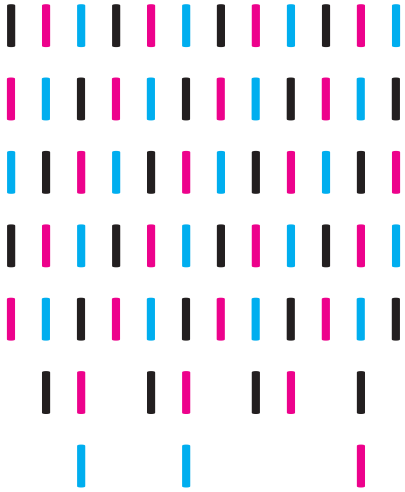

NEPS SURVEY PAPERS

Christian Lorenz, Karin Berendes, and Sabine Weinert

MEASURING RECEPTIVE GRAMMAR IN KINDERGARTEN AND ELEMENTARY SCHOOL CHILDREN IN THE GERMAN NATIONAL EDUCATIONAL PANEL STUDY

NEPS Survey Paper No. 24
Bamberg, July 2017

LIfBi
LEIBNIZ INSTITUTE FOR EDUCATIONAL TRAJECTORIES

NEPS
National Educational Panel Study

Survey Papers of the German National Educational Panel Study (NEPS)
at the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg

The NEPS Survey Paper Series provides articles with a focus on methodological aspects and data handling issues related to the German National Educational Panel Study (NEPS).

The NEPS Survey Papers are edited by a review board consisting of the scientific management of LIfBi and NEPS.

They are of particular relevance for the analysis of NEPS data as they describe data editing and data collection procedures as well as instruments or tests used in the NEPS survey. Papers that appear in this series fall into the category of 'grey literature' and may also appear elsewhere.

The NEPS Survey Papers are available at https://www.neps-data.de (see section “Publications“).

Editor-in-Chief: Corinna Kleinert, LIfBi/University of Bamberg/IAB Nuremberg

Contact: German National Educational Panel Study (NEPS) – Leibniz Institute for Educational Trajectories – Wilhelmsplatz 3 – 96047 Bamberg – Germany – contact@lifbi.de

Measuring Receptive Grammar in Kindergarten and Elementary School Children in the German National Educational Panel Study

Christian Lorenz, Leibniz Institute for Educational Trajectories, Bamberg
Karin Berendes, Hector Research Institute of Education Sciences and Psychology, Tübingen
Sabine Weinert, University of Bamberg

E-mail address of lead author:

grundschule@lifbi.de

Bibliographic data:

Lorenz, C., Berendes, K., & Weinert, S. (2017). *Measuring Receptive Grammar in Kindergarten and Elementary School Children in the National Educational Panel Study* (NEPS Survey Paper No. 24). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study. doi:10.5157/NEPS:SP24:1.0

Measuring Receptive Grammar in Kindergarten and Elementary School Children in the German National Educational Panel Study

Abstract

Listening comprehension, the ability to understand the meaning of words, sentences and texts, is an important skill for participation in educational processes and a significant predictor of learning outcomes. Therefore, listening comprehension is measured in different starting cohorts within the German National Educational Panel Study (NEPS). In the NEPS Starting Cohort 2, the kindergarten cohort, listening comprehension in the majority language (German) is measured on the word (receptive vocabulary) and sentence level (receptive grammar). This paper presents an overview of the measurement of receptive grammar, i. e. the comprehension of sentences and words with varying grammatical characteristics. In Starting Cohort 2, receptive grammar skills were assessed at two measurement points, namely in the penultimate year of kindergarten, when children were 4 to 5 years old, and the first year of elementary school, when children were 6 to 7 years old. The paper reports the test procedures and describes common factors as well as differences between the two assessments. Furthermore, the detailed test administration processes are described and descriptive results are succinctly presented.

Keywords

Receptive Grammar, Starting Cohort 2, Kindergarten, Elementary School, Testing Procedure

1. Receptive Grammar as an Important Skill for the Educational Processes

Basic interpersonal communication in general, and academic language in particular, are core competencies for successful educational participation (Berendes, Dragon, Weinert, Heppt, & Stanat, 2013). Central to the mastery of academic language are grammatical competencies (Dehn, 2011). During school lessons, receptive grammar skills are especially important. Restricted grammatical competencies may negatively affect learning processes. Depending on the extent of the restriction, the comprehension of teacher explanations and instructions for worksheets, texts, and class discussions is more or less limited or even impaired. In particular, children with low socio-economic family backgrounds (Weinert & Ebert, 2013) and children who do not speak German at home (Dubowy, Ebert, von Maurice, & Weinert, 2008) are often characterized by low (receptive and productive) grammar in the majority language. Furthermore, some grammatical structures have been found to pose a special burden for kindergarten children who do not speak German at home (Berendes, Wagner, Meurers, & Trautwein, 2015). Besides, grammatical competencies are also a reliable predictor of later literacy skills, e.g. in reading comprehension in elementary school (Ebert & Weinert, 2013; Ennemoser, Marx, Weber, & Schneider, 2012; von Goldammer, Mähler, & Hasselhorn, 2011). For these reasons, receptive grammar is assessed twice in the Starting Cohort 2 of the National Educational Panel Study (NEPS) (Blossfeld, Roßbach, & von Maurice, 2011), once in kindergarten and once at the beginning of elementary school. For detailed information on grammar development and its interrelations with other language acquisition processes as well as the assessments in the NEPS, see Berendes, Weinert, Zimmermann, & Artelt (2013).

