Word definition skills in elementary school children – The contribution of bilingualism, cognitive factors, and social factors

IDA ROSQVIST1, KETTY ANDERSSON1, OLOF SANDGREN1, VIVEKA LYBERG-AHLANDER1,2, KRISTINA HANSSON3,3 & BIRGITTA SAHLÉN1

1Logopedics, Phoniatrics and Audiology, Department of Clinical Sciences, Lund, Lund University, Lund, Sweden, 2Speech Language Pathology, Faculty of Arts, Psychology and Theology, Abo Akademi University, Turku, Finland

Abstract

Purpose: Vocabulary relates to overall language proficiency and is important for academic success. Word definition (WD) tasks can be used to assess vocabulary depth and definition skills. We investigate monolingual and bilingual children’s performances on a WD task, and how bilingualism, level of parental education, school characteristics (proportion of students with Swedish as second language and proportion of parents with tertiary education), CELF-4 Core Language Score, and non-verbal IQ contribute to their performance. We also evaluate the level of difficulty of the test items and the test’s internal consistency.

Method: Two hundred and eight children (mean age 7:8, range 6:8–9:0) were assessed with a 10-item WD task. Amount of information included in the definitions gave the WD score and number of words with at least partially correct information gave a Word knowledge score.

Result: The bilingual group had lower scores on both measures. In isolation bilingualism explained 15% of the variance of the WD score. With all background factors included, the only significant predictor was CELF-4 Core Language Score, uniquely explaining 24.3% of the variance. Response patterns on the WD score were similar between groups. Internal consistency was > α = 0.7 for both measurements.

Conclusion: Bilingual children performed lower than monolingual children on a WD task, but bilingualism alone cannot explain poor results.

Keywords: word definition; language development; vocabulary development; bilingualism; language skills

Introduction

Word definition (WD) tasks require both linguistic (vocabulary and grammar) and metalinguistic (using language to talk about language) skills (Marinellie & Johnson, 2004). It is also a pragmatic task where a clear, concise and complete expression can aid the listener in processing the speaker’s intention (Gutierrez-Clellen & DeCurtis, 1999). WD skills are required for preventing or repairing miscommunication in everyday conversation and are essential for academic success (Marinellie & Johnson, 2004). In this study, we examine the performance of elementary school children, age 6–9 years, on a WD task with respect to their knowledge of the words and how much information is included in their oral definitions. We also examine how the results are explained by the background factors bilingualism, level of parental education (LPE), school characteristics, general language ability, and non-verbal IQ. Finally, we evaluate the level of difficulty and the internal consistency of the WD task constructed for this study.

A well-developed vocabulary is strongly related to overall language proficiency (Schmitt, 2010) and vocabulary skills are viewed as the most important factor for academic success (Vermeer, 2001). Vocabulary knowledge is multifaceted and can be assessed in different ways (e.g. Read, 2000), focussing on breadth or depth. Vocabulary breadth refers to the total size of a person’s lexicon. Vocabulary depth refers to the knowledge a person has of individual words (Schoonen & Verhallen, 2008). Breadth is often measured with tasks requiring picture naming
or pointing to a pictured referent and the result is interpreted in a binary manner with right or wrong answers. Depth is assessed with tasks such as providing synonyms, word associations or definitions and the result is interpreted in terms of the precision of the answers, how much information is included or what developmental level the responses reflect (McGregor et al., 2012). This perspective highlights that word knowledge is a continuum with different levels (Verhallen & Schoonen, 1993).

**Development of WD skills and influencing factors**

Children’s WD skills improve gradually with respect to both content and form during the school years (McGregor et al., 2012). Young children’s definitions are often contextualised examples based on personal experience. For example, if asked to define car the response could be *We have a blue car.* Later on children might describe observable functional or perceptual features of the word, without mentioning key defining attributes, such as *A car has wheels and doors.* The use of abstract, generalised expressions with core defining attributes shows further development of WD skills, such as *A car is a vehicle that is used to transport people* (Johnson & Anglin, 1995). Nouns are often defined using the so-called Aristotelian, or formal, definition, with a superordinate term combined with distinguishing characteristics in a modifying clause (Marinellie & Johnson, 2004). This is possible due to the hierarchical organisation of nouns in the internal lexicon, with sub- and superordinate connections to other nouns. Definitions of verbs and adjectives instead use synonyms (Nippold, 2006), possibly because the lexical relations of verbs and adjectives are less structured and predictable (Marinellie & Johnson, 2004). Most research on word definitions has focussed on definitions of nouns and it is important to study children’s definition skills for other parts of speech as well (Nippold, 2006). The task in the present study therefore includes verbs and adverbs as well as nouns. The level of word definitions is also affected by concreteness. McGregor et al. (2012) found that children obtain lower results for abstract than for concrete words, regardless of whether they had typical language development, autism spectrum disorder (ASD), or developmental language disorder (DLD). In a study investigating the written definitions by college students Sadoski, Kealy, Goetz, and Paivio (1997) found that concrete nouns generated definitions of higher quality than abstract nouns. Their interpretation is that concrete language to a higher extent evokes mental imagery, which can enhance performance in language tasks. The task used in the present study includes both concrete and abstract words.

Another influencing factor is socioeconomic status (SES). Lower SES is acknowledged to have a detrimental effect on the development and wellbeing of children and adolescents (Letourneau, Duffett-Leger, Levac, Watson, & Young-Morris, 2013). More on SES and language development is found in Supplementary materials. Vocabulary measures, in particular, tend to be lower in children from low SES backgrounds, (e.g. Meir & Armon-Lotem, 2017; Spencer, Clegg, & Stackhouse, 2012) but other language skills have also been shown to be affected (e.g. Meir & Armon-Lotem, 2017). Dickinson and Snow (1987) found that children from higher SES backgrounds gave formal definitions of higher quality than children from lower SES backgrounds, based on the use of superordinate terms, relative clauses, and amount of definition features included in the definitions of 10 nouns. However, when comparing “communicative adequacy” (providing enough information for the listener to be able to identify the described word) no differences were found between the groups. The differences in definition skills were neither correlated with size of vocabulary, nor with “communicative adequacy”.

