

Informationen zur Kompetenztestung

NEPS Startkohorte 8 — Klasse 5
Bildung für die Welt von morgen

2. Welle: 6. Jahrgangsstufe

Urheberrechtlich geschütztes Material
Leibniz-Institut für Bildungsverläufe (LifBi)
Wilhelmsplatz 3, 96047 Bamberg
Direktorin: Prof. Dr. Cordula Artelt
Administrativer Direktor: Dr. Stefan Echinger
Bamberg; 14. August 2026

Informationen zur Testung an Regelschulen	
Stichprobe	Studie A105, Startkohorte 8, Klasse 6, Erhebungswelle 2, Jahr 2023
Testsituation	Gruppenerhebung in Räumlichkeiten der Schulen, in der Regel mit zwei Erhebungsleitungen und maximal 24 Schülerinnen und Schülern pro Erhebungsgruppe. Der Ablauf der Testung und die automatisierte Zeitmessung wurde von einem Master-PC aus durch die Erhebungsleitung gesteuert.
Ablauf der Testung	Die Erhebung fand an einem Tag statt. Die Schülerinnen und Schüler bearbeiteten zunächst Aufgaben aus den Bereichen Digitale Kompetenz und ICT Literacy. Nach einer 15-minütigen Pause folgte Hörverstehen auf Wortebene. Direkt im Anschluss erhielten Schülerinnen und Schüler mit entsprechendem sprachlichen Hintergrund Aufgaben zum Hörverstehen Russisch bzw. Türkisch, für die anderen Schülerinnen und Schüler gab es Aufgaben zum Allgemeinwissen. Nach jedem dieser einzelnen Tests wurde den Schülerinnen und Schülern jeweils eine zusätzliche Frage zur prozeduralen Metakognition gestellt. Alle Aufgaben wurden am Computer bearbeitet; für die Aufgaben zum Hörverstehen auf Wortebene, zum Hörverstehen Russisch bzw. Türkisch und zum Allgemeinwissen wurden Kopfhörer zur Verfügung gestellt. Nach der Aufgabenbearbeitung gab es wieder eine 15-minütige Pause. Abschließend füllten die Schülerinnen und Schüler einen computerbasierten Schülerfragebogen selbstständig aus (CASI). Testreihenfolge Regelschule - Digitale Kompetenz - ICT Literacy - Hörverstehen auf Wortebene - Allgemeinwissen ODER Hörverstehen Russisch bzw. Türkisch - Schülerfragebogen (CASI)
Testdauer (reine Bearbeitungszeit)	92 min bzw. 100 min (+ 40 min Schülerfragebogen)
Pausen	2 x 15 min
Gesamtadministration	ca. 207 min
Informationen zu den einzelnen Tests	

Konstrukt	Anzahl der Items	Vorgegebene Bearbeitungszeit	Erhebungsmodus	Nächste Messung
Digitale Kompetenz	23	20 min	TBT	2025
ICT Literacy	23 (Stage 1: 8, Stage 2:15) Pro Rotation 18)	28 min	TBT	2025
Hörverstehen auf Wortebene (Wortschatz)	max. 19 Sets mit jeweils 12 Aufgaben (mit Abbruchkriterium)	20 min	TBT	
Allgemeinwissen	64	20 min	TBT	2028
Hörverstehen Russisch bzw. Türkisch (L1-Test)	48	28 min	TBT	-
Prozedurale Metakognition	1 pro Konstrukt	jeweils 1 min	TBT	

Informationen zur Testung an Förderschulen				
Stichprobe	Studie A105, Startkohorte 8, Klasse 6, Erhebungswelle 2, Jahr 2023			
Testsituation	Gruppenerhebung in Räumlichkeiten der Schulen, in der Regel mit zwei Erhebungsleitungen und maximal 12 Schülerinnen und Schülern pro Erhebungsgruppe. Der Ablauf der Testung und die automatisierte Zeitmessung wurde von einem Master-PC aus durch die Erhebungsleitung gesteuert.			
Ablauf der Testung	Die Erhebung fand an einem Tag statt. Die Schülerinnen und Schüler bearbeiteten zunächst Aufgaben aus dem Bereich Digitale Kompetenz. Nach einer 15-minütigen Pause folgten Aufgaben zu Hörverstehen auf Wortebene und zu Allgemeinwissen. Alle Aufgaben wurden am Computer bearbeitet; für die Aufgaben zum Hörverstehen auf Wortebene und zum Allgemeinwissen wurden Kopfhörer zur Verfügung gestellt. Nach jedem dieser einzelnen Tests wurde den Schülerinnen und Schülern jeweils eine zusätzliche Frage zur prozeduralen Metakognition gestellt. Im Anschluss an die Aufgabenbearbeitung gab es eine 15-minütige Pause. Abschließend füllten die Schülerinnen und Schüler einen computerbasierten Schülerfragebogen selbstständig aus (CASI). Testreihenfolge - Digitale Kompetenz - Hörverstehen auf Wortebene - Allgemeinwissen - Schülerfragebogen (CASI)			
Testdauer (reine Bearbeitungszeit)	63 min (+ 30 min Schülerfragebogen (CASI))			
Pausen	2 x 15 min			
Gesamtadministration	ca. 145 min			
Informationen zu den einzelnen Tests				
Konstrukt	Anzahl der Items	Vorgegebene Bearbeitungszeit	Erhebungsmodus	Nächste Messung
Digitale Kompetenz	19	20 min	TBT	2025

