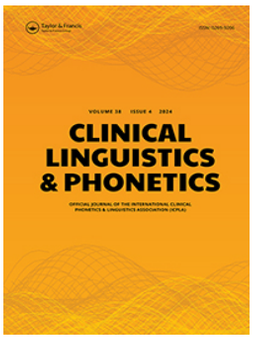

Clinical Linguistics & Phonetics

ISSN: (Print) (Online) Journal homepage: www.informahealthcare.com/journals/iclp20

Exploring grammatical development in children aged 2;6 to 7: a novel approach using elicited production

Christina Kauschke, Kim Lawatsch, Anne Tenhagen & Tobias Dörfler

To cite this article: Christina Kauschke, Kim Lawatsch, Anne Tenhagen & Tobias Dörfler (25 Mar 2024): Exploring grammatical development in children aged 2;6 to 7: a novel approach using elicited production, Clinical Linguistics & Phonetics, DOI: 10.1080/02699206.2024.2328200

To link to this article: https://doi.org/10.1080/02699206.2024.2328200

Published online: 25 Mar 2024.

CLINICAL LINGUISTICS & PHONETICS
https://doi.org/10.1080/02699206.2024.2328200

Taylor & Francis
Taylor & Francis Group

Exploring grammatical development in children aged 2;6 to 7: a novel approach using elicited production

Christina Kauschke[a], Kim Lawatsch[a,b], Anne Tenhagen[c], and Tobias Dörfler[d]

[a]Clinical Linguistics, Department of German Linguistics, University of Marburg, Marburg, Germany; [b]Clinic for Cognitive Neurology, University Hospital & Faculty of Medicine, University of Leipzig, Leipzig, Germany; [c]Faculty of Applied Health Sciences, European University of Applied Sciences, Cologne, Germany; [d]Psychological Assessment, Institute of Psychology, University of Education Heidelberg, Heidelberg, Germany

ABSTRACT

The assessment of children's grammatical skills is a crucial component of diagnosing language disorders. Elicited production is a commonly used method for obtaining data on a child's productive language abilities. We introduce a new instrument developed as part of the third edition of a standardised test battery for German. This instrument utilises elicited production, wherein participants describe coloured pictures depicting everyday situations, in order to generate four test scores: mean length of utterances, completeness of utterances, and two grammar scores comprising relevant target structures. The construction of the grammar scores was inspired by the Index of Productive Syntax (IPSyn), modified for German and computerised. The detailed results provide a comprehensive profile of a child's syntactic and morphological strengths and weaknesses. Analysis of data collected from 348 monolingual German children who formed part of the norming sample, aged between 2;6 and 6;11 years of age, revealed age-related changes in these scores. Additionally, the age range was determined for both grammatical milestones and 'red flags', which may indicate potential problems in language development. In conclusion, the newly developed, time-efficient instrument allows for a detailed assessment of grammatical skills, identification of potential intervention targets, and facilitates various research objectives.

ARTICLE HISTORY
Received 14 October 2023
Revised 23 February 2024
Accepted 2 March 2024

KEYWORDS
Assessment; elicited production; IPSyn; grammar; DLD; German

Introduction

In both language acquisition research and clinical practice, a range of methods have been employed to evaluate children's proficiency in the domain of grammar (Ambridge & Rowland, 2013; Blume & Lust, 2017; Lund, 2000). For younger age groups, questionnaires ask parents to report on their children's use of grammatical structures. Many tests on direct tasks such as sentence repetition, sentence comprehension, or judgement tasks. Another influential methodological approach is the collection and analysis of elicited or spontaneous speech productions.

Collecting children's grammatical production

Elicitation tasks are specially designed tasks often included in standardised language tests. The child is stimulated by selected materials such as pictures, videos, or animations,

CONTACT Christina Kauschke kauschke@uni-marburg.de Philipps-Universität Marburg, Institut für Germanistische Sprachwissenschaft, AG Klinische Linguistik, Pilgrimstein 16, Marburg 35037, Germany

sometimes with the introduction of characters like a 'talking dog' (Räsänen et al., 2016) to evoke specific target structures. The level of prompting of these target structures can be more or less explicit, ranging from neutral questions like 'What is happening?' to constraining questions like 'Which cow is Max looking at?' (Zukowski, 2009). Compared to natural language samples, elicitation tasks yield more examples of certain target structures (Steel et al., 2013). On the other hand, not all target structures are elicited easily, implying some linguistic features may remain underrepresented. The advantages are that responses can easily be analysed as correct or incorrect, and that test scores can be derived from the children's responses, allowing for quantifiable assessments. Taken together, elicitation tasks are a time-efficient way to assess the mastery of selected target structures, but they may not provide a comprehensive assessment of a child's overall grammatical competence or actual grammatical production (Klatte et al., 2022).

Spontaneous speech production, also referred to as natural speech samples, involves collecting speech samples in real-life, naturalistic situations. This informal measure yields a more ecologically valid representation of linguistic performance in daily life. Typically, around 100 utterances are transcribed for analysis, but shorter samples of 50 or even 25 words can yield reliable results (Klatte et al., 2022). This approach offers an overview of the child's current grammatical abilities, although specific target structures may not be present in the sample. However, transcription and sentence-by-sentence analysis of spontaneous speech production is time-consuming. As a result, research findings are often limited to a small number of participants. Proper linguistic expertise and adequate training are needed to conduct accurate and reliable analyses. Although natural language samples are considered as the 'gold standard to identify gaps in children's use of grammar for setting therapy goals' (Klatte et al., 2022, p. 2), clinicians are often reluctant to use this method routinely in clinical practice because it is time-consuming and demanding (Klatte et al., 2022; Pezold et al., 2020). In addition, in the case of natural language sampling the results are usually not transformed into test scores.

The present study explores an intermediate approach: children are asked to describe pictures in their own words, without being directed to specific target structures from the outset. The utterances produced are then analysed using language sample analysis.

Analysing children's grammatical production

Once speech production data has been collected and transcribed, a number of measures can be derived from the child's utterances. A well-known and long-standing index of language development is MLU, which measures the mean length of utterances (MLU) in words or morphemes (Blume & Lust, 2017; Brown, 1973; Rice et al., 2010). This calculation can be done relatively quickly and easily and serves as a valuable estimate of language acquisition and productivity, especially in young children. However, utterance length only gives a broad, superficial description of the language level and does not provide an in-depth grammatical analysis (Klatte et al., 2022, p. 13). In addition to length, utterances can also be analysed in terms of completeness (Lund, 2000) or diversity (Hadley et al., 2018). To gain a more comprehensive understanding of grammatical strengths and weaknesses, it is necessary to evaluate specific grammatical features of the target language (Pezold et al., 2020). One such instrument used for this purpose is the Index of Productive Syntax (IPSyn).

The IPSyn was developed in 1990 as a research tool for English (Scarborough, 1990). Briefly, a speech sample is reviewed for instances of the production of 59 syntactic structures within 4 subscales. Each item is scored on a scale of 0, 1, or 2 points depending on how frequently a child uses the given structure. Following the original version, recent efforts have been made to further develop the tool. Altenberg and colleagues introduced a slightly revised version (IPSyn-R: Altenberg et al., 2018). Yang and colleagues proposed a modification and reduced the number of variables to 42 (IPSyn-C: Yang et al., 2022). Furthermore, researchers have been working on the implementation of automated coding procedures to improve time-efficiency (AC-IPSyn: Altenberg & Roberts, 2016; Hassanali et al., 2014; CLAN-IPSyn: MacWhinney et al., 2020). Based on these recent developments, an IPSyn score can provide:

- An overview snapshot of syntactic forms that are and are not yet in the child's sample, serving as a measure of grammatical emergence rather than mastery. (Altenberg et al., 2018, p. 996)
- A normed device capable of discriminating typical from atypical performance. (Yang et al., 2022, p. 240)
- A tangible list of structures that, if not observed, (. . .) make excellent targets for intervention. (Yang et al., 2022, p. 250)

Building upon these advantages, our objective was to develop a comparable tool specifically designed for assessing grammatical abilities in German-speaking children that fulfils the following criteria: The instrument should be based on the collection of elicitation data, deliver norms for grammatical performance that enable a comparison of a child's values with their age group, identify potential areas of concern and detect signs of grammatical impairment or developmental delays. Above all, the tool should serve as a resource for planning appropriate and targeted interventions for children with grammatical difficulties. Klatte et al. (2022) observed a gap between reported advantages of using natural sample analyses and the actual use of this method in clinical practice. The authors identified the time investment as well as negative beliefs about the clinicians' own knowledge and skills as the main barriers. We therefore sought to overcome those barriers by automatising the coding procedures to enhance the overall efficiency of the process, and by developing an approach that requires a manageable amount of linguistic expertise.