Bilingual children from language minorities often lag behind peers from the majority language group and reach lower levels of academic achievements. Factors explaining this include for example SES and level of proficiency in the language of instruction (Verhallen & Schoonen, 1998). Language exposure has also been shown to have strong effects on the rate of language development in both simultaneous and sequential bilinguals (Thordardottir, 2019). The core of the language difficulties seen in bilingual children is vocabulary acquisition, with a smaller vocabulary as well as shallower word knowledge, and less evolved oral word definitions (Verhallen & Schoonen, 1993; 1998; Vermeer, 2001). More information on WD skills in bilingual children is provided in Supplementary materials. In studies investigating the link between lexical knowledge and bilingualism, it is important be aware to what extent the study takes the heterogeneity of the participants into account with respect to exposure to languages, socio-cultural and socioeconomic background, type of education system, and language dominance and preference (Schwartz & Katzir, 2012).

**The Swedish schooling system and student composition**

In Sweden, children start grade 1 the fall semester the year they turn seven. Prior to grade 1, they attend one year of preparatory school. Sweden currently thus has 10 years of mandatory schooling.

In recent decades, Sweden has received an increasing number of immigrants. Today approximately 25% of Swedish school children were either born abroad or in Sweden with both parents born abroad. Furthermore, most children with special education needs are typically integrated within the mainstream setting and Swedish classrooms tend to be heterogeneous in terms of disabilities and educational needs.
This places high demands on teachers to accommodate the needs of all children. For more information on the Swedish schooling system and student composition, see Supplementary materials.

Research questions

Our first research question is “How does a monolingual group perform on a WD test compared to a bilingual group, and how do bilingualism, level of parental education, school characteristics (proportion of students with Swedish as second language (L2) and proportion of parents with tertiary education), general language ability, and non-verbal IQ contribute to the performance?”

Our second research question is “What is the level of difficulty of the ten test items and the internal consistency of the WD task used in this study?”

Method

Participants

Two hundred and eight children, attending mainstream education with Swedish as the primary curricular language, participated. A total of 399 children from six grade 1 and 2 classes, in six different schools from two different municipalities in southern Sweden, were invited. Of these, 229 children accepted. Fifteen children moved before completing the study, three children were excluded due to difficulties participating in the assessment and two children were newly arrived in Sweden and were judged by their teachers to have insufficient Swedish language skills to understand instructions and participate in the tasks. One child completed the assessments but was excluded from further analysis due to being an influential outlier. The data in the current study consist of baseline data from a larger study on the effect of in-service training for teachers, led by speech-language pathologists (SLPs), on children’s language development. The participants in the present study were students in the classes of the participating teachers.

The participating schools distributed written and oral information about the study together with a written consent form to the parents of all children in the participating classes. The parents were asked to provide information about language exposure and use, former or current SLP and/or special education services for the child and LPE in a written questionnaire. In accordance with ethical guidelines, informed consent was required to participate. The project was approved by the Regional Ethical Review Board in Lund (2016/8).

The mean age of the participating children was 7:8 (years:months), ranging from 6:8 to 9:0, SD = 7.1 months. Data were collected at the beginning of the school year. Including the preparatory school year, the participants in grade 1 had one year of formal schooling at data collection, and the participants in grade 2 had 2 years. The gender distribution of the sample was 112 girls (53.8%) and 96 boys (46.2%). The written parental questionnaire mentioned above was completed for 182 of the 208 participants.

Ninety-three (44.7%) of the participants were bilingual as defined by Grosjean (2008), i.e. using two or more languages in everyday life. Seventy-nine (85%) of these used two languages in their everyday life, and 14 (15%) used three languages. According to the parental questionnaire the participants used 24 different languages apart from Swedish. The bilingual group is heterogeneous in terms of at what age the child was first exposed to the languages used and amount of exposure to the languages used. Both simultaneous and sequential bilinguals were included. Information on relative use of languages was acquired for 71 of the 93 bilingual participants in the parental report. Eighteen (25%) reported that they predominantly used Swedish (>60% of the time on a daily basis), 29 (41%) predominantly used a language other than Swedish (>60% of the time), and 24 (34%) used the languages roughly equivalent amounts of time (40–60% of the time).

A 3-point rating scale (1 = mandatory schooling (equals 9 years of schooling in Sweden), 2 = high school (equals 12 years of schooling in Sweden) or 3 = university level) was used to assess LPE. In accordance with Hurks et al. (2010), the highest level of completed parental education was chosen as LPE for each child (see Supplementary materials, Table 1). For example, if one parent had completed mandatory schooling (=1 point), and the other parent had completed high school (=2 points), high school (=2 points) was chosen as LPE for the child. An independent samples t-test with bootstrapping for the LPE in the monolingual (M = 2.72) and bilingual

| Whole group (n = 208) | Monolingual group (n = 115) | Bilingual group (n = 93) |
|-----------------------|-----------------------------|--------------------------|
| **WKn score**         |                             |                          |
| Range                 | 1–10                        | 3–10                     | 1–10                     |
| Mean                  | 7.02                        | 7.85***                  | 5.99***                  |
| SD                    | 2.213                       | 1.529                    | 2.487                    |
| **WD score**          |                             |                          |
| Range                 | 2–21                        | 4–21                     | 2–18                     |
| Mean                  | 9.86                        | 11.00***                 | 8.13***                  |
| SD                    | 4.023                       | 3.3308                   | 4.171                    |