2. Assessment of Receptive Grammar in the National Educational Panel Study

Receptive grammar can be measured using a picture selection task. The child has to listen to a sentence and must choose the picture that fits that sentence (generally out of four options). The grammatical structure of the sentences varies systematically while the words used in the sentences, i.e. the vocabulary, are held rather constant and easy. This is an economical and suitable approach for assessing listening comprehension of grammatical structures in large-scale studies. One grammar test based on this approach is the Test for Reception of Grammar (TROG, Bishop, 1989) and its German version TROG-D (Fox, 2006). A shortened version of the TROG-D is used in the NEPS.[1] All or nearly all syntactic category groups of the original test were maintained, but only approximately half of the items were used. Sentences with negation, prepositions, passive constructions, subordinate clauses, and relative clauses are examples that were included. This test was administered twice: in the first wave, when children attended kindergarten at age 4 to 5, the test was conducted in an individual setting, and in the third wave, when children were 6 to 7 years of age and attended the first grade of primary school, it was administered in a group setting. The shortened kindergarten version comprises 48 and the first-grade version 40 multiple choice items. Compared to the kindergarten version, in the school version, easy items were left out and difficult items were added to the test. 24 items were used as anchor items at both measurement points. Thus, it is possible to analyze the impact and effects of individual differences in receptive grammar, their interrelation with other variables, as well as their influencing variables. Note that only a

[1] Licensed edition from Annette V. Fox: TROG-D with permission of Schulz-Kirchner Verlag GmbH, Idstein, Germany.

subgroup of the children (i.e. about 471 children) who were tested in kindergarten or in primary school was tested on both measurement points (see Steinhauer, Zinn, Gaasch, & Goßmann, 2016).

2.1 Survey Instrument: Concept and Structure of the Receptive Grammar Test

Each item of the TROG-D, and thus each item of the adapted NEPS version, consists of a 2x2-picture-matrix (cf. Figure 1). One picture represents the correct answer while the other three show false answer options, so-called distractors. These three distractor-pictures alternate in a way that the particular meaning is slightly different with respect to their grammatical or lexical structure or characteristics. For each item, a word, a word group, or a sentence is presented to the child using a CD. The child must select the matching picture for the content of the oral stimulus just heard. He/she must point at the correct picture or indicate the matching number, respectively. The items and distractors comprise limited vocabulary from the basic interpersonal language register which should be well known to the children (e.g. boy, girl, cat, ball).