Acquisition and assessment of grammatical skills in German

German is an inflected language with nouns, pronouns, or articles marked for three genders, four cases, and two numbers. Verb placement depends on sentence type: the verb must be placed in the final position (VF) in most subordinate clauses, in the first position (V1) in yes-no-questions, and in the second position in main clauses (V2). Since the vast majority of declarative main clauses are characterised by V2, German has been called a verb-second language (Müller, 2015; Tracy & Thoma, 2009) with the V2 position restricted to finite verbs (Schulz & Schwarze, 2017). Overall, the fundamental syntactic structures of German are established by a child's third birthday, whereas the acquisition of nominal inflection, especially plural and case marking, tends to be more prolonged. In this section, we will highlight some relevant syntactic and morphological milestones.

The process of acquiring syntax starts with word combinations. When children's multi-word combinations include a verb, it is typically nonfinite (verb stem or infinitive) and positioned at the end of the utterance (Rothweiler, 2015; Tracy, 2008). During the third year of life, there is a shift towards declarative sentences with finite verbs in V2 (Clahsen et al., 1996). In a transitional phase, sentences with finite verbs in the final position may occur, while nonfinite verbs in the second position are much less common. Schulz and Schwarze (2017) even identify a 'ban' on these non-target language structures in typical language acquisition. These two crucial developmental milestones – V2 and subject-verb-agreement – are usually acquired concurrently. Around the age of three, subordinate clauses are produced. In transitional phases, the conjunction may be omitted or replaced by a placeholder (Rothweiler, 2015). Regarding morphology, the suffixes needed to mark verbs for person are acquired gradually (Bittner, 2013). Initially, children often produce full verbs either as bare verb stems or as infinitives. The suffix *-st* for the 2nd person singular is the last to be acquired (Clahsen et al., 1993). Although children have largely acquired verb inflection for person by the age of three, a subsequent period of consolidation may be necessary before they fully and confidently master subject-verb agreement in direct elicited production (Hasselaar et al., 2020). Another aspect of verb morphology involves the formation of participles to indicate perfect tense. First participles emerge in spontaneous speech around the age of two (Szagun, 2011), although they may still be prone to errors. Preschoolers show few difficulties with regular verbs with or without the *ge-* prefix: In an elicitation experiment, 87% of those stimuli were accurately produced at the age of five, increasing to 96% at the age of seven (Kauschke et al., 2017). However, overregularizations of irregular participles occur in approximately 10% of all participles during the third and fourth years of life (Rothweiler, 2015) and may still be noticed until school age.

As German noun inflection often lacks transparency, regularity, and systematicity, grasping the intricate rules of German noun morphology is challenging. Gender, visible at the article and determined by the noun, generally poses minimal difficulties for typically developing monolingual children (Szagun, 2004). The complex German plural system is acquired gradually (Kauschke et al., 2011; Kauschke et al., 2013; Laaha et al., 2006; Szagun, 2001; Thater & Ulrich, 2018). While plural marking of nouns begins early around 1;4 years, errors like the omission or overgeneralisation of suffixes, additions, or omission of umlauts, may persist for an extended period. Case marking on articles may be considered as the most demanding aspect of German morphology. The mastery of the case system unfolds through several intermediate steps, with the accusative and then the dative case acquired subsequent to the unmarked nominative (Clahsen, 1984; Eisenbeiss et al., 2006; Hasselaar et al., 2019; Scherger, 2015; Scherger et al., 2023; Ulrich et al., 2016). The acquisition of the dative case, in particular, has been demonstrated to be prolonged and error-prone, persisting up to primary school age (Ulrich et al., 2016).

Various assessment tools were designed to explore children's grammatical skills in German, either as components of comprehensive test batteries or as specific tests (for a review, see Spreer, 2018). Comprehension tests such as the TROG-D (Fox-Boyer, 2023), measure sentence comprehension through sentence-picture-matching tasks. Several test batteries, e.g. LiSE-DaZ (Schulz & Tracy, 2011), SETK 3–5 (Grimm, 2015), SET 3–5 (Petermann, 2016), SET 5–10 (Petermann, 2018), PDSS (Kauschke et al., 2023), include subtests that assess comprehension and/or production of selected syntactic and/or morphological target structures. In the ESGRAF 4–8 (Motsch, 2023), grammatical production

in children aged 4;0 to 8;11 years is evoked by explicit prompting, within a motivating context (a circus). Elicited target structures comprise V2 placement in main clauses, subject-verb agreement, VF placement in subordinate clauses, gender, plural and case marking (accusative, dative). Older children can also be tested on genitive case and passive constructions.

The PDSS (Kauschke et al., 2023) is a standardised language assessment battery for German-speaking children aged 2;6 to 7 years, with its theoretical foundation rooted in clinical linguistics. To facilitate the assessment process, the tool offers automated procedures implemented by browser-based software. The battery presents a comprehensive profile of speech and language abilities, spanning phonological and phonetic skills, lexical and semantic abilities, grammatical abilities, and narrative competence. Within the domain of grammar, the battery encompasses five subtests: sentence comprehension, production of definite articles and grammatical gender marking, elicitation of noun plurals, elicitation of case markings on definite articles (accusative and dative), and sentence production based on picture description. The present study focuses on the sentence production subtest, which is described in detail in the Methods section. The objective of this subtest was to employ a procedure that takes advantage of natural language sampling while also ensuring efficiency for application in clinical practice. It emphasises grammatical variables not covered by specific elicitation subtests within the battery, which can be better assessed in a more natural context.

The present study introduces this novel instrument for assessing grammatical skills in German children and addresses the following research questions: Are the measures derived from our instrument age-sensitive, i.e. do their means differ between age groups? Can the results be used to determine developmental milestones for the acquisition of grammar?

Materials and methods

Material

The tool for grammatical analysis is part of the recently revised third edition of the test battery PDSS (translated: Patholinguistic Diagnosis of Developmental Language Disorders, Kauschke et al., 2023). To construct the material for this subtest, a set of coloured pictures was created that depict everyday situations and events. These pictures were selected from a larger set of 20 during a piloting phase. Those pictures that elicited the most utterances were deemed particularly stimulating and were included in the final sets. The set for children under 3 years of age consists of 10 pictures, while children older than 3 are presented with 15 pictures. To encourage speech production, each picture is introduced by a neutral, open question: 'What is happening here?'. This gives the child the opportunity to describe the picture in their own words. In cases where the child responds very briefly, another open question may be used: 'What else can you tell me about it?'. For children older than 3, the experimenter then poses one or two specific, predefined questions to each picture. We observed in our pilot studies that directive questions, designed to evoke specific structures, tend to have an overwhelming and demotivating effect on young children. Direct questions are therefore only used in the version for children aged three and above.

Prior analyses showed children over the age of three (15 pictures with questions) can be expected to produce about 50 utterances, while children under the age of three (10 pictures without questions) can be expected to produce about 15 utterances. Within the software, the child's utterances are recorded and manually transcribed for analysis.

The analysis involves a multi-step procedure: first, non-analysable utterances (yes/no responses, hesitation signals and attention-getting devices, direct repetitions, interjections, 'I don't know', parts of songs or rhymes, interruptions, and incomprehensible utterances) are excluded; then, the software automatically calculates MLU in words.[1] Next, a computer-assisted analysis is conducted to assess the completeness of utterances: the experimenter classifies each utterance as complete, incomplete (omission of obligatory constituent and/or function word), or ellipsis following a detailed manual (Kauschke et al., 2023). Finally, two Grammar Scores are determined. To this end, we developed a computer-assisted scoring procedure inspired by the IPSyn and modified for German. Each transcript is examined for instances of listed structures, awarding points as soon as the required number of occurrences is identified. For children under the age of 3, the examiner searches for 13 variables across 5 subscales (see Appendix A). The scoring procedure for older children includes 7 subscales with 39 variables (see Appendix B). For examples of the children's utterances, see Appendix C.

Prior analyses revealed that focusing solely on scoring the emergence of target structures may not be sufficient for detecting grammatical weaknesses. For example, a child easily may produce articles correctly in two noun phrases but omit them in many other contexts. Therefore, our goal was not only to capture grammatical emergence but also to identify signs of immature or impaired grammar. Consequently, we created two scores: Score A lists structures that indicate grammatical phenomena expected to emerge in the course of grammar acquisition, while Score B includes non-target structures that reflect intermediate stages of language development. Persistent use of these non-target structures may indicate grammatical impairment.

Depending on the variable, one, two or three exemplars of each given structure are required (see Appendices A and B). By screening the speech sample, the examiner identifies the necessary number of adequate exemplars for each variable. Subsequently, the software calculates the points. Structures that reflect important milestones of grammatical impairment in German (e.g. main clause with finite verb in second position, correct subject-verb agreement, see Clahsen et al., 1993, 1996; Tracy, 2008) as well as structures that indicate known symptoms of grammatical delay or impairment (e.g. verb-final position in main clause, errors in subject-verb agreement, omissions of subjects, objects, verbs or articles, see Clahsen et al., 1997; Hamann et al., 1998; Hasselaar et al., 2020; Rothweiler et al., 2012; Ruberg et al., 2020) are weighted twice (see Appendices A and B). As a result, Score A has a maximum of 10 points (children <3;0 years) or 25 points (children >3;0 years; the higher, the better), while Score B has a maximum of 11 points (children <3;0 years) or 24 points (children >3;0 years; the lower, the better). The software output displays the T-values and percentile ranks, enabling a comparison of a child's abilities with their age group. Finally, an overview of all attested variables allows for a detailed profile of a child's syntactic and morphological strengths and weaknesses.

