

The logo for NEPS (National Educational Panel Study) features the acronym 'NEPS' in a bold, blue, sans-serif font. To the left of the text is a vertical orange bar with a white bracket-like shape on its right side, which frames the text.

National Educational Panel Study

Information on Competence Testing

NEPS Starting Cohort 3 — Grade 5

*Paths Through Lower Secondary School —
Educational Pathways of Students in Grade 5
and Higher*

Wave 6: Grade 9

Research Data

The logo for LifBi (Leibniz Institute for Educational Trajectories) consists of the letters 'LifBi' in a bold, black, sans-serif font. A vertical blue bar is positioned to the left of the 'i', and a vertical pink bar is positioned to the left of the 'B'. The bars are of equal height and are separated by a small gap.

LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES

Copyrighted Material
Leibniz Institute for Educational Trajectories (LifBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Sabine Weinert
Executive Director of Research: Dr. Jutta von Maurice
Executive Director of Administration: Dr. Robert Polgar
Bamberg; September 25, 2018

Information on testing				
Sample	Study A98, Grade 9, Starting Cohort 3, Year 2015			
Test situation	Group testing, normally taking place in the classroom, 1 test instructor, normally 1 supervisory teaching staff			
Test sequence	<p>The tests were held on two days. On the first test day, all students were presented with tests in the domains reading competence (2 difficulty levels), declarative metacognition, domain general cognitive function (DGCF) and listening comprehension. The difficulty levels of the reading test (“reading competence 1” or “reading competence 2”) were assigned to the students depending on their performance in the reading competence test in Grade 7. On the second test day, immigrant students from Former Soviet Union or Turkey participated in the screening test, followed by the listening proficiency test in Russian or Turkish (assignment based on information regarding immigrant background of preload data in grade 7 and grade 8).</p> <p>Sequence test booklet 1 on test day 1: reading competence 1 (easy) + declarative metacognition Sequence test booklet 2 on test day 1: reading competence 2 (difficult) + declarative metacognition Sequence test booklet 3 on test day 1: domain general cognitive function + listening comprehension Sequence test booklet 1 on test day 2: Russian or Turkish screening + listening proficiency Russian or Turkish (L1)</p>			
Test duration (net processing time)	Test day 1: 125,5 min (including student questionnaire 40 min) Test day 2: 34,5 min			
Breaks	Test day 1: 2x 15 min Test day 2: approx. 5 min break after screening test			
Administration time	Test day 1: approx. 186 min Test day 2: approx. 47 min			
Information on the individual tests				
Construct	Number of Items	Allowed Processing Time	Survey Mode	Next Measurement (until 2015)
Reading competence 1 or	32	28 min	paper-pencil	Grade 12
Reading competence 2	33	28 min	paper-pencil	
Declarative metacognition	8	15 min	paper-pencil	

<i>Cognitive basic skills</i>				
Perceptual speed	3 x 31 = 93	3 x 30 sec	paper-pencil	
Reasoning	3 x 4 = 12	3 x 3 min	paper-pencil	
Listening comprehension	16	28 min	paper-pencil, auditory presentation	
<i>Listening Comprehension Russian and Turkish</i>				
Russian or Turkish screening test	8	2,5 min	paper-pencil, auditory presentation	
Russian or Turkish L1 test	32	30 min	paper-pencil, auditory presentation	
<i>Stage-specific procedural metacognition</i>				
Regarding the reading competence domain	6	3	paper-pencil	See above
Regarding the listening comprehension domain	3	2	paper-pencil	
Regarding the Russian or Turkish L1 test	1	1	paper-pencil	See above

Preliminary note

The development of the individual tests is based on framework concepts. They constitute overarching concepts on the basis of which education-relevant competences are to be shown consistently and coherently over the entire personal history. Therefore, the following framework concepts that served as a basis for the development of the test tools to measure the above-mentioned constructs are identical in the different studies.

