NEPS

National Educational Panel Study

Research Data

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 9: Grade 12

LIfBi

LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES

Leibniz Institute for Educational Trajectories (LIfBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Cordula Artelt
Executive Director of Research: Dr. Jutta von Maurice
Executive Director of Administration: Dr. Robert Polgar
Bamberg; December 19, 2019

	Information on testing		
Test setting	Assessment in schools (Main Study A101), students attending Grade 12 in general schools including students who repeated or skipped a grade level.		Assessment at the study participant´s home for students in the individual field (Main Study B108); including students who changed or left school after Grade 9, 10 and 11 (without students formerly attending schools for students with special educational needs (SEN)[1]).
Test situation	Paper based group testing in classrooms of the regular school, normally with 1 test instructor and 1 supervisory teacher.		Paper and tablet based testing, personal interview held at the study participant´s home, 1 test instructor
Tests	Paper based tests in the domains reading competence, mathematical competence , ICT literacy, English and scientific thinking (+ procedural metacognition, one or more judgements after each test)		Paper based tests in the domains reading competence, mathematical competence and tablet based test in the domain ICT literacy (+ procedural metacognition, one or more judgements after each test)
Test sequence	The tests were administered on one day. All five tests were presented to all participants. The tests were predetermined in four different sequences. The sequence of the domains ICT literacy - reading competence was the same as in Grade 9. The order of the domains English - scientific thinking was random. After the tests, all students filled out a questionnaire. Sequence 1: ICT literacy, reading competence, mathematical competence, English, scientific thinking Sequence 2: reading competence, ICT literacy, mathematical competence, scientific thinking , English Sequence 3: ICT literacy, reading competence, mathematical competence, scientific thinking , English		The tests were administered on one day. All participants received two tests. After the test, all participants filled out a questionnaire. Students of the school type Grammar school („Gymnasium“) and apprentice in the last year of training were asked to participate in a follow-up questionnaire after the survey. Sequences : Mathematical competence or/and reading competence or/and ICT literacy. The test on mathematic and reading competence have two difficulty levels and the test on ICT literacy has three difficulty levels. The rotations were assigned to the

	Sequence 4: reading competence, ICT literacy, mathematical competence, English, scientific thinking The difficulty levels of the test on reading competence (“reading competence 1 easy” or “reading competence 2 difficult”) were assigned to the students depending on their performance in the previous reading test in Grade 9. The assignment of sequence 1 or 2 was made according to the sequence of the survey in Grade 9. All in all, there were eight rotations regarding difficulty and sequences.		students depending on their performance in the previous competence test in their schooldays. All in all, there were 32 rotations/booklets regarding difficulty and sequences.
Test duration (net processing time)	150 min + 40 min questionnaire		60 min test + 30 min questionnaire
Breaks	10 min break between first and second test; 15 min break after third test and 15 min break before the questionnaire		Between two tests maximum 10 min breaks possible
	Information on the individual tests		

Construct	Number of items	Allowed processing time	Survey method	Next Measurement
Mathematical competence	22	28 min	paper pencil	
Domain specific procedural metacognition regarding mathematical competence	1	1 min	paper pencil	
Reading competence	29	28 min	paper pencil	
Domain specific procedural metacognition regarding reading competence	6	2 min	paper pencil	

ICT literacy	32 / 20[2]	28 / 26 min	paper pencil / TBT	
Domain specific procedural metacognition regarding ICT literacy	1 / 2	1 / 2 min	paper pencil / TBT	
English reading competence	30	29 min	paper pencil	
Domain specific procedural metacognition regarding English reading competence	1	1 min	paper pencil	
Scientific thinking	32	29 min	paper pencil	
Domain specific procedural metacognition regarding scientific thinking	1	1 min	paper pencil	

[1] Former participants in schools for students with special educational needs (SEN) did not complete the tests, but did the questionnaire by telephone interview (CATI).

[2] The first number refers to the ICT literacy test in main study A101, the second number to the test in sthe main study B108. In the main study B108, the test included simulation-based items in addition to MC items.

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.

