NEPS
National Educational Panel Study

Research Data

Information on Competence Testing

NEPS Starting Cohort 1 — Newborns

Education From the Very Beginning

Wave 8: 7 years

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: N. N.
Bamberg; May 25, 2021

<table>
<tr><td colspan="5">Test information</td></tr>
<tr><td>Test information</td><td colspan="4">Seven-year-old children were tested individually in their homes in the presence of the anchor person and the interviewer.</td></tr>
<tr><td>Test sequence</td><td colspan="4">The four competence measures were administered in the following sequence:

1. Vocabulary: Listening comprehension at word level (Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). Peabody Picture Vocabulary Test-Revision IV German Adaption, PPVT-IV. Frankfurt am Main, Germany: Pearson Assessment) + Procedural metacognition
2. Phonological working memory: Backward number span
3. Scientific literacy + Procedural metacognition
4. Delay of gratification: Executive control

The children either used the tablet to solve the tasks on their own (vocabulary and scientific literacy) or answered verbally (phonological working memory and delay of gratification).</td></tr>
<tr><td>Test duration (excluding setup)</td><td colspan="4">approx. 40 minutes</td></tr>
<tr><td colspan="5">Information about the administered competence measures</td></tr>
<tr><td>Construct</td><td>Number of tasks</td><td>Duration (approx.)</td><td>Mode of administration</td><td>Next assessment</td></tr>
<tr><td>Vocabulary: Listening comprehension at word level</td><td>19 sets with 12 tasks each (max.), with a stopping rule</td><td>13 minutes</td><td>visual stimuli presented on a tablet; each task featured one word and four possible selections; administered on a tablet</td><td>Wave 10 (2021)</td></tr>
<tr><td>Domain-specific procedural metacognition
for the domain vocabulary: Listening comprehension at word level</td><td>1</td><td>1 minute</td><td>technology-based testing; image-based multiple selection; administered on tablet PC</td><td>Wave 10 (2021)</td></tr>
<tr><td>Phonological working memory: Backward number span</td><td>18 tasks (max.) , with a stopping rule</td><td>5 minutes</td><td>oral reply; administered on a tablet</td><td>-</td></tr>
</table>

Scientific literacy	21 tasks (max.)	20 minutes	technology-based assessment; picture-based multiple choice and multiple true-false; administered on a tablet	Wave 10 (2021)
Domain-specific procedural metacognition for the domain scientific literacy	1	1 minute	technology-based testing; image-based multiple selection; administered on a tablet	Wave 10 (2021)
Delay of gratification: Executive control	-	1 minute	physical objects; administered on a tablet	-

Preface

The development of the individual tests is based on framework concepts. They are 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, which served as a basis for the development of the test tools to measure the above-mentioned constructs, are identical in the different studies.

In addition to the competence measures, which are coherently assessed across the lifespan, stage-specific measures are assessed at specific points in time at which these measures are especially meaningful (cf. Berendes, Weinert, Zimmermann, & Artelt, 2013[1]). Usually, these assessments are not repeated.

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

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). In Starting Cohort 1 of the NEPS, from children aged 3 years, listening comprehension is solely captured at word level and later on, in primary school, via indicators of reading ability.

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)[2], 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 1.

In the NEPS, the test was administered via a tablet PC. The children's task was to select one picture out of four, matching it to the word presented as an auditory cue by the tablet.

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 children. The practice unit at the start of the test also varies depending on the age and performance of the children. Once a child 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 performance in the practice phase as well as on the age of the children. If the child 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 child has made more than seven mistakes – has been identified; the sets that have already been processed are left out.

[2] 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.

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. Kindergarten and elementary school children are shown a 5-point smiley scale to give their judgments.

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.

Procedural metacognition of 7-year-old children in Starting Cohort 1 of the NEPS was assessed using tablets.

References

Brown, A. L. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. E. Weinert & 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 (Ed.), *The psychology of learning and motivation* (pp. 125–141). New York: Academic Press.