[1]We chose to calculate MLU based on words instead of morphemes because this is the common and recommended measure for German data (Clahsen et al., 1993) and the automatic calculation is easier as no annotation of morphemes is required.

Participants

The following analyses were conducted on a subsample from the larger sample recruited to establish norms for the PDSS (for a description of the full sample see Kauschke et al., 2023). We targeted 50 participants per age group, randomly chosen from the overall sample, provided that the child generated a minimum of five intelligible utterances. We also ensured a balanced age distribution within each age group. The final subsample comprised 348 monolingual children (50% female), divided into 7 age groups from 2;0 to 6;11 years. Table 1 shows the number of participants per age group. As language development is particularly dynamic in the first years of life, we had four semi-annual age groups from 2;0 until 3;11 years, and three annual groups for the older children.

Participants were recruited from all over Germany through cooperating institutions (mainly day care-centres). Parents gave their informed consent before their children participated in the study. The study was approved by the Ethical Committee of the University of Education Heidelberg, Germany. Parent questionnaires were used to exclude the presence of uncorrected hearing or vision disorders, neurological disorders such as epilepsy, general developmental delays, or autism spectrum disorders. Data collection took place in a separate room of the cooperating institution, guided by trained and supervised examiners.

Psychometric properties

Overall, the PDSS test battery demonstrates robust psychometric properties for all subtests with a mean Cronbach's $\alpha = 0.82$ and adequate discrimination parameters as well as item difficulties (see Kauschke et al., 2023). As simple scoring (correct/incorrect) is not possible for the grammar analysis focused on in the present study, additional steps have been taken to ensure the validity of our procedure. First, the PDSS can significantly verify different language development of children undergoing speech and language therapy compared to typically developing children in 15 out of 19 subscales at $p < 0.05$, including MLU ($t(297) = 2.3$; $p < 0.05$). In addition, the grammar variables are related to results from other established language tests. There are substantial small to large convergent correlations coefficients

Table 1. Descriptive statistics for the grammar analysis.

Age group	2;0–2;6	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;11	5;0–5;11	6;0–6;11
Number	*N* = 48	*N* = 50	*N* = 50	*N* = 50	*N* = 50	*N* = 50	*N* = 50
MLU in words							
mean	2.62	3.14	3.65	4.16	4.28	5.01	4.85
SD	1.05	0.938	0.698	0.567	0.592	0.899	0.521
Proportion of complete utterances							
mean	27.00%	45.70%	59.30%	72.80%	75.70%	86.10%	92.30%
SD	0.29	0.27	0.22	0.14	0.17	0.12	0.06
Summed Score A*							
mean	7.23	8.88	20.7	22.4	21.8	22.4	23.1
SD	2.72	2.08	3.06	2.11	3.01	1.87	1.83
Summed Score B*							
mean	5.31	3.86	8.26	5.86	5.62	2.82	2.08
SD	2.69	2.42	4.17	3.52	3.75	3.08	2.65

*Note that the maximum number of points for scores A and B differs for children over and under the age of three (A: max. 10/25 points, B: 11/24 points).

(MLU: 0.16, $p > 0.05$/n.s.; Score A: 0.47, $p < 0.05$; Score B: −0.52, $p < 0.05$) to the pseudo word repetition subtest from the SET 5–10 (Petermann, 2018) as well as small to large substantial convergent correlations (MLU: 0.17, $p < 0.05$; Score A: 0.52, $p < 0.01$; Score B: −0.30, $p < 0.01$) to the sentence repetition subtest of HASE (Schöler & Brunner, 2008).

To assess reliability, intraclass correlations (ICC) were calculated using the SPSS software package (Version 27). ICC was determined separately for the two versions of the scoring procedure (children under and over three years of age). As the procedures for children under and over 3 years of age differ in terms of the number and weighting of the variables, two independent data sets were created. Consequently, two different calculation steps were completed. For this purpose, two-way mixed models were calculated in each case. It was assumed that the four raters are not independent of each because of their joint project work. Therefore, a systematic bias effect was assumed and the 'absolute match' setting was chosen in the software.

Five transcripts from children in the third year of life were analysed independently by four trained raters. Intraclass correlation coefficients (ICC) were calculated following the procedure of Koo and Li (2016). The ICC for determining the completeness of the utterances was excellent (0.996 with a 95% confidence interval between 0.980 and 0.999). Good to excellent interrater reliability was also found for the grammar scores (Score A: ICC 0.955, 95% confidence interval between 0.830 and 0.995; Score B: ICC value 0.965, 95% confidence interval between 0.861 and 0.996).

Ten transcripts of children older than three were analysed by four trained raters as for the younger children. Again, the ICC for completeness of the utterance was excellent (0.916 with a 95% confidence interval between 0.903 and 0.928). Grammar Score A (0.925, with a 95% confidence interval between 0.806 and 0.979) was rated with good to excellent reliability as well as Score B (0.910, 95% confidence interval between 0.763 and 0.975). In sum, inter-rater agreement was excellent for both age groups under investigation.

Thus, the forthcoming analyses rely on sound data and can be interpreted as follows.

Analyses

To analyse general age effects, ANOVAs were run with the four main variables (MLU, completeness, Score A, Score B) across the seven age groups; and post-hoc analyses were performed between age groups using jamovi software (The Jamovi Project, 2023).

In addition, developmental milestones for single variables were determined (see Rudolph & Leonard, 2016, for a discussion of early language milestones). According to a paediatric definition commonly used in Germany (Michaelis et al., 2013), a milestone is established at the typical age when approximately 50% of children show a specific behavioural sign of development. If a child does not demonstrate the expected behaviour at an age when 90% of other children do, this is considered a 'red flag'[2] and indicates potential developmental problems (Jenni, 2022; Michaelis et al., 2013). We applied these criteria to the grammar

[2]The terminology used in German paediatrics distinguishes between 'milestones' (germ. *Meilensteine*) and 'border stones' (germ. *Grenzsteine*), the latter defining the age at which most (90%) of typically developing children have reached a developmental step. Here, we translate the German term *Grenzstein* ('border stone') as 'red flag'. In contrast, the American Academy of Pediatrics does not differentiate between milestones and warning signs. Instead, they use the term 'milestone' to signify the age at which ≥ 75% of children would typically achieve a specific developmental milestone in natural settings (Zubler et al., 2022). Children who do not reach a milestone warrant close monitoring.

scores and examined the proportion of children in each age group who produced the required number of instances for each grammatical variable (see Visser-Bochane et al., 2020 for a similar analysis of questionnaire data). For Score A, a milestone is reached when at least 50% of the children in the sample show instances of the specific variable. When 90% or more of the children produce the target structure, the age for a red flag is reached. On the other hand, for Score B, a milestone is reached when no more than 50% of the children still produce a non-target structure, and a red flag can be raised when only 10% or less do.

Results

Age-group comparisons

Table 1 shows the descriptive statistics for the four grammatical values that emerged from the analysis (MLU, completeness, score sums).

Analyses of variance reveal significant effects of age for MLU in ($F(6,150) = 53.0$, $p < .001$) as well as for the completeness of utterances ($F(6,144) = 80.8$, $p < 0.001$), with rates increasing from less than 30 to more than 90%. The post-hoc group comparisons shown in Appendix C point to developmental patterns: After a strong initial increase in MLU, there seems to be a plateau from 3;6 to 4;11 years, followed by a further increase and another plateau between 5;0 and 6;11 years. The plateau between 3;6 and 4;11 years is also evident in the proportion of complete utterances.

Turning to the grammar scores, Table 1 demonstrates that Score A shows a significant upward trend with age (<3 years: $F(1,87.9) = 11.32$, $p < 0.001$; > 3 years: $F(4,121) = 5.98$, $p < 0.001$), whereas Score B decreases significantly (<3 years: $F(1,93.9) = 7.88$, $p < 0.01$; > 3 years: $F(4,122) = 26.47$, $p < 0.001$). The post-hoc comparisons for the five age groups above three years of age are shown in Appendix C. From 3 years onwards, Score A maintains at a relatively stable and high level above 20 (see Table 1), despite the significant overall age-related trend. The post-hoc group comparisons for Score A (see Appendix D) reveal significant differences only between the children aged 3;0–3;6 and older children, whereas no significant group differences were found beyond 3;6 years. Nevertheless, a closer look at the outliers plotted in Figure 1 points to individual cases where children of different age groups achieved very low scores. Score B declines in the third year of life, falls sharply between 3;0 and 3;6 years of age, and remains at a plateau between 3;6 and 5 years of age, before falling sharply again at 5 years of age.