Mathematical competence

In the National Education Panel Study, the construct of mathematical competence is based on the idea of mathematical literacy as was defined, for example, in PISA. Thus, the construct describes "[...] an individual's capacity to identify and understand the role that mathematics plays in the world, to make well-founded mathematical judgments and to use and engage with mathematics in ways that meet the needs of that individual's life as a constructive, concerned and reflective citizen." (OECD, 2003, 24). Regarding younger children, this idea refers to competent handling of mathematical problems in age-specific contexts.
Accordingly, mathematical competence in NEPS is operationalized by items assessing more than pure mathematical knowledge; instead, solving the items requires recognizing and flexibly applying mathematics in realistic, mainly extra-mathematical situations.

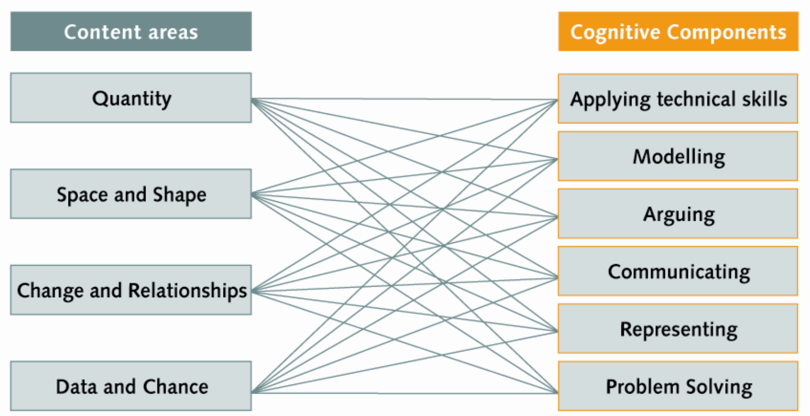

Fig. 1: Framework of mathematical competence in NEPS

The NEPS framework of mathematical competence distinguishes between content-related and process-related components (cf. Fig. 1). In detail, the content areas are characterized as follows:

- Quantity comprises all kinds of quantifications when numbers are used to organize and describe situations.
  Examples from the elementary sector: comparisons of sets, counting (ordinal/cardinal aspects of numbers), simple operations (e.g., adding)
  Examples from the adult sector: calculations of percentages and interests, calculations of area and volume, use of different units, simple equation systems
- Space and Shape includes all types of planar and spatial configurations, shapes or patterns.
  Examples from the elementary sector: recognizing geometric shapes, simple properties of shapes, perspective
  Examples from the adult sector: three-dimensional mathematical objects, geometric mappings, elementary geometric theorems
- Change and Relationships includes all kinds of (functional) relationships and patterns.
  Examples from the elementary sector: recognizing and continuing patterns, relationships among numbers, proportionality
  Examples from the adult sector: interpreting curves or function graphs, properties of linear, quadratic, and exponential functions, extremum problems

- Data and Chance comprises all situations involving statistical data or chance.
  Examples from the elementary sector: intuitively assessing probabilities, collecting and structuring data
  Examples from the adult sector: interpreting statistics, basic statistical methods, calculating probabilities

The cognitive components of mathematical thinking processes are distinguished as follows:

- Applying technical skills includes using known algorithms and remembering mathematical knowledge or calculation methods.
- Modelling includes the representation in a situation model and in a mathematical model as well as interpreting and validating results in real-life situations.
- Arguing includes assessing explanations and proofs, but also developing own explanations or proofs.
- Communicating requires communication on mathematical contents and includes, among other things, the correct and adequate use of mathematical technical terms.
- Representing comprises the use and interpretation of mathematical representations such as tables, charts or graphs.
- Problem Solving takes place, when there is no obvious approach, and, therefore, includes systematic trying, generalizing or examining special cases.

The test items used in NEPS refer to one content area that is mainly addressed by the item, but may well contain several cognitive components (further description of the framework in Neumann et al., 2013). This differentiation renders the framework concept of mathematical competence in NEPS compatible with both the PISA studies and the German National Mathematics Education Standards. Some literature also show a high correlation between NEPS, the PISA studies and federal states comparisons from the Institute of Educational Quality Improvement (IQB): r = .89 for NEPS-PISA and r = .91 for NEPS-IQB (van den Ham, 2016).