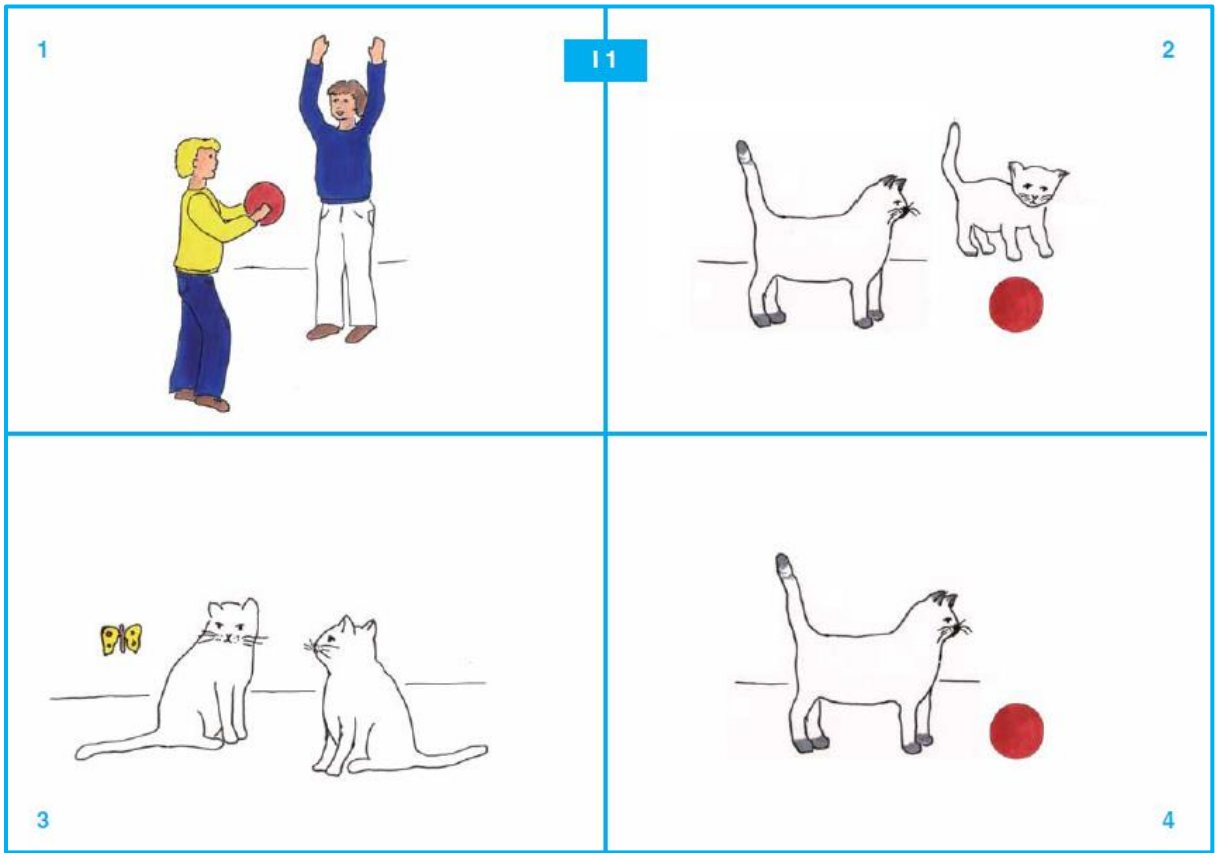

Figure 1: TROG-D Example for the target sentence: „Die Katzen schauen den Ball an.“ (“The cats are looking at the ball.”)

21 grammatical structures (Table 1) are tested by using either two or four items.[2] The items concerning the same grammatical structure are called a “set.” A set is considered to be answered correctly if all exercises of the set are answered correctly. According to the test manual (Fox, 2006, p. 9), the sets are organized hierarchically in terms of their degree of

[2] In the original test instrument, all structures are represented by four items.

difficulty, which means that the exercises' difficulty is supposed to increase incrementally. However, that does not apply for our data and will be reported in a future technical report.

Table 1: Items of the TROG-D which have been applied at the first and/or third measurement point of the NEPS Starting Cohort 2 (items used at both measurement points are colored grey)

	Measurement point 1 (Kindergarten)		Measurement point 3 (1st grade)		Item label in the original
Study number	**A12**		**A14**		
Sets on grammatical structure	**Item no.**	**Variable**	**Item no.**	**Variable**	**TROG-D**
Nouns	1	grk1a101_c			A1
	2	grk1a202_c			A2
	3	grk1a303_c			A3
	4	grk1a404_c			A4
Verbs	5	grk1b105_c			B1
	6	grk1b206_c			B2
	7	grk1b307_c			B3
	8	grk1b408_c			B4
Adjectives	9	grk1c109_c			C1
	10	grk1c210_c			C2
	11	grk1c311_c			C3
	12	grk1c412_c			C4
Two Element Sentences	13	grk1d113_c			D1
			1	grg1d101_c	D2
	14	grk1d214_c			D3
			2	grg1d202_c	D4
Three Element Sentences	15	grk1e115_c			E3
	16	grk1e216_c			E4
Negation	17	grk1f117_c			F3
	18	grk1f218_c			F4
Prepositions „in“ and „auf“					G1
(“in”, “on”)	19	grk1g119_c			
	20	grk1g220_c			G4
Perfect tense	21	grk1h121_c			H1
	22	grk1h222_c			H3
Plural	23	grk1i123_c			I1
	24	grk1i224_c			I3
Prepositions „über“ and „unter“	25	grk1j125_c	3	grg1j103_c	J2
(“over“ and “under”)	26	grk1j226_c	4	grg1j204_c	J3
Passive Voice	27	grk1k127_c	5	grg1k105_c	K1
	28	grk1k228_c	6	grg1k206_c	K4
Personal Pronouns Nominative	29	grk1l129_c	7	grg1l107_c	L1
	30	grk1l230_c	8	grg1l208_c	L2
Relative Clauses	31	grk1m131_c	9	grg1m109_c	M2
	32	grk1m232_c	10	grg1m210_c	M4
Personal Pronouns	33	grk1n133_c	11	grg1n111_c	N2
Accusative/Dative	34	grk1n234_c	12	grg1n212_c	N3
Two Objects	35	grk1o135_c	13	grg1o113_c	O1
			14	grg1o214_c	O2
			15	grg1o315_c	O3
	36	grk1o236_c	16	grg1o416_c	O4
Subordination with			17	grg1p117_c	P1
„während/nachdem“	37	grk1p137_c	18	grg1p218_c	P2
(“while“/“after“)	38	grk1p238_c	19	grg1p319_c	P3
			20	grg1p420_c	P4
Topicalization			21	grg1q121_c	Q1
	39	grk1q139_c	22	grg1q222_c	Q2
	40	grk1q240_c	23	grg1q323_c	Q3
			24	grg1q424_c	Q4
Conjunction „weder-noch“			25	grg1r125_c	R1
(“neither-nor”)	41	grk1r141_c	26	grg1r226_c	R2