Determination of milestones

Next, we looked at individual variables from the grammar scores and determined the proportion of children who produced exemplars of each variable. The following figures show the frequency distribution for some selected variables relevant to the acquisition of German grammar.

In main clauses, the finite verb has to appear in the second position. Already at the age of 2;6 years, more than 70% of the children in the sample were producing these target structures (Figure 2), and the proportion increased to 90% by the age of 3. During a typical intermediate stage of grammatical development children place nonfinite verbs (infinitives or bare verb stems) in the sentence-final position before fully mastering correct main clauses. The proportion of children producing those non-target structures decreased

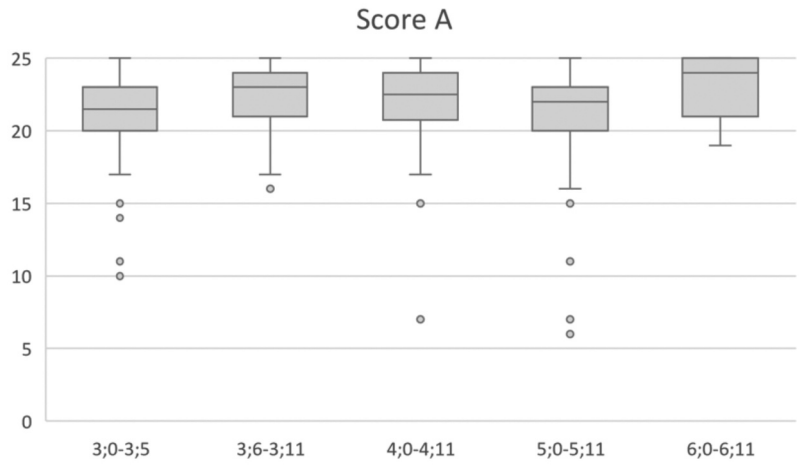

Figure 1. Boxplot of score a results.

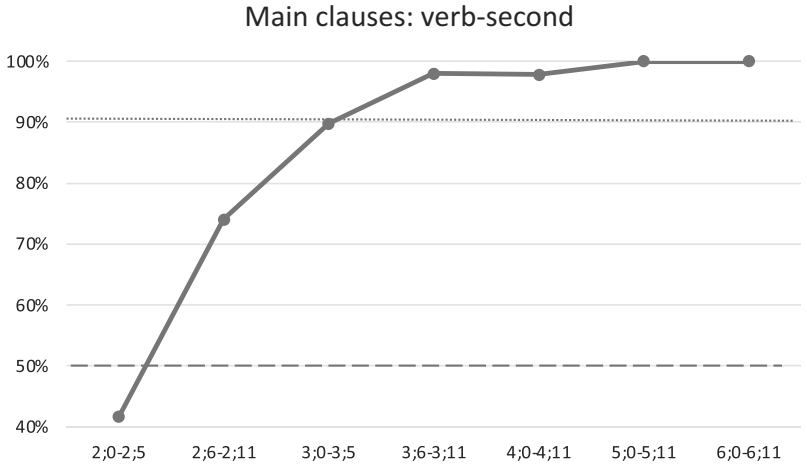

Figure 2. Proportion of children producing main clauses with finite verbs in second position (Score A). Dashed line: 50% milestone, dotted line: 90% red flag.

rapidly during the third year of life and was slightly above 10% by the age of 3 (Figure 3). The emergence of finite verbs in final position was much less frequent (Figure 3) and did never exceed 10%.

Regarding verb inflection to mark subject-verb-agreement, children began producing correctly finite verbs early, and from the age of three onwards, almost all children in the sample demonstrated this ability (Figure 4). At the same time, instances of verb inflection errors (use of infinitives, verb stems or substitutions of inflectional suffixes) were still observed in more than 10% of the children until the age of four (Figure 4).

Apart from verb inflection and verb placement, the use of obligatory articles in noun phrases is another relevant feature of German. By age 2;6, at least 90% of the children were

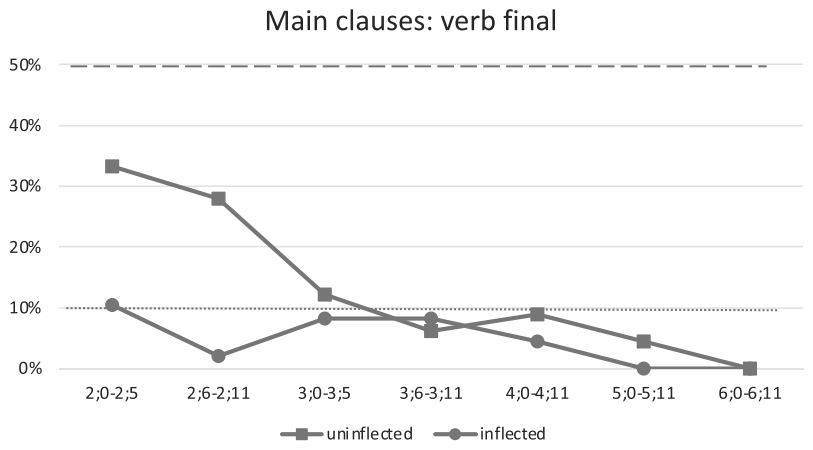

Figure 3. Proportion of children producing main clauses with nonfinite/finite verbs in final position (Score B). Dashed line: 50% milestone, dotted line: 10% red flag.

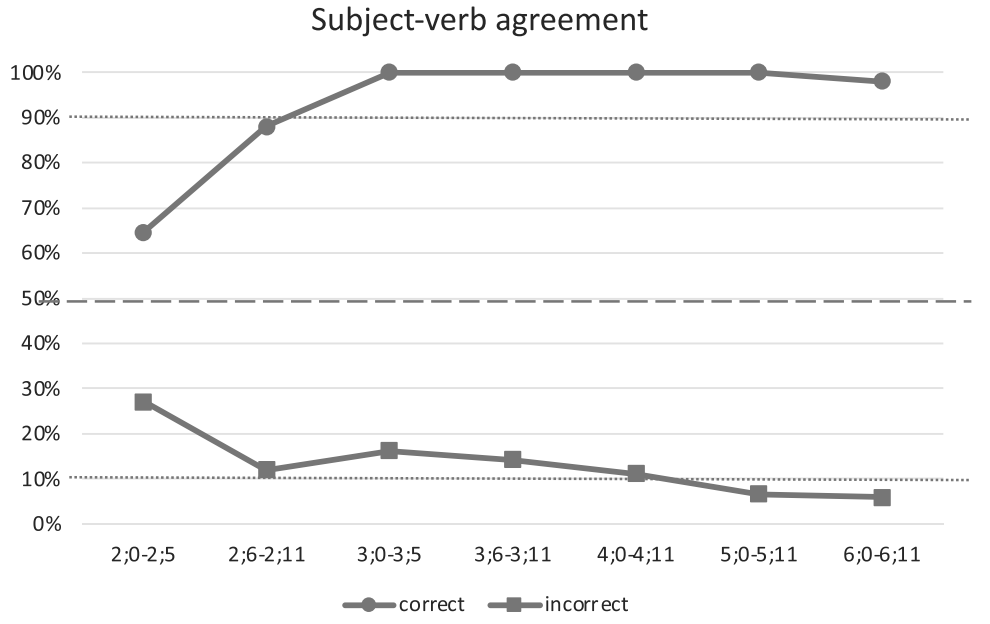

Figure 4. Proportion of children producing correct subject-verb agreement correctly (Score A)/incorrectly (Score B). Dashed line: 50% milestone, dotted line: 90% resp.10% red flag.

already producing articles in obligatory contexts. However, article omissions were not uncommon in the speech samples. Throughout the observation period, more than 10% of the children occasionally omitted articles.

Table 2. Scoring example of Anni, 4;1 years.

	MLU (w)	completeness	Score A	Score B
Raw Score	3.68	69%	21	10
T-value	38	43	43	37
Percentile	11	24	23	10

Scoring example

Finally, we present the case of Anni, a girl aged 4;1 years. During the picture description task Anni produced 66 utterances with an MLU of 3.68. 69% of Anni's utterances in the speech sample were complete, while the remaining 31% lacked obligatory constituents and/or function words. The scoring analysis yielded a raw score of 21 for Score A and 10 for Score B. Table 2 displays the T-scores and percentile ranks obtained in comparison with her age cohort, demonstrating that completeness and Score A are within the normal range and reflect average performance. However, a closer examination of the profile reveals that Anni did not produce any subordinate sentences, resulting in the absence of variables S2, VS4 (finite verb in sentence-final position) and W5 (conjunction), thus reducing score A.

