NEPS
National Educational Panel Study

Research Data

Information on Competence Testing

NEPS Starting Cohort 1 — Newborns

Education From the Very Beginning

Wave 4: 37-39 months

LIfBi
LEIBNIZ INSTITUTE FOR
EDUCATIONAL TRAJECTORIES

Test information				
Test information	Three-year-old children were tested individually in their homes in the presence of the anchor person and the interviewer.			
Test sequence	The 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, Germany: Pearson) 2. Digit span: phonological working memory 3. Categorization (nonverbal): SON-R subtest (SON categories) (Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (1998). *Snijders-Oomen nonverbal intelligence test, SON-R 2.5-7*. Manual & Research Report. Lisse, Netherlands: Swets & Zeitlinger.) 4. Delay of gratification: executive control The children either used the tablet to solve the tasks on their own (vocabulary and categorization) or answered verbally (digit span and delay gratification). All tasks and instructions were administered on a tablet.			
Test duration (excluding setup)	approx. 35 minutes			
Information about the administered competence measures				
Construct	**Number of items**	**Duration (approx.)**	**Mode of administration**	**Next assessment(s)**
Vocabulary: listening comprehension at word level	19 sets with 12 tasks each (max.), with a stopping rule	15 minutes	visual stimuli presented on a tablet; each task featured one word and four possible selections	Wave 6 (2017) Wave 8 (2019)
Digit span: phonological working memory	15 tasks (max.) , with a stopping rule	5 minutes	oral reply; a tablet was used for data entry	Wave 7 (2018)
Categorization: SON-R subtest	2 sets with a total of 15 tasks (max.) , with a stopping rule	10 minutes	visual stimuli presented on a tablet	-
Delay of gratification: executive control	-	5 minutes	oral reply; a tablet was used for data entry; additional physical objects	Wave 6 (2017) Wave 8 (2019)

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 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 the 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 a good and internationally applicable indicator 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 four tasks. The testing phase begins with Set 1 for all children (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; this represents the highest difficulty level achieved.

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

The Scientific Use File contains the number of administered practice items, the correctness of the answers for each test item (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.* Pearson.

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

Digit span – phonological working memory

The short-term memory or working memory is regarded as the bottleneck of information processing because it has a limited capacity. On the one hand, people can store an almost unlimited amount of information over the long term; 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 interindividually and generally increases during childhood into adolescence (for a brief overview, see Weinert, 2010).

In the National Educational Panel Study, the construct “digit span” is based on the theoretical framework of the working memory model, for example by Baddeley and Hitch (1974). The performance in so-called span tasks is taken as an indicator of the phonological working memory’s capacity (Baddeley, 1992). In span tasks, sequences of numbers (or digits) are presented in auditory form, and the test person is instructed to reproduce them in the same order (i.e., “digit span”). Span tasks usually present digit spans of increasing length until the child cannot reproduce them correctly anymore; the result is the longest digit span the child is able to reproduce immediately and correctly

after hearing it once (Baddeley, Gathercole, & Papagno, 1998). The short-term storage and immediate reproduction of auditory information is associated with the phonological loop, which is a passive subsystem in the working memory model (Baddeley & Hitch, Working memory, 1974). Due to the fast presentation rate of the digit spans, differences in the usage of memory strategies are minimized. Therefore, individual performance can be interpreted as an indicator for the capacity of the respective person's phonological working memory. The reproduction of the digit span is not only influenced by the structural capacity of the phonological loop but also by the speed of articulation and item identification, which in turn is associated with prior knowledge aspects, such as linguistic knowledge.

The individual capacity of the phonological short-term or working memory has been shown to be predictive of the development of linguistic skills, in particular of the acquisition of vocabulary (for an overview, see Baddeley, Gathercole, & Papagno, 1998; Gathercole & Baddeley, 1993; Weinert, 2010) and of the acquisition of reading skills (e.g., Berendes, Weinert, Zimmermann, & Artelt, 2013; Gathercole & Baddeley, 1993). In addition, a study by Krajewski and Schneider (2009) found an association between preschool phonological loop capacity and mathematical development in school.

In Starting Cohort 1 of the NEPS, the digit span task is based on the German version of the "Kaufman Assessment Battery for Children" (K-ABC; Melchers & Preuß, 2009). The task tests the ability to immediately reproduce a verbally presented digit span in the correct order (numerical memory). Digits between 1 and 10 are used, except for the multi-syllabic digit 7 (cf., Melchers & Preuß, 2009). The task and the auditory cues are presented in a standardized and age-appropriate way on a tablet PC; the instructional language is German. The children's task is to immediately reproduce the respective digit span in the presented order.