Hörverstehen auf Wortebene (Wortschatz)	max. 19 Sets mit jeweils 12 Aufgaben (mit Abbruchkriterium)	20 min	TBT	
Allgemeinwissen	64	20 min	TBT	2028
Prozedurale Metakognition (Meta-P)	1 pro Konstrukt	jeweils 1 min	TBT	

Vorbemerkung

Der Entwicklung der einzelnen Tests liegen Rahmenkonzeptionen zugrunde. Dabei handelt es sich um übergeordnete Konzeptionen, auf deren Basis bildungsrelevante Kompetenzen über den gesamten Lebenslauf in konsistenter und kohärenter Weise abgebildet werden sollen. Die Rahmenkonzeptionen, auf deren Grundlage die Testinstrumente zur Messung der oben genannten Konstrukte entwickelt wurden, sind deshalb in den verschiedenen Studien identisch. Die Aufgabenformate können sich dabei unterscheiden.

Im Vergleich zur Regelschule wurde für die Erhebungen an Förderschulen ein Testheft für Digitale Kompetenz mit teilweise leichteren Items zusammengestellt.

Digitale Kompetenz

Vor dem Hintergrund der fortschreitenden Digitalisierung gewinnt ein kompetenter Umgang mit digitalen Medien zunehmend an Bedeutung für eine erfolgreiche gesellschaftliche Teilhabe. Um dieser Entwicklung Rechnung zu tragen, wird im NEPS ergänzend zu ICT Literacy (siehe Beschreibung in diesem Dokument; für ausführlichere Informationen siehe Senkbeil et al., 2013) die Digitale Kompetenz erfasst (Artelt, 2025).

Das NEPS definiert Digitale Kompetenz als den verantwortungsvollen und reflektierten Umgang mit digitalen Technologien, der auf umfassenden Kenntnissen sowie einer kritischen Auseinandersetzung mit den Inhalten, Mechanismen und Auswirkungen digitaler Medien und Technologien beruht (Lockl et al., in Vorbereitung; Tural & Schwaß, 2023). In diesem Rahmenkonzept wird Digitale Kompetenz in vier Facetten gegliedert (siehe Abbildung 1). (1) Informationen bewerten: Dazu gehört die Beurteilung der Glaubwürdigkeit von Quellen sowie das Verständnis dafür, wie algorithmische Filter den Zugang zu und die Sichtbarkeit von Informationen beeinflussen. Außerdem umfasst die Facette ein grundlegendes Verständnis der Funktionsweise von Systemen, die auf künstlicher Intelligenz basieren. (2) Absichten und Strategien erkennen: Dabei geht es um das Erkennen von Absichten, die darauf abzielen, Verhalten zu beeinflussen, darunter kommerzielle, soziale oder politische Absichten. Dazu gehört auch das Wissen über Strategien und Mechanismen der Beeinflussung, einschließlich über Phänomene der Manipulation und Beeinflussung von Meinungen sowie der Einschränkung der Meinungsvielfalt durch Filterblasen und Echokammern. (3) Kommunizieren und Interagieren: Dazu gehören eine effektive und verantwortungsvolle Zusammenarbeit sowie soziales Engagement in digitalen Kontexten, einschließlich des Bewusstseins für mögliche soziale Folgen, wie beispielsweise Phänomene wie Hashtag-Debatten, Fake-Profile, Cybermobbing sowie Shitstorms und Candystorms. (4) Mit Daten umgehen: Diese Facette befasst sich mit dem verantwortungsvollen Umgang mit den eigenen Daten, einschließlich Privatsphäre-Einstellungen, Fragen der Datensammlung, Möglichkeiten zur Verringerung digitaler Spuren sowie Problemstellungen im Zusammenhang mit der Weitergabe persönlicher Daten. Die Facette umfasst auch den Umgang mit den Daten anderer Personen und geistigen Eigentum, einschließlich ethischer und rechtlicher Aspekte.