Bibliography

Neumann, I., Duchhardt, C., Grüßing, M., Heinze, A., Knopp, E., & Ehmke, T. (2013). Modeling and assessing mathematical competence over the lifespan. Journal for Educational Research Online, 5(2), 80–109. Retrieved http://journal-for-educational-research-online.com/index.php/jero/article/view/362.

Organisation for Economic Co-Operation and Development [OECD] (2003). The PISA 2003 assessment framework – mathematics, reading, science and problem solving knowledge and skills. Paris: OECD.

Van den Ham, A.-K. (2016). Ein Validitätsargument für den Mathematiktest der National Educational Panel Study für die neunte Klassenstufe. Unpublished doctoral dissertation, Leuphana University Lüneburg, Lüneburg.

ICT Literacy

The ability to effectively use information and communication technologies (ICT) not only plays an important role in many workplace settings, but is also becoming increasingly important in people's everyday lives (ETS, 2002; Wittwer & Senkbeil, 2008). Because of the rapidly changing technological environment, self-regulated and continuous life-long learning is a key or meta competence for successfully keeping pace with recent developments in the area of ICT (e.g., Blossfeld, Doll, & Schneider, 2008).

More recent conceptualizations are not exclusively confined to technological literacy, that is, knowledge of hardware and software applications and understanding technological concepts. Instead, information literacy, that is, 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 to be understood as a functional literacy that helps people to acquire other important competencies and skills for professional success (educational and work settings) and to achieve private goals across the lifespan. 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" (Katz, 2007, p. 6).

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 & 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.

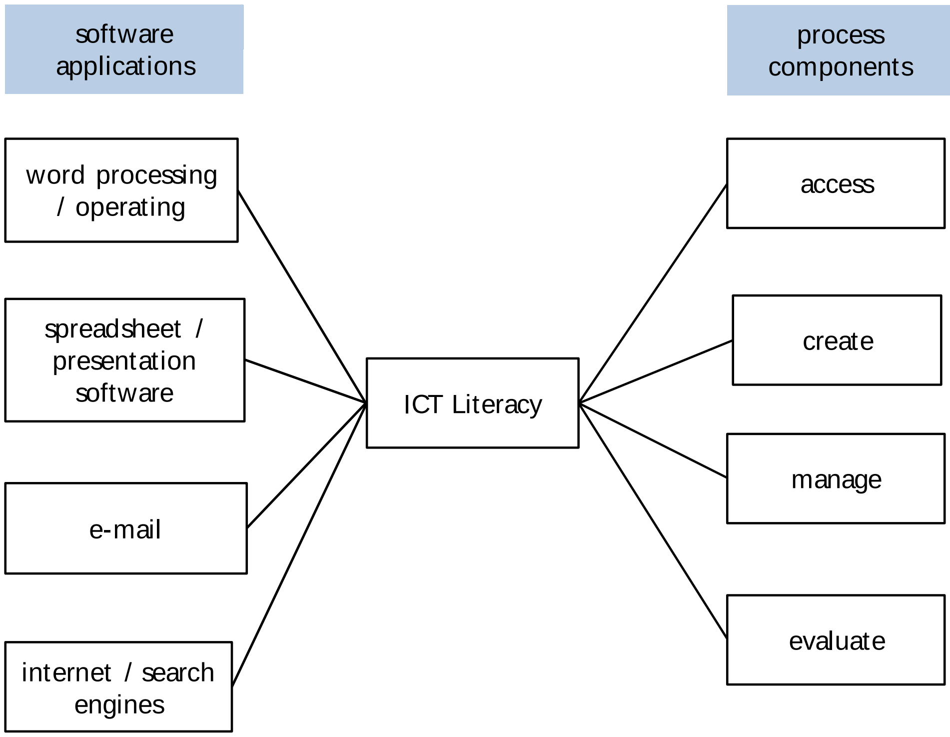

Figure. 1: Assessment framework for ICT Literacy in the German National Educational Panel Study

Item types and formats

In the first project stage (until 2017), ICT literacy tests were administered paper-based in the form of multiple-choice tasks. Since 2018, additional computer-based and interactive tasks are used for an extended and optimized construct representation. This extension in task format can be understood as a transition from knowledge-based (based on MC tasks) to knowledge- and behavior-based construct representation (based on additional simulation tasks) (Senkbeil & Ihme, 2019).