	42	grk1r242_c	27	grg1r327_c	R3
			28	grg1r428_c	R4
Relative Clause with Pronoun (Accusative/Dative)			29	grg1s129_c	S1
			30	grg1s230_c	S2
	43	grk1s143_c	31	grg1s331_c	S3
	44	grk1s244_c	32	grg1s432_c	S4
Coordination with „und“ (“and“)	45	grk1t145_c	33	grg1t133_c	T1
			34	grg1t234_c	T2
			35	grg1t335_c	T3
	46	grk1t246_c	36	grg1t436_c	T4
Subordination with „dass“ (“that“)	47	grk1u147_c	37	grg1u137_c	U1
			38	grg1u238_c	U2
			39	grg1u339_c	U3
	48	grk1u248_c	40	grg1u440_c	U4

2.2 Procedure

2.2.1 Measurement point 1, penultimate year of kindergarten, individual setting

At the first measurement point, the penultimate year of kindergarten, children (*N* = 2,915) were 60.07 months old on average (*SD* = 4.24). The tests were scheduled on two days with each child. 49.7 percent of the children were female. Children were tested individually and all tests were instructed as playful games by well-trained test administrators.[3] During the first day, the children were tested on natural sciences. The second day was split in half, the first part of the survey was used for items dealing with receptive grammar and the second half for testing receptive vocabulary. This order was the same for all children. The grammar test was introduced to the child as a “searching game.” For each item, the test administrator showed the corresponding four pictures presented on one page of a test booklet (A4 landscape format, the four pictures filling the page completely; printed in black and white). Due to a ring binding, the pages could easily be turned to go to the next item. The items were presented auditorily using a CD player. To minimize distracting sounds and to sharpen the child’s concentration, both the child and the test administrator listened to the items via headset. If the child refused to wear headphones, the items were played on loudspeakers.

The children received the following introduction in German:

> *„Wir machen jetzt ein Suchspiel. Dazu zeige ich dir ein Buch mit vielen Bildern. Das ist ein ‚Findebuch‘, ein Buch, wo du Sachen suchen sollst. Du sollst Dir immer alle Bilder genau anschauen. Dann hörst du, was du suchen sollst und du zeigst mir dann das richtige Bild.*
>
> *Schau dir jetzt alle Bilder auf dieser Seite an. Zeige mir … SCHUH.”*

The English translation is as follows:

> *”Now, we are going to play a searching game. I will show you a booklet that consists of lots of pictures. This is a “finding booklet”, a booklet in which you have to find things. Look at every picture carefully. Then you will hear what you need to be looking for. Afterwards, point at the right picture.*

[3] For detailed information on test administration and test administrator training (“train-the-trainer program”) in the main studies of the NEPS kindergarten cohort, see Weinert and Berendes (2012). To acquire this poster, please contact karin.berendes@uni-tuebingen.de.

Please look at every picture on that page and show me … SHOE."

The test administrator played all the items one after another and paused after each item to observe and protocol the child's reaction, i.e. which picture has been selected by the child to match the presented item. The item had to be replayed if the child did not understand it clearly, if he/she pointed on several pictures simultaneously, or if he/she did not react for a period of more than five seconds.

All items were presented in a predetermined order until the termination criterion was met (see below for further information). The child did not receive any aid or feedback on correctness. The child was supposed to be motivated solely on the basis of comments on her/his cooperation while the comments should not give any hints concerning whether the item was solved correctly or not.

The test administrator recorded the child's answer for each item (cf. *Figure 2*). Furthermore, information on the child's reaction, whether there was no reaction or whether the item had to be replayed (both not included in the Scientific Use File, however), was noted down as well. This also applied when an item was answered incorrectly. The assessment with the TROG-D-items took about ten minutes.

Track	Item	Answer					Repeat	Wrong
		1	2	3	4	NR		
Set I								
I1	The cats are looking at the ball.	☐	* ☐	☐	☐	☐	⊔	☐
I2	*[Second item of set I]*	☐	☐	*☐	☐	☐	⊔	☐
							Set wrong	☐

Note. NR = no reaction. ** Asterisk marks the correct answer.*

Figure 2: Excerpt of the protocol sheet for measurement point 1.

Immediately after the test, each child was asked to estimate his/her own achievement in the "searching game" using a smiley-based procedural metacognition task (see Lockl, 2013).

Termination criterion

The test had to be interrupted once five consecutive sets have been classified as wrong. One set was classified as wrong as soon as one exercise of the set was solved incorrectly. Furthermore, the test had to be interrupted if the child did not show any reaction during two successive sets. In the Scientific Use File, all items which were not reached due to the termination criterion are denoted as -94.