Abb. 1: Facetten der Rahmenkonzeption von Digitale Kompetenz im NEPS

Aufgabentypen und -formate

Diese Facetten werden durch szenariobasierte Items operationalisiert, die typische digitale Situationen widerspiegeln, die Schülerinnen und Schüler der unteren Sekundarstufe erleben. Die Aufgaben werden sowohl in text- als auch in bildbasierter Form präsentiert, wobei beide Typen als einfache und komplexe Multiple-Choice-Aufgaben umgesetzt wurden. Das Bildmaterial in den bildbasierten Aufgaben stellt realistische digitale Umgebungen dar, wie beispielsweise nachgebildete Websites oder digitale Plattformen, und bieten so authentische Kontexte, in denen das Wissen und das Verständnis der Schülerinnen und Schüler für digitale Prozesse und Praktiken gemessen werden können.

Bei der Zusammenstellung der Testhefte für eine Erhebung wird jeweils angestrebt, die vier Facetten des Rahmenkonzepts balanciert abzubilden. Innerhalb der Facetten erfolgt die Auswahl und Gewichtung der hier beschriebenen Aspekte jedoch in jeder Erhebung in Abhängigkeit der jeweiligen Zielgruppe.

Skalierung der Items

Die Testitems werden raschskaliert und über Link-Studien längsschnittlich miteinander verknüpft. Die Ergebnisse der Skalierung für die Klassenstufe 6 sind in den Technical Reports zu finden (Kutscher, 2026a, 2026b).

Literatur

- Artelt, C. (2025). Bildung in Zeiten von Digitalität und Künstlicher Intelligenz. In: N. McElvany, C. Dignath, M. Becker, u. a. (Hrsg.), *Welche Kompetenzen soll die Schule von heute für die Gesellschaft von morgen vermitteln?* (S. 41–55). Münster: Waxmann. doi: 10.31244/9783818850111.04.
- Kutscher, T. (2026a). NEPS technical report for digital competence: scaling results of starting cohort 8 for grade 6 (NEPS Survey Paper No. 137). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/NEPS:SP137:1.0>.
- Kutscher, T. (2026b). NEPS technical report for digital competence: scaling results of starting cohort 8 for grade 6 in special schools (NEPS Survey Paper No. 136). Leibniz Institute for Educational Trajectories. <https://doi.org/10.5157/NEPS:SP136:1.0>.
- Lockl, K., Tural, S., Schwaß, M., Kutscher, T., Hornberger, M., Artelt, C., & Wolter, I. (in Vorbereitung). Measuring digital competence in the NEPS - Conceptual framework and findings from a pilot study.
- Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013). The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*, 5, 139–161.
- Tural, S., & Schwaß, M. (2023, February 28–March 3). Digital natives = digital experts? Designing a novel instrument to measure digital competence in NEPS-SC8 secondary school students

[Posterpräsentation]. 10. Tagung der Gesellschaft für Empirische Bildungsforschung (GEBF), Universität Duisburg-Essen, Deutschland.

ICT Literacy

Die Fähigkeit, Informations- und Kommunikationstechnologien effektiv zu nutzen (ICT Literacy), spielt in Schulen, am Arbeitsplatz und im täglichen Leben der Menschen eine wichtige Rolle (Fraillon, Ainley, Schulz, Friedman, & Duckworth, 2019). Daher haben Forscher und Organisationen (z.B. die Europäische Kommission und die International Society for Technology in Education) Rahmenkonzeptionen zur Förderung der ICT Literacy entwickelt, in denen Kompetenzen und Fähigkeiten beschrieben werden, die für die Wissensgesellschaft als wichtig erachtet werden (Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016). Neuere Konzeptualisierungen von ICT Literacy integrieren technologische und kognitive Aspekte zur Definition dieser Kompetenz. Technologische Aspekte umfassen die Kenntnis von Hardware- und Softwareanwendungen und das Verständnis technologischer Konzepte. Kognitive Aspekte, die als Informationskompetenz bezeichnet werden und die Fähigkeit umfassen, digitale Medien zu nutzen, um auf Informationen zuzugreifen, sie zu erstellen, zu verwalten und kritisch zu bewerten und sie effektiv für die eigenen Zwecke zu nutzen, spielen ebenfalls eine wichtige Rolle (ETS, 2002; Fraillon et al., 2019).