Phonological working memory: Backward digit span

Short-term memory or working memory is regarded as the bottleneck of information processing because of its limited capacity. On the one hand, people can store an almost unlimited amount of information in long-term memory; on the other hand, their ability to immediately reproduce unrelated information (e.g., a telephone number) after hearing it once is limited. Short-term or working memory performance (functional capacity) differs between individuals and generally increases during the transition from childhood to adolescence (for a brief overview, see Weinert, 2010).

In the National Educational Panel Study (NEPS), the assessment of the so-called "digit span" as well as the "backward digit span" is based on the theoretical model of working memory suggested, for example, by Baddeley and Hitch (1974). The performance in span tasks is taken as an indicator of phonological working memory's capacity (Baddeley, 1992). In span tasks, sequences of numbers (or digits) are presented auditorily, and the test person is instructed to immediately reproduce each sequence in the same order (resulting in an indicator of the individual's "digit span"). Span tasks usually present digit sequences of increasing length until the child cannot reproduce them correctly anymore; the indicator ("digit span") is the longest digit sequence the child is able to reproduce immediately and correctly after hearing it once (Baddeley, Gathercole, & Papagno, 1998). While span tasks requiring the same reproduction order test the performance (or capacity) of the (passive) "Phonological Loop" (Baddeley et al., 1998), span tasks requiring a reversed reproduction order capture the performance of the "Central Executive" (Baddeley, 1986). In the theoretical model used here, the Central Executive is conceptualized as a functional control unit of working memory that allows a limited amount of information to be stored and actively manipulated (Baddeley, 2012). Central Executive performance in this context is negatively related to attention deficits (Rapport, Alderson, Kofler, Sarver, Bolden, & Sims, 2008) and positively related to reading/spelling skills (Andersson, 2008) and mathematical operations, such as addition or multiplication (Raghubar, Barnes, & Hecht, 2010), among others.

In NEPS Starting Cohort 1, the backward digit span task is based on a subtest of the German version of the Kaufman Assessment Battery for Children (K-ABC; Melchers & Preuß, 2009). The task assesses the ability to directly reproduce a verbally presented series of digits in reverse order. Digits between 1 and 10 are used, whereby the digit 7 was omitted due to polysyllabicity (Melchers & Preuß, 2009). The test was administered in a playful way on a tablet (administration language: German) and the standardized digit sequences (items) were presented auditorily. The children's task was to reproduce the respective items in reverse order immediately after presentation. The interviewers protocolled the spoken sequence on the tablet. If the child did not respond, NR (non-response) was entered, which counted as incorrect. A digit sequence could only be repeated if the child could not actually hear the numbers, otherwise repetition was not allowed.

The test consists of a practice phase and a subsequent testing phase. In both phases, two items form a task unit. The practice phase comprises one task unit, namely two items with two digits each. If one of the two practice items is not answered or answered incorrectly, it is repeated to ensure that the instruction was understood correctly. The practice items are not included in the total score. Regardless of whether the practice items were answered correctly or incorrectly, the testing phase follows. The testing phase consists of eight task units and accordingly 16 items. The first two task units feature digit sequences with two digits each and thereafter the number of digits increases by one additional digit per task unit. Only the items of the testing phase are relevant for test termination. The test is terminated if the child answers both items of a task unit incorrectly or not at all. The child's answer is only considered to be correct if the child reproduces the numbers in the correct reverse order. Each correct answer equals one point; thus, a maximum total score of 16 points is possible.

The Scientific Use File contains: the total number of administered practice items; all scored test items; the sum score (i.e., the number of all correctly solved test items); the maximum digit span correctly reproduced; as well as a variable indicating at which digit span the test was terminated.

References

Andersson, U. (2008). Working memory as a predictor of written arithmetical skills in children: The importance of central executive functions. *British Journal of Educational Psychology*, *78*(2), 181–203. https://doi.org/10.1348/000709907X209854

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Clarendon Press.