In contrast to Anni's age-appropriate scores for completeness and Score A, her performance on MLU and Score B was more than one standard deviation below the mean for her age cohort. The low MLU can be attributed to simple, short utterances and to omissions of subjects (variable NP-S3, e.g. *regnet aber*, lit. *but rains*), obligatory objects (SP-O4, e.g. *die Oma gibt eine Puppe*, lit. *the grandma gives a doll*), and verbs (V12, e.g. *der blau*, lit. *he blue*). In addition, Anni produced several non-target structures. Regarding morphology, utterances with incorrect subject-verb agreement were observed (V7, e.g. *der habt ein Helm auf*, lit. *he haves a helmet on*). Anni also produced sentences with incorrect verb placement,

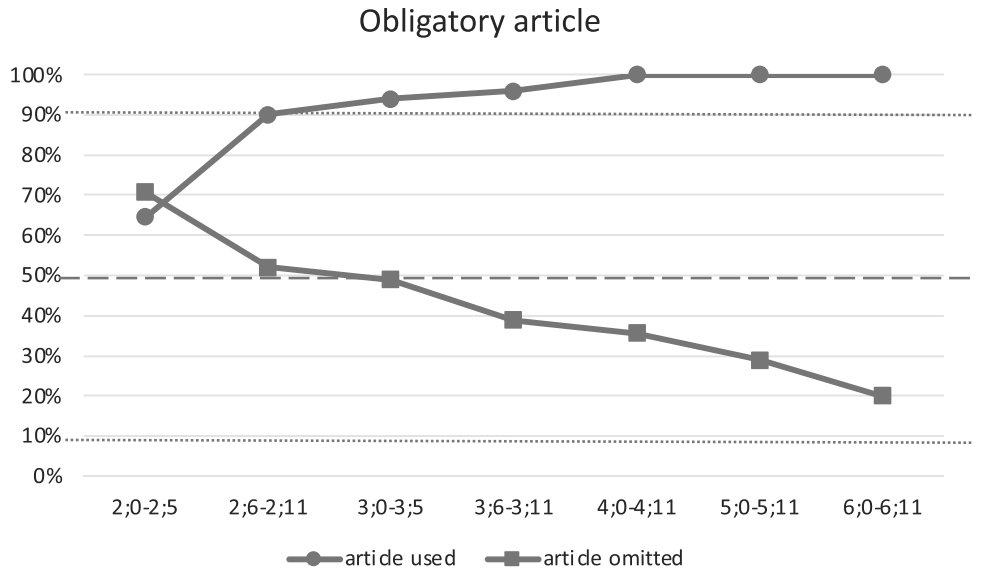

Figure 5. Proportion of children producing (Score A)/omitting (Score B) obligatory articles in noun phrases. Dashed line: 50% milestone, dotted line: 90% resp. 10% red flag.

where the verb was moved to sentence-initial position (VS6, e.g. *kommt ein Auto*, lit. *comes a car*). Overall, the profile highlights specific weaknesses in Anni's speech production, such as a lack of subordinate clauses, omission of sentence constituents, problems with subject-verb agreement, and some difficulties with verb placement. These areas present potential targets for intervention.

Discussion

The aim of the present study was to develop an effective tool for assessing the grammatical abilities of German-speaking children using elicited production in a picture-description task. This tool is part of a larger, standardised language assessment battery (PDSS, Kauschke et al., 2023). Beyond traditional measures such as MLU and utterance completeness, we sought to analyse specific syntactic and morphological aspects of the target language to identify grammatical strengths and weaknesses. In order to find a time-efficient procedure for this endeavour, we built upon the IPSyn approach (Scarborough, 1990) by screening a child's speech sample for the emergence of relevant target structures. As a modification, we also looked for the occurrence of non-target utterances, which may indicate problems in grammatical development. Results from 348 monolingual children from the norming sample across a wide age range (2;0–6;11 years) suggest that the instrument is able to reflect age-related changes in grammatical performance.

First, the observed significant increase in MLU is consistent with research findings across numerous languages, including German. However, data on utterance length German language acquisition are mainly confined to early ages up to three years and/or to a limited number of participants (e.g. Clahsen et al., 1993, 1996; Kauschke, 2013; Szagun & Schramm, 2019). Thus, our results extend existing findings by adding data from a larger number of participants covering a wider age range, and by demonstrating a pattern of strong early increases, followed by periods of plateau and subsequent growth. Of course, conclusions about developmental patterns have to be treated with caution when using cross-sectional data.

The results also showed a corresponding age-related increase regarding the completeness of utterances. So far, no data are available on the completeness of utterances in German child language. The finding of growth in MLU as well as in completeness is somewhat expectable as utterances become longer when omissions of constituents or function words decline. In this respect, MLU and completeness are not independent of each other. However, the determination of completeness goes beyond a mere measure of length by considering whether obligatory elements are omitted in sentences or phrases. Especially in the later stages of language development, when large leaps in MLU are no longer observed, completeness may serve as a more sensitive measure, capable of detecting grammatical difficulties even when MLU falls within the expected age range. To confirm this hypothesis, further research would be required in future studies.

We also found age-related effects for the newly developed scores, suggesting that as children get older, they produce more of the target structures (Score A), while their use of non-target structures decreases (Score B). Score A, which reflects the emergence of target structures, increased significantly during the third year of life. The age group of children from 3;0–3;6 differed from the older age groups, but after 3;6, this score remained stable and did not differ between groups. This stability indicates that the basic syntactic and

morphological structures of German are present in typically developing, monolingual children by the age of 3;6. Nevertheless, the outliers highlight the effectiveness of the assessment procedure, which aims to identify children who may encounter difficulties in acquiring the grammar of their language. Score B follows a decreasing pattern with a similar plateau from 3;6 to 5 years of age as seen in the MLU and completeness results. For clinical purposes, the two scores should be considered complementary in drawing conclusions about a child's grammatical performance.

In addition, we calculated the proportion of children who produced the required number of exemplars of the target or non-target structures that were part of the scores. This measure served as the basis for determining the age of grammatical milestones and red flags. The rationale for this approach was as follows: when a child fails to exhibit a target structure while at least 50% of their peers are already using it, this indicates that the child has not reached that milestone yet. If, even at an age when 90% of their peers are using the target structure, the child still does not produce it, this may be an indication of delayed or impaired grammatical development. On the other hand, if a child employs a non-target structure at a time when 90% of their peers have moved past such usage, this can also serve as a warning sign. The results (see Figures 2–5) point to a strong dynamic with sharp increases or decreases during the third year of life. Variables related to verb placement in main clauses are attested by 90% or more of the participants by the age of 3. The vast majority of children place the finite verb in second position after a period in which nonfinite or – less frequently – finite verbs occur in final position. These findings imply that a basic feature of German grammar emerges in the third year of life, and they are in line with previous research showing that young children acquiring German tend to demonstrate intermediate production of nonfinite verb-final structures, while finite verb-final structures are less commonly observed (e.g. Clahsen et al., 1996). Our results suggest that if a child continues to produce nonfinite or finite verbs in sentence-final position after the age of 3, this can be seen as a warning sign. Attention is also warranted if a child does not produce any correct verb inflections to mark subject-verb agreement after the age of 3, or still makes agreement errors. Subject-verb agreement has been consistently identified as a vulnerable area in (monolingual as well as bilingual) children with Developmental Language Disorder (DLD, Hasselaar et al., 2020; Rothweiler et al., 2012; Ruberg et al., 2020). Finally, articles in noun phrases should be attested by the age of 2;6, but some article omissions are to be expected in 20–50% of children throughout the preschool years.

The present cross-sectional study illustrates the linguistic behaviour of children across age groups. In order to utilise the tool for diagnostic purposes, T-scores and percentile ranks with confidence intervals for MLU, completeness, and the grammar scores can be obtained for individual raw scores from the test manual, as illustrated in the scoring example above.

Limitations

A word of caution is needed with regard to the interpretation of the results: The mere appearance of structures as recorded in the grammar scores does not, of course, mean that these structures have been fully mastered. While these scores provide insights into the emergence of target language structures and the decline of intermediate-level structures, they do not allow for a quantitative assessment of the accuracy of specific structures. The development of these emergent skills continues over time, requiring consolidation and refinement as language proficiency evolves. For example, in a study by Hasselaar et al. (2020) focusing on verb

inflection, typically developing children at an average age of 3;3 years were able to correctly mark verbs in the second and third person singular at 75% (81% using softer criteria), while at an average age of 4;0 years, their accuracy improved to 85% (98% with softer criteria). If we consider 90% accuracy as the criterion for complete mastery of a structure, it becomes evident that children may need a more extended period to comprehensively, confidently, and flexibly master verb placement or subject-verb agreement (Ulrich, 2017). As Altenberg and colleagues have noted for the IPSyn (Altenberg et al., 2018, p. 996), such scores provide an 'overview snapshot' of a child's actual grammatical production, rather than reflecting full mastery. A comprehensive individual assessment requires a detailed examination of the accuracy of specific syntactic structures or morphological paradigms. Therefore, the grammatical analysis of elicited production based on picture description as described in the present study is only one component integrated into a more extensive test battery (PDSS, Kauschke et al., 2023), which includes specially designed elicitation tasks to evaluate gender, plural, and case marking, along with a sentence comprehension task.

The present study relies on cross-sectional data from a wide age range. Future research should include longitudinal studies to follow developmental trajectories and provide more robust insights into the course of development.