ICT Literacy stellt somit eine Metakompetenz dar, die Menschen unterstützt, wichtige Kompetenzen und Fähigkeiten für Bildungs- und Arbeitssituationen zu erwerben und private Zielstellungen über die gesamte Lebensspanne umzusetzen (van Laar, van Deursen, van Dijk, & de Haan, 2017). Da sich Menschen über die Lebensspanne hinweg selbstgesteuert neues Wissen und Fähigkeiten aneignen müssen und diese zunehmend durch digitale Medien vermittelt werden, ist ICT Literacy zudem eine notwendige Voraussetzung, um mit den aktuellen Entwicklungen im Bereich der digitalen Medien erfolgreich Schritt zu halten (Goldhammer, Gniewosz, & Zylka, 2017).

Eine weit verbreitete Definition von ICT Literacy, auf die auch wir uns beziehen, wurde vom *ICT Literacy Panel* formuliert: „ICT literacy is the ability to appropriately use digital technology, communication tools, and/or networks to solve information problems in order to function in an information society. This includes having the ability to use technology as a tool to research, organize, and communicate information“ (ETS, 2002, p. 16).

In Anlehnung an die Definition des *International ICT Literacy Panel* wird ICT Literacy im Nationalen Bildungspanel als eindimensionales Konstrukt konzeptualisiert, das sich gemäß der oben erwähnten Definition in verschiedene Prozess- und Softwarekomponenten differenzieren und als Strukturmodell darstellen lässt (Senkbeil, Ihme & Wittwer, 2013a,b). Die Prozesskomponenten beschreiben die benötigten Wissensbestände und Fertigkeiten, die für einen zielorientierten Umgang mit Informations- und Kommunikationstechnologien notwendig sind. Jede Prozesskomponente wird als das Zusammenwirken von kognitiven und technologischen Fähigkeiten verstanden. Folgende Prozesskomponenten werden differenziert

Anwenden und Verstehen: Grundkenntnisse über das Betriebssystem und relevante Programmanwendungen (z. B. Textverarbeitung) sowie Kenntnisse grundlegender Operationen, um auf Informationen zugreifen zu können (z. B. ein Dokument öffnen und speichern)

Erzeugen: Fähigkeit, Dokumente und Dateien zu bearbeiten oder zu erstellen (z. B. Tabellen anlegen, Formeln setzen)

Suchen und Organisieren: Fähigkeit, Informationen effizient zu ermitteln (z. B. adäquate Suchbegriffe eingeben) oder Informationen miteinander nach spezifischen Kriterien zu vergleichen (z. B. Datensätze sortieren)

Bewerten: Fähigkeit, Informationen (z. B. hinsichtlich Glaubwürdigkeit) zu bewerten und auf dieser Grundlage Entscheidungen zu treffen.

Die inhaltsbezogenen Komponenten beinhalten diejenigen Programmanwendungen, die für die Bewältigung informationsbezogener Anforderungen notwendig sind: Differenziert werden (a) *Betriebssystem / Textverarbeitung*, (b) *Tabellenkalkulation und Präsentationsprogramme*, (c) *E-Mail- und andere Kommunikationsanwendungen* wie z. B. Foren und (d) *Internet bzw. internetgestützte Suchmaschinen und Datenbanken*. Alle Testaufgaben wurden so konstruiert, dass sie jeweils genau einer Prozesskomponente und einer Programmanwendung zugeordnet sind.