Baddeley, A. D. (1992). Working memory. *Science*, *255*(5044), 556–559. https://doi.org/10.1126/science.1736359

Baddeley, A. D. (2012). Working memory: theories, models, and controversies. *Annual review of psychology*, *63*, 1–29. https://doi.org/10.1146/annurev-psych-120710-100422

Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, *105*(1), 158–173. https://doi.org/10.1037/0033-295X.105.1.158

Baddeley, A. D. & Hitch, G. (1974). Working memory. *Psychology of Learning and Motivation*, *8*, 47–89. https://doi.org/10.1016/S0079-7421(08)60452-1

Melchers, P. & Preuß, U. (2009). *Kaufman Assessment Battery for Children (K-ABC), dt. Version (8., unveränderte Aufl.)*. Frankfurt, Germany: Pearson Assessment.

Raghubar, K. P., Barnes, M. A., & Hecht, S. A. (2010). Working memory and mathematics: A review of developmental, individual difference, and cognitive approaches. *Learning and individual differences*, *20*(2), 110–122. https://doi.org/10.1016/j.lindif.2009.10.005

Rapport, M. D., Alderson, R. M., Kofler, M. J., Sarver, D. E., Bolden, J., & Sims, V. (2008). Working memory deficits in boys with attention-deficit/hyperactivity disorder (ADHD): The contribution of central executive and subsystem processes. *Journal of abnormal child psychology*, *36*(6), 825–837. https://doi.org/10.1007/s10802-008-9215-y

Weinert, S. (2010). Beziehungen zwischen Sprachentwicklung und Gedächtnisentwicklung. In H.-P. Trolldenier, W. Lenhard, & P. Marx (Eds.), *Brennpunkte der Gedächtnisforschung: Entwicklungs- und pädagogisch-psychologische Perspektiven* (pp. 147–170). Göttingen, Germany: Hogrefe.

Scientific literacy

Scientific literacy is the precondition for participating in world affairs marked by science and technology (Prenzel, 2000; Prenzel et al., 2001; Rost et al., 2004) and is viewed as a predictor for an economically, socially and culturally successful life. Many problems and issues we encounter in our daily life require an understanding of natural sciences and technology. Scientific topics and problems affect all people. Therefore, current discussions on the goals of scientific education focus on the concept of scientific literacy for all people (Osborne & Dillon, 2008). Such literacy is the basis for lifelong learning, serves as a connection for further learning (OECD, 2006; Prenzel et al., 2007) and, thus, also influences professional careers.

Based on this, the NEPS definition of scientific literacy follows the Anglo-Saxon literacy concept (Bybee, 1997; Gräber, Nentwig, Koballa & Evans, 2002; OECD, 2006) which does not regard scientific competence as the simple reproduction of acquired knowledge but rather as the flexible use of acquired knowledge in different situations and contexts of daily life.

In the NEPS, scientific literacy is understood as the use of scientific knowledge in the environmental, technological and health contexts (Hahn et al., 2013). In addition, the concept distinguishes between content-related and process-related elements (see Fig. 1). Knowledge of science comprises content-related *matter*, *systems*, *development* and *interactions*. Knowledge about science includes *enquiry and scientific reasoning* which involve, among other things, checking hypotheses, interpreting findings as well as measuring principles and measuring error control.