Using MC tasks, ICT-related declarative and procedural knowledge can be measured as well as procedural, application-oriented knowledge. The test persons have to deal with realistic problems embedded in a range of authentic situations. To do so, most items used screenshots, for example, an internet browser, an electronic database, or a spreadsheet as prompts. Participants were asked to choose (out of four to eight) the best fitting answer options. Often, realistic answer alternatives are given as distractors in the form of buttons or menus which are integrated into the respective screenshots, or real program applications are used to construct the answer options (Senkbeil & Ihme, 2014).

Using simulation-based tasks, additional multimedia content (e.g. text, graphics, video) can be integrated into the stimulus material, allowing a high level of interactivity between response behavior and stimulus presentation. The acquired fact and concept knowledge has to be applied to specific problems in realistic situations. In this way procedural and meta-cognitive skills (goal-oriented thinking and action sequences, which are required for more complex activities, e.g. the preparation of a presentation) can be captured using simulation-based tasks.

Since MC tasks and simulation-based tasks correlate very highly and substantially (Ihme, Senkbeil Goldhammer & Gerick, 2017; Senkbeil & Ihme, 2019), ICT literacy in NEPS is further conceptualized as a unidimensional construct.

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

Blossfeld, H.-P., Doll, J., & Schneider, T. (2008). Bildungsprozesse im Lebenslauf: Grundzüge der zukünftigen Bildungspanelstudie für die Bundesrepublik Deutschland. Recht der Jugend und des Bildungswesens, 3/2008, 321-328.

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.

Ihme, J.M., Senkbeil, M., Goldhammer, F. & Gerick, J. (2017). Assessment of computer and information literacy in ICILS 2013: Do different item types measure the same construct? European Educational Research Journal, 16, 716–732.

Katz, I. R. (2007). Testing ICT literacy in digital environments: The ETS iskills Assessment. Information Technology and Libraries, 26, 1-19.

Senkbeil, M. & Ihme, J. M. (2019). Diagnsotik von ICT Literacy: Messen Multiple-Choice-Aufgaben und simulationsbasierte Aufgaben vergleichbare Konstrukte? Diagnostica.

Senkbeil, M., & Ihme, J. M. (2014). Wie valide sind Papier-und-Bleistift-Tests zur Erfassung computerbezogener Kompetenzen? Diagnostica, 60, 22-34.

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. Zeitschrift für Erziehungswissenschaft, 16, 671-691.

Wittwer, J., & Senkbeil, M. (2008). Is students' computer use at home related to their mathematical performance at school? Computers & Education, 50, 1558-1571.

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.).