The task consists of a practice phase and subsequent learning and test items. The practice phase contains one item which is repeated if the answer is incorrect or missing to ensure that the child has understood the instructions; the practice phase is not included in the total score. The phase is followed by five sets, each consisting of three items. The first two items of the first set are learning items. This means that the children receive feedback on the correctness of their answers from the tablet PC, and the item is repeated if necessary. The learning items are only included in the total score if they are reproduced correctly at the first attempt. The first set features items with two digits and the number of digits increases by one per set; hence, Set 5 features items with six digits each. The practice items are not relevant for the termination of the test; after the practice phase and up to Set 3, the test ends if all items of a respective set have been answered incorrectly. After Set 3, the test ends immediately after a single incorrect answer. The theoretically achievable maximum score is 15 (13 test items and two learning items); each correct item is scored with one point.

In the Scientific Use File[3], the following variables are published: the number of administered practice items; the correctness of each learning and test item; the total score of all correctly solved learning and test items; the longest digit span achieved (at least one correct answer); a variable that displays the set in which the test was terminated. For the learning items, only the child's first attempt was scored.

[3] Note: The described data refer to SUF version SC1:7.0.0.

References

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

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

Baddeley, A., 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

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

Gathercole, S. E. & Baddeley, A. D. (1993). Phonological working memory: A critical building block for reading development and vocabulary acquisition? *European Journal of Psychology of Education*, *8*(3), 259–272. https://doi.org/10.1007/BF03174081

Krajewski, K. & Schneider, W. (2009). Exploring the impact of phonological awareness, visual-spatial working memory, and preschool quantity-number competencies on mathematics achievement in elementary school: Findings from a 3-year longitudinal study. *Journal of experimental child psychology*, *103*(4), 516–531. https://doi.org/10.1016/j.jecp.2009.03.009

Melchers, P. & Preuß, U. (2009). *Kaufman Assessment Battery for Children (K-ABC), German Version (8th unchanged edition)*. Frankfurt, Germany: Pearson Assessment.

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.

Categorization: SON-R subtest

The Snijders-Oomen Nonverbal Intelligence Test (SON-R 2½-7; Tellegen, Winkel, Wijnberg-Williams, & Laros, 2007) is a standardized individual test for the assessment of the non-verbal intelligence of children aged between 2.5 and 7 years, with various subtests. In Starting Cohort 1 of the NEPS, the subtest “Categories” of the SON-R 2½-7 was carried out; it measures abstract thinking or reasoning abilities (for an overview, e.g., Quinn, 2011). The categorization tasks mainly require the test person to derive principles of order or to establish connections between (abstract) terms or objects. The ability to make abstractions, i.e. to recognize connections and to draw conclusions, is central to almost all concepts of intelligence (e.g., Cattell, 1963; Spearman, 1923; Wechsler, 1950). The subtest “Categories” measures a facet of such capabilities. It also proved to be suitable for the NEPS because it could be adapted comparatively easily to tablet computer administration. It is a brief indicator of non-verbal basic cognitive skills in early childhood.

The subtest “Categories” consists of a total of 15 test items, which are divided into two parts, as well as an additional practice item to ensure that the children understand the test items in the second part

of the test. In the first part (Items 1 to 7), the child has to sort four to six pictures, each based on specific characteristics. For example, flowers and candies have to be put together. In the items of the second part (Items 8 to 15), the child is shown three pictures that have a common characteristic, e.g., three different kinds of fruit. The child has to choose two pictures with the same characteristics as the first three pictures out of a series of five or more pictures.

In Starting Cohort 1, the test was administered in an age-appropriate way on a tablet computer. As stated in the test manual of the SON-R 2½-7, the items were implemented as a learning test; i.e., after each item the children received feedback on the correctness of their answers (Tellegen et al., 2007). After three mistakes the test automatically ended.

The following variables are included in the Scientific Use File: all scored test and practice items (not solved/solved); an estimator of the individual test performance (Weighted Likelihood Estimator, WLE); a standard error for the estimator to relate the individual performance to the overall group (Hoijtink & Boomsma, 1995; Warm, 1989). No further background variables were included in the estimation model.