Reading competence

The ability to understand and use written texts is an important precondition for further developing personal knowledge and personal skills and a prerequisite for participating in cultural and social life. Manifold areas of knowledge and life are made accessible through reading. The range of reading occasions is very wide, and reading fulfills many different functions (cf. Groeben & Hurrelmann, 2004). They range from reading for expanding knowledge, which is crucial for further education, to lifelong learning as well as literary-esthetic reading. Not only do texts convey information and facts, but they also transfer ideas, moral concepts, and cultural contents. Accordingly, the concept of reading competence in the National Education Panel incorporates functional understanding as a basis for reading competence, as is also reflected in the Anglo-Saxon *Literacy Concept* (see also OECD, 2009), with a focus on competent handling of texts in different typical everyday situations.

In order to represent the concept of reading competence over the entire life span as coherently as possible, three characteristic features are specified in the framework concepts of the NEPS reading competence tests. They are considered in the following age- and stage-specific test forms:

1. text functions, text types,
2. comprehension requirements,
3. task formats.

1. Text functions/text types

The NEPS distinguishes between five text functions and associated text types, which are represented in each version of the test: a) factual texts, b) commenting texts, c) literary texts, d) instructions, and e) advertising texts (Gehrer, Zimmermann, Artelt, & Weinert, 2013). This selection is based on the assumption that these five text functions have practical relevance for the various age backgrounds of the participants. The text functions and/or text types (see Gehrer & Artelt, 2013) can be characterized as follows:

Texts conveying factual information represent basic texts for learning, fundamental acquisition of knowledge, and extraction of information; examples of these are: articles, reports, reportages, and announcements. Texts with a commenting function are texts in which a stand is taken or contradictive arguments are discussed and in which reflection is integrated. Examples of such texts are cleverly worded essays or humorous comments, which are implemented in tests for college students and adult cohorts. In school cohorts, a text with a discussion about the pleasures and disadvantages of smoking may be used, for example. The literary-esthetic function of texts is included in the third category, which encompasses short stories and extracts from novels or stories. Specific literary text types such as stage plays, satires, or poems are excluded as a result of their specific reception, which is presumably strongly dependent on educational track and curriculum. The fourth category comprises text types that are product inserts such as building and assembly instructions, package inserts for medication, work instructions, and cooking recipes. The fifth category (appeals,

advertisements, notifications) includes text types such as job advertisements and recreation programs.

The five selected text functions and their associated text types are implemented in each test booklet over the life span as a longitudinal concept, which means that each test/each test booklet for measuring reading competence contains five texts corresponding to the five text functions. Unlike the PISA studies, the NEPS does not include discontinuous texts such as graphs, tables, and road maps. Discontinuous texts are excluded from the NEPS concept as they place special demands on readers, which are not always meaningful for each age group in which reading competence is measured.

Age-specific selection (text complexity, topic selection/task requirements):

For each age cohort, texts are selected according to their thematic orientation as well as their lexical, semantic, and grammatical properties which have to be appropriate for the respective group of readers.

The growth of reading competence from childhood to early adulthood is taken into account by increasing the text complexity (larger vocabulary, longer words, foreign words, higher complexity of sentence structures) and the basic length of texts. In addition, texts are selected on topics that correspond to and are appropriate for the environment of the respective age group. They cover a wide spectrum of topics ranging from animals (for children) to social and philosophical questions related to the meaning of life for adults. Additionally, the test material is adjusted to the respective age group through age-adapted phrasing of the questions, the answer options, and the comprehension requirements of the tasks.

2. Comprehension requirements / task types

From the literature on reading competence and text comprehension (e.g., Kintsch, 1998; Richter & Christmann, 2002), it is possible to derive different types of comprehension requirement which are reflected in the NEPS concept in three specific requirement types of tasks (task types). The variants are called *types* as there is no explicit assumption that the tasks of one type are necessarily more difficult or easier than tasks of another type (Gehrer, Zimmermann, Artelt, & Weinert, 2013).

For tasks of the first type (“finding information in the text”), detailed information must be identified at sentence level; in other words, the reader is required to decipher words and recognize statements or propositions. For tasks on this requirement cluster, the wording of the information needed to solve the respective tasks is either contained in the text and identical with the task itself, or the phrasing varies slightly.

In the case of the second task type (“drawing text-related conclusions”), conclusions have to be drawn from several sentences that have to be related to each other in order to extract local or global coherence. In some cases, the relevant sentences are located closely together. In others, several sentences are spread over entire sections. In another form of this task type, the reader has to understand the thoughts expressed in the entire text, which requires the comprehension and integration of larger and more complex text portions.

For the third type, the main requirement involves “reflecting and assessing”, which is often linked to the mental representation of the text in a situation model in literature. In one version of this task type, the task is to understand the central idea, the main events, or the core message of text, whereas in another version the purpose and intention of a text have to be recognized or the readers are asked to assess the credibility of a text.

The different comprehension requirements can be found in all text functions and are considered in the respective test versions in a well-proportioned ratio. (cf. Fig. 1.).

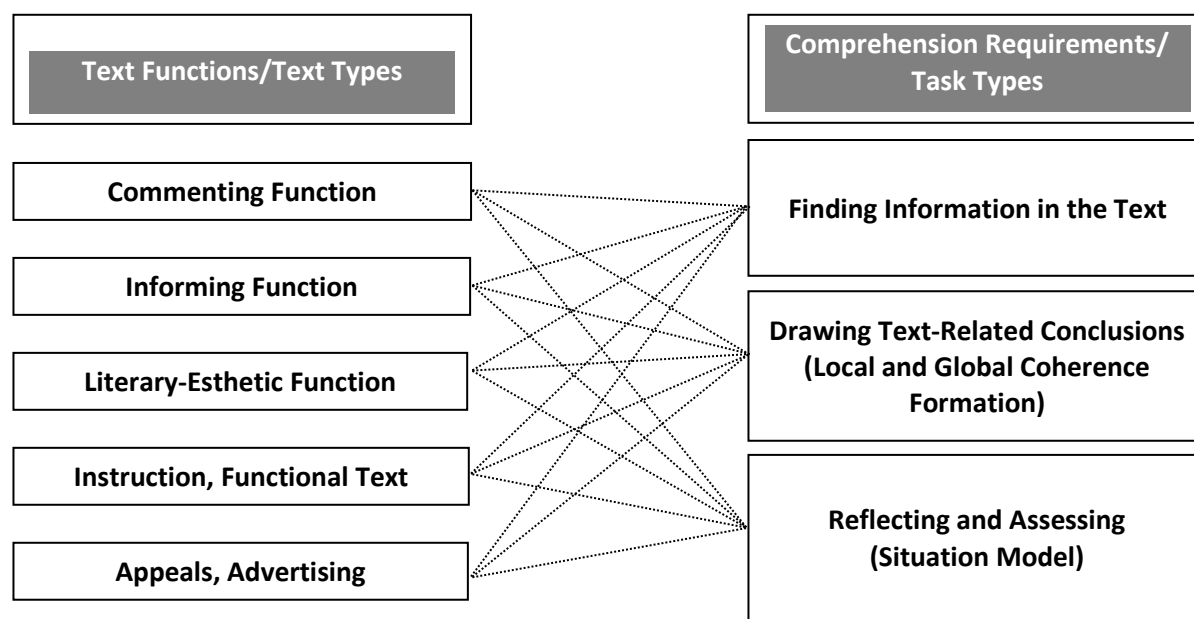


Fig. 1: Text functions and comprehension requirements (cf. Gehrler, Zimmermann, Artelt, & Weinert, 2013, p. 63)

3. Task formats

The majority of tasks have a multiple-choice format. This tasks format consists of a question/assignment about a text for which four answers are offered, one of which is the correct answer. As another task format, decision-making tasks are used, which require readers to judge individual statements and state whether they are right or wrong according to the text. So-called matching tasks represent a third format in which, for example, a subtitle must be chosen and assigned to different sections of a text. For tasks of the second and third formats, summaries are made, if necessary, thus creating answers with partly correct solutions (partial-credit items).

By systematically considering different text functions which are implemented in different age groups in realistic and age-adapted texts with appropriate text themes and different comprehension requirements, it is possible to operationalize reading competence as a comprehensive ability construct.

Scaling of items

Items of several task formats have been Rasch-scaled and longitudinally linked (Fischer, Rohm, Gnabms, & Carstensen, 2016). In addition, partial-credit items have been calculated based on the answers on decision-making tasks and matching tasks. Therefore, subjects' answers to the tasks are aggregated in one score and are not used as single items. The quality criteria and psychometric characteristics of the items are presented in the technical reports of the different starting cohorts (*for SC3*: Scharl, Fischer, Gnabms, & Rohm, 2017).

Bibliography

- Fischer, L., Rohm, T., Gnabms, T., & Carstensen, C. H. (2016). Linking the data of the competence tests (NEPS Survey Paper No. 1). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Gehrer, K. & Artelt, C. (2013). *Literalität und Bildungslaufbahn: Das Bildungspanel NEPS*. In A. Bertschi-Kaufmann, & C. Rosebrock (Hrsg.). *Literalität erfassen: bildungspolitisch, kulturell, individuell*. S. 168-187. Weinheim, Germany: Juventa.
- Gehrer, K., Zimmermann, S., Artelt, C. & Weinert, S. (2013). NEPS framework for assessing reading competence and results from an adult pilot study. *Journal for Educational Research Online* 5(2), 50-79.
- Groeben, N. & Hurrelmann, B. (Hrsg.) (2004). *Lesesozialisation in der Mediengesellschaft: Ein Forschungsüberblick*. Weinheim: Juventa.
- Kintsch, W. (1998). *Comprehension. A paradigm for cognition*. Cambridge: University Press.
- OECD (2009). *PISA 2009 assessment framework – Key competencies in reading, mathematics, and science*. Paris: OECD
- Richter, T. & Christmann, U. (2002). Lesekompetenz: Prozessebenen und interindividuelle Unterschiede. In N. Groeben, B. Hurrelmann (Hrsg.), *Lesekompetenz: Bedingungen, Dimensionen, Funktionen* (S. 25-58). Weinheim: Juventa.

Technical Report

- Scharl, A., Fischer, L., Gnabms, T., & Rohm, T. (2017). NEPS Technical Report for reading: Scaling results of starting cohort 3 for grade 9 (NEPS Survey Paper No. 20). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Metacognition

Metacognition is the knowledge and control of the own cognitive system. According to Flavell (1979) und Brown (1987), declarative and procedural aspects of metacognition are differentiated which are both covered in the National Education Panel.

Declarative Metacognition

Declarative metacognition refers to knowledge about person, task and strategy variables that an individual can verbalize (Flavell, 1979). This includes, for example, knowledge about the strengths and weaknesses of one's own memory and learning, knowledge about cognitive requirements of tasks (i.e., their difficulty), as well as knowledge about strategies of attaining cognitive learning and achievement goals. It is assumed that the declarative aspect of metacognition constitutes a necessary prerequisite for strategic learning. Knowledge about different kinds of strategies can again be divided into declarative, procedural, and conditional strategy knowledge. Declarative strategy knowledge is the awareness of strategies, that is, the awareness that a certain strategy exists. Procedural knowledge describes how a strategy works effectively and conditional knowledge helps to understand which strategies are more useful for solving a certain task than others (Borkowski, Milstead, & Hale, 1988; Paris, Lipson, & Wixson, 1983).

In the National Educational Panel Study (NEPS), the declarative aspect of metacognition is measured by scenario-based knowledge tests. The construction of the tests is based on existing test instruments that refer to domain-specific knowledge (mostly in the domain of reading, e.g., the test on knowledge about reading strategies, Schlagmüller & Schneider, 2007) or to domain-general knowledge (Neuenhaus, Artelt, Lingel, & Schneider, 2011). These test instruments have been proven to be reliable and economic in use, they refer to concrete learning situations, and are interpretable against a clear benchmark.

The tests on declarative metacognition that are administered in the NEPS include several scenarios describing different school and leisure-time activities. For each scenario, a list of approaches of differing strategic quality is presented and participants are asked to rate the usefulness of each alternative. In order to be appropriate for the different age groups some characteristics of the tests (e.g., the number of the presented alternatives or the context in which the scenarios are embedded) are modified.

Test scoring is done with reference to the relative usefulness of the presented alternatives. Thus, the test instrument can be characterized as a test assessing conditional and relational knowledge about strategies (cf. Händel et al., 2013). The evaluation of the relative usefulness of the strategies is based on the ratings of experts who are scientists in the field of educational psychology and learning strategies. Accordingly, a pair comparison is scored as correct if the judgment on a strategy pair concurs with the expert ratings, and as incorrect if the judgment on a strategy pair contradicts the expert ratings.

Procedural metacognition

Procedural metacognition includes the regulation of the learning process through activities of planning, monitoring and controlling. Within the framework of NEPS in combination with the competence tests of the individual domains, the procedural aspect of metacognition is not assessed as a direct measure of such planning, monitoring and controlling activities but as a metacognitive judgement that refers to the control of the learning performance during (and/or shortly after) the learning phase (also see Nelson & Narens, 1990). After the study participants have taken their competence tests, they are requested to rate their own performance. They are asked to state the portion of questions presumably answered correctly.

Usually, one question is asked per domain. For competence domains that can be divided into coherent individual parts (e.g. reading competence referring to different texts), the inquiry of procedural metacognition is referred to these parts as well, which, of course, leads to a longer processing time.

Bibliography

- Borkowski, J. G., Milstead, M., & Hale, C. (1988). Components of children's metamemory: Implications for strategy generalization. In F. E. Weinert & M. Perlmutter (Eds.), *Memory development: Universal changes und individual differences* (pp. 73–100). Hillsdale, NJ: Erlbaum.
- Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert and R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 65-116). Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Flavell, J. H. (1979). Metacognition and Cognitive Monitoring: A New Area of Cognitive-Developmental Inquiry. *American Psychologist*, *34*, 906-911.
- Händel, M., Artelt, C., & Weinert, S. (2013). Assessing metacognitive knowledge: Development and evaluation of a test instrument. *Journal of Educational Research Online*, *5*(2), 162-188.
- Nelson, T.O. & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G.H. Bower (Hrsg.), *The psychology of learning and motivation* (pp. 125-141). New York: Academic Press.
- Neuenhaus, N., Artelt, C., Lingel, K., & Schneider, W. (2011). Fifth graders metacognitive knowledge: general or domain specific? *European Journal of Psychology of Education*, *26*, 163–178.
- Paris, S. G., Lipson, M. Y., & Wixson, K. K. (1983). Becoming a strategic reader. *Contemporary Educational Psychology*, *8*(3), 293–316.
- Schlagmüller, M., & Schneider, W. (2007). *WLST 7-12. Würzburger Lesestrategie-Wissenstest für die Klassen 7 bis 12*. Göttingen: Hogrefe.

Domain General Cognitive Functions (non-verbal) – Perceptual speed and reasoning

In NEPS, domain general cognitive functions (DGCF) are measured based on the differentiation between “cognitive mechanics” and “cognitive pragmatics” following Baltes, Staudinger and Lindenberger (1999). While the former is measured using task contents as education-independent, new and domain-unspecific as possible, the tasks for measuring cognitive pragmatics are based on acquired skills and knowledge (Ackerman, 1987). Consequently, some of the domain-specific performance tests used within the framework of NEPS may serve as indicators of cognitive pragmatics.

In contrast to this, the tests of DGCF aim at assessing individual differences of fluid cognitive abilities (cognitive mechanics). While these are subject to age-related changes, in comparison to the education- and knowledge-related competences they prove to be less culture-, experience- and language-dependent. In this context, these tests provide an individual basis and differentiating basic function for the acquisition of education-dependent competences.

Among the facets of cognitive mechanics, two common marker variables stand out: perceptual speed and reasoning.

Perceptual speed marks the basal speed of information processing (“*speed*”). In NEPS, this is measured by the Picture Symbol Test (NEPS-BZT). This is based on an improved version of the Digit-Symbol Test (DST) from the tests of the Wechsler family by Lang, Weiss, Stocker and von Rosenblatt (2007). Analogously to this improved version, the NEPS-BZT requires the performance to write the correct number for the given symbols according to an answer key.

Reasoning serves as key marker of mental performance (Baltes et al., 1999). The NEPS reasoning test (NEPS-MAT) is designed as a matrices test in the tradition of the typical reasoning tests. Each item of the matrices test consists of several horizontally and vertically arranged fields in which different geometrical elements are shown – with only one field remaining free. The logical rules on which the pattern of the geometrical elements is based have to be deduced in order to be able to select the right complement for the free field from the offered solutions.

Both tests have been designed in such a way that they can be effectively used without changes to the item sets across as many age groups as possible and relatively independent from the subjects’ first language. Currently, they are administered as paper-and-pencil tests, while computer-assisted administration is generally possible.

The results of both tests provide an estimator of DGCF which, however, is not directly comparable to the overall result of a classical intelligence test (IQ). It rather serves as control of differential initial capacities in the competence acquisition process.

Bibliography

- Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bulletin*, *102*, 3-27.
- Baltes, P. B., Staudinger, U. M. & Lindenberger, U. (1999). Lifespan psychology: Theory and application to intellectual functioning. *Annual Review of Psychology*, *50*, 471-507.
- Lang, F. R., Weiss, D., Stocker, A. & Rosenblatt, B. v. (2007). Assessing cognitive capacities in computer-assisted survey research: Two ultra-short tests of intellectual ability in the Germany Socio-Economic Panel (SOEP). *Schmollers Jahrbuch. Journal of Applied Social Science Studies*, *127*, 183-192.

Listening comprehension at text/discourse level as indicators of linguistic competence in German

The importance of linguistic competence for learning in school as well as for explaining social disparities during school careers is largely undisputed.

In NEPS, linguistic competences in German are measured through listening comprehension at word, sentence, and text/discourse level, on the one hand, and – from 2nd grade elementary school onwards – through reading ability indicators (reading competence, reading speed), on the other hand. However, not all indicators are measured at each survey, and the focus is placed on the coherent assessment of reading competence over the life span.

Listening comprehension at text/discourse level

Beyond listening comprehension at word and sentence levels, listening comprehension at text/discourse level can be regarded as essential for keeping track of classroom activities and thus for acquiring competence at school. Within the NEPS, an indicator of listening comprehension at text level is assessed in Grade 9. The aim is to measure listening comprehension in the majority language according to the literacy approach and on the basis of functional models (cf. Hecker, Südkamp, Leser, & Weinert, 2015).

Listening comprehension at text level is rarely measured in large-scale studies (exception: the National Assessment Studies in Germany, Böhme & Bremerich-Vos, 2012). In NEPS, an important goal is to measure listening comprehension at text level independently of reading competence. This ensures that specific subgroups, such as persons with a migration background or persons with insufficient alphabetization due to a lack of reading competence, are not disadvantaged. Another goal is to implement a valid instrument for assessing listening comprehension. Listening comprehension, according to Buck (2004), is defined as the ability to process shorter and longer sections of real-world spoken language in real time, to extract the main linguistic information from an acoustically presented text, and to draw conclusions that are clearly implied by the content of an aural text. Following this view and the literacy perspective – analogous to the assessment of reading competence in NEPS – listening comprehension is examined in everyday contexts and thus in a manner that is as authentic as possible. In contrast to reading literacy, the aim of listening comprehension is to cover everyday language skills in particular, so that the concept of orality and the conceptual writing of texts are systematically taken into account.

The concept for assessing listening comprehension at text level therefore distinguishes between various aural texts on a continuum of spontaneity of spoken language moving from conceptually more oral language (e.g., discussion) to conceptually more written listening texts (e.g., speech), as shown in Figure 1.

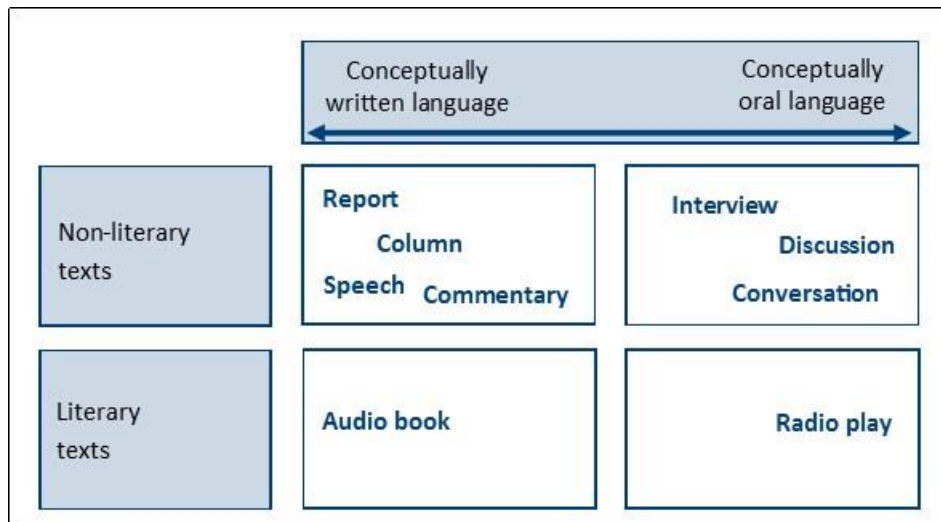


Figure 1: Conceptual classification of listening texts

Two aural texts, an interview and an audio book, were used in the main survey in Grade 9 (Starting Cohort 3). Based on the literature on the processes of text comprehension, the questions asked about the content of the aural texts distinguish between three different cognitive requirements: Questions of Type 1 require the literal understanding of explicit utterances and are directed towards the understanding of shorter utterances on a literal level. Type 2 questions require the listener to make inferences and draw conclusions based on the aural text. Type 3 questions aim at testing the understanding of implicit meanings and statements. These questions can only be answered if the information in the text is combined across different text passages and if the test person has reflected on and evaluated the content of the text (cf. Hecker et al., 2015; for the scaling of the test, see Rohm, Freund, Gnambs, & Fischer, 2017).

During the test, an aural text is presented once audiotively on CD. Subsequently, the first question about the aural text and the corresponding answer alternatives are given twice audiotively on the CD. Participants note their answers in the test booklet by ticking "correct" or "not correct" for the respective statements of the aural text and can check and, if necessary, correct their answers during the second trial. Then, further questions about the aural text follow in the same way. In order to measure listening comprehension at text level independently of reading skills, no text is printed in the test booklets.

Bibliography

- Böhme, K. & Bremerich-Vos, A. (2012). Beschreibung der im Fach Deutsch untersuchten Kompetenzen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik, Ergebnisse des IQB-Ländervergleichs 2011*. Münster: Waxmann.
- Buck, G. (2004). *Assessing listening*. New York: Cambridge University Press.

- Hecker, K., Südkamp, A., Leser, C. & Weinert, S. (2015). Entwicklung eines Tests zur Erfassung von Hörverstehen auf Textebene bei Schülerinnen und Schülern der Klassenstufe 9 (NEPS Working Paper No. 53). Bamberg: Leibniz-Institut für Bildungsverläufe, Nationales Bildungspanel.
- Rohm, T., Freund, M., Gnams, T., & Fischer, L. (2017). NEPS Technical Report for Listening Comprehension: Scaling Results of Starting Cohort 3 for Grade 9 (NEPS Survey Paper No. 21). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Listening Comprehension in the First Languages (L1¹) Russian and Turkish

The effects of immigrant students' first language proficiency on their educational success are still highly disputed. On the one hand, theoretical perspectives and empirical evidence suggest positive effects of L1 proficiency on second language acquisition and on educational success within the country of residence (e.g., Cummins, 1979). On the other hand, neutral and negative effects of L1 proficiency are proposed (e.g., Esser, 2006). The empirical evidence of this controversy is, however, unsatisfactory because there is a lack of investigations systematically assessing L1 proficiency with objective tests (cf. Kristen et al., 2010).

In order to elucidate this controversy within the NEPS, the L1 proficiency of students from the two largest immigrant groups in Germany—that is students whose families immigrated from the area of the Former Soviet Union or from Turkey—is measured with objective tests. The NEPS assesses L1 proficiency at three measure points that are particularly relevant for educational trajectories: at secondary school level in Grade 9 and Grade 7 as well as at elementary school level in Grade 2. The proficiency in Russian and Turkish at these three measure points is assessed with listening comprehension tests specifically developed for this purpose (for Grade 9: Edele, Schotte, Hecht, & Stanat, 2012; Edele, Schotte, & Stanat, 2015; for Grade 7: Taraszow, Schotte, Edele, & Stanat, in preparation). The assessment of listening comprehension was chosen as a dimension of language proficiency because children of immigrants typically acquire the L1 within their family context and do not necessarily read or write their L1.

The L1-tests at secondary school level consist of several independent text units with a length of 100 to 150 words each. Every text unit is followed by four to five questions in a multiple-choice format, which the students were requested to answer. Both, text units and subsequent questions were audio recorded by native speakers of Russian or Turkish and presented to the students in a standardized way from CD. The construction of the L1-tests was based on the aim to assess a broad range of language proficiency. Therefore, texts representing written literary language (expositions and narrations) as well as texts involving oral features (dialogues) were used. In order to ensure that the L1-tests measure language proficiency rather than prior knowledge, the texts either cover topics that should be equally familiar to all students (e.g., everyday situations in school) or topics that are likely to be equally unfamiliar to all students (e.g., an explanation of the living conditions of a rare mammal). All text units were tested and validated by extensive pilot studies.

In order to allow for comparisons of students' first language proficiency between the Grades at secondary school level, two text units are part of both the L1-tests of Grade 7 and the L1-tests of Grade 9 (Taraszow, Schotte, Edele, & Stanat, in preparation). When using these 'anchor items', the objectively assessed L1 proficiency becomes comparable at different measure points of secondary school level.

¹ The term first language (L1) is used interchangeably with the language of the family's country of origin, irrespective of whether the student acquired this language prior to German, as the labeling L1 suggests, or simultaneously.

Bibliography

- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research*, 49(2), 222-251.
- Edele, A., Schotte, K., & Stanat, P. (2015). Listening Comprehension Tests of Immigrant Students' First Languages (L1) Russian and Turkish in Grade 9: Extended Report of Test Construction and Validation (NEPS Working Paper No. 57). Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten* [Language and Integration: Social determinants and consequences of language acquisition of immigrants]. Frankfurt am Main: Campus Verlag.
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., & Will, G. (2011). The education of migrants and their children across the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2), 121-137. doi: 10.1007/s11618-011-0194-3