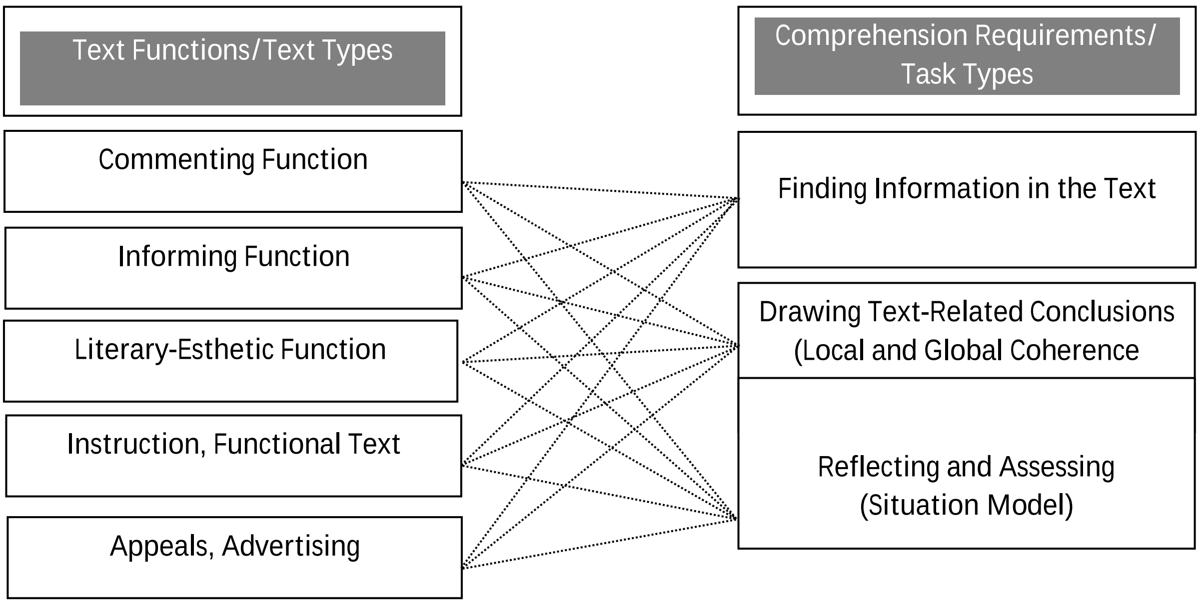

Fig. 1: Text functions and comprehension requirements (cf. Gehrer, Zimmermann, Artelt, & Weinert, 2013, p. 63)

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.

4. Scaling of items

Items of several task formats have been Rasch-scaled and longitudinally linked (Fischer, Rohm, Gnambs, & 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 (Krannich, Jost, Rohm, Koller, Carstensen, Fischer & Gnambs, 2017; Pohl, Haberkorn, Hardt & Wiegand, 2012; Scharl, Fischer, Gnambs, & Rohm, 2017).

Bibliography

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.

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. (2012). The assessment of reading competence (including sample items for grade 5 and 9). Scientific Use File 2012, Version 1.0.0. Bamberg: University of Bamberg, National Educational Panel Study.

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 Reports

Krannich, M., Jost, O., Rohm, T., Koller, I., Carstensen, C. H., Fischer, L., & Gnambs, T. (2017). NEPS Technical Report for reading: Scaling results of Starting Cohort 3 for grade 7 (NEPS Survey Paper No. 14). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

Pohl, S., Haberkorn, K., Hardt, K., & Wiegand, E. (2012). NEPS Technical Report for Reading – Scaling Results of Starting Cohort 3 in Fifth Grade (NEPS Working Paper No. 15). Bamberg: Otto-Friedrich-Universität, Nationales Bildungspanel.

Scharl, A., Fischer, L., Gnambs, 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.

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

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.

Competence in English Reading

The reading competence tasks for English developed by the Institute of Quality Development in Education (Institut für Qualitätsentwicklung im Bildungswesen (IQB)) take into account the different aspects of written texts listed in the National Educational Standards (Nationale Bildungsstandards (KMK, 2003, 2004)) and the Common European Framework of References (Gemeinsamen Europäischen Referenzrahmen (GER; Europarat, 2001)). The task texts are characterized by a high degree of authenticity in relation to English-speaking cultures, i.e. in the sense of representative expository and narrative texts from English-speaking societies.

Based on the National Educational Standards and the GER, the IQB developed test specifications that served as a basis for item development by trained experts. In order to ensure most effective recording of reading competence, maximum attention was paid to perfect fit in terms of text, item and answer format in the further development of tasks.

The tasks used in this study can be allocated to the levels B1 through C1 of the GER that are described as follows (Europarat, 2001, p. 227):

B1: [...] At this level, it is possible to understand texts containing every day or job-related language. [...]

B2: [...] At this level, it is possible to understand articles and reports on current topics if the author gives his opinion on a problem or expresses a certain perspective. [...]

C1: [...] At this level, it is possible to understand complex technical and literary texts as well as recognize differences in style. One can understand technical language in articles and technical instructions, even if they are outside one's own subject.

A detailed description of the English competence test developed by the IQB, including the reading competence test, is contained in Rupp, Vock, Harsch, and Köller (2008).