Possible challenges during individual testing

The test is rather easy to administer and challenges are rare. Nevertheless, the following possible challenges and their handling were thoroughly instructed to guarantee standardization in test administration:

- The child does not answer.

The test administrator had to wait five seconds and replay the sentence another time (in case of technical problems, the test administrator repeats the sentence orally). If the child did not answer another time, the test administrator presented the sentence again. If the child, however, did not answer again, the lack of reaction had to be noted down on the protocol sheet and the test administrator had to go on with the next sentence. In cases of no reaction during two ensuing sets, the test had to be interrupted.

- The child refers to the correct picture but does not point at it.

This can happen especially at the beginning of the test. In this case, the test administrator had to explain the procedure again: The child was supposed to listen and then to point at the respective picture. Afterwards, the current sentence could be repeated. If this did not prove to be any assistance, the test administrator had to guide the child's finger to the correct picture. "NR" for **n**o **r**eaction had to be noted down in the protocol sheet. If it was not possible to make the child point at the right picture within the next items, the test had to be interrupted.

- The child is pointing to more than one picture simultaneously.

In this case, the test administrator had to say: "Look carefully, only one picture is correct. Please show me … ." Then, the entire item had to be replayed and the repetition had to be noted down on the protocol sheet. If the child was still pointing at multiple pictures, the respective numbers of the picture had to be noted down; the answer, however, had to be classified as wrong.

- The child points to a picture prior to the end of the sentence (either played from CD or read in case of technical problems).

The test administrator then had to say: "Please wait until the speaker has finished reading out. Please show me the picture after she has finished speaking." or "Please wait, I have not finished reading out yet. Please show me the picture after I have finished speaking," respectively. Afterwards, the test administrator had to repeat the sentence regardless of whether the hasty answer was right or wrong. To prevent the child from pointing to one picture before having heard the whole sentence, the test administrator may cover the entire page with a blank paper. The repetition had to be noted down on the protocol sheet.

- The child changes his/her answer.

If the child corrected himself/herself (e.g. *"Wait, this answer is correct."*), the test administrator classified the final answer.

- The child always points to the same position on the page.

If the test administrator realized that the child was continuously pointing to the same position on the page, he advised the child to look on every picture carefully and guided the child's finger to all four pictures once again. In case of impairments (e.g. limited range of vision), the

test administrator had to make a respective note on the protocol sheet, which is not included in the Scientific Use File.

- The child repeats the sentence but in a divergent manner.

It might occur that a child repeats the respective sentences, although this is not envisaged. If a child repeated a sentence incorrectly, the test administrator had to repeat the entire sentence again and note down the answer. Furthermore, the test administrator was always supposed to observe whether the picture shown by the child corresponded to his/her incorrect sentence repetition.

- The child gives an answer but remains uncertain.

Occasionally, a child is not sure about his/her answer. This might occur between two items or directly after giving an answer. The child's reaction can be: *"Oh, that was wrong!"* He/she might also say that a picture does not correspond with a particular sentence. In this case, the following reaction was appropriate:

CD/Test administrator: "*The knife is on the shoe."*

Child*:* "*No, the knife is in the shoe."* (The child is referring to a wrong picture.)

Test administrator*:* "*Exactly, but listen once again - show me: The knife is on the shoe."*

- The child is distracted.

If the child was distracted, the respective sentence had to be repeated. The repetition had to be noted down on the protocol sheet.

- For the items C3 or C4, where the child has to point at completely black or white colored pictures, the child points to a black and/or white area on the distracting pictures.

The test administrator had to note down the picture at which the child was pointing. However, the test administrator should not classify it as a mistake by not ticking the respective bracket. This means that pointing at a white or black space was not supposed to lead to a false set. This applies for the termination criterion as well.

2.2.2 Measurement point 3, grade 1 in primary schools, group setting

In contrast to measurement point 1, in school receptive grammar was assessed in a group setting with an average of 9.8 children per group in the first graders' classrooms in primary schools. This was the third measurement point for children already tested in kindergarten and the first measurement point for their classmates who were additionally sampled in first grade. Only 471 children who were tested in kindergarten were also tested in elementary school, because most children did not move over to an elementary school that participated in the NEPS study and therefore could not be tested (see Steinhauer et al., 2016). The average age of the children (N = 6,442) was 85.18 months (SD = 4.89). 50.7 percent of them were female. Competence tests in the areas of mathematics, natural sciences, vocabulary, declarative metacognition, and grammar were taking place on two days and were administered according to the following rotation pattern (cf. *Table 2*). The grammar test was issued in one test booklet, printed in black and white, and showing two items on one page. On each right-hand page, two tasks were presented; each 4x4 picture matrix was 14.6cm wide and 10.4cm high. The left pages of the booklet were blank. The same booklet also contained the vocabulary and declarative metacognition tests. Analogous to measurement point 1, this was followed by a

procedural metacognition task in which the children were asked to assess their own achievement on the grammar test (cf. Lockl, 2013). In addition to this, this test booklet was distributed in each testing group in two variants. In order to preempt cheating within the group setting, the distractors in one of the test booklet versions were ordered differently by switching picture No. 1 (top left) and picture No. 4 (bottom right) in each item. An ANOVA indicates no significant differences in the WLE scores for the booklet versions: F (7,6435) = 1.526, p = .153, cf.