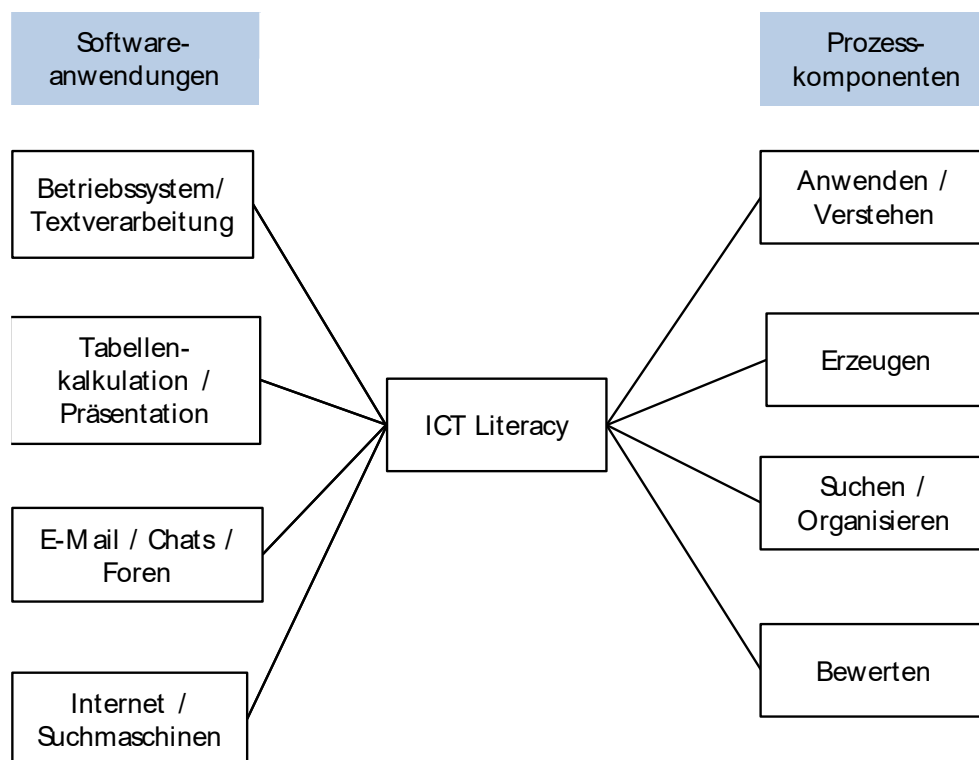


Abb. 2: Rahmenkonzeption von ICT Literacy im Nationalen Bildungspanel

Aufgabentypen und -formate

Die Erfassung von ICT Literacy beinhaltete zwei Arten von Aufgaben.

Der erste Aufgabentyp, die wissensbasierten und statischen Items, die etwas mehr als die Hälfte der Aufgaben ausmachen, waren Bleistift-und-Papier-Fragen. Diese Aufgaben beschreiben realistische Problemstellungen in einer Fülle authentischer Situationen und erfassen im Sinne deklarativer und

prozeduraler Wissensbestände, ob die Testpersonen angemessen mit bestimmten Aufgabenstellungen umgehen können (z.B. eine Datei auf einem bestimmten Laufwerk speichern). Dabei werden sie gefragt, was sie in der betreffenden Situation tun würden. Für eine möglichst realitätsnahe Gestaltung werden in den Aufgabenstimulus Screenshots integriert (z. B. von einem Internet-Browser oder Tabellenkalkulationen). Häufig werden als Distraktoren realistische Antwortalternativen in Form von Schaltflächen oder Menüs vorgegeben, die in die jeweiligen Screenshots integriert sind, oder es werden reale Programmanwendungen herangezogen und für die Konstruktion der Antwortmöglichkeiten verwendet (Senkbeil et al., 2013b). Bei den statischen Aufgaben gab es zwei Arten von Antwortformaten: einfache Multiple-Choice-Aufgaben (MC) und komplexe Multiple-Choice-Aufgaben (CMC). Bei MC-Items musste der Testteilnehmer die richtige Antwort aus vier bis sechs Antwortmöglichkeiten identifizieren, wobei eine Option richtig war und drei bis fünf Antwortmöglichkeiten als Distraktoren fungierten (d. h. sie waren falsch). Bei CMC-Aufgaben wurden mehrere Teilaufgaben mit jeweils zwei Antwortmöglichkeiten (richtig/falsch) gestellt.

Bei der zweiten Art von Aufgaben handelte es sich um interaktive Aufgaben, bei denen generische Software oder universelle Anwendungen simuliert wurden, um eine Aktion durchzuführen. Diese interaktiven Aufgaben bezogen sich zusätzlich auf prozedurales und strategisches Wissen (z. B. Planung, Ausführung und Überwachung des Problemlösungsprozesses). Die Befragten mussten bestimmte Aufgaben und Probleme lösen, indem sie diese Simulationen anwendeten und mit ihnen interagierten: Dabei kann es sich um Einzelaufgaben handeln (z. B. das Öffnen eines Webbrowsers) oder um eine Abfolge von Schritten (z. B. "Speichern unter" mit einem bestimmten Dateinamen, Sortieren oder Filtern einer webbasierten Datenbank nach einem oder mehreren Kriterien). Die interaktiven Aufgaben enthielten lineare und nichtlineare Aufgaben. Lineare Aufgaben erforderten die Ausführung von mindestens zwei Befehlen in einer vorgeschriebenen Reihenfolge (z. B. Öffnen einer Datei vom Desktop, "Speichern unter" mit einem bestimmten Dateinamen und Verschieben der Datei auf ein anderes Laufwerk). Bei nichtlinearen Aufgaben mussten die Befragten ein gewünschtes Ergebnis durch die Ausführung einer Reihe von Unterbefehlen erreichen, wobei die Reihenfolge der Befehle variabel war.