Bibliography

Europarat (2001). Gemeinsamer europäischer Referenzrahmen für Sprachen: lernen, lehren, beurteilen. Berlin: Langenscheidt.

KMK (2003). Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Mittleren Abschluss [National educational standards for the first foreign language (English/French) for the Mittlerer Schulabschluss]. München: Luchterhand.

KMK (2004). Bildungsstandards für die erste Fremdsprache (Englisch/Französisch) für den Hauptschulabschluss [National educational standards for the first foreign language (English/French) for the Hauptschulabschluss]. München: Luchterhand.

Rupp, A. A., Vock, M., Harsch, C. & Köller, O. (2008). Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and out-comes in Germany. Münster: Waxmann.

Scientific Thinking

The aim of scientific thinking (Wissenschaftspropädeutik) is one of the three main goals of higher secondary school in Germany (KMK, 1995). The German concept Wissenschaftspropädeutik can be best translated as preparation for a life with and in science and comprises of a highly complex theoretical construct that is deeply rooted in German Bildungstheorie (systematic, scientific examination of education, its goals personal, social and economic relevance; sometimes described as philosophy of education) and a humanistic understanding of education.

In its core the term Wissenschaftspropädeutik means the preparation of young individuals for a sophisticated handling of science itself as well as the preparation for a lifetime of learning and operating in a society, deeply dependent on science and its outcomes (Huber, 2000). Therefore, the term scientific thinking was chosen as a translation of the German concept Wissenschaftspropädeutik. Few studies exist on scientific thinking and its impact on personal and professional development like career choice or achievement. One reason for this lack of research roots is an absence of adequate instruments to measure Wissenschaftspropädeutik/scientific thinking (Dettmers, Trautwein, Neuman & Lüdtke, 2010). Therefore, the assessment of scientific thinking was incorporated as one objective of the National Educational Panel Study (NEPS) for pupils in their final year at higher secondary school (Gymnasium).

The conceptualisation of scientific thinking competence in NEPS was based on the understanding that this competence does not only serve as a preparation of future scientists for their academic experiences but also as a preparation of future laymen for life (Huber, 2005). In this modern understanding, scientific thinking pronounces the ability to orientate oneself in the system ‘science’. Here a substantial overlap with research areas like Nature of Science, Scientific Inquiry or Scientific Reasoning are visible. Traditionally, scientific thinking (Wissenschaftspropädeutik) not only drew on natural sciences but on the comparison of all academic disciplines (Hahn, 2013). Therefore, it is often stressed that only through their comparison the potentials and limitations of different scientific perspectives become visible. In credit of its broad and complex nature, scientific thinking (Wissenschaftspropädeutik) has been described as “a concept without edges” (Griese, 1983, p. 257).

Due to the broadness of the concept and with regard to limited testing time, the NEPS approach concentrates on one key component of scientific thinking. Based on a popular structure implemented by Huber (1997) scientific thinking can be divided into three subsequent tiers: “the learning and practicing of science (basic terms, methods)”, the learning and practicing “through science (a habitus of questioning and reasoning)”, and the learning “about science (critical reflection of the bigger picture)” (Huber, 1997, p. 348, translated by the authors). Even though these three tiers handle science from different angles, they are subsequent and interrelated. Huber located the reflection of scientific ideas regarding their genesis, limitations, and consequences on the third tier, which can be understood as the most complex level. This tier was chosen to be the centre of the NEPS assessment of scientific thinking competence as “metascientific reflection”.