Conclusion

To conclude, the new procedure offers a valuable tool for assessing grammatical abilities in German for several reasons: it is suitable for use across the pre-school age range, it is age-sensitive and reflects key developmental milestones. It also serves clinical purposes by identifying children with grammatical problems. When used for individual diagnostic purposes, the components of the new tool (namely MLU, completeness of utterances, and the two grammar scores) reliably evaluate a child's grammatical abilities in comparison with their peers. In particular, the scoring analysis provides a comprehensive profile of syntactic and morphological strengths and weaknesses, offering a deeper insight for targeted interventions.

Acknowledgments

The authors would like to thank Julia Siegmüller and Steffi Sachse, who are part of the PDSS team together with the first and last author. We are grateful to Lea Hundertmark for her cooperation in piloting the materials and to Hanna Schmidt and Judith Hollnagel for their efforts in analysing the transcripts for reliability. We would also like to thank the student assistants who carried out the data collection and all the children, families and institutions who participated.

Disclosure statement

The first and last authors are co-authors of the language test referred to in this article (PDSS: Kauschke, Dörfler, Sachse & Siegmüller, 2023).

Funding

The author(s) reported there is no funding associated with the work featured in this article.

References

Altenberg, E. P., & Roberts, J. A. (2016). Promises and pitfalls of machine scoring of the index of productive syntax. *Clinical Linguistics & Phonetics*, *30*(6), 433–448. https://doi.org/10.3109/02699206.2016.1139184

Altenberg, E. P., Roberts, J. A., & Scarborough, H. S. (2018). Young children's structure production: A revision of the index of productive syntax. *Language, Speech, and Hearing Services in Schools*, *49*(4), 995–1008. https://doi.org/10.1044/2018_LSHSS-17-0092

Ambridge, B., & Rowland, C. F. (2013). Experimental methods in studying child language acquisition. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 149–168. https://doi.org/10.1002/wcs.1215

Bittner, D. (2013). Grammatische Entwicklung. In S. Ringmann & J. Siegmüller (Eds.), *Handbuch Spracherwerb und Sprachentwicklungsstörungen – Schuleingangsphase* (pp. 51–76). Elsevier.

Blume, M., & Lust, B. (2017). *Research methods in language acquisition. Principles, procedures, and practices*. De Gruyter.

Brown, R. (1973). *A first language: The early stages*. Harvard University Press.

Clahsen, H. (1984). Der Erwerb von Kasusmarkierungen in der deutschen Kindersprache. *Linguistische Berichte*, *89*, 1–31.

Clahsen, H., Bartke, S., & Göllner, S. (1997). Formal features in impaired grammars: A comparison of English and German SLI children. *Journal of Neurolinguistics*, *10*(2–3), 151–171. https://doi.org/10.1016/S0911-6044(97)00006-7

Clahsen, H., Eisenbeiss, S., & Penke, M. (1996). Lexical learning in early syntactic development. In H. Clahsen (Ed.), *Generative perspectives on language acquisition. Empirical findings, theoretical considerations and crosslinguistic comparisons* (pp. 129–160). John Benjamins. https://doi.org/10.1515/9783110363685-012

Clahsen, H., Penke, M., & Parodi, T. (1993). Functional categories in early child German. *Language Acquisition*, *3*(4), 395–429. https://doi.org/10.1207/s15327817la0304_3

Eisenbeiss, S., Bartke, S., & Clahsen, H. (2006). Structural and lexical case in child German: Evidence from language-impaired and typically developing children. *Language Acquisition*, *13*(1), 3–32. https://doi.org/10.1207/s15327817la1301_2

Fox-Boyer, A. V. (2023). *TROG-D. Test zur Überprüfung des Grammatikverständnisses*. Schulz-Kirchner.

Grimm, H. (2015). *SETK 3-5: Sprachentwicklungstest für drei - bis fünfjährige Kinder*. Hogrefe.

Hadley, P. A., McKenna, M. M., & Rispoli, M. (2018). Sentence diversity in early language development: Recommendations for target selection and progress monitoring. *American Journal of Speech-Language Pathology*, *27*(2), 553–565. https://doi.org/10.1044/2017_AJSLP-17-0098

Hamann, C., Penner, Z., & Linder, K. (1998). German impaired grammar: The clause structure revisited. *Language Acquisition*, *7*(2–4), 193–245. https://doi.org/10.1207/s15327817la0702-4_5

Hassanali, K., Liu, Y., Iglesias, A., Solorio, T., & Dollaghan, C. (2014). Automatic generation of the index of productive syntax for child language transcripts. *Behavior Research Methods*, *46*(1), 254–262. https://doi.org/10.3758/s13428-013-0354-x

Hasselaar, J., Letts, C., & McKean, C. (2019). Case marking in German-speaking children with specific language impairment and with phonological impairment. *Clinical Linguistics & Phonetics*, *33*(1–2), 117–134. https://doi.org/10.1080/02699206.2018.1505955

Hasselaar, J., Letts, C., & McKean, C. (2020). Verb morphology in German-speaking children with developmental language disorder and phonological impairment. *Clinical Linguistics & Phonetics*, *34*(7), 671–691. https://doi.org/10.1080/02699206.2019.1692076

The Jamovi Project. (2023). *Jamovi (Version 2.3)* [Computer Software]. https://www.jamovi.org

Jenni, O. (2022). Meilen- und Grenzsteine der Entwicklung. *Monatsschrift für Kinderheilkunde*, *170*(7), 651–662. https://doi.org/10.1007/s00112-022-01547-z

Kauschke, C. (2013). The interrelation between lexical and grammatical abilities in early language acquisition. In S. Bartsch & N. Ruhlig (Eds.), *Lexical bootstrapping – the role of lexis and semantics in child language development* (pp. 143–164). Mouton de Gruyter. https://doi.org/10.1515/9783110308693.143

Kauschke, C., Dörfler, T., Sachse, S., & Siegmüller, J. (2023). *Patholinguistische Diagnostik bei Sprachentwicklungsstörungen (PDSS)* (3rd ed.). Elsevier.

Kauschke, C., Kurth, A., & Domahs, U. (2011). Acquisition of German noun plurals in typically developing children and children with specific language impairment. *Child Development Research*, *2011*, 1–17. https://doi.org/10.1155/2011/718925

Kauschke, C., Renner, L., & Domahs, U. (2013). Prosodic constraints on inflected words: An area of difficulty for German-speaking children with specific language impairment? *Clinical Linguistics & Phonetics*, *27*(8), 574–593. https://doi.org/10.3109/02699206.2013.798033

Kauschke, C., Renner, L., & Domahs, U. (2017). Past participle formation in specific language impairment. *International Journal of Language and Communication Disorders*, *52*(2), 168–183. https://doi.org/10.1111/1460-6984.12255

Klatte, I. S., Van Heugten, V., Zwitserlood, R., & Gerrits, E. (2022). Language sample analysis in clinical practice: Speech-language pathologists' barriers, facilitators, and needs. *Language, Speech, and Hearing Services in Schools*, *53*(1), 1–16. https://doi.org/10.1044/2021_LSHSS-21-00026

Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, *15*(2), 155–16. https://doi.org/10.1016/j.jcm.2016.02.012

Laaha, S., Ravid, D., Korecky-Kröll, K., LAAHA, G., & Dressler, W. U. (2006). Early noun plurals in German: regularity, productivity or default? *Journal of Child Language*, *33*(2), 271–302. https://doi.org/10.1017/S0305000906007379

Lund, N. J. (2000). Assessment of language structure: From syntax to event-based analysis. *Seminars in Speech and Language*, *21*(3), 267–274. https://doi.org/10.1055/s-2000-13200

MacWhinney, B., Roberts, J. A., Altenberg, E. P., & Hunter, M. (2020). Improving automatic IPSyn coding. *Language, Speech, and Hearing Services in Schools*, *51*(4), 1187–1189. https://doi.org/10.1044/2020_LSHSS-20-00090

Michaelis, R., Berger, R., Nennstiel-Ratzel, U., & Krägeloh-Mann, I. (2013). Validierte und teilvalidierte Grenzsteine der Entwicklung. *Monatsschrift Kinderheilkunde*, *161*(10), 898–910. https://doi.org/10.1007/s00112-012-2751-0

Motsch, H.-J. (2023). *ESGRAF 4-8: Grammatiktest für 4-8-jährige Kinder - Testmanual*. Reinhardt.

Müller, S. (2015). German: A grammatical sketch. In Kiss, T. A. Alexiadou (Ed.), *Handbücher Zur Sprach- und Kommunikationswissenschaft/Handbooks of linguistics and communication science (HSK)* (pp. 1447–1478). De Gruyter Mouton.

Petermann, F. (2016). Sprachstandserhebungstest für Kinder im Alter zwischen 3 und 5 Jahren *(SET 3-5)*. Hogrefe.

Petermann, F. (2018). *Sprachstandserhebungstest für Kinder im Alter zwischen 5 und10 Jahren* (SET 5-10, 3. Ed. ed.). Hogrefe.