Difference between monolingual and bilingual group: *** p = < 0.001.
Table II. Unstandardised beta (\(B\)), the standard error for the unstandardised beta (\(SE\) \(B\)), the standardised beta (\(b\)), and the semi-partial correlation squared (\(sr^2\)) from the hierarchical linear regression model explaining WD score with the independent variables Bilingualism, LPE, School characteristics, CELF-4 CLS, and RCPM

| Measure                          | Model 1 | Model 2 | Model 3 | Model 4 | Model 5 |
|---------------------------------|---------|---------|---------|---------|---------|
| Bilingualism                    | -1.123  | 0.576   | -0.387  | 0.150   | -2.592  |
| Level of parental education     | 0.150   | 0.572   | -0.321  | 0.093   | -1.857  |
| School characteristic           | 1.264   | 0.431   | 0.213   | 0.041   | 0.769   |
| CELF-4 CLS                      | 0.230   | 0.142   | 0.014   | 0.796   | 0.341   |
| RCPM score                      | -0.094  | -0.331  | 0.036   | 0.769   | 0.468   |

\(R^2_{model} = 0.150\), \(R^2_{adj} = 0.145\)

---

from the hierarchical linear regression model explaining WD score with the independent variables Bilingualism, LPE, School characteristics, CELF-4 CLS, and RCPM.

---

As previously mentioned, most children with special education needs are integrated within the mainstream setting in Sweden and no participants were excluded from participation in this study due to special needs. Fifty-two children had previous or ongoing SLP and/or special education services according to the parental questionnaire. Five of the 25 six-year-olds, 29 of the 100 seven-year-olds, 17 of the 81 eight-years-olds, and 1 of the 2 nine-year-olds had previous or ongoing SLP and/or special education services. Reasons for receiving SLP and/or special educations services were mainly speech, language or reading difficulties, attention deficit hyperactivity disorder (ADHD), autism, hearing or visual impairment, or difficulties with concentration or impulse control. See Supplementary materials, Table 1 for an overview of demographic information on the participants.

To assess the language ability of the children four subtests of the Swedish adaptation of the Clinical Evaluation of Language Fundamentals – Fourth Edition (CELF-4) (Semel, Wiig, & Secord, 2013) were administered and scored according to the test manual. The four subtests Concepts and Following Directions, Word Structure, Recalling Sentences, and Formulated Sentences together give the Core Language score (CLS). One participant did not complete this test. Standard scores ranged from 40 to 123, with a mean of 78.7, SD 23.6 (n=207). To assess the non-verbal IQ of the children the Raven’s Coloured Progressive Matrices Test (RCPM) (Raven, 2008) was administered and scored according to the test manual. The RCPM gives an estimation of the non-verbal component of Spearman’s g-factor and consists of 36 items with increasing difficulty where the child is required to select the missing piece among six elements to complete a pattern. Standard scores ranged from < 60 to 135 with a mean of 93.88, SD = 17.793 (n = 208). Both instruments have an expected mean score of 100 with a SD of 15. The lower-than-expected results for the CELF-4 CLS for the sample are discussed in Andersson et al. (2019). The norms for the test are based on a smaller sample than in the current study. In addition, the normative sample consists of only monolingual participants and a smaller proportion of children from lower SES backgrounds. When using the test with participants from low SES backgrounds, with Swedish as L2, and/or attending schools with many other students with the same circumstances, the risk of low results on the CELF-4 CLS is increased. The scores of the sample for the RCPM are also somewhat lower than expected, especially for the bilingual group. However, when RCPM is used...
on children from a minority group, if the test language is their L2, and/or if they have a language disorder the results compared to normative information must be interpreted with caution (Raven, 2008).

The two participating municipalities present different student bases in terms of LPE and the proportion of children with foreign background (i.e. the child has two parents born in a country other than Sweden or the child is born outside of Sweden). The schools in municipality A have a lower proportion of parents with tertiary education and a higher proportion of students with foreign background compared to the national average (Swedish National Agency for Education, 2019). In Municipality B the situation is reversed (see Supplementary materials, Table 2). An index of school characteristics (possible range 2–12) was calculated based on rank scores (from the lowest (1) to the highest (6)) of the proportion of students with Swedish as first language and the proportion of parents with tertiary education in the participating schools based on publicly available data (Swedish National Agency for Education, 2019), in accordance with Andersson et al. (2019).

**Procedure**

Five certified SLPs administered all tests individually at the schools in a separate room during school-time. All tasks had written instructions to ensure procedural fidelity. The children were assessed with Raven’s CPM, CELF-4 CLS subtests and a WD task in a fixed order for all participants. The assessment took approximately 45 minutes.

**Word definition task**

To ensure that the children understood the task they first heard an example of a definition of the word break/recess (as in “taking a short break during school hours” Swedish: rast) and were then asked to define the word draw (Swedish: rita) as a practice item. They later gave oral definitions of ten stimulus words in response to the question “What does ‘XXX’ mean?” The approved prompt to elicit more information from the child was ‘Can you say something else?’ The 10 stimulus words in the WD task were administered in the same order for all the participating children and the task took approximately 5 minutes to complete. The mobile application RecUp, Irradiated Software, LLC was used for audio recordings of the responses and, when possible, responses were also transcribed orthographically in real-time.