Skalierung der Items

Die Testitems werden raschskaliert und über Link-Studien längsschnittlich miteinander verknüpft (Fischer, Rohm, Gnams & Carstensen, 2016). Die Gütekriterien und psychometrischen Kennwerte der Items werden in den Technical Reports der jeweiligen Startkohorten ausgewiesen.

Literatur

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., Gnams, T., & Carstensen, C. H. (2016). *Linking the data of the competence tests (NEPS Survey Paper No. 1)*. Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study.

- Fraillon, J., Ainley, J., Schulz, W., Friedman, T. & Duckworth, D. (2019). *Preparing for life in a digital world: IEA International Computer and Information Literacy Study 2018 International Report*. Amsterdam: International Association for the Evaluation of Educational Achievement (IEA).
- Goldhammer, F., Gniewosz, G., & Zylka, J. (2017). ICT engagement in learning environments. In S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Eds.), *Assessing contexts of learning. An international perspective* (pp. 331–354). Springer International Publishing.
- Senkbeil, M. & Ihme, J. M. (2020). Diagnostik von ICT Literacy: Messen Multiple-Choice-Aufgaben und simulationsbasierte Aufgaben vergleichbare Konstrukte? Vergleich der Testergebnisse zweier Instrumente aus den aktuellen Large-Scale-Studien ICILS 2013 und NEPS. *Diagnostica*, 66, 147–157.
- Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013a). The Test of Technological and Information Literacy (TILT) in the National Educational Panel Study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online*, 5, 139–161.
- Senkbeil, M., Ihme, J. M., & Wittwer, J. (2013b). Entwicklung und erste Validierung eines Tests zur Erfassung technologischer und informationsbezogener Literacy (TILT) für Jugendliche am Ende der Sekundarstufe I. *Zeitschrift für Erziehungswissenschaft*, 16, 671-691.
- Siddiq, F., Hatlevik, O. E., Olsen, R. V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning from the past – A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, 19, 58–84.
- van Laar, E., van Deursen, A.J.A.M., van Dijk, J.A.G.M., & de Haan, J. (2017). The relation between 21st-century skills and digital skills: A systematic literature review. *Computers in Human Behavior*, 72, 577–588.

Wortschatz: rezeptives Hörverständnis auf Wortebene

Hörverstehen auf Wort-, Satz- und Text-/Diskursebene als Indikatoren der Sprachkompetenz im Deutschen

Die Bedeutung sprachlicher Kompetenzen für schulisches Lernen sowie für die Erklärung sozialer Disparitäten in den Schulkarrieren ist weitgehend unbestritten.

Die sprachlichen Kompetenzen im Deutschen werden in NEPS einerseits über das Hörverstehen auf Wort-, Satz- und Text-/Diskursebene sowie andererseits – ab der 2. Grundschulklasse – über Indikatoren der Lesefähigkeiten (Lesekompetenz [Textverständnis], Lesegeschwindigkeit) erfasst. Das Hörverstehen auf Wortebene wird in der Startkohorte 8 einmalig in der Klassenstufe 6 erfasst.

Hörverstehen auf Wortebene (rezeptiver Wortschatz)

Maße des rezeptiven Wortschatzes stellen einen guten, international anschlussfähigen Indikator für die erworbenen sprachlichen Fähigkeiten und Fertigkeiten von Kindern und Erwachsenen dar. In zahlreichen großen internationalen Panelstudien wie zum Beispiel dem Head Start Family and Child Experiences Survey – FACES (USA)¹, dem National Longitudinal Survey of Children and Youth – NLCSY (Kanada; u.a. Lipps & Yiptong-Avila, 1999), der British Cohort Study – BCS70 (z.B. Bynner, 2004) oder der European Child Care and Education (ECCE)-Study, die in Deutschland, Österreich, Spanien und Portugal durchgeführt wurde (z.B. European Child Care and Education (ECCE)-Study Group, 1997), wird der passive Wortschatz als zentraler, manchmal sogar als alleiniger Indikator der vor dem Hintergrund individueller Grundfähigkeiten (z.B. Arbeitsgedächtniskapazität, Geschwindigkeitsvariablen) und Umweltanregung kumulativ erworbenen sprachlich-kognitiven Fähigkeiten erhoben.