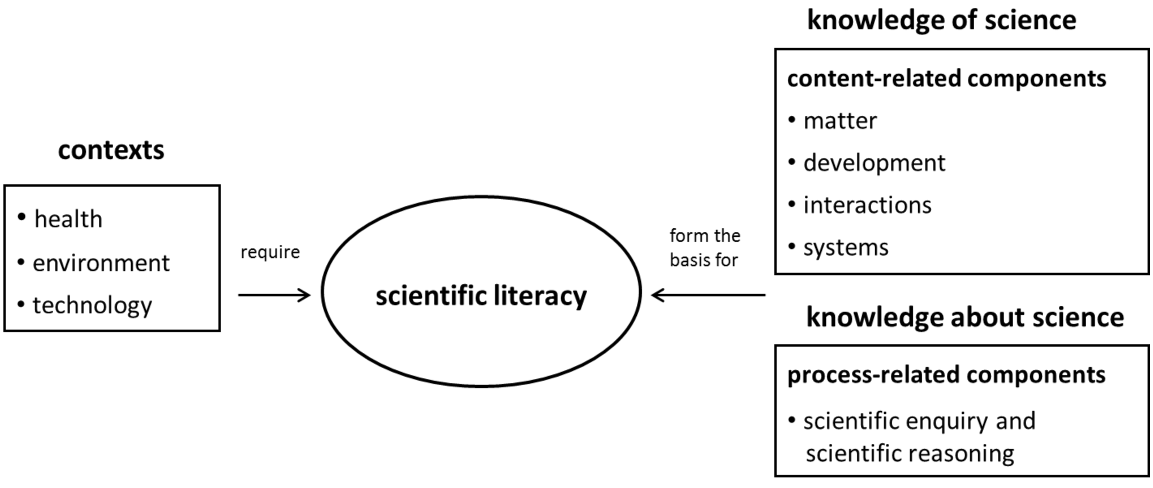

Fig.1. Application contexts as well as content-related and process-related elements of the NEPS scientific literacy test (Hahn et al., 2013).

To select its contexts as well as the content-related and process-related elements, the NEPS uses PISA (OECD, 2006), the *Benchmarks for Scientific Literacy* of the *American Association for the Advancement of Science* (AAAS, 2009) and the education standards of the Conference of Ministers of Education for the medium-level school-leaving qualification (KMK, 2005a, 2005b, 2005c) as a guideline. The selected contexts are of personal, social and global relevance. Considering current scientific research and the general events of the day, it is assumed that they will remain important across the entire life span of the test persons. Figure 2 provides an overview of the overlap of content-related components between PISA, the German educational standards and the NEPS. The selected content-related and process-related elements cover central concepts of all scientific disciplines.