A list of words used in the teaching context in the current grades was composed by the authors. It was based on publicly available video recordings, produced by the Swedish National Agency for Education, showing everyday teaching situations in Swedish schools in the current grades as well as on the authors’ experiences from the school setting. From that word list, 10 cross-curricular words were selected. The words were considered by the authors, based on their extensive experience from the school setting, to be frequently used in the teaching situation and judged to be of medium difficulty for children in the current grades to avoid floor and ceiling effects. The WD task consisted of the following 10 stimulus words: jump (Swedish: hoppa), play (Swedish: spela), headline (Swedish: rubrik), choose (Swedish: välja), task (Swedish: uppgift), tell (Swedish: berätta), together (Swedish: tillsammans), ponder (Swedish: fundera), difference (Swedish: skillnad), and adult (Swedish: vuxen). The task was piloted on two children recruited outside of the project to ensure it was feasible for children in the same age span as the participants in the current study.

The responses on the WD task were scored by the first author based on transcriptions of the children’s responses entered into a Microsoft Excel spreadsheet. Each word was scored for two aspects: “Word knowledge” (WKn) and depth of Definition. Scoring of Definition was based on amount of information included, in accordance with McGregor et al. (2012). Inaccurate definitions received 0 points. Inaccuracies were defined in accordance with Storck and Looft (1973): misinterpretations (Difference = It’s the same thing, isn’t it? Like this chair and this chair is the same thing), incorrect definition (Headline = A store), clang association (Difference = Divorced. Swedish: Difference = Skillnad, Divorced = Skilda), repetition without explanation (Tell = Tell) or omission (no response). Definitions that bore some meaningful relationship to the target word but did not define it received 1 point. Definitions that were minimal but conventional received 2 points. Conventional definitions with accurate and well-elaborated amount of information received 3 points. Specific criteria for what type of information the definitions should include for each scoring level, as well as examples for each scoring level, were set up for each test item. Criteria and examples were based on definitions in two Swedish dictionaries (Ernby, Gellerstam, Malmgren, & Axelsson, 2001; Svenska Akademin [Swedish Academy], 2009), and a preliminary scoring system was created by two final year SLP students (Erlandsson & Yhlen, 2019) and was further developed by the first and the fifth author. For example, for the word adult the criteria for definitions that received 1 point included either: A. (A person that is) big or B. examples of adults (Teachers and parents are adults). For 2 points, the definitions included either: A. (A person that is) over 18, B. Older than a child, or C. (A person that is) big together with examples of what it means to be an adult (e.g. rights and responsibilities). Definitions receiving 3 points included: (A person that is) over 18 together with examples of what it means to be an adult. For the WKn measure the children received 1 if the WD score was at least 1 point, and 0 if the WD score was 0. Uncertainties in scoring were agreed upon through discussion. The possible range for
WKn was 0–10, and the possible range for WD was 0–30.

Task instructions, models, prompt, and scoring guidelines with examples for the WD task appear in Appendix A.

**Inter-rater reliability of WD**

A random selection of 10% of the transcripts of the word definitions was scored by the fifth author to calculate inter-rater reliability. Inter-rater reliability for scoring was calculated using Cohen’s Kappa ($\kappa = 0.779$ for WKn, $p = <0.001$ and 0.558 for WD, $p = <0.001$ which are labelled ‘Substantial’ and ‘Moderate’, respectively by Landis and Koch (1977). Moderate strength of agreement for WD would in many cases be viewed as too low. However, it should be noted that Cohen’s kappa statistics are based on categories and are too stringent for numeric data such as the scale used to measure the quality of word definitions (Kurland & Snow, 1997).

**Data analysis**

IBM SPSS Statistics 25 (IBM SPSS Statistics, Armonk, NY) for Windows was used to perform all statistical analyses. Prior to performing any statistical analysis, all assumptions relevant to the calculations were checked.

An independent samples $t$-test was used to explore any significant differences between the monolingual and the bilingual group on the WD score. The assumptions of independent observation and normality were met, but the assumption of homogeneity was violated, hence, values for “equal variances not assumed” are reported. For the WKn score, both the assumptions of normality and homogeneity were violated. Therefore, an independent samples $t$-test with bootstrapping, using 2000 samples, was used to explore any significant differences between the monolingual and the bilingual group on the WKn score. Values for “equal variances not assumed” are reported.

A hierarchical regression was conducted to investigate the contribution to the WD score of Bilingualism, LPE, School characteristics, CELF-4 CLS and RCPM score. The independent variables Sex ($r = 0.048$, $p = 0.247$), Grade ($r = 0.219$, $p = < 0.01$), and current or previous SLP and/or special education services ($r = -0.174$, $p = < 0.01$) were included in a preliminary analysis but were excluded from the final analysis due to too weak relationship with the outcome variable. The sample size of 208 was presumed sufficient with the five independent variables included in the analysis. The assumptions of singularity, multicollinearity, normality, linearity, homoscedasticity, and independence of residuals were also met. Initial screening of the data revealed one influential outlier who had a higher result than the rest of the sample. This participant was excluded from further analysis. A five-stage hierarchical regression was conducted with WD score as the dependent variable. At stage 1 of the regression Bilingualism was entered. At stage 2, the LPE variable was entered, at stage 3 School characteristics, at stage 4 CELF-4 CLS, and, at stage 5, RCPM score was entered.

The alpha level was set at 0.05 for dependent variables. To explore the level of difficulty of the ten test items the distribution of scores (0/1/2/3) for the whole sample, as well as for the monolingual and bilingual group separately was analysed. Cronbach’s alpha coefficient was used to investigate the internal consistency of the test.

**Result**

**Group comparison and contribution of background factors**

The monolingual group outperformed the bilingual group on the WKn score as well as on the WD score (see Table I). An independent samples $t$-test with bootstrapping for the WKn score revealed that this difference was significant $t(146) = 6.47$, $p = <0.001$, and an independent samples $t$-test for the WD score revealed that this difference was also significant, $t(173) = 5.88$, $p = <0.001$.