Table 3. The same holds true for the position of the grammar test: t (6441) = -1.564, p = .118, cf. *Table 4*.

Table 2: Test design, testing variants and rotations

	Day 1	**Day 2**
Test group 1	Mathematical competence Scientific competence	Receptive vocabulary Receptive grammar Declarative metacognition
Test group 2	Mathematical competence Scientific competence	Receptive vocabulary Receptive grammar Declarative metacognition
Test group 3	Mathematical competence Scientific competence	Declarative metacognition Receptive vocabulary Receptive grammar
Test group 4	Scientific competence Mathematical competence	Declarative metacognition Receptive vocabulary Receptive grammar
Test group 5	Receptive vocabulary Receptive grammar Declarative metacognition	Mathematical competence Scientific competence
Test group 6	Receptive vocabulary Receptive grammar Declarative metacognition	Scientific competence Mathematical competence
Test group 7	Declarative metacognition Receptive vocabulary Receptive grammar	Mathematical competence Scientific competence
Test group 8	Declarative metacognition Receptive vocabulary Receptive grammar	Scientific competence Mathematical competence

Table 3: Average WLE scores per booklet version

Booklet no.	N	Position of TROG-D test	WLE	S.D.	S.E.
1	808	Last	1.720	1.211	.043
2	771	Last	1.712	1.173	.043
3	843	Last	1.621	1.235	.043
4	783	Last	1.648	1.207	.043
5	866	Penultimate	1.664	1.140	.039
6	826	Penultimate	1.768	1.195	.042
7	799	Penultimate	1.755	1.189	.042
8	792	Penultimate	1.695	1.162	.041
Total	6,443		1.697	1.189	.015

Table 4: Average WLE scores per position of the TROG-D test

Position of TROG test	N	WLE	S.D.	S.E.
Last	3,162	1.673	1.208	.021
Penultimate	3,281	1.720	1.171	.020
Total	6,443	1.697	1.189	.015

The items were played via CD, using speakers from a portable CD player. The concept of the grammar test was illustrated on posters showing two exemplary items.

During the test, the items were played consecutively. Playback was stopped between each item, guaranteeing that the children had enough time to mark the corresponding picture in their exercise books. Once the test administrator was reassured that all children were listening carefully, the presentation of the next item could be resumed. Items were repeated after technical problems, distractions etc.

Each page of the booklet bore a symbol on its top so that all children knew on which page to continue the test. By saying: *"Please turn to the page with the apple,"* the test administrator indicated the page, which always consisted of two items with four pictures each.

Contrary to the first measurement point, children were presented with all items in the group setting, i.e. there was no predetermined termination criterion. The process time for the grammar test was planned to amount to ten minutes. Since some groups could not finish the test within the given amount of time, the time limit for particular parts of the test was extended by five minutes. Protocols of 660 tests show that only five testing groups (0.8 %) needed less than the predetermined process time. The precise predetermined process time was maintained by 74 groups (11.2 %), whereas the time total frame was extended by 581 groups (88.0 % of the total number of groups).

3. Data Analysis

3.1 Descriptive Statistics for Measurement Point 1

At measurement point 1 the correct answers of each child were added up, resulting in a sum score between 0 and 48. 45.5 percent of the children finished the entire test, while the other children reached the termination criterion before the end of the test. The maximum score reached was 47, meaning that no child solved all items correctly.

As the descriptive results show, the test is relatively easy, reflected in a left-skewed distribution (cf. *Table 5* and *Figure 3*).

Table 5: Descriptive statistics of the receptive grammar sum score, measurement point 1

Number of valid cases	2,915
Mean	31.01
Median	32.00
Mode	37.00
Standard deviation	7.54
Variance	56.87
Skewness	-.71
Kurtosis	.22
Minimum	0
Maximum	47

