening tools available for the early ages. Developing a short parent-report instrument is a relatively low-cost way of creating such a tool.

## CRediT authorship contribution statement

**Filip Smolík:** Conceptualization, Data curation, Formal analysis, Methodology, Funding acquisition, Supervision, Writing – original draft, Writing – review & editing. **Ilona Bytešníková:** Conceptualization, Investigation, Project administration, Resources, Writing – review & editing.

## Acknowledgments

[withdrawn for blinding]

## Appendix A

List of words and approximate translations

| | | | |
|---|---|---|---|
| *Animal sounds* | | nebe | sky |
| kykyryký | cock-a-doodle-doo | strom | tree |
| *Greetings* | | *Action words* | |
| dobrou noc | good night | chtít | want |
| *Nominals* | | jít | go, walk |
| auto | car | dát | give |
| pastelka | color pencil | zlomit | break, snap |
| marmeláda | fruit jam | brečet | cry |
| rohlík | breadroll | házet | throw |
| hrášek | pea | opravovat | fix |
| oblečení | clothes | bydlet | stay, live |
| tričko | t-shirt | skončit | end, finish |
| nos | nose | *Time, quality etc.* | |
| nehet | finger-/toenail | večer | evening |
| sprcha | shower | tady | here |
| okno | window | hodně | a lot |
| šuplík | drawer | žádný | none |
| kartáček | brush | moje | my, mine |
| kamarád/ka | friend (male/female) | kde | where |
| babička | grandma | dlouhý | long |
| pes | dog | suchý | dry |
| želva | turtle | malý | small |
| liška | fox | | |

## References

Archibald, L. M. D., & Joanisse, M. F. (2009). On the sensitivity and specificity of nonword repetition and sentence recall to language and memory impairments in children. *Journal of Speech, Language and Hearing Research, 52*, 899–914. https://doi.org/10.1044/1092-4388(2009/08-0099.

Bishop, D. V. M., Snowling, M. J., Thompson, P. A., Greenhalgh, T., Adams, C., Archibald, L., et al. (2017). Phase 2 of CATALISE: A multinational and multidisciplinary Delphi consensus study of problems with language development: Terminology. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 58*(10), 1068–1080. https://doi.org/10.1111/jcpp.12721.

Singleton, N. C. (2018). Late talkers: Why the wait-and-see approach is outdated. *Pediatric Clinics, 65*(1), 13–29.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment, 6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284.

Dale, P. S., and Penfold, M. (2011). Adaptations of the MacArthur-Bates CDI into non-US English languages. Available online: https://mb-cdi.stanford.edu/documents/AdaptationsSurvey2011.pdf (last accessed March 10, 2021).

Dale, P. S., Price, T. S., Bishop, D. V. M., & Plomin, R. (2003). Outcomes of early language delay: I. Predicting persistent and transient language difficulties at 3 and 4 years. *Journal of Speech, Language, and Hearing Research, 46*(3), 544–560. https://doi.org/10.1044/1092-4388(2003/044.