Das international zur Erfassung des rezeptiven Wortschatzes am meisten eingesetzte Instrument ist der inzwischen in verschiedenen Versionen vorliegende Peabody Picture Vocabulary Test (PPVT; Dunn, 1959; Dunn & Dunn, 1981, 1997, 2007). Der PPVT ist im Grundsatz über einen sehr großen Altersbereich (von 2,5 Jahren bis ins hohe Erwachsenenalter) hinweg einsetzbar und zugleich einfach in der Durchführung und Auswertung. In Startkohorte 8 wird eine deutschsprachige Version des PPVT-IV (Dunn & Dunn, 2007; deutsche Bearbeitung von Lenhard, Lenhard, Segerer & Suggate, 2015) eingesetzt. Die Administrierung erfolgte in der Startkohorte 8 über einen Laptop mit Kopfhörern. Aufgabe der Schülerinnen und Schüler war es, zu jedem einzeln vorgegebenen Wort, welches auditiv vom Laptop vorgegeben wird, aus jeweils vier Bildern das zu dem Wort gehörige Bild auszuwählen.

Nach den Vorgaben des PPVT-IV variiert der Schwierigkeitsgrad (Testeinstieg und -abbruch) in Abhängigkeit vom Alter und der Leistung der Schülerinnen und Schüler. Der Test beginnt mit einer Übungseinheit, die je nach Alter und Leistung des Kindes gesteuert wird. Wurden in der Übungsphase mindestens 2 Aufgaben richtig gelöst, beginnt die Testphase. Insgesamt besteht der Test aus 19 schwierigkeitsgestaffelten Sets mit jeweils 12 Items.

Testablauf in dieser Welle: Der Test beginnt mit einer Übungsphase, die aus mind. 2 und max. 6 Aufgaben besteht. Das Startset ist abhängig vom Alter der Schülerinnen und Schüler. Wurde im Startset mehr als 1 Fehler gemacht, wird zum nächstniedrigeren Set gesprungen, bis in einem Set maximal 1 Fehler gemacht wurde (Bodenset). Danach wird der Test – unter Auslassung der bereits

¹ <http://www.acf.hhs.gov/programs/opre/hs/faces/>

bearbeiteten Sets – so lange durchgeführt, bis das Deckenset mit mehr als 7 Fehlern identifiziert wurde.

Im Scientific Use File finden sich die Anzahl administrierter Übungsisems, Korrektheit der Antworten für jedes Testitem (richtig, falsch), das Bodenset und das Deckenset. Zudem wird der Summenwert angegeben, der die Anzahl der richtig gelösten Items angibt. Dabei werden alle Items, die sich in niedrigeren Sets als das Bodenset befinden, als richtig gelöst angenommen.

Literatur

- 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). Adelaide, Australia: National Centre for Vocational Education Research. Dunn, L. M. (1959). *Peabody Picture Vocabulary Test (PPVT): Manual of directions and forms*. Nashville, TN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1981). *Peabody Picture Vocabulary Test-Revised (PPVT-R)*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. M. (1997). *Peabody Picture Vocabulary Test, Third Edition (PPVT-III)*. Circle Pines, MN: American Guidance Service.
- Dunn, L. M., & Dunn, L. 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, Fachbereich Erziehungswissenschaft, .
- Lenhard, A., Lenhard, W., Segerer, R., & Suggate, S. (2015). *Peabody picture vocabulary test-4*. Frankfurt, Germany: Pearson.
- Lipps, G., & Yiptong-Avila, J. (1999). From home to school: how Canadian children cope. *Education Quarterly Review*, 6(2), 51–57.

Allgemeinwissen

Ergänzend zur Erfassung funktionaler, domänenspezifischer Kompetenzen sowie weiterer Kompetenzbereiche wird im NEPS auch das Allgemeinwissen der Schülerinnen und Schüler berücksichtigt. Basierend auf der Theorie der fluiden und kristallinen Intelligenz von Cattell und Horn (Horn, 1965, 1988; Cattell, 1971) und dem CHC-Modell (McGrew, 2009) zählt das Allgemeinwissen zu den kristallinen Fähigkeiten, deren Ausprägung maßgeblich auf zuvor erworbenem Wissen und erlernten Fertigkeiten beruht.

Zur Erfassung des Allgemeinwissens von Schülerinnen und Schülern der 6. Klassenstufe in der Startkohorte 8 wird eine Parallelversion des Subtests „Allgemeinwissen“ des Berliner Tests zur Erfassung fluider und kristalliner Intelligenz für die 5. bis 7. Jahrgangsstufe (BEFKI 5–7; Schroeders, Schipolowski & Wilhelm, 2020) eingesetzt.