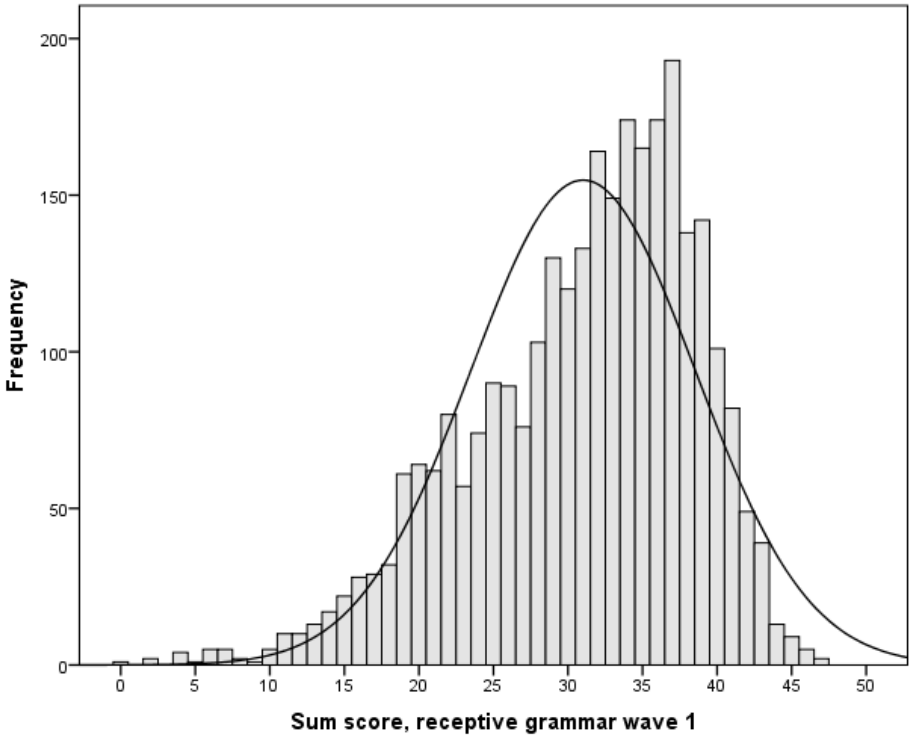

Figure 3: Histogram with normal distribution curve of the grammar sum score, measurement point 1.

3.2 Descriptive Statistics for Measurement Point 3

At measurement point 3, the test was shortened to include 40 items; compared to measurement point 1 some more difficult items were added, so that the skewness of the distribution was reduced (cf. *Table 6* and *Figure 4*). Data users should be aware that the sum score of this measurement point is not suitable for statistical analysis. This is because the test was aborted in 87 of 672 test groups by the administrator when the time limit was reached or when the test took too much time, respectively. Therefore, children in those test groups did not have the chance to complete the test, resulting in 12.9 percent of test groups with group-specific not-reached items, encoded as -94 in the Scientific Use File. Across all test groups, there is an average of 1.6 not-reached items. Looking only at the groups that had to terminate the test prematurely, the average is 12.5 not-reached items. Thus, the sum scores are not comparable with each other. Therefore, we calculated weighted likelihood estimates (WLE) that account for the test terminations. More detailed information on the data, especially regarding the scaling conducted for measurement point 3 to deal with this challenge, will be a central issue in a separate technical report. Furthermore, it is planned to also calculate WLEs for measurement point 1 that will then be published in a future update of the scientific use file.

Table 6: Descriptive statistics of the receptive grammar WLE score, measurement point 3

Number of valid cases	6,443
Mean	1.70
Median	1.64
Mode	2.49
Standard deviation	1.19
Variance	1.42
Skewness	.23
Kurtosis	.57
Minimum	-5.64
Maximum	5.91

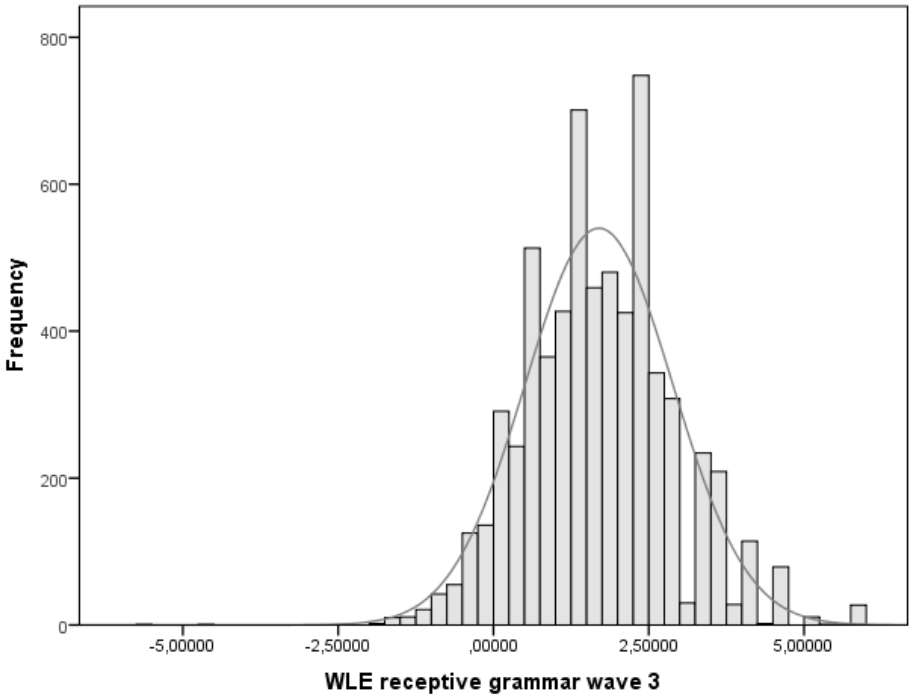

Figure 4: Histogram with normal distribution curve of the grammar WLE score, measurement point 3.

4. References

Berendes, K., Dragon, N., Weinert, S., Heppt, B. & Stanat, P. (2013). Hürde Bildungssprache? Eine Annäherung an das Konzept Bildungssprache und aktuelle empirische Forschungsergebnisse. In A. Redder & S. Weinert (Hrsg.), *Sprachförderung und Sprachdiagnostik. Perspektiven aus Psychologie, Sprachwissenschaft und empirischer Bildungsforschung* (S. 17-41). Münster: Waxmann.

Berendes, K., Wagner, W., Meurers, D. & Trautwein, U. (2015). Grammatikverständnis von Kindern unterschiedlicher sprachlicher und sozioökonomischer Herkunft. *Frühe Bildung, 4*(3), 126–134.

Berendes, K., Weinert, S., Zimmermann, S., & Artelt, C. (2013). Assessing language indicators across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online*, *5*(2), 15–49.

Bishop, D. V. M. (1989). *Test for Reception of Grammar (TROG)*. Oxford, UK: Medical Research Council.

Blossfeld, H.-P., H.-G. Roßbach und J. von Maurice (Hrsg.) (2011). Education as a Lifelong Process – The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft: Sonderheft 14.*

Dehn, M. (2011). Elementare Schriftkultur und Bildungssprache. In S. Fürstenau & M. Gomolla (Hrsg.), *Migration und schulischer Wandel: Mehrsprachigkeit* (S. 129–151). Wiesbaden: VS Verlag für Sozialwissenschaften.

Dubowy, M., Ebert, S., von Maurice, J. & Weinert, S. (2008). Sprachlich-kognitive Kompetenzen beim Eintritt in den Kindergarten. Ein Vergleich von Kindern mit und ohne Migrationshintergrund. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 40*(3), 124–134.

Ebert, S. & Weinert, S. (2013). Predicting reading literacy in primary school: The contribution of various language indicators in preschool. In M. Pfost, C. Artelt & S. Weinert (Hrsg.), *The development of reading literacy from early childhood to adolescence* (S. 93-149). Bamberg: University of Bamberg Press.

Ennemoser, M., Marx, P., Weber, J., & Schneider, W. (2012). Spezifische Vorläuferfertigkeiten der Lesegeschwindigkeit, des Leseverständnisses und des Rechtschreibens. Evidenz aus zwei Längsschnittstudien vom Kindergarten bis zur 4. Klasse. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, *44*(2), 53–67.

Fox, A. (2006). *Test zur Überprüfung des Grammatikverständnisses (TROG-D)*. Idstein: Schulz-Kirchner.

Lockl, K. (2013). *Assessment of procedural metacognition: Scientific Use File 2013.* Bamberg: University of Bamberg, National Educational Panel Study.

Steinhauer, H. W., Zinn, S., Gaasch, C. & Goßmann, S. (2016). NEPS Technical Report for Weighting: Weighting the sample of Kindergarten children and Grade 1 students of the National Educational Panel Study (Wave 1 to 3) (NEPS Working Paper No. 66). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

von Goldammer, A., Mähler, C. & Hasselhorn, M. (2011). Vorhersage von Lese- und Rechtschreibleistungen durch Kompetenzen der phonologischen Verarbeitung und der Sprache im Vorschulalter. In M. Hasselhorn & W. Schneider (Hrsg.), *Frühprognose schulischer Kompetenzen* (S. 32-50). Göttingen u.a.: Hogrefe.

Weinert, S., & Berendes, K. (2012). *Competence measurement and test administrator training in the Kindergarten cohort of the National Educational Panel Study.* Poster presentation at the EUCCONET (European Child Cohort Network) & SLLS (Society for Longitudinal and Life Course Studies) International Conference, Paris, France.

Weinert, S., & Ebert, S. (2013). Spracherwerb im Vorschulalter: Soziale Disparitäten und Einflussvariablen auf den Grammatikerwerb. *Zeitschrift für Erziehungswissenschaft*, *16*, 303–332.

Acknowledgement

This paper uses data from the National Educational Panel Study (NEPS): Starting Cohort Kindergarten, doi:10.5157/NEPS:SC2:3.0.0. From 2008 to 2013, NEPS data was collected as part of the Framework Programme for the Promotion of Empirical Educational Research funded by the German Federal Ministry of Education and Research (BMBF). As of 2014, the NEPS survey is carried out by the Leibniz Institute for Educational Trajectories (LIfBi) at the University of Bamberg in cooperation with a nationwide network.