Pezold, M. J., Imgrund, C. M., & Storkel, H. L. (2020). Using computer programs for language sample analysis. *Language, Speech, and Hearing Services in Schools*, *51*(1), 103–114. https://doi.org/10.1044/2019_LSHSS-18-0148

Räsänen, S. H. M., Ambridge, B., & Pine, J. M. (2016). An elicited-production study of inflectional verb morphology in child finnish. *Cognitive Science*, *40*(7), 1704–1738. https://doi.org/10.1111/cogs.12305

Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech, Language, and Hearing Research*, *53*(2), 333–349. https://doi.org/10.1044/1092-4388(2009/08-0183)

Rothweiler, M. (2015). Spracherwerb. In J. Meibauer, U. Demske, J. Geilfuß-Wolfgang, J. Pafel, K. H. Ramers, M. Rothweiler, & M. Steinbach (Eds.), *Einführung in die germanistische Linguistik* (3. ed., pp. 255–293). J.B. Metzler.

Rothweiler, M., Chilla, S., & Clahsen, H. (2012). Subject–verb agreement in specific language impairment: A study of monolingual and bilingual German-speaking children. *Bilingualism: Language and Cognition*, *15*(1), 39–57. https://doi.org/10.1017/S136672891100037X

Ruberg, T., Rothweiler, M., Veríssimo, J., & Clahsen, H. (2020). Childhood bilingualism and specific language impairment: A study of the CP-domain in German SLI. *Bilingualism: Language and Cognition*, *23*(3), 668–680. https://doi.org/10.1017/S1366728919000580

Rudolph, J. M., & Leonard, L. B. (2016). Early language milestones and specific language impairment. *Journal of Early Intervention*, *38*(1), 41–58. https://doi.org/10.1177/1053815116633861

Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguistics*, *11*(1), 1–22. https://doi.org/10.1017/S0142716400008262

Scherger, A.-L. (2015). Kasus als klinischer Marker im Deutschen. *Logos*, *23*(3), 164–175. https://doi.org/10.7345/prolog-1503164

Scherger, A., Kizilirmak, J., & Folta-Schoofs, K. (2023). Ditransitive structures in child language acquisition: An investigation of production and comprehension in children aged five to seven. *Journal of Child Language*, *50*(4), 1022–1039. https://doi.org/10.1017/S0305000922000174

Schöler, H., & Brunner, M. (2008). *Auditives Screening in der Einschulungsuntersuchung (HASE)* (2nd ed.). Hogrefe.

Schulz, P., & Schwarze, R. (2017). How strong is the ban on non-finite verbs in V2? Evidence from early second language learners of German with and without SLI. *Zeitschrift für Sprachwissenschaft*, *36*(1), 51–78. https://doi.org/10.1515/zfs-2017-0004

Schulz, P., & Tracy, R. (2011). *Linguistische Sprachstandserhebung - Deutsch als Zweitsprache (LiSe-DaZ)*. Hogrefe.

Spreer, M. (2018). *Diagnostik von Sprach- und Kommunikationsstörungen im Kindesalter*. Ernst Reinhard Verlag.

Steel, G., Rose, M., Eadie, P., & Thornton, R. (2013). Assessment of complement clauses: A comparison between elicitation tasks and language sample data. *International Journal of Speech-Language Pathology*, *15*(3), 286–295. https://doi.org/10.3109/17549507.2013.777852

Szagun, G. (2001). Learning different regularities: The acquisition of noun plurals by German-speaking children. *First Language*, *21*(62), 109–141. https://doi.org/10.1177/014272370102106201

Szagun, G. (2004). Learning by ear: On the acquisition of case and gender marking by German-speaking children with normal hearing and with cochlear implants. *Journal of Child Language*, *31*(1), 1–30. https://doi.org/10.1017/S0305000903005889

Szagun, G. (2011). Regular/Irregular is not the whole story: The role of frequency and generalization in the acquisition of German past participle inflection. *Journal of Child Language*, *38*(4), 731–762. https://doi.org/10.1017/S0305000910000255

Szagun, G., & Schramm, S. A. (2019). Lexically driven or early structure building? Constructing an early grammar in German child language. *First Language*, *39*(1), 61–79. https://doi.org/10.1177/0142723718761414

Thater, S., & Ulrich, T. (2018). Pluralmarkierung bei deutschsprachigen Kindern zwischen 4 und 9 Jahren. *Forschung Sprache*, *8*(1), 44–61. https://www.forschung-sprache.eu/fileadmin/user_upload/Dateien/Heftausgaben/2018-1/5-70-2018-01-03.pdf

Tracy, R. (2008). *Wie Kinder Sprachen lernen: Und wie wir sie dabei unterstützen können*. Gunter Narr Verlag.

Tracy, R., & Thoma, D. (2009). Convergence on finite V2 clauses in L1, bilingual L1, and early L2 acquisition. In C. Dimroth (Ed.), *Studies on language acquisition: Functional categories in learner language* (Vol. 37, pp. 1–44). Mouton de Gruyter.

Ulrich, T. (2017). *Grammatikerwerb und grammatische Störungen im Kindesalter. Ergebnisse des Forschungsprojekts GED 4-9 und ihre Implikationen für sprachdiagnostische und -therapeutische Methoden* [Postdoctoral thesis]. Universität zu Köln, Kölner UniversitätsPublikationsServer (KUBS). https://kups.ub.uni-koeln.de/9011/

Ulrich, T., Berg, M., Penke, M., Ludtke, U., & Motsch, H.-J. (2016). Der Dativerwerb - Forschungsergebnisse und ihre therapeutischen Konsequenzen. *Logos*, *24*(3), 176–190. https://doi.org/10.7345/prolog-1603176

Visser-Bochane, M. I., Reijneveld, S. A., Krijnen, W. P., van der Schans, C. P., & Luinge, M. R. (2020). Identifying milestones in language development for young children ages 1 to 6 years. *Academic Pediatrics*, *20*(3), 421–429. https://doi.org/10.1016/j.acap.2019.07.003

Yang, J. S., MacWhinney, B., & Ratner, N. B. (2022). The index of productive syntax: Psychometric properties and suggested modifications. *American Journal of Speech-Language Pathology*, *31*(1), 239–256. https://doi.org/10.1044/2021_AJSLP-21-00084

Zubler, J. M., Wiggins, L. D., Macias, M. M., Whitaker, T. M., Shaw, J. S., Squires, J. K., Pajek, J. A., Wolf, R. B., Slaughter, K. S., Broughton, A. S., Gerndt, K. L., Mlodoch, B. J., & Lipkin, P. H. (2022). Evidence-informed milestones for developmental surveillance tools. *Pediatrics*, *149*(3), e2021052138. https://doi.org/10.1542/peds.2021-052138

Zukowski, A. (2009). Elicited production of relative clauses in children with Williams syndrome. *Language and Cognitive Processes*, *24*(1), 1–43. https://doi.org/10.1080/01690960801966118

Appendices

Appendix A. Grammar scores for children < 3 years

Subscale		Variable	Number of instances	Score	weight
Verbs	V1	Use of different main verbs	2	A	2
	V2	Subject-verb-agreement correct	2	A	2
	V3	Subject-verb-agreement incorrect	2	B	2
	V4	Omission of verb	2	B	2
Verb placement	VS1	Finite verb in second position of main clause	2	A	2
	VS2	Nonfinite verb in final position of main clause	2	B	2
	VS3	Finite verb in final position of main clause	2	B	1
Subject	S1	Subject realised	2	A	2
	S2	Omission of subject	2	B	2
Object	O1	Object realised	2	A	1
	O2	Omission of obligatory object	2	B	1
Articles	A1	Article used (independent of gender or case marking)	2	A	1
	A2	Omission of obligatory article	2	B	1

Appendix B. Grammar scores for children > 3 years

Subscale		Variable	Number of instances	Score	weight
Sentence structures	S1	Main clauses	3	A	1
	S2	Subordinate clauses	3	A	1
Verbs	V1	Use of different main verbs	3	A	2
	V2	Auxiliary verbs	2	A	1
	V3	Modal verbs	2	A	1
	V4	Particle verb (particle separated from main verb)	2	A	1
	V5	Particle not separated from main verb	2	B	1
	V6	Subject-verb-agreement correct	2	A	2
	V7	Subject-verb-agreement incorrect	2	B	2
	V8	Past tense (completed action) expressed	2	A	1
	V9	Past tense (completed action) not expressed	2	B	1
	V10	Correct participle	2	A	1
	V11	Participle error	2	B	1
	V12	Omission of verb	2	B	2

(*Continued*)