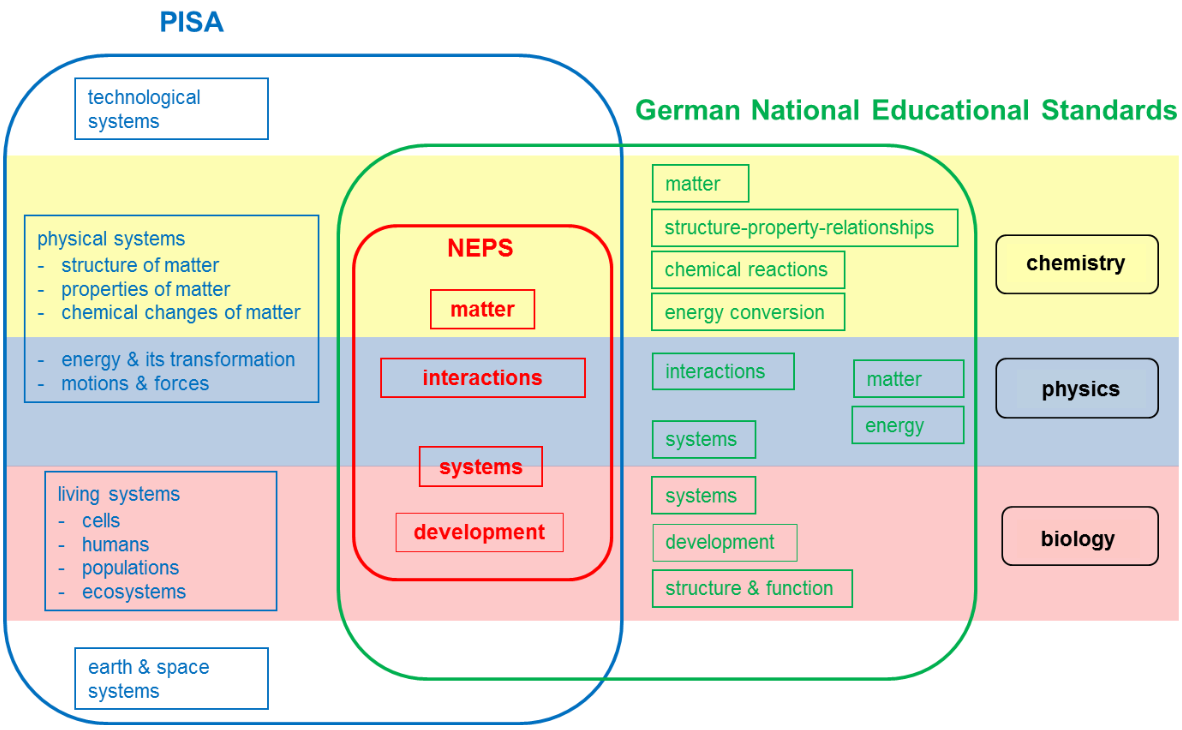

Fig.2. Overview of the overlap of content-related components between PISA, the German educational standards and the NEPS (Hahn et al., 2013).

The scientific literacy of 7-year-old children in Starting Cohort 1 of the NEPS is assessed using a tablet-based test. The test items are embedded in a “science and technology game”, and a little dragon called “Nepsi” guides the children through the game. He reads the picture-based items and possible answers to them and then asks the children either to pick the right answer out of four pictures (multiple choice) or to judge whether successively shown pictures present a right or wrong answer (multiple true-false).

In the end, one scientific literacy score is computed and published in the Scientific Use File for Starting Cohort 1.

References

American Association for the Advancement of Science. (1989). *Science for all Americans: A Project 2061 Report on goals in science, mathematics and technology*. Washington, DC: American Association for the Advancement of Science.

American Association for the Advancement of Science. (AAAS). (2009). *Benchmarks for science literacy*. *Project 2061.* Retrieved from http://www.project2061.org/publications/bsl/online/index.php

Bybee, R. W. (1997). Towards an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.). *Scientific literacy – An international symposium* (pp. 37–68). Kiel, Germany: Institute for Science and Mathematics Education (IPN).

Gräber, W., Nentwig, P., Koballa, T. & Evans, R. (Eds.). (2002). *Scientific Literacy. Der Beitrag der Naturwissenschaften zur Allgemeinen Bildung*. Opladen: Leske + Budrich.

Hahn, I., Schöps, K., Rönnebeck, S., Martensen, M., Hansen, S., Saß, S., Dalehefte, I. M., & Prenzel, M. (2013). Assessing scientific literacy over the lifespan – A description of the NEPS science framework and the test development. *Journal for Educational Research Online*, *5*(2), 110–138.

KMK (2005a). *Resolutions of the Conference of Ministers of Education: Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Resolution of 16.12.2004. Munich, Germany: Luchterhand.

KMK (2005b). *Resolutions of the Conference of Ministers of Education: Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss*. Resolution of 16.12.2004. Munich, Germany: Luchterhand.

KMK (2005c). Resolutions of the Conference of Ministers of Education*: Bildungsstandards im Fach Physik für den Mittleren Schulabschluss.* Resolution of 16.12.2004. Munich, Germany: Luchterhand

OECD (2006). *Assessing scientific, reading and mathematical literacy. A framework for PISA 2006*. Paris: Organisation for Economic Co-Operation and Development.

Osborne, J. & Dillon, J. (2008). *Science education in Europe: Critical reflections. A report to The Nuffield Foundation*. London, United Kingdom: King's College.