To further explore how the background factors Bilingualism, LPE, School characteristics, CELF-4 CLS, and RCPM score contribute to the WD score a five-stage hierarchical linear regression was conducted.

Correlation calculations between the dependent variable (WD score) and the independent variables (Bilingualism, LPE, School characteristics, CELF-4 CLS, and RCPM score) revealed significant correlations ($p < 0.001$), indicating associations between all variables, ranging from 0.265 to 0.731.

The hierarchical multiple regression conducted to investigate the unique and shared variance explained by the independent variables (see Table II) revealed that at stage 1, Bilingualism added as a single predictor contributed significantly to the regression model, $F(1,179) = 31.52$, $p = < 0.001$ and accounted for 15% of the variation in WD scores. Adding LPE at stage 2 explained an additional 4.1% of the variation. The unique contribution of Bilingualism now decreased to 9.3% and shared variance was 5.7%. The change in $R^2$ was significant, $F(2,178) = 20.971$, $p = < 0.01$. Adding School characteristics at stage 3 explained an additional 2.4% of the variation in WD scores and the change in $R^2$ was significant, $F(3,177) = 16.115$, $p = < 0.05$. However, now LPE was no longer a significant predictor. The unique contribution of Bilingualism decreased to 3.6%. The shared variance was 14.3%. The addition of CELF-4 CLS at stage 4 of the regression model explained an additional 33.1% of the variation in WD scores and this change in $R^2$ was also
significant, $F(4,176) = 52.92, p = < 0.001$. In stage 4, the only significant predictor of WD score was CELF-4 CLS and this predictor uniquely explained 33.2% of the variation. The shared variance was 20.28%. Together the four independent variables accounted for 54.6% of the variance in WD score. Finally, the addition of RCPM score at the fifth, and final, stage of the regression model explained an additional 0.2% of the variation in WD scores but this change in $R^2$ was not significant, $F(5,175) = 42.4, p = 0.406$. At this stage, the adjusted $R^2$ decreased with 0.001, also indicating that adding the variable RCPM score does not substantially improve the model. However, the model as a whole was still significant, $p = < 0.001$. With all five independent variables included in stage 5 the only significant predictor of WD score was still CELF-4 CLS. Together the five independent variables accounted for 54.8% of the variance and the shared variance was 29.29%. For an overview of unique, shared and unexplained variance in Definition score based on the five-stage hierarchical linear regression model, see Table 3 in Supplementary materials.

**Level of difficulty of the test items and the internal consistency of the WD test**

The test administrators reported that following the preparatory instruction, providing an answer for the test items did not cause difficulty for most participants. The item with the highest occurrence of score 0 answers was headline (86.9%). The item with the lowest occurrence of score 0 answers was together (9.1%). For the WKn measure, the participants received 1 point when the WD score was at least 1 point. Hence, 0 points for WD corresponds to 0 points for Word knowledge.

A comparison between the patterns for the WD score in the monolingual and bilingual group shows an overall similar pattern (see Figure 1, which shows the distribution of different WD scores for each word). However, the bilingual group has a higher occurrence of answers rendering 0 points.

One test item, task, shows a different pattern than the other items with a higher proportion of answers rendering 0 points than expected for the bilingual group.

For the WKn score Cronbach’s alpha $\alpha$ was 0.737. The value would have been slightly higher if the test item headline (Cronbach’s alpha $\alpha = 0.741$) or tell (Cronbach’s alpha $\alpha = 0.744$) had been removed, but this was not done. Cronbach’s alpha $\alpha$ was 0.739 for the WD score. Internal consistency for the 10 test items is above 70% on both measures. Removing test items would not substantially increase internal consistency.

**Discussion**

We examined the performances of 208 6–9-year-old children on a WD task. We compared how the monolingual and bilingual participants performed on two measures, WKn and WD. We also examined to what extent the background factors Bilingualism, LPE, School characteristics, CELF-4 CLS, and RCPM score explained the participants’ performance on the WD score. Furthermore, we analysed the level of difficulty of the 10 test items and the internal consistency of the WKn and WD measures.

The bilingual group performed lower on both the WKn and WD score. This is in line with previous studies showing that bilingualism is associated with a smaller vocabulary, shallower word knowledge, and lower quality of word definitions in the L2 (Verhallen & Schoonen, 1993). Higher SES (Dickinson & Snow, 1987) and more school experience (Gini, Benelli, & Belacchi, 2004) have also been shown to be associated with higher results on WD tasks. When adding LPE and School characteristics to the regression model, the unique contribution of bilingualism is reduced from 15% to 3.6% and shared variance increases to 14.3%. However, with all factors added, the only significant predictor of WD skills was CELF-4 CLS, which uniquely explained a quarter of the variance. The CELF-4 CLS is itself explained by

![Figure 1. Mean proportion of WD scores, range 0–3, for the ten test items for the monolingual group of participants, $n = 115$ (left), and the bilingual group of participants, $n = 93$ (right).](image-url)
high levels of shared variance among bilingualism, LPE, and school characteristics (Andersson et al., 2019). Thus, the CELF-4 CLS reflects not only language skills but also the bilingualism and SES factors. Furthermore, although the CELF-4 CLS does not include any specific lexical measure, vocabulary skills are strongly associated with various measures of language skills (Schmitt, 2010). The task of providing oral word definitions places demands on language, requiring both vocabulary breadth and depth as well as knowledge of, and ability to express definitions in a clear and conventional way. The strength of CELF-4 CLS as a predictor of WD skills is therefore not surprising. The RCPM score does not contribute significantly to the model, but mainly decreases the unique contribution of the CELF-4 CLS and increases the shared variance, indicating that the RCPM and CELF-4 CLS are associated, which is also seen in the correlation \( r = 0.57 \) between the two predictors. The CELF-4 CLS can thus be viewed as a complex measure, which incorporates different cognitive, linguistic as well as contextual factors. Bilingualism and SES contribute to the WD score through CELF-4 CLS.