*Diagnostic and statistical manual of mental disorders (vol.3)*. (2013). Washington, DC: American Psychiatric Association.
Dunn, L. M., & Dunn, D. M. (2007). *Peabody picture vocabulary test iv*. Circle Pines, MN: American Guidance Service.
Eriksson, M., Westerlund, M., & Berglund, E. (2002). A screening version of the Swedish communicative development inventories designed for use with 18-month-old children. *Journal of Speech, Language, and Hearing Research, 45*(5), 948–960. https://doi.org/10.1044/1092-4388(2002/077.
Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., et al. (1993). *The macarthur communicative development inventories: Users guide and technical manual*. Baltimore: Paul H. Brookes Publishing Co.
Fenson, Larry, Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J. (2007). *MacArthur-Bates communicative development inventories*. Baltimore: Paul H. Brookes Publishing Company.
Fenson, Larry, Pethick, S., Renda, C., Cox, J. L., Dale, P. S., & Reznick, J. S. (2000). Short-form versions of the MacArthur communicative development inventories. *Applied Psycholinguistics, 21*(1), 95–115. https://doi.org/10.1017/s0142716400001053.
Feldman, H. M., Dale, P. S., Campbell, T. F., Colborn, D. K., Kurs-Lasky, M., Rockette, H. E., & Paradise, J. L. (2005). Concurrent and predictive validity of parent reports of child language at ages 2 and 3 years. *Child development, 76*, 856–868.
Haman, E., Łuniewska, M., Hansen, P., Simonsen, H. G., Chiat, S., Bjekić, J., et al. (2017). Noun and verb knowledge in monolingual preschool children across 17 languages: Data from Cross-linguistic Lexical Tasks (LITMUS-CLT). *Clinical Linguistics & Phonetics, 31*(11-12), 818–843. https://doi.org/10.1080/02699206.2017.1308553.
Haman, E., Łuniewska, M., & Pomiechowska, B. (2015). Designing cross-linguistic lexical tasks (CLTs) for bilingual preschool children. In S. Armon-Lotem, J. de Jong, & N. Meir (Eds.), *Assessing multilingual children: Disentangling bilingualism from language impairment* (pp. 196–240). Bristol: Multilingual Matters.
Heilmann, J., Weismer, S. E., Evans, J., & Hollar, C. (2005). Utility of the MacArthur-Bates Communicative Development Inventory in identifying language abilities of late-talking and typically developing toddlers. *American Journal of Speech-Language Pathology, 14*(1), 40–51. https://doi.org/10.1044/1058-0360(2005/006.
Hughes, N., Chitsabesan, P., Bryan, K., Borschmann, R., Swain, N., Lennox, C., et al. (2017). Language impairment and comorbid vulnerabilities among young people in custody. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 58*(10), 1106–1113. https://doi.org/10.1111/jcpp.12791.
Humphreys, L. E., & Ciminero, A. R. (1979). Parent report measures of child behavior: A review. *Journal of Clinical Child Psychology, 8*(1), 56–63. https://doi.org/10.1080/15374417909532884.
Jackson-Maldonado, D., Marchman, V. A., & Fernald, L. C. H. (2013). Short-form versions of the Spanish MacArthur-Bates Communicative Development Inventories. *Applied Psycholinguistics, 34*(4), 837–868. https://doi.org/10.1017/S0142716412000045.
Libertus, M. E., Odic, D., Feigenson, L., & Halberda, J. (2015). A Developmental Vocabulary Assessment for Parents (DVAP): Validating parental report of vocabulary size in 2- to 7-year-old children. *Journal of Cognition and Development, 16*(3), 442–454. https://doi.org/10.1080/15248372.2013.835312.
Łuniewska, M., & Smolík, F. (2016). *Cross-linguistic lexical tasks: Czech version (CLT-Cz)*.
McCabe, P. C. (2005). Social and behavioral correlates of preschoolers with specific language impairment. *Psychology in the Schools, 42*(4), 373–387. https://doi.org/10.1002/pits.20064.
McKean, C., Reilly, S., Bavin, E. L., Bretherton, L., Cini, E., Conway, L., & Mensah, F. (2017). Language outcomes at 7 years: Early predictors and co-occurring difficulties. *Pediatrics, 139*(3).
Morgan, P. L., Farkas, G., Hillemeier, M. M., Hammer, C. S., & Maczuga, S. (2015). 24-Month-Old Children with larger oral vocabularies display greater academic and behavioral functioning at Kindergarten entry. *Child Development, 86*, 1351–1370.
Newton, P. E., & Shaw, S. D. (2013). Standards for talking and thinking about validity. *Psychological Methods, 18*(3), 301–319. https://doi.org/10.1037/a0032969.
Plante, E., & Vance, R. (1994). Selection of preschool language tests: A data-based approach. *Language, Speech, and Hearing Services in Schools, 25*(1), 15–24. https://doi.org/10.1044/0161-1461.2501.15.
R Core Team. (2020). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for statistical computing. Computer software available online http://www.R-project.org/.
Redmond, S. M., & Rice, M. L. (1998). The socioemotional behaviors of children with SLI: Social adaptation or social deviance? *Jounal of Speech, Language, and Hearing Research, 41*, 688–700.