Prenzel, M. (2000). Lernen über die Lebensspanne aus einer domänenspezifischen Perspektive: Naturwissenschaften als Beispiel. In F. Achtenhagen & W. Lempert (Eds.), *Lebenslanges Lernen im Beruf - seine Grundlegung im Kindes- und Jugendalter. Volume IV. Formen und Inhalte von Lernprozessen* (pp. 175–192). Opladen, Germany: Leske + Budrich.

Prenzel, M., Rost, J., Senkbeil, M., Häußler, P., & Klopp, A. (2001). Naturwissenschaftliche Grundbildung: Testkonzeption und Ergebnisse. In J. Baumert, E. Klieme, M. Neubrand, M. Prenzel, U. Schiefele, W. Schneider, P. Stanat, K.-J. Tillmann & Weiß, M. (Eds.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (pp. 191–248). Opladen, Germany: Leske + Budrich.

Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O., Carstensen, C. H., & Hammann, M. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In M. Prenzel, C. Artelt, J. Baumert, W. Blum, M. Hammann, E. Klieme & R. Pekrun (Eds.), *PISA 2006 – Die Ergebnisse der dritten internationalen Vergleichsstudie* (pp. 63–105). Münster, Germany: Waxmann.

Rost, J., Prenzel, M., Carstensen, C.-H., Senkbeil, M., & Groß, K. (Ed.). (2004). *Naturwissenschaftliche Bildung in Deutschland. Methoden und Ergebnisse von PISA 2000*. Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Delay of gratification: Executive control

Self-regulation is defined as the ability to control and manage one's own thinking, feeling, and actions (Neubauer, Gawrilow, & Hasselhorn, 2011), as well as to plan, pursue, and consequently achieve personal goals (Zimmermann, 2000). Self-regulation skills encompass several quite different facets that are measured via a wide variety of assessment methods (e.g., self-assessment, peer-, parent-, educator-assessment/judgement, direct standardized observations, experimental tasks). The National Educational Panel Study (NEPS) captures different aspects and facets of self-regulation, such as cognitive self-regulation in terms of procedural and declarative metacognition (Weinert et al., 2019), various survey items related to emotional and behavioral self-control, as well as tasks measuring "delay of gratification".

The ability to self-regulate and to engage in cognitive deliberation processes regarding potential behaviors (Mischel, 1974) is considered significant for child development. A number of studies indicate that well-developed abilities of self-regulation at preschool age are predictive of later academic performance (Watts, Duncan, & Quan, 2018), the ability to cope with stress (stress resistance), the development of socio-emotional competencies, and the ability to focus and maintain attention (Baumeister & Vohs, 2004; Kochanska, Murray, & Coy, 1997; Tangney, Baumeister, & Boone, 2004; Wulfert, Block, Ana, Rodriquez, & Colsman, 2002).

NEPS Starting Cohort 1 (Newborns) and NEPS Starting Cohort 2 (Kindergarten), amongst others, include delay of gratification[3] tasks (for Starting Cohort 2, see Luplow, Schönmoser, Lorenz, & Schmitt, 2019). To assess delay of gratification, two procedures are used, namely the waiting paradigm and the choice paradigm (Mischel, 1974; Mischel, 2015). In previous waves (waves 4 and 6; at ages 3 and 5, respectively), a delay of gratification task was conducted using a waiting paradigm. This means that the child had to wait for a certain (for the child unspecified) amount of time (wave 4: 181 seconds; wave 6: 301 seconds) in order to receive an even larger incentive in addition to a small immediate incentive (i.e., a present). The child could choose to stop waiting, with the consequence of receiving only the small incentive. Waiting time (waves 4 and 6) and waiting behavior (wave 6) were recorded.