Dieser Test erfasst deklaratives Faktenwissen aus 16 verschiedenen Inhaltsbereichen, die den drei breiten Domänen Natur-, Geistes- und Sozialwissenschaften zugeordnet werden können (vgl. Tabelle 1). Die Auswahl der Inhaltsbereiche durch die Testautoren und deren Zuordnung zu den breiten Domänen orientierte sich an der empirischen Klassifikation von Ackerman (2000) und einer Sichtung bestehender Testverfahren (Amthauer, Brocke, Liepmann & Beauducel, 2001; Hossiep und Schulte, 2007). Ziel der Testkonstruktion war es, deklaratives Faktenwissen möglichst breit zu erfassen und dabei sowohl schulisch vermitteltes als auch außerschulisch erworbenes Wissen zu berücksichtigen, das in der Gesellschaft gemeinhin honoriert und zum kulturellen Gut gezählt wird und sich somit von unbeständigem oder nur eingeschränkt gültigem Wissen wie etwa aktuellen Fußballergebnissen oder dem regionalen Busfahrplan abgrenzt. Ausgeklammert wurden mit Blick auf das Alter der Zielpopulationen solche Wissensbereiche, die sich auf akademisches Spezialwissen beziehen, das typischerweise erst im tertiären Bildungsbereich erworben wird (vgl. etwa den Bereich Psychologie bei Ackerman, 2000).

Tabelle 1

Erfasste Wissensbereiche, deren Inhalte und Zuordnung zu den drei breiten Wissensdomänen

Domäne	Inhaltsbereich	Inhalte (beispielhaft)
Natur- wissen- schaften	Physik	physikalische Größen und Phänomene (z. B. aus Energielehre und Optik)
	Chemie	Organische und anorganische Chemie (z. B. Eigenschaften von Stoffen, chemische Symbole)
	Biologie	Biologische Vorgänge, Tierreich und menschlicher Körper (z. B. Anatomie, Merkmale bestimmter Tierarten)
	Medizin	Erkrankungen des Menschen und ihre Behandlung (z. B. Symptome und Übertragungswege von Krankheiten)
	Geografie	Erdkunde und Geologie (z. B. Lage bestimmter Gewässer, Merkmale von Gesteinen)

	Technologie	Technologien, technische Geräte und Kommunikation (z. B. Energieerzeugung, Internet)
Geistes- wissen- schaften	Literatur	Literarische Werke und Textsorten
	Kunst	künstlerische Techniken und Farbenlehre
	Musik	Instrumente und Musikstile, Werkkenntnisse
	Religion	Konzepte und bedeutende Vertreter verschiedener Weltreligionen
	Philosophie	Bedeutende Philosophen und Fragestellungen der Philosophie
Sozial- wissen- schaften	Geschichte	Historische Ereignisse und Personen aus mehreren Jahrtausenden
	Recht	Rechtliche Institutionen und Gesetze in Deutschland (z. B. Verbraucherrechte, Urheberrecht)
	Politik	Politische Institutionen und Wahlen (z. B. Bundesregierung, Europäische Union)
	Wirtschaft	Prinzipien und Begriffe aus Volks- und Betriebswirtschaft (z. B. Warenverkehr, Gewerkschaften)
	Finanzen	Begriffe aus der Finanzwelt (z. B. Börse, Währungen)

Der Test umfasst insgesamt 64 Items, wobei jeder der 16 Wissensbereiche durch vier Items repräsentiert wird. Für die Bearbeitung des Tests stehen insgesamt 20 Minuten zur Verfügung. Sämtliche Aufgaben sind als Multiple-Choice-Items mit jeweils vier Antwortalternativen konzipiert, von denen genau eine Antwort korrekt ist.

Literatur

- Ackerman, P. L. (2000). Domain-specific knowledge as the „dark matter“ of adult intelligence: Gf /Gc, personality and interest correlates. *Journal of Gerontology: Psychological Sciences*, 55B, 69-84. <https://doi.org/10.1093/geronb/55.2.P69>
- Amthauer, R., Brocke, B., Liepmann, D., & Beauducel, A. (2001). *Intelligenz-Struktur-Test 2000 R*. Göttingen: Hogrefe.
- Cattell, R. B. (1971). *Abilities: Their Structure, Growth, and Action*. Boston, MA: Houghton Mifflin.
- Horn, J. L. (1965). *Fluid and crystallized intelligence: A factor analytic study of the structure among primary mental abilities*. Dissertation, University of Illinois.
- Horn, J. L. (1988). *Thinking about human abilities*. In J. R. Nesselroade (Ed.), *Handbook of multivariate psychology* (pp. 645-685). New York: Academic Press.

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

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

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

Hörverstehen in den Herkunftssprachen Russisch und Türkisch (L1²)

Der Zusammenhang von herkunftssprachlichen Kompