NEPS
National Educational Panel Study

Research Data

Information on Competence Testing

NEPS Starting Cohort 8 — Grade 5

Education for Tomorrow's World

Wave 2: Grade 6

LIfBi
LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES

Leibniz Institute for Educational Trajectories (LIfBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Cordula Artelt
Administrative Director: Dr. Stefan Echinger
Bamberg; August 14, 2026

Information on testing at mainstream schools	
Sample	Study A105, Starting Cohort 8, Grade 6, Survey wave 2, Year 2023
Test situation	Group testing, normally taking place in the classroom with two test instructors and 24 students at most for each group. The test sequence and the automated time measurement were controlled by the test administrator by using a master PC.
Test sequence	The survey took place on one day. The students first completed tasks in the areas of digital competencies and ICT literacy. After a 15-minute break, a word-level listening comprehension test was administered. Immediately afterwards, students with a corresponding language background received listening comprehension tasks in Russian or Turkish, while the remaining students worked on a general knowledge test. After each of these individual tests, students were asked an additional question regarding procedural metacognition. All tasks were completed on the computer; for the tasks regarding word-level listening comprehension, listening comprehension in Russian or Turkish and general knowledge, headphones were provided. Following the tests, there was another 15-minute break. Finally, the students completed a computer-assisted self-administered student questionnaire (CASI). **Test sequence regular school** - Digital Competence - ICT Literacy - Listening comprehension at word level - General knowledge OR Listening comprehension Russian or Turkish - Student questionnaire (CASI)
Test duration (net processing time)	92 min or 100 min (+ 40 min student questionnaire)
Breaks	2 x 15 min
Administration time	approx. 207 min
Information on the individual tests	

Construct	Number of items	Allowed Processing Time	Survey Mode	Next Measurement
Digital Competence	23	20 min	TBT	2025
ICT Literacy	23 (Stage 1: 8, Stage 2:15) per rotation 18)	28 min	TBT	2025
Listening comprehension at word level	19 sets with 12 tasks each (max.), with a stopping rule	20 min	TBT	
General knowledge	64	20 min	TBT	2028
Listening comprehension Russian or Turkish (L1 Test)	48	28 min	TBT	-
Procedural Metacognition	1 per test	1 min each	TBT	

Information on testing at special schools	
Sample	Study A105, Starting Cohort 8, Grade 6, Survey wave 2, Year 2023
Test situation	Group testing, normally taking place in the classroom with two test instructors and 12 students at most for each group. The test sequence and the automated time measurement were controlled by the test administrator by using a master PC.
Test sequence	The survey took place on one day. The students first completed tasks in digital competencies. After a 15-minute break, tasks on word-level listening comprehension and general knowledge followed. All tasks were completed on the computer; for the tasks regarding word-level listening comprehension and general knowledge, headphones were provided. After each of these individual tests, students were asked an additional question regarding procedural metacognition. Following the tests, there was a 15-minute break. Finally, the students completed a computer-assisted self-administered student questionnaire (CASI). **Test sequence special school** - Digital Competence - Listening comprehension at word level - General knowledge - Student questionnaire (CASI)
Test duration (net processing time)	63 min (+ 30 min student questionnaire)
Breaks	2 x 15 min
Administration time	approx. 145 min

Information on the individual tests

Construct	**Number of items**	**Allowed Processing Time**	**Survey Mode**	**Next Measurement**
Digital Competence	19	20 min	TBT	2025

Listening comprehension at word level	19 sets with 12 tasks each (max.), with a stopping rule	20 min	TBT	
General knowledge	64	20 min	TBT	2028
Procedural Metacognition (Meta-P)	1 per Construct	1 min each	TBT	

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.

Compared to regular schools, the digital competence test used a different booklet for special schools with partially easier items.

Digital Competence

In light of the ongoing digital transformation of society, the competent use of digital media is becoming increasingly important for successful social participation. To account for this development, NEPS assesses Digital Competence in addition to ICT literacy (see the description in this document; for more detailed information, see Senkbeil et al., 2013), with the aim of both constructs complementing one another (Artelt, 2025).

The NEPS conceptualizes Digital Competence as the responsible and reflective use of digital technologies, grounded in comprehensive knowledge of and critical reflection on the content, mechanisms, and effects of digital media and technologies (Lockl et al., in prep.; Tural & Schwaß; 2023). Within this framework, digital competence is organized into four facets (see Figure 1). (1) Evaluating Information: involves assessing the credibility of sources and understanding of how algorithmic filtering shapes the access to and visibility of information. It also includes a basic understanding of how systems based on artificial intelligence (AI) operate. (2) Detecting Intentions and Strategies: focuses on recognizing intentions that are designed to influence behavior, including commercial, social, or political intentions. It also encompasses knowledge on strategies and mechanisms of influence, including phenomena of manipulation of and influence on opinions and the restriction of diversity of opinion through filter bubbles and echo chambers. (3) Communicating and Interacting: focuses on effective and responsible collaboration and social engagement in digital contexts, including the awareness of potential social consequences, such as phenomena like hashtag debates, fake profiles, cyberbullying, as well as shitstorms and candystorms. (4) Handling Data: addresses the responsible management of one's own data including privacy settings, questions of data collection, ways to reduce digital traces, and issues related to sharing personal information. It also refers to the consideration of other people's data and intellectual property, including ethical and legal aspects.

Fig. 1: Facets of the Framework for Digital Competence in the NEPS

Item types and formats

These facets are operationalized through scenario-based items that reflect typical digital situations experienced by lower secondary students. The items were presented in both text- and image-based formats, with both types implemented as simple and complex multiple-choice items. The visual material used in the image-based items depicts realistic digital environments, such as reconstructed websites or digital platforms, providing authentic contexts in which students' knowledge and understanding of digital processes and practices can be assessed.

When compiling test booklets for an assessment, the aim is to ensure a balanced representation of the four facets of the framework. Within each facet, however, the selection and weighting of the aspects described here may vary across assessments depending on the target population.

Scaling of the items

For estimating item parameters a Rasch model is used. In order to compare competencies across different measurement occasions, the different measurements are linked. The scaling results of the test booklets and items for Grade 6 are described in the technical reports (Kutscher, 2026a, 2026b).

References:

Artelt, C. (2025). Bildung in Zeiten von Digitalität und Künstlicher Intelligenz. In: N. McElvany, C. Dignath, M. Becker, u. a. (Hrsg.), *Welche Kompetenzen soll die Schule von heute für die Gesellschaft von morgen vermitteln?* (S. 41–55). Münster: Waxmann. doi: 10.31244/9783818850111.04.

Kutscher, T. (2026a). NEPS technical report for digital competence: scaling results of starting cohort 8 for grade 6 (NEPS Survey Paper No. 137). Leibniz Institute for Educational Trajectories. https://doi.org/10.5157/NEPS:SP137:1.0.

Kutscher, T. (2026b). NEPS technical report for digital competence: scaling results of starting cohort 8 for grade 6 in special schools (NEPS Survey Paper No. 136). Leibniz Institute for Educational Trajectories. https://doi.org/10.5157/NEPS:SP136:1.0.

Lockl, K., Tural, S., Schwaß, M., Kutscher, T., Hornberger, M., Artelt, C., & Wolter, I. (in prep.). Measuring digital competence in the NEPS - Conceptual framework and findings from a pilot study.

Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013). The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*, *5*, 139–161.

Tural, S., & Schwaß, M. (2023, February 28–March 3). Digital natives = digital experts? Designing a novel instrument to measure digital competence in NEPS-SC8 secondary school students [Posterpräsentation]. 10. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Universität Duisburg-Essen, Deutschland.

ICT Literacy

The ability to effectively use ICT (information and communication technologies) plays an important role in schools, many workplaces and in people's everyday lives (Fraillon, AInley, Schulz, Friedman, & Duckworth, 2019). Therefore, researchers as well as organizations (e.g., European Commission, International Society for Technology in Education) have developed frameworks to promote ICT literacy by describing competencies and skills that are considered important for the knowledge society (Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016). More recent conceptualizations of ICT literacy integrate technological and cognitive aspects to define this competence. Technological aspects encompass the knowledge of hardware and software applications and understanding technological concepts. Cognitive aspects, labeled as information literacy, encompass the ability to use digital media to access, create, manage and critically evaluate information and to use it effectively for one's own purposes, also plays an important role (ETS, 2002).

Thus, ICT literacy is understood as a meta-competence that helps people to acquire important competencies and skills for educational and work situations and to achieve private goals over the entire life span (van Laar, van Deursen, van Dijk, & de Haan, 2017). Furthermore, since people need to acquire new knowledge and skills in a self-regulated way over the life span that are increasingly mediated through digital media, ICT literacy is a necessary prerequisite for successfully keeping pace with recent developments in the area of digital media (Goldhammer, Gniewosz, & Zylka, 2017).

One widely used definition of ICT literacy to which we also refer was formulated by the *ICT Literacy Panel*: „ICT literacy is the ability to appropriately use digital technology, communication tools, and/or networks to solve information problems in order to function in an information society. This includes having the ability to use technology as a tool to research, organize, and communicate information" (ETS, 2002, p. 16).

In the context of NEPS, *ICT Literacy* is conceptualized as a unidimensional construct comprising the facets of process components and software applications (see Figure 1; Senkbeil & Ihme, 2020; Senkbeil, Ihme & Wittwer, 2013a, b). As a basis for constructing the instrument assessing computer literacy in NEPS, we use a framework that identifies four process components of computer literacy representing the knowledge and skills needed for a problem-oriented use of modern information and communication technology. Each process component integrates technological and cognitive aspects of the construct. The process components are defined as follows:

Access: knowledge of basic operations used to retrieve information (e.g., entering a search term in an internet browser, opening and saving a document);

Create: the ability to create and edit documents and files (e.g., setting up tables, creating formulas);

Manage: the ability to find information within a program (e.g., retrieving information from tables, processing the hits returned by a search engine);

Evaluate: the ability to assess information and to use it as the basis for informed decisions (e.g., assessing the credibility of the information retrieved).

Apart from the process components, the construction of the NEPS tests for ICT Literacy is guided by a categorization of software applications that are used to locate, process, present, and communicate information: (a) word processing and operating systems, (b) spreadsheet and presentation software,

(c) e-mail and other communication applications, and (d) internet and internet-based search engines. Each item in the test refers to one process component and one software application.

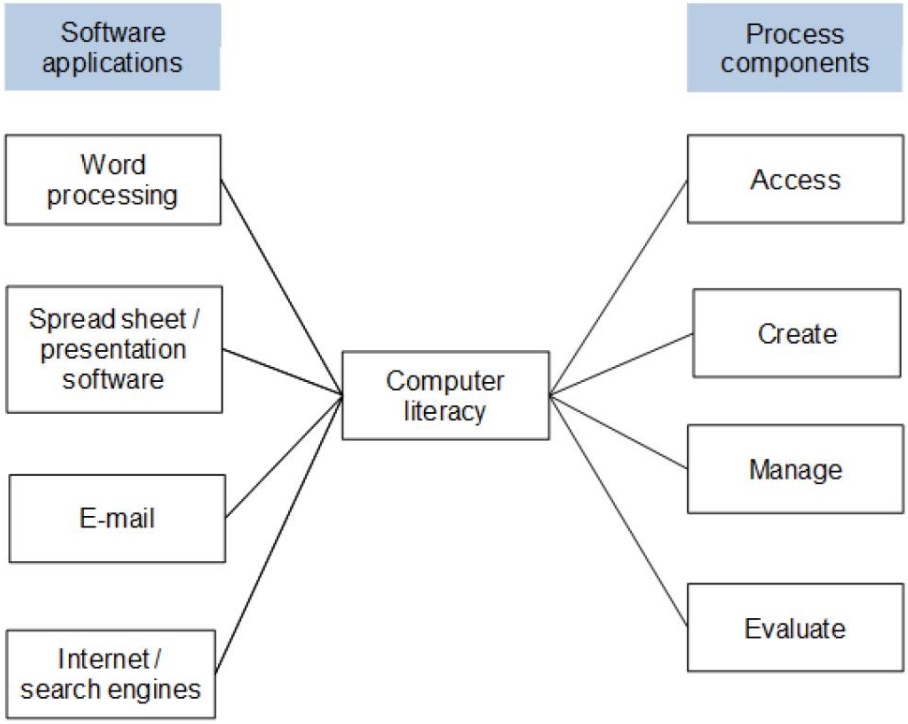

Fig. 2: Assessment framework for ICT Literacy in the German National Educational Panel Study

Item format

The assessment of computer literacy contains two types of tasks.

The first type of tasks, knowledge-based and *static items*, which account for about half of the tasks, were pencil-and-paper questions. These task addressed factual (e.g., computer terminology) and conceptual knowledge (classifications, and principles), whether the test subjects can deal appropriately with certain computer-based tasks (e.g. saving a file on a specific drive). To do so, participants were presented with realistic problems embedded in a range of authentic situations. About half of the items used screenshots, for example, of an internet browser, an electronic database, or a spreadsheet as prompts (see Senkbeil et al., 2013). The static items included two types of response formats: simple multiple choice (MC) and complex multiple choice (CMC) items. In MC items the test taker had to identify the correct answer out of four to six response options with one option being correct and three to five response items functioning as distractors (i.e., they were incorrect). In CMC items several subtasks with two response options each (true / false) were presented.

The second type of tasks were *interactive items* of simulations of generic software or universal applications to complete an action. These interactive items additionally addressed procedural and strategic knowledge (e.g., planning, executing, and monitoring the problem-solving process). Respondents were required to solve specific tasks and problems by using and interacting with these

simulations: These may be single-action tasks (such as opening a web-browser) or may contain a sequence of steps (such as "Save As" with a specific file name, sorting or filtering a web-based database according to one or more criteria). The interactive items contained linear and nonlinear tasks. Linear tasks required the execution of at least two commands executed in a necessary prescribed sequence (e.g., opening a file from the desktop, "Save As" with a specific file name, and moving the file to another drive). Nonlinear tasks required respondents to reach a desired outcome by executing various subcommands, with the order of the commands variable.

Scaling of the tests

For estimating item and person parameters for ICT literacy a Rasch model is used. In order to compare competencies across different measurement occasions and examine competence development over time the different measurements are linked (Fischer, Rohm, Gnambs & Carstensen, 2016). The psychometric quality and the scaling results of the tests and items are described in the technical reports of each starting cohort.

Bibliography

ETS [Educational Testing Service] (2002). *Digital transformation. A framework for ICT literacy. A report of the International ICT Literacy Panel*. Princeton, NJ: ETS.

Fischer, L., Rohm, T., Gnambs, 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.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T. & Duckworth, D. (2019). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 International Report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).

Goldhammer, F., Gniewosz, G., & Zylka, J. (2017). ICT engagement in learning environments. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning. An international perspective* (pp. 331–354). Springer International Publishing.

Senkbeil, M. & Ihme, J. M. (2020). Diagnostik von ICT Literacy: Messen Multiple-Choice-Aufgaben und simulationsbasierte Aufgaben vergleichbare Konstrukte? Vergleich der Testergebnisse zweier Instrumente aus den aktuellen Large-Scale-Studien ICILS 2013 und NEPS [Assessment of ICT literacy: Do multiple-choice tasks and simulation-based tasks measure comparable constructs? Comparison of the test results of two large-scale instruments: ICILS 2013 versus NEPS]. *Diagnostica, 66*, 147–*157.*

Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013a). The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*, *5*, 139--161.

Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013b). Entwicklung und erste Validierung eines Tests zur Erfassung technologischer und informationsbezogener Literacy (TILT) für Jugendliche am Ende der Sekundarstufe I. [Development and validation of a technological and informational literacy test (TILT) for end of secondary school] *Zeitschrift für Erziehungswissenschaft, 16,* 671-691.

Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review, 19*, 58–84.

van Laar, E., van Deursen, A.J.A.M., van Dijk, J.A.G.M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior, 72*, 577−588.

Vocabulary: Listening comprehension at word level

Listening comprehension at word, sentence and text/discourse level as indicators of linguistic competence in German

The importance of linguistic competences for school learning and for explaining social disparities in school careers is largely undisputed.

In the National Educational Panel Study (NEPS), German linguistic competences are captured, on the one hand, via listening comprehension at word, sentence and text/discourse level and, on the other hand, (from 2nd primary school class onwards) via indicators of reading skills (reading competence [text comprehension], reading speed). Listening comprehension at word level is assessed once in grade 6 for starting cohort 8.

Listening comprehension at word level (receptive vocabulary)

Measures of receptive vocabulary are good and internationally applicable indicators of language skills and abilities acquired by children and adults. In numerous large international panel studies, passive vocabulary is collected as the central and sometimes sole indicator of cumulatively acquired linguistic-cognitive abilities taking into consideration individual basic skills (e.g., working memory capacity, speed variables) and environmental stimuli. Examples of such studies are the Head Start Family and Child Experiences Survey – FACES (USA)[1], the National Longitudinal Survey of Children and Youth – NLCSY (Canada; among others Lipps & Yiptong-Avila, 1999), the British Cohort Study – BCS70 (e.g., Bynner, 2004) or the European Child Care and Education (ECCE) Study, which is conducted in Germany, Austria, Spain and Portugal (e.g., European Child Care and Education (ECCE) Study Group, 1997).

The internationally most frequently used instrument for assessing receptive vocabulary is the Peabody Picture Vocabulary Test (PPVT), which is available in various versions (Dunn, 1959; Dunn & Dunn, 1981, 1997, 2007). Generally, the PPVT can be used for a very large age range (from 2.5 years to late adulthood) and is both easy to carry out and to analyze. A German version of the PPVT-IV (Dunn & Dunn, 2007; German version by Lenhard, Lenhard, Segerer, & Suggate, 2015) was used in Starting Cohort 8. In the NEPS, the test was administered via a PC and headphones. The student's task was to select one picture out of four, matching it to the word presented as an auditory cue by the PC.

According to the guidelines of the PPVT-IV, the level of difficulty (test entry and termination) varies depending on the age and performance of the students. The practice unit at the start of the test also varies depending on the age and performance of the students. Once a student has solved at least two tasks correctly during the practice phase, he or she moves on to the test phase. The test consists of a total of 19 sets with staggered levels of difficulty, each set consisting of 12 items.

Testing procedure in this wave: The test begins with a practice phase consisting of at least two and a maximum of six tasks. The starting set depends on the age of the students. If the student makes more than one mistake in the starting set, the next lower set follows until a maximum of one mistake is made in one set (basal set). The test is then carried out until the ceiling set – the set in which the student has made more than seven mistakes – has been identified; the sets that have already been processed are left out.

[1] http://www.acf.hhs.gov/programs/opre/hs/faces/

The Scientific Use File contains the number of administered practice items, all scored test items (correct/false), the basal set, and the ceiling set. In addition, the sum score is included, indicating the number of correctly solved items. All items that are in lower sets than the basal set are assumed to be correct.

References

Bynner, J. (2004). Participation and progression: use of British Cohort Study data in illuminating the role of basic skills and other factors. *Nuffield Review of 14-19 Education and Training*, Working Paper 9.

Dunn, L. M. (1959). *Peabody Picture Vocabulary Test (PPVT). Manual of directions and forms.* Nashville, TN: American Guidance Service.

Dunn, L. M. & Dunn, D. M. (1981). *Peabody Picture Vocabulary Test-Revised (PPVT-R).* Circle Pines, MN: American Guidance Service.

Dunn, L. M. & Dunn, D. M. (1997). *Peabody Picture Vocabulary Test, Third Edition (PPVT-III).* Circle Pines, MN: American Guidance Service.

Dunn, L. M. & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test, Fourth Edition (PPVT-4).* Upper Saddle River, NJ: Pearson.

European Child Care and Education (ECCE) Study Group. (1997). *European Child Care and Education.* Berlin, Germany: Freie Universität Berlin, Department of Education.

Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody Picture Vocabulary Test-4.* Frankfurt am Main, Germany: Pearson Assessment.

Lipps, G. & Yiptong-Avila, J. (1999). From home to school: how Canadian children cope. *Education Quarterly Review, 6*(2), 51–57.

General Knowledge

In addition to assessing functional, domain-specific competencies and other areas of competence, the National Educational Panel Study (NEPS) also takes students' general knowledge into account. According to Cattell and Horn's theory of fluid and crystallized intelligence (Horn, 1965, 1988; Cattell, 1971) and the CHC model (McGrew, 2009), general knowledge is considered a "crystallized" ability whose development depends substantially on previously acquired knowledge, learning and education.

To assess the general knowledge of sixth-grade students in Starting Cohort 8, a parallel version of the "General Knowledge" subtest from the *Berlin Test of Fluid and Crystallized Intelligence for Grades 5 to 7* (BEFKI 5–7; Schroeders, Schipolowski, & Wilhelm, 2020) is administered.

This test measures declarative factual knowledge across 16 different content areas, which can be assigned to the three broad domains of natural sciences, humanities, and social sciences (see Table 1). The selection of content areas by the test developers and their assignment to these broad domains were guided by Ackerman's (2000) empirical classification and a review of existing assessment instruments (Amthauer, Brocke, Liepmann, & Beauducel, 2001; Hossiep & Schulte, 2007). The aim of the test development was to assess declarative knowledge as broadly as possible while taking into account both knowledge typically acquired through formal schooling and knowledge gained outside the school context. This includes knowledge that is generally valued in society and regarded as part of the cultural heritage. Consequently, it is distinguished from transient or only partially relevant knowledge, such as current football scores or regional bus schedules. Given the age of the target population, areas of knowledge related to highly specialized academic expertise, which is typically acquired only in tertiary education, were excluded (for example, the domain of psychology in Ackerman, 2000).

Table 1

Knowledge areas assessed, their content, and assignment to the three broad knowledge domains

Domain	Content area	Content (examples)
Natural Sciences	Physics	Physical quantities and phenomena (e.g., energy and optics)
	Chemistry	Organic and inorganic chemistry (e.g., properties of substances, chemical symbols)
	Biology	Biological processes, the animal kingdom, and the human body (e.g., anatomy, characteristics of specific animal species)
	Medicine	Human diseases and their treatment (e.g., symptoms and modes of disease transmission)
	Geography	Physical geography and geology (e.g., location of specific bodies of water, characteristics of rocks)
	Technology	Technologies, technical devices, and communication (e.g., energy generation, the Internet)
Humanities	Literature	Literary works and text genres

	Art	Artistic techniques and color theory
	Music	Musical instruments and styles, knowledge of musical works
	Religion	Concepts and key representatives of major world religions
	Philosophy	Influential philosophers and central philosophical questions
Social Sciences	History	Historical events and figures spanning several millennia
	Law	Legal institutions and laws in Germany (e.g., consumer rights, copyright law)
	Politics	Political institutions and elections (e.g., the Federal Government of Germany, the European Union)
	Economics	Principles and concepts of macroeconomics and business administration (e.g., trade, labor unions)
	Finance	Financial terms (e.g., stock markets, currencies)

The test consists of a total of 64 items, with each of the 16 knowledge areas represented by four items. Participants are given 20 minutes to complete the test. All questions are designed as multiple-choice items, each providing four response options, of which exactly one answer is correct.

Bibliography

Ackerman, P. L. (2000). Domain-specific knowledge as the „dark matter" of adult intelligence: Gf /Gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences*, *55B*, 69-84. https://doi.org/10.1093/geronb/55.2.P69

Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.

Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston, MA: Houghton Mifflin.

Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Dissertation, University of Illinois.

Horn, J. L. (1988). *Thinking about human abilities. In J. R. Nesselroade (Ed.), Handbook of multivariate psychology* (pp. 645-685). New York: Academic Press.

Hossiep, R. & Schulte, M. (2007). *BOWIT – Bochumer Wissenstest*. Göttingen: Hogrefe.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*, 1–10. https://dx.doi.org/10.1016/j.intell.2008.08.004

Schroeders, U., Schipolowski, S., & Wilhelm, O. (2020). *BEFKI 5-7: Berliner Test zur Erfassung fluider und kristalliner Intelligenz für die 5. bis 7. Jahrgangsstufe*. Hogrefe.

Listening comprehension in Russian and Turkish as Heritage Languages (L1[2])

The relationship between proficiency in the heritage language (L1) and the educational success of individuals from immigrant families remains contested. On the one hand, theoretical approaches and empirical findings suggest that L1 proficiency may positively affect second language acquisition and educational success in the host country (e.g., Cummins, 1979; Edele et al., 2023; Edele & Stanat, 2016; Kempert et al., 2016). On the other hand, negative or nonsignificant associations have also been reported (e.g., Edele et al., 2023; Esser, 2006). The empirical evidence remains insufficient partly because L1 proficiency has rarely been assessed systematically using objective measures (see Edele, Seuring, et al., 2015; Kristen et al., 2011).

To investigate these assumptions within NEPS, receptive proficiency in the heritage languages Russian and Turkish is assessed. The tests are administered to students who speak either Russian or Turkish in addition to German and whose families immigrated to Germany from Turkey or countries of the former Soviet Union. The listening comprehension tests administered in Starting Cohort 8 constitute a further development of the L1 tests previously used in Starting Cohorts 3 and 4 (Tinkl et al., in preparation). Listening comprehension was selected as the assessment modality because the L1 is often acquired primarily within the family and its acquisition does not necessarily include the development of literacy skills.

The revised L1 tests comprise independent text units adapted to students in Grade 6. Because the previous tests were comparatively easy, particularly for students with high levels of L1 proficiency, and provided limited differentiation at the upper end of the proficiency distribution (Edele et al., 2012; Edele, Schotte, et al., 2015), additional, more demanding text units were developed. To assess a broad proficiency range while maintaining an efficient testing time, the SC8 L1 tests were implemented using a *branched-adaptive design*. Unlike fully adaptive tests, in which individual items are continuously selected on the basis of previous responses, a branched-adaptive design assigns students to a test path after they have completed a predefined set of items (Kubinger & Wild, 1989). In the first stage, all students completed a common *routing stage* comprising three text units and a total of 14 items. Students who answered at least nine items correctly were assigned to the *difficult branch*; all other students were assigned to the *easy branch*. They subsequently completed five additional text units, meaning that each student completed eight text units in total.

The tests were implemented as technology-based assessments (TBA) and completed individually on a computer using headphones. The listening texts were recorded by professional speakers whose first language is Russian or Turkish, and were presented as standardized audio recordings. The corresponding questions and response options were also presented auditorily while simultaneously being displayed on the screen. The listening texts themselves were not presented in written form. Students answered the multiple-choice items directly on the computer. The tests were designed to assess listening comprehension as broadly as possible. They therefore included both conceptually written texts in the form of narratives and expository texts and conceptually oral texts in the form of dialogues. The listening texts addressed both everyday and school-related topics. To minimize confounding between test performance and students' prior knowledge, the selected content was

[2] The term L1 („first language") is used here synonymously with the language of the family's country of origin, irrespective of whether it was in act the first language acquired by the target person.

either expected to be similarly familiar to all students or equally unfamiliar to all. The newly developed text units and items were evaluated in pilot studies.

References

Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children *Review of Educational Research 49*(2), 222–251. https://doi.org/https://doi.org/10.3102/00346543049002222

Edele, A., Schotte, K., Hecht, M., & Stanat, P. (2012). Listening comprehension tests of immigrant students' first language (L1) Russian and Turkish in grade 9: Scaling procedure and results. *NEPS Working Paper No. 13*. https://doi.org/10.5157/NEPS:WP13:1.0

Edele, A., Schotte, K., & Stanat, P. (2015). Assessment of immigrant students' listening comprehension in their first languages (L1) Russian and Turkish in grade 9: Extended report of test construction and validation. *NEPS Working Paper No. 57*. https://doi.org/10.5157/NEPS:WP57:1.0

Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, *52*, 99–123. https://doi.org/org/10.1016/j.ssresearch.2014.12.017

Edele, A., Seuring, J., Schotte, K., Kristen, C., & Stanat, P. (2023). Is the first language a resource, an obstacle, or irrelevant for language minority students' education? In S. Weinert, G. J. Blossfeld, & H.-P. Blossfeld (Eds.), *Education, Competence Development and Career Trajectories: Analysing Data of the National Educational Panel Study (NEPS)* (pp. 349–367). Springer. https://doi.org/10.1007/978-3-031-27007-9_16

Edele, A., & Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, *108*(2), 163–180. https://doi.org/10.1037/edu0000060

Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten* (1. Aufl. ed.). Campus Verlag.

Kempert, S., Edele, A., Rauch, D., Wolf, K. M., Paetsch, J., Darsow, A., Maluch, J., & Stanat, P. (2016). Die Rolle der Sprache für zuwanderungsbezogene Ungleichheiten im Bildungserfolg [The role of language for immigration-related disparities in academic achievement]. In C. Diehl, C. Hunkler, & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten* (pp. 157–241). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-04322-3_5

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. https://doi.org/10.1007/s11618-011-0194-3

Kubinger, K. D., & Wild, B. (1989). Die Optimierung der Meßgenauigkeit beim "branched"-adaptiven Testen. In K. D. Kubinger (Ed.), *Moderne Testtheorie* (pp. 187–218). Psychologie Verlags Union.

Tinkl, L. A., Jindra, C., & Edele, A. (in preparation). *Assessment of immigrant students' listening comprehension in their first language (L1) Russian and Turkish in Grade 6 in Starting Cohort 8: further development, test construction and validation.* (NEPS Working Paper)

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.

Procedural metacognition

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

In Starting Cohort 8, this self-assessment is conducted using an 11-point scale. The ratings are coded from 0 (presumably no tasks solved correctly) to 10 (presumably all tasks solved correctly).

Bibliography

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.

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.