The proportion of test scores equal to score 0 (no correct information) and score 1 (partially correct information but not defined in a conventional way) is large (most often 60% or more). Many of the definitions that were awarded 1 point were contextualised examples based on personal experience, as typically seen in young children, such as Task: Task is almost like, if I have a paper and the teacher says “you shall write here" then that’s your task, to write. There were also examples of definitions describing observable functional or perceptual features of the word, but without mentioning key defining attributes, typically seen later in development, such as Adult: Adult, that is that someone is bigger than you, sort of. Hence, many of the participants have yet to master the use of abstract, generalised expressions with core defining attributes in their further development of definition skills (Johnson & Anglin, 1995).

Teachers may be surprised when they discover how little knowledge some children have of words they seemingly know (Schoonen & Verhallen, 2008). Children might be able to use a word in a superficial manner in some contexts, but without deep knowledge of how the word is used in other contexts. Some of the participants gave examples indicating that their understanding of the word adult is associated with school staff, a response such as That you’re a teacher. That you work with children, without an understanding of what it means to be an adult in terms of rights and responsibilities. Teachers should be aware that lexical deficits can be evident not just in terms of vocabulary size but also in vocabulary depth.

School experience also influences children’s definition skills (Gini et al., 2004). Classroom interactions offer an opportunity to practice formal definitions (Gutierrez-Cleflen & DeCurtis, 1999) and learning to combine adequate content and conventional form is influenced by formal instruction and schooling (Benelli, Belacchi, Gini, & Lucangeli, 2006). Teachers can support the development of WD skills by providing explicit examples of conventional language use and expand and elaborate children’s own spontaneous language production, but also by working on the depth of word knowledge. When teachers work on developing new vocabulary during activities in school, formal definitions are emphasised (Gini et al., 2004) and children are often asked to provide definitions for words they read. This requires that they think about language and demonstrate their knowledge of the words while using specific features in combination with the appropriate semantic category for the defined word (Gutierrez-Cleflen & DeCurtis, 1999). Preliminary analysis of the data in the present study showed that the relationship between the amount of school experience (measured by grade) and WD skill was too weak to be included in the model for explaining variance in performance. The difference in school experience between groups was only one year. With a larger time span, for example 2 years of school experience instead of one, it is possible that performance would be explained to a larger extent by the amount of school experience.

A possible explanation for the generally low WD scores could be that the test did not focus on concrete nouns used in everyday language. WD tasks for children often consist of concrete nouns such as hat, dog, or umbrella. Words of this type, which are common in everyday language, would be considered Tier 1 words in the three-tiered model for vocabulary developed by Beck, McKeown, and Kucan (2013). Tier 3 words are domain or content specific, such as molecule, photosynthesis, or filibuster, which are often explicitly defined by teachers or in textbooks. For the task in the present study, we chose cross-curricular words from different areas, such as difference, task and headline. They would be defined as Tier 2 words by Beck et al. (2013). Children may seldom encounter Tier 2 words before starting school and such words are seldom explicitly defined by teachers. Still, Tier 2 words are essential in the school setting and should be targeted for explicit instruction (Beck et al., 2013). The study did include one concrete noun, headline, which had the largest occurrence of 0 points responses. Thus, a concrete noun does not automatically generate definitions of higher quality than abstract words as also shown by Sadoski et al. (1997). The incorrect definitions of the word headline were misinterpretations, clang associations, or repetition without explanations indicating that many children did not understand the meaning of the word. The word with the lowest occurrence of 0 points was together, which is an abstract adverb. The word with the highest occurrence of 3 points was ponder (Swedish: fundera) which is an abstract verb. The many high-quality
definitions of the word *ponder* may be due to the fact that *ponder* has a seemingly easily accessible synonym, *think* (Swedish: *tänka*) and verbs are often defined using synonym verbs (Nippold, 2006). This increased the amount of conventional and elaborated definitions. Hence, not only part of speech and concreteness affect the quality of definitions but also factors such as exposure to the word and easily accessible synonyms.

The internal consistency of the task is above Cronbach’s α = 0.7 for both WKn and WD scores and removing any test items would not substantially increase internal consistency. The 10 test items can be regarded as representing one factor.

The five independent variables in our final model (Table 3) left 45.2% of the variance of the WD score unexplained. In addition to placing demands on language, WD skills are also related to cognitive development, for example, metacognition and executive functions. Adding metacognitive and executive function measurements in studies of WD skills could possibly contribute to further explaining some of the variance in performances. Adding a language measure that is less affected by background factors could give a clearer picture of how language skills *per se* contribute to the variance in WD scores. A measure that might be less affected by SES is nonword repetition (Meir & Armon-Lotem, 2017). Such a task could give a more appropriate reflection of the child’s potential for language learning. Nonword repetition, preferably using quasi-universal nonwords (Boerma et al., 2015), could be included in future studies of how background factors explain performance on WD tasks and to disentangle the influence of contextual factors from the child’s internal capacity.