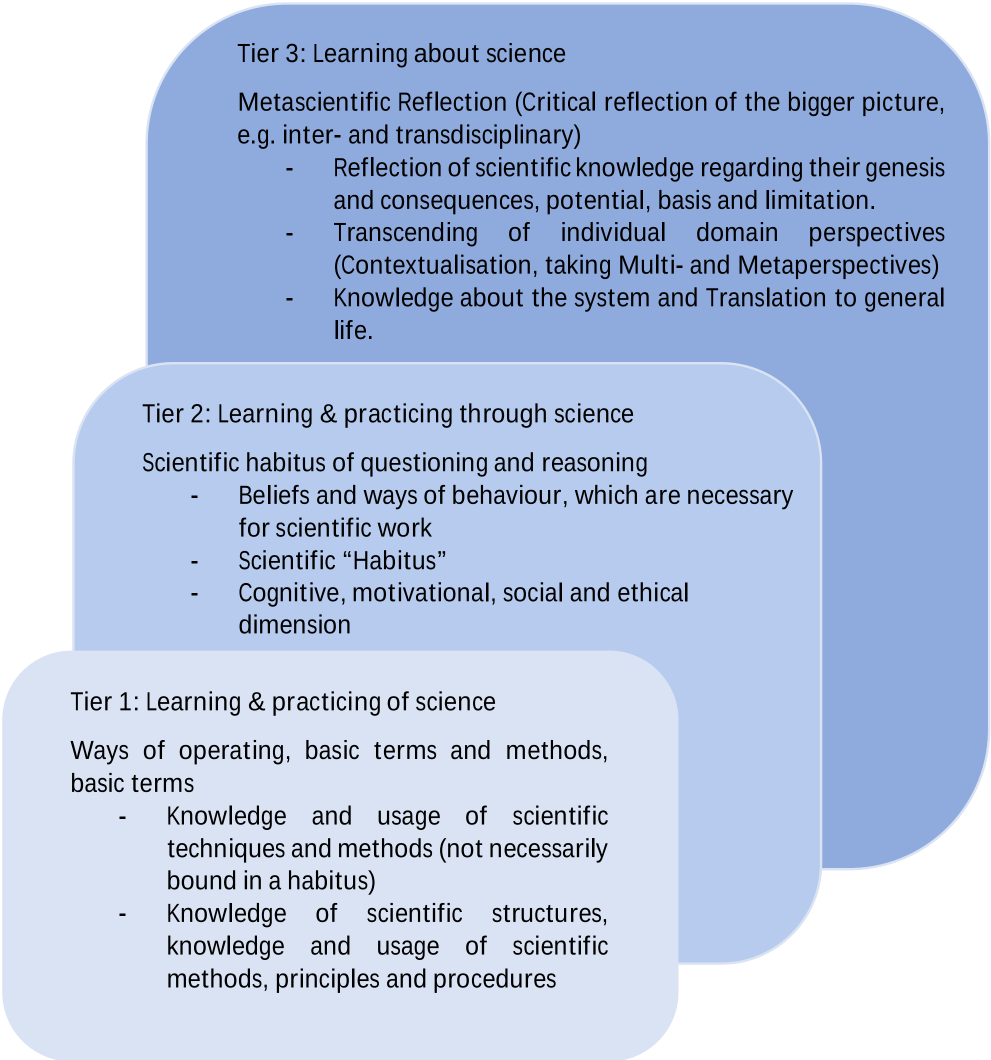

Figure 1. Tiers of scientific thinking (Wissenschaftspropädeutik) (Müsche 2009, modified of Huber (1997), translated by the authors)

During test development, the NEPS framework was based on a translation of the popular three tiers into a normative structural model with three dimensions by Müsche (2009). Drawing on Huber, Müsche defined the dimensions (1) metascientific knowledge, (2) understanding of methods, and (3) metascientific reflection. The third dimension is comprised of five sub competences:

1. To contextualise scientific ways of generating knowledge, scientific claims and results in a bigger picture
2. To reflect on scientific ideas regarding their foundation, potential, the circumstances of their development and consequences
3. To judge scientific processes of knowledge generation and potential using methodological knowledge
4. To question and test the validity, explanation power and limitations of scientific claims
5. To contrast inconsistent knowledge and contradictory theoretical approaches (Müsche, 2009, p. 78).

Item format

The test is composed of stimulus texts, describing the scientific controversy, and claims regarding central aspects of the controversy. Items have to be answered in a forced choice format (correct vs. incorrect). All controversies focus relevant, complex, and multicausal problems of public interest. The length of the texts varies between 300 and 400 words and can incorporate graphs or tables. In the stimulus texts a short introduction into the topic is given, the controversy between two scientists is explained and the positions of the researchers are described. The texts give all necessary information for solving the items and pose a judgement on the claims made by two fictional scientists in the controversy. All five subcomponents named by Müsche (2009) are covered across the items. To be adequate for testing the abilities of future competent laypeople, the items do not require it to plan complicated scientific studies or evaluate scientific decisions, but to reflect on scientific theories and stances.