In the present wave, the child could choose between two options (choice paradigm). The child was told that he/she could either receive one incentive today or two incentives tomorrow. Thus, the child was asked to make a decision. If the immediate small incentive was chosen, the interviewer handed over one incentive directly; if the delayed two incentives were chosen, the interviewer gave the child both incentives at the end of the parental questionnaire. Up until this time, however, the child believed he/she would receive both incentives the next day. Thus, a situation of decision-making was created, as is common for the choice paradigm. The interviewer administered the delay of gratification task and the child's decision was protocolled on a tablet. The child's decision-making process was not influenced. Each incentive was wrapped in small incentive bag. The child could not see the incentive bags, as both were inside a transportation bag. Thus, there were no further clues regarding incentive characteristics (e.g., sort, size, or value). In each case, the child was allowed to take the individually wrapped incentive(s) out of the transport bag themselves. The two incentives were a Frisbee and a card game.

[3] Note: In the NEPS, the terms "delay of gratification" and "delayed gratification" are used synonymously.

References

Baumeister, R. F. & Vohs, K. D. (2004). *Handbook of self-regulation: Research, theory, and applications.* New York: The Guilford Press.

Garon, N., Bryson S. E., & Smith I. M. (2008). Executive Function in Preschoolers: A Review Using an Integrative Framework. *Psychological Bulletin*, *134*(1), 31–60.

Kochanska, G., Murray, K. T., & Coy, K. C. (1997). Inhibitory control as a contributor to conscience in childhood: From toddler to early school age. *Child Development, 68*(2), 263–277.

Lemmon, K., & Moore, C. (2007). The development of prudence in the face of varying future rewards. *Developmental Science, 10*(4), 502–511.

Lorenz, C., Schmitt, M., Luplow, N., & Schönmoser, C. (2016). Soziale Disparitäten im Vorschulalter und der Einfluss der Selbstregulation. *Zeitschrift für Grundschulforschung, 9*(1), 65–77.

Luplow, N., Schönmoser, C., Lorenz, C., & Schmitt, M. (2019). *Die Messung des Belohnungsaufschubes in der Startkohorte 2 des Nationalen Bildungspanels (NEPS) im Kindergarten und der Grundschule (NEPS Survey Paper No. 54).* Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study (NEPS).

Mischel, W. (1974). Processes in delay of gratification. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology* (Vol. 7, pp. 249–292). New York, NY: Academic Press.

Mischel, W. (2015). *Der Marshmallow Test.* Munich, Germany: Siedler Verlag.

Mischel, W. & Gilligan, C. (1964). Delay of gratification, motivation for the prohibited gratification, and responses to temptation. *The Journal of Abnormal and Social Psychology, 69*(4), 411–417.

Neubauer, A., Gawrilow, C., & Hasselhorn, M. (2011). Belohnungsaufschub: Ein Ansatz zur Frühprognose volitionaler Kompetenzen. In M. Hasselhorn & W. Schneider (Eds.), *Frühprognose schulischer Kompetenzen* (pp. 202–220). Göttingen, Germany: Hogrefe.

Tangney, J. P., Baumeister, R. F., & Boone, A. (2004). High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. *Journal of Personality, 72*(2), 271–324.

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development, 12*(2), 199–212.

Watts, T. W., Duncan, G. J., & Quan, H. (2018). Revisiting the Marshmallow Test: A Conceptual Replication Investigating Links Between Early Delay of Gratification and Later Outcomes. *Psychological Science, 29*(7), 1159–1177.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., Carstensen, C. H., & Lockl, K. (2019). Development of Competencies Across the Life Course. In H.-P. Blossfeld & H.-G. Roßbach (Eds.), *Education as a Lifelong Process: The German National Educational Panel Study (NEPS)* (pp. 57–82). Wiesbaden, Germany: Springer VS.

Wulfert, E., Block, J. A., Ana, E. S., Rodriguez, M. L., & Colsman, M. (2002). Delay of Gratification: Impulsive Choices and Problem Behaviors in Early and Late Adolescence. *Journal of Personality, 70*(4), 533–552.

Zimmermann, B. J. (2000). Attaining Self-Regulation: A Social Cognitive Perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of Self-Regulation* (pp. 13–39). San Diego, CA: Academic Press.