Reese, E., & Read, S. (2000). Predictive validity of the New Zealand MacArthur communicative development inventory: Words and sentences. *Journal of Child Language, 27*(2), 255–266. https://doi.org/10.1017/S0305000900004098.
Rescorla, L. (1989). The Language Development Survey: A screening tool for delayed language in toddlers. *Journal of Speech and Hearing Disorders, 54*(4), 587–599. https://doi.org/10.1044/jshd.5404.587.
Rescorla, L. (2002). Language and reading outcomes to age 9 in late-talking toddlers. *Journal of Speech, Language, and Hearing Research, 45*, 360–371.
Rescorla, L. (2009). Age 17 language and reading outcomes in late-talking toddlers: Support for a dimensional perspective on language delay. *Journal of Speech, Language, and Hearing Research, 52*(1), 16–30. https://doi.org/10.1044/1092-4388(2008/07-0171.
Rescorla, L. (2011). Late talkers: Do good predictors of outcome exist? *Developmental Disabilities Research Reviews, 17*(2), 141–150. https://doi.org/10.1002/ddrr.1108.
Rescorla, Leslie, Constants, H., Bialecka-Pikul, M., Stepien-Nycz, M., & Ochal, A. (2017). Polish Vocabulary Development in 2-Year-Olds: Comparisons With English Using the Language Development Survey. *Journal of Speech Language and Hearing Research, 60*(4), 1029–1035. https://doi.org/10.1044/2016_JSLHR-L-15-0385.
Rice, M. L., Taylor, C. L., & Zubrick, S. R. (2008). Language outcomes of 7-year-old children with or without a history of late language emergence at 24 months. *Journal of Speech, Language, and Hearing Research, 51*(2), 394–407. https://doi.org/10.1044/1092-4388(2008/029.
Seymour, H. N., Roeper, T., & de Villiers, J. G. (2003). *Diagnostic evaluation of language variation–screening test (DELV–Screening test)*. San Antonio, TX: The Psychological Corporation.
Sim, F., Thompson, L., Marryat, L., Ramparsad, N., & Wilson, P. (2019). Predictive validity of preschool screening tools for language and behavioural difficulties: A PRISMA systematic review. *PloS one, 14*(2). e0211409.
Smolík, F., & Bytešníková, I. (2017). Stručný dotazník dětského slovníku: Vývoj a normy nástroje pro screening vývoje jazyka v raném věku. *Československá Psychologie, 66*, 460–473.
Smolík, F., Turková, J., Marušincová, K., & Malechová, V. (2017). *Dotazník vývoje komunikace dovyko ii*. Filozofická fakulta UK: Praha.
Sparrow, S. S., Cicchetti, D. V., & Balla, D. A. (2005). The Vineland adaptive behavior scales. *Major psychological assessment instruments* (2nd ed). Circle Pines, MN: American Guidance Service.
Squires, J., & Bricker, D. (2009). Ages and Stages Questionnaires- Third Edition. *ASQ-3 user's guide*. Baltimore,MD: Brookes.
St Clair, M. C., Pickles, A., Durkin, K., & Conti-Ramsden, G. (2011). A longitudinal study of behavioral, emotional and social difficulties in individuals with a history of specific language impairment (SLI). *Journal of Communication Disorders, 44*, 186–199. https://doi.org/10.1016/j.jcomdis.2010.09.004.
*Standards for Educational and Psychological Testing*. (2014). Washington, DC: American Educational Research Association.
Stasinopoulos, M., & Rigby, R. A. (2007). Generalized Additive Models for Location Scale and Shape (GAMLSS) in R. *Journal of Statistical Software, 23*(7), 1–46.
Thal, D., Tobias, S., & Morrison, D. (1991). Language and gesture in late talkers. *Journal of Speech Language and Hearing Research, 31*(1), 115–123. https://doi.org/10.1044/jshr.3403.604.
Tomblin, J. B., Smith, E., & Zhang, X. (1997). Epidemiology of specific language impairment: Prenatal and perinatal risk factors. *Journal of Communication Disorders, 30*, 325–344.
Uilenburg, N., Wiefferink, K., Verkerk, P., van Denderen, M., van Schie, C., & Oudesluys-Murphy, A. M. (2018). Accuracy of a screening tool for early identification of language impairment. *Journal of Speech, Language, and Hearing Research, 61*(1), 104–113. https://doi.org/10.1044/2017_JSLHR-L-16-0173.
Whitehouse, A. J. O., Watt, H. J., Line, E. A., & Bishop, D. V. M. (2009). Adult psychosocial outcomes of children with specific language impairment, pragmatic language impairment and autism. *International Journal of Language and Communication Disorders, 44*(4), 511–528. https://doi.org/10.1080/13682820802708098.
Wiig, E. H., Secord, W. A., & Semel, E. (2013). *Clinical evaluation of language fundamentals–fifth edition (CELF-5)*. Bloomington, MN: Pearson Assessment.

Williams, K. (2007). *EVT-2: Expressive vocabulary test* (second edition). Minneapolis, MN: NCS Pearson.
Zambrana, I. M., Pons, F., Eadie, P., & Ystrom, E. (2014). Trajectories of language delay from age 3 to 5: Persistence, recovery and late onset. *International Journal of Language and Communication Disorders, 49*(3), 304–316. https://doi.org/10.1111/1460-6984.12073.
Zimmerman, I., Steiner, V., & Pond, R. (1992). *Preschool language scale—Third edition*. Chicago: Psychological Corporation.