**Limitations and future research**

One limitation of the study is the heterogeneity in the bilingual group in terms of language background and degree of exposure to Swedish. Access to detailed information on previous language exposure would have enabled a greater understanding of differences in performance within this group. The formulation of test items is another factor that may have influenced the results. In many cases, the children did not understand the word they were asked to define, as reflected in the WKn score. A task with test items of which all participants show at least basic knowledge could provide a clearer image of their WD skills. Furthermore, the scoring method we used mainly focussed on content. Analysis of form, as well as pragmatic aspects would also be interesting and relevant to relate to different background factors. The sample mirrors linguistic and culturally diverse classrooms that we see in many Swedish schools, with children with and without special education needs, and a large distribution in terms of skills and knowledge. The participants’ performance on CELF-4 CLS and Raven’s CPM is likely to reflect the children’s language and special education needs to some extent. Future studies could investigate how different subgroups with more homogenous samples perform on the WD task. It would also be interesting to add more language-independent measures to give a clearer picture of how SES, bilingualism and language skills interact.

**Conclusion and clinical implications**

Bilingual children with Swedish as L2 perform lower on a WD task than monolingual children with Swedish as mother tongue at the group level. However, the strongest predictor for WD skills was CELF-4 CLS, which, in turn, to a high extent is explained by shared variance among bilingualism, LPE and school characteristics. It is important that SLPs are aware that test results should be considered in conjunction with the child’s background since contextual factors contribute to a risk for language learning difficulties. At-risk children should be offered targeted intervention to support their language development.

Children, in particular children with Swedish as L2 in the age range 6–9, may have superficial knowledge of words commonly used in the teaching situation, and need support to deepen their vocabulary knowledge. Teachers play a crucial role in explicitly teaching word meanings and supporting children’s WD skills.

WD tasks are useful to assess depth of vocabulary knowledge and should be included in assessments of vocabulary skills, carried out by for example school-based SLPs. Type of words chosen may need to be adapted to the age of the children. Factors such as exposure to the word and the existence of easily accessible synonyms together contribute to how difficult a word is to define.

**Acknowledgements**

The authors thank the children, their parents and staff from the schools participating in the study for their co-operation, Sara Gustavsson for her help in data collection and Julia Erlandsson and Tor Yhlen for their help with transcriptions, initial analysis and preliminary scoring system.

**Declaration of interest**

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

**Social media handles**

Ida Rosqvist: https://twitter.com/logopida @logopida

**Supplementary material**

Supplemental data for this article can be accessed at https://doi.org/10.1080/17549507.2021.2000027.
Funding
This work was supported by The Swedish Institute for Educational Research under Grant 2016/46.

ORCID
Ida Rosqvist http://orcid.org/0000-0002-8496-2084
Ketty Anderson http://orcid.org/0000-0003-2680-9660
Olof Sandgren http://orcid.org/0000-0002-5098-6096
Viveka Lyberg-Åhländer http://orcid.org/0000-0002-7178-1604
Kristina Hansson http://orcid.org/0000-0002-0467-5190
Birgitta Sahlén http://orcid.org/0000-0002-8468-0546

References
Andersson, K., Hansson, K., Rosqvist, I., Lyberg Åhländer, V., Sahlén, B., & Sandgren, O. (2019). The contribution of bilingualism, parental education, and school characteristics to performance on the Clinical Evaluation of Language Fundamentals: Fourth Edition, Swedish. Frontiers in Psychology, 10, 1586. doi:10.3389/fpsyg.2019.01586
Beck, I.L., McKeown, M.G., & Kucan, L. (2013). Bringing words to life: robust vocabulary instruction. New York, NY: The Guilford Press.
Benelli, B., Belacchi, C., Gini, G., & Lucangeli, D. (2006). ‘To define means to say what you know about things’: The development of definitional skills as metalinguistic acquisition. Journal of Child Language, 33, 71–97.
Boerma, T., Chiat, S., Leseman, P., Timmermeister, M., Wijnen, F., & Blom, E. (2015). A quasi-universal nonword repetition task as a diagnostic tool for bilingual children learning Dutch as a second language. Journal of Speech, Language, and Hearing Research, 58, 1747–1760. http://dx.doi.org/10.1044/2015_JSLHR-L-15-0058
Dickinson, D.K., & Snow, C.E. (1987). Interrelationships among prereading and oral language skills in kindergartners from two social classes. Early Childhood Research Quarterly, 2, 1–25. doi:10.1016/0885-2006(87)90010-X
Erlandsson, J., & Yhlen, T. (2019). “Skilnaden betyder att det kan vara en skillnad mellan en person och en annan person” En jämförelse mellan en och flerspråkiga skolbarns definitioner av shokoladetäcke ord [‘Difference means it can be a difference between a person and another person’ A comparison between mono- and bilingual school children’s definitions of school-related words]. Student Paper. Lund: Lund University.
Ermby, B., Gellerstam, M., Malmgren, S.-G., & Axelsson, P. (2001). Norstedts första svenska ordbok: 18.000 ord och fraser [Norstedt’s first Swedish wordbook: 18.000 words and phrases]. Stockholm: Norstedts ordbok.
Gini, G., Benelli, B., & Belacchi, C. (2004). Children’s definitional skills and their relations with metalinguistic awareness and school achievement. European Journal of School Psychology, 2, 239–267.
Grosjean, F. (2008). Studying bilinguals. Oxford: Oxford University Press.
Gutierrez-Clellen, V.F., & DeCurtis, L. (1999). Word definition skills in Spanish-speaking children with language impairment. Communication Disorders Quarterly, 21, 23–31. doi:10.1177/152574019902100104
Hurks, P.P.M., Jolles, J., Schrans, D., Meij, C., Wassenberg, R., & Feron, F.J.M. (2010). Developmental changes in semantic verbal fluency: Analyses of word productivity as a function of time, clustering, and switching. Child Neuropsychology, 16, 366–387. doi:10.1080/09297041003671184
Johnson, C.J., & Anglin, J.M. (1995). Qualitative developments in the content and form of children’s definitions. Journal of Speech and Hearing Research, 38, 612–629.
Kurland, B.F., & Snow, C.E. (1997). Longitudinal measurement of growth in definitional skill. Journal of Child Language, 24, 603–625.
Landis, R.J., & Koch, G.G. (1977). The measurement of observer agreement for categorical data. Biometrics, 33, 159–174. doi:10.2307/2529310
Letourneau, N.L., FIFF, L., Levac, L., Watson, B., & Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. Journal of Emotional & Behavioral Disorders, 21, 211–224. doi:10.1177/1063426611421007
Marinellie, S.A., & Johnson, C.J. (2004). Nouns and verbs: A comparison of definitional style. Journal of Psycholinguistic Research, 33, 217–235.
McGregor, K., Berns, A., Owen, A., Michels, S., Duff, D., Bahnsen, A., & Lloyd, M. (2012). Associations between syntax and the lexicon among children with or without ASD and language impairment. Journal of Autism & Developmental Disorders, 42, 35–47. doi:10.1007/s10803-011-1210-4
Meir, N., & Armon-Lotem, S. (2017). Independent and combined effects of socioeconomic status (SES) and bilingualism on Children’s Vocabulary and Verbal Short-Term Memory. Frontiers in Psychology, 8, 1442. doi:10.3389/fpsyg.2017.01442
Nippold, M.A. (2006). Later language development: School-age children, adolescents, and young adults. Austin, TX: Pro-Ed.
Raven, J. (2008). Raven’s Educational: Coloured Progressive Matrices and Crichton Vocabulary Scale. Manual. London: Pearson Inc.
Read, J. (2000). Assessing vocabulary. Cambridge: Cambridge University Press.
Sadoski, M., Kealy, W.A., Goetz, E.T., & Piavio, A. (1997). Concreteness and imagery effects in the written composition of definitions. Journal of Educational Psychology, 89, 518–526. doi:10.1037/0022-0663.89.3.518
Schmitt, N. (2010). Researching vocabulary: A vocabulary research manual. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
Schoonen, R., & Verhallen, M. (2008). The assessment of deep word knowledge in young first and second language learners. Language Testing, 25, 211–236. doi:10.1177/0265532207086782
Schwartz, M., & Katzir, T. (2012). Depth of lexical knowledge among bilingual children: The impact of schooling. Reading and Writing: An Interdisciplinary Journal, 25, 1947–1971. doi:10.1007/s11145-011-9308-9
Semel, E.M., Wiig, E.H., & Secord, W.A. (2013). Clinical evaluation of language fundamentals (4th ed.). Swedish Standardised Edition. Bromma: Pearson Assessment.
Spencer, S., Clegg, J., & Stackhouse, J. (2012). Language and disadvantage: A comparison of the language abilities of adolescents from two different socioeconomic areas. International Journal of Language & Communication Disorders, 47, 274–284. doi:10.1111/j.1460-6984.2011.00104.x
Storck, P.A., & Looft, W. (1973). Development of definitional skill. Journal of Language & Communication Disorders, 21, 274–284. doi:10.3389/fpsyg.2017.00340
Svenska Akademien [Swedish Academy]. (2009). Svensk ordbok: utgiven av Svenska Akademien [Swedish wordbook: published by the Swedish Academy]. Stockholm: Norstedt.
Swedish National Agency for Education. (2019). Information on pupil characteristics. https://www.skolverket.se/skolutveckling/statistik. Accessed on April 3, 2019.
Thordardottir, E. (2019). Amount trumps timing in bilingual children, adolescents, and young adults. Austin, TX: Pro-Ed.
Young-Morris, C. (2013). Socioeconomic status and child development: A meta-analysis. Journal of Emotional & Behavioral Disorders, 21, 211–224. doi:10.1177/1063426611421007
Verhallen, M., & Schoonen, R. (1993). Lexical knowledge of monolingual and bilingual children. Applied Linguistics, 14, 344–363. doi:10.1093/applin/14.4.344
Verhallen, M., & Schoonen, R. (1998). Lexical knowledge in L1 and L2 of third and fifth graders. *Applied Linguistics, 19*, 452–470. doi:10.1093/applin/19.4.452

Vermeer, A. (2001). Breadth and depth of vocabulary in relation to L1/L2 acquisition and frequency of input. *Applied Psycholinguistics, 22*, 217–234. doi:10.1017/S0142716401002041

**Appendix A**

**Task instructions, models and prompt for the WD task**

If someone asked me: ‘What does *break* mean?’, then I could say ‘It’s when you’re outside playing, you’re free from class and you take a pause’. Now I will say a word to you! What does *draw* mean?

Approved prompt to elicit more information: ‘Can you say something else?’

**Scoring guidelines with examples for the WD task, in accordance with McGregor et al. (2012)**

0 – Inaccurate definition.

**Headline:** If you’re singing there are some people watching.

**Difference:** It’s the same thing, isn’t it? Like this chair and this chair is the same thing.

1 – Information is at least partially correct but definition is not conventional.

Choose: You can choose something you want and choose clothes and choose teddy bears.

Tell: When you’re going to tell a secret. The teacher can tell you what to do as homework. When the teacher says ‘today I will tell a story’.

2 – Information is correct and conventional but minimal.

Task: You have a task you have to do. You have to do something someone told you to do.

Together: You are, you are two people instead of just one, or you’re more than two.

3 – Information is correct, conventional, and well elaborated.

**Adult:** When you’re over 18 you’re an adult. Then you’re an adult. Over 18. You start working, take responsibility over yourself and those kinds of things. Move away from home.

**Ponder:** Ponder can mean that if, if I’m working in my math book and I encounter a difficult assignment, then I have to ponder on it. Sort of think about it.