Scaling of the tests

In order to estimate the item and person parameters for scientific thinking competence, a Rasch model was used and estimated in ConQuest 4.2. Item parameters are estimated difficulties for dichotomous variables in the Rasch model. Ability estimates for scientific thinking competence were estimated as weighted maximum likelihood estimates. The psychometric quality and the scaling results of the tests and items are described in the technical reports for the Starting Cohort 4 in 12th Grade.

Bibliography

Dettmers, S., Trautwein, U., Neumann, M., & Lüdtke, O.,(2010). Aspekte von Wissenschaftspropädeutik [Aspects of Wissenschaftpropädeutik]. In: U. Trautwein, M. Neumann, G. Nagy, O. Lüdtke & K. Maaz (Ed..), Schulleistungen von Abiturienten. Die neu geordnete gymnasiale Oberstufe auf dem Prüfstand (pp. 243–266). Wiesbaden: VS Verl. für Sozialwissenschaften. http://dx.doi.org/10.1007/978-3-531-92037-5_9.

Griese, W. (1983). Wissenschaftspropädeutik in der gymnasialen Oberstufe [Research prodedeutics in the upper secondary level of academic track schools] (Doctoral Dissertation). Universität Oldenburg.

Hahn, S. (2013). Wissenschaftspropädeutik in der gymnasialen Oberstufe [Wissenschaftspropädeutik in the upper-secondary level of academic track schools]. In: Dorit Bosse, Franz Eberle und Barbara Schneider-Taylor (Hg.): Standardisierung in der gymnasialen Oberstufe (pp.161-174). Wiesbaden: Springer VS.

Huber, L. (1997). Fähigkeit zum Studieren- Bildung durch Wissenschaft. Zum Problem der Passung zwischen Gymnasialer Oberstufe und Hochschule [Ability to study - education through science. With respect to the fit between academic track schools and higher education institutions]. In E. Liebau, W. Mack & C. T. Scheilke (Hrsg.), Das Gymnasium. Alltag, Reform, Geschichte, Theorie (pp. 333-351). Weinheim: Juventa.

Huber, L. (2000). Wissenschaftspropädeutik, allgemeine Studierfähigkeit und ihre unterrichtliche Umsetzung in Grundkursen [Wissenschaftpropädeutik, general ability to study and the instructional implementation in basic courses]. In Hessisches Landesinstitut für Pädagogik (Hrsg.), Bildung braucht guten Grund: Beiträge zur Reform der Grundkurse (pp. 17–46). Wiesbaden: HeLP

Huber, L. (2005). Warum fächerübergreifender Unterricht in der Sekundarstufe II? [Why interdisciplinary teaching in upper-secondary level II?] In. Author & Tillmann, K.-J. (Hrsg.),

Versuchsschulen und das Regelschulsystem – Bielefelder Erfahrungen (pp. 109-124). Bielefeld: Laborschule Bielefeld.
KMK - Kultusministerkonferenz (1995). Weiterentwicklung der Prinzipien der gymnasialen Oberstufe und des Abiturs. Abschlussbericht der von der Kultusministerkonferenz eingesetzten Expertenkommission [Further development of principles for academic track schools. Final report of the expert committee appointed by the Kultusministerkonferenz]. Bonn: Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland.
Müsche, H. (2009). Wissenschaftspropädeutik aus psychologischer Perspektive – Zur Dimensionierung und Konkretisierung eines bildungstheoretischen Konzeptes [Wissenschaftpropedeutik from a psychological perspective - Regarding the dimension and ascertainment of the theoretical based concept]. TriOS, 4(2), 61– 109.

Short Version of
Oschatz, K., Kramer, J. & Wagner, W. (2017). The assessment of Wissenschaftspropädeutik as metascientific reflection. Scientific Use File 2017, Version X.X.X. Bamberg: Leibniz Institute for Educational Trajectories, National Educational Panel Study.