Subscale		Variable	Number of instances	Score	weight
Verb placement	VS1	Finite verb in second position of main clause	3	A	2
	VS2	Nonfinite verb in final position of main clause	2	B	2
	VS3	Finite verb in final position of main clause	2	B	2
	VS4	Finite verb in final position of subordinate clause	2	A	1
	VS5	Verb in second position in subordinate clause	2	B	1
	VS6	Other incorrect verb placement	1	B	1
Subject	NP-S1	Noun phrase as subject	3	A	1
	NP-S2	Personal pronoun as subject	3	A	1
	NP-S3	Omission of subject	2	B	2
Object	NP-O1	Noun phrase as object	3	A	1
	NP-O2	Personal pronoun as object	3	A	1
	NP-O3	Object realised as prepositional phrase instead of noun phrase	2	B	1
	NP-O4	Omission of obligatory object	2	B	2
Other constituents	W1	Adverbial realised	3	A	1
	W2	Attributive adjective realised	1	A	1
	W3	Preposition correctly realised	3	A	1
	W4	Preposition error or omission	3	B	1
	W5	Conjunction realised	2	A	1
	W6	Conjunction error or omission	2	B	1
Articles	A1	Article used (independent of gender or case marking)	3	A	1
	A2	Definite article + noun: gender correct	3	A	1
	A3	Indefinite article + noun: gender correct	3	A	1
	A4	Definite article + noun: gender incorrect	3	B	1
	A5	Indefinite article + noun: gender incorrect	3	B	1
	A6	Omission of obligatory article	3	B	2

Appendix C. Examples for children's utterances

Score	Variable >3 y	Variable <3 y	Variable	Example for child's utterance with literal translation	Correct target	Translation of target utterance
A	S1		Main clauses	Ein Junge nimmt eine Tasche *A boy takes a bag*		A boy is taking a bag
A	S2		Subordinate clauses	Wenn das Schwein alles aufisst *When the pig everything eats*		When the pig eats everything
A	V1	V1	different main verbs	Das Kind weint *The child cries* Der hat sein Schlüssel verloren *He has his key lost*		The child is crying He lost his key
A	V2		auxiliary verbs	Und die haben gesehen *And they have seen.*		And they have seen
A	V3		modal verbs	Die kann nicht das Auto aufmachen *She can not the car open*		She can't open the car
A	V4		Particle verb (particle separated from main verb)	Der hebt das dann wieder auf *He picks it then again up*		Then he picks it up again
B	V5		Particle not separated from main verb	Der aufesst *He upeats*	Der isst auf	He eats up
A	V6	V2	Subject-verb agreement correct	Der braucht auch ein Krankenhaus *He needs also a hospital*		He also needs a hospital
B	V7	V3	Subject-verb agreement incorrect	Der nicht reiten$_{inf}$ *He not ride*	reitet	He doesn't ride
A	V8		Past tense (completed action) expressed	Weil der Kofferraum offen war *Because the trunk open was*		Because the trunk was open
B	V9		Past tense (completed action) not expressed	Der ist auf ein Pferd reiten$_{inf}$ *He is on a horse ride*	Der ist auf ein(em) Pferd geritten	He was riding a horse
A	V10		Correct participle	geweint *cried*		cried
B	V11		Participle error	gefallt *falled*	gefallen	fell
B	V12	V4	omission of verb	Ein Schweinchen groß *A piggy big*	Ein Schweinchen ist groß	A (little) pig is big

(*Continued*)

(Continued).

Score	Variable >3 y	Variable <3 y	Variable	Example for child's utterance with literal translation	Correct target	Translation of target utterance
A	VS 1	VS1	Finite verb in V2 in main clause	Der läuft über die Straße *He walks across the road*		He walks across the road
B	VS 2	VS2	Nonfinite verb in VF in main clause	Ein Schwein Blatt essen *A pig leave eat*	Ein Schwein isst (ein) Blatt	A pig is eating a leaf
B	VS3	VS3	finite verb in VF in main clause	Junge in Schule kommt *Boy in school comes*	(Der) Junge kommt in (die) Schule	The boy is coming to school
A	VS4		VF in subordinate clause	wenn man bei der Oma ist *when one at the grandma is*		When you are at grandma's
B	VS5		V2 in subordinate clause	dass sie liebt Süßigkeiten *that she loves sweets*	dass sie Süßigkeiten liebt	That she loves sweets
B	VS6		other incorrect verb placement	Geht sie da rauf auf die Wiese *Goes she there up on the meadow*	Sie geht da rauf auf die Wiese	She is going up there to the meadow
A	NP-S1	NP-S1	Noun phrase as subject	Die Kinder gehen schnell rein The children go quickly inside		The children are going inside quickly
A	NP-S2		Personal pronoun as subject	Er schützt sich vor dem Regen *He protects himself from the rain*		He protects himself from the rain
B	NP-S3	NP-S2	Omission of subject	Essen Möhren *Eat carrots*	Sie essen Möhren	They are eating carrots
A	NP-O1	NP-O1	Noun phrase as object	Der verliert seinen Schlüssel *He loses his key.*		He loses his key
A	NP-O2		Personal pronoun as object	Die Schwester hilft ihm einfach *The sister helps him just*		The sister just helps him
B	NP-O3		Object realised as prepositional phrase instead of noun phrase	Weil er hat von dem Spiel verloren *Because he has from the game lost*	Weil er das Spiel verloren hat	Because he lost the game
B	NP-O4	NP-O2	Omission of obligatory object	Weil der mag *Because he likes*	Weil der Süßigkeiten mag	Because he likes sweets
A	W1		Adverbial realised	Es regnen heute viel *It rain today a lot*		Today it's raining a lot
A	W2		Attributive adjective realised	Die Oma schenkt ihnen einen neuen Teddy *Grandma gives them a new teddy*		Grandma gives them a new teddy bear

(*Continued*)

(Continued).

Score	Variable >3 y	Variable <3 y	Variable	Example for child's utterance with literal translation	Correct target	Translation of target utterance
A	W3		Preposition correctly realised	Die raufen sich um den Ball *They fight for the ball*		They are fighting for the ball
B	W4		Preposition error or omission	Die geht nach Auto *She goes after car*	Die geht zum Auto	She is going to the car
A	W5		Conjunction realised	Damit die die essen kann *So that she them eat can*		So that she can eat them
B	W6		Conjunction error or omission	Wegen es regnet *Due to it rains*	Weil es regnet.	because it's raining
A	A1	A1	Article used (regardless of gender or case marking)	Wieder das Schwein *Again the$_{neut}$ pig*		Again, the pig
A	A2		Definite article + noun: gender correct	Das Schwein *The$_{neut}$ pig*		The$_{neut}$ pig
A	A3		Indefinite article + noun: gender correct	Eine Maus *A$_{fem}$ mouse*		A$_{fem}$ mouse
B	A4		Definite article + noun: gender incorrect	Der Mädchen *The$_{mask}$ girl*	Das Mädchen	The$_{neut}$ girl
B	A5		Indefinite article + noun: gender incorrect	Eine Teddybär *A$_{fem}$ teddy bear*	Ein Teddybär	A$_{mask/neut}$ teddy bear
B	A6	A2	Omission of obligatory article	Pferd ist groß *Horse is big*	Das Pferd ist groß	The$_{neut}$ horse is big

Appendix D. Post-hoc comparisons

Tukey Post-hoc Test for MLU:

	2;0–2;6	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;11	5;0–5;11	6;0–6;11
2;0–2;6	–	−0.519*	−1.027***	−1.545***	−1.657***	−2.391***	−2.235***
2;6–2;11		–	−0.508*	−1.026***	−1.138***	−1.872***	−1.716***
3;0–3;5			–	−0.518*	−0.630**	−1.364***	−1.208***
3;6–3;11				–	−0.112	−0.846***	−0.690***
4;0–4;11					–	−0.734***	−0.578**
5;0–5;11						–	0.156
6;0–6;11							–

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Tukey Post-hoc Test for completeness:

	2;0–2;6	2;6–2;11	3;0–3;5	3;6–3;11	4;0–4;11	5;0–5;11	6;0–6;11
2;0–2;6	–	−0.187***	−0.323***	−0.458***	−0.4868***	−0.590***	−0.6530***
2;6–2;11		–	−0.136*	−0.271***	−0.2996***	−0.403***	−0.4658***
3;0–3;5			–	−0.135*	−0.1638***	−0.267***	−0.3300***
3;6–3;11				–	−0.0286	−0.132*	−0.1948***
4;0–4;11					–	−0.104	−0.1662***
5;0–5;11						–	−0.0626***
6;0–6;11							–

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Tukey Post-hoc Test for Score A:

	3;0–3;5	3;6–3;11	4;0–4;11	5;0–5;11	6;0–6;11
3;0–3;5	–	−1.68**	−1.080	−1.7000**	−2.380***
3;6–3;11		–	0.600	−0.0200	−0.700
4;0–4;11			–	−0.6200	−1.300
5;0–5;11				–	−0.680
6;0–6;11					–

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

Tukey Post-hoc Test for Score B:

	3;0–3;5	3;6–3;11	4;0–4;11	5;0–5;11	6;0–6;11
3;0–3;5	–	2.40**	2.640**	5.44***	6.18***
3;6–3;11		–	0.240	3.04***	3.78***
4;0–4;11			–	2.80***	3.54***
5;0–5;11				–	0.740
6;0–6;11					–

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.