References

Cattell, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1–22. https://doi.org/10.1037/h0046743

Hoijtink, H. & Boomsma, A. (1995). On Person Parameter Estimation in the Dichotomous Rasch Model. In G. H. Fischer & I. W. Molenaar (Eds.), *Rasch Models* (pp. 53–68). New York, NY: Springer-Verlag. https://doi.org/10.1007/978-1-4612-4230-7_4

Quinn, P. C. (2011). Born to categorize. In U. Goswami (Ed.), *The Wiley-Blackwell Handbook of Childhood Cogniive Development* (pp. 129–153). Chichester, United Kingdom: Wiley-Blackwell. https://doi.org/10.1002/9781444325485.ch5

Spearman, C. (1923). *The nature of "intelligence" and the principles of cognition.* Oxford, United Kingdom: Macmillan.

Tellegen, P. J., Winkel, M., Wijnberg-Williams, B. J., & Laros, J. A. (2007). *Snijders-Oomen Non-verbaler Intelligenztest (SON-R2 1/2-7; German version).* Göttingen, Germany: Hogrefe.

Warm, T. A. (1989). Weighted likelihood estimation of ability in item response theory. *Psychometrika, 54*(3), 427–450. https://doi.org/10.1007/BF02294627

Wechsler, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist, 5*(3), 78–83. https://doi.org/10.1037/h0063112

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 abilities include different and diverse facets,

which are measured accordingly (e.g., self-/external judgement, direct standardized observations, experimental tasks). The NEPS covers different aspects and facets of self-regulation, aspects of cognitive self-regulation in the form of procedural and declarative metacognition (Weinert, Artelt, Prenzel, Senkbeil, Ehmke, Carstensen, & Lockl, 2019) as well as various survey items designed to capture emotional and behavioral self-control.

The NEPS Starting Cohort 1 (Newborns) and Starting Cohort 2 (Kindergarten) also use delay of gratification[4] tasks (for Starting Cohort 2, see Luplow, Schönmoser, Lorenz, & Schmitt, 2019). To assess delay of gratification, two procedures are implemented: the waiting paradigm and the choice paradigm (Mischel, 1974; Mischel, 2015). The waiting paradigm, which is used in Starting Cohort 1 for children aged 3 and 5 years, is used to examine the child's ability to delay an immediate small reward in favor of a bigger time-delayed reward (Mischel & Gilligan, 1964). Thus, the paradigm constitutes a reaction inhibition task (Garon, Bryson, & Smith, 2008), which can conceptually be assigned to executive functions, which are significant for controlling and planning actions and behavior (Neubauer et al., 2011). In Starting Cohort 1, executive functions are also measured at the age of 4 years with a flanker task.

The ability to self-regulate and cognitively process different types of behavior (Mischel, 1974) is considered to be important for child development. A number of studies indicate that well-developed abilities of self-regulation at preschool age are predictive of later academic performance, the ability to cope with stress, the development of socio-emotional competencies and the ability to concentrate (Baumeister & Vohs, 2004; Kochanska, Murray, & Coy, 1997; Tangney, Baumeister, & Boone, 2004; Wulfert, Block, Ana, Rodriquez, & Colsman, 2002), although these findings have also been the subject of criticism (Watts, Duncan, & Quan, 2018). Furthermore, there is evidence that the decision in favor of an immediate small gift or a bigger gift at a later time is related to the age of children, to social background characteristics and other child measures (Lemmon & Moore, 2007; Thompson, Barresi, & Moore, 1997; Watts et al., 2018). The tendency to wait usually rises with increasing age (Lorenz, Schmitt, Luplow, & Schönmoser, 2016).

In this wave, both the setting and the waiting conditions (Mischel, 1974) were standardized: two wrapped gifts – one big and one small – were presented to the child and a USB button was placed between the gifts. The child could either wait for a certain amount of time, which was unknown to the child, to receive the big gift or press the button to receive the small gift immediately. Hence, this decision determined which gift was selected and the waiting time. The maximum waiting time was 3 minutes (181 seconds), after which the child automatically received the big gift.

The following variables are included in the Scientific Use File: the waiting time (in seconds) and the child's decision (big vs. small gift).

References

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

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

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. https://doi.org/10.1037/0033-2209.134.1.31

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. https://doi.org/10.1111/j.1467-8624.1997.tb01939.x

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). https://doi.org/10.5157/NEPS:SC2:7.0.0

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. https://doi.org/10.1016/S0065-2601(08)60039-8

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. https://doi.org/10.1111/j.0022-3506.2004.00263.x

Thompson, C., Barresi, J., & Moore, C. (1997). The development of future-oriented prudence and altruism in preschoolers. *Cognitive Development, 12*(2), 199–212. https://doi.org/10.1111/j.0022-3506.2004.00263.x

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. https://doi.org/10.1177/0956797618761661

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. https://doi.org/10.1007/978-3-658-23162-0_4

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. https://doi.org/10.1111/1467-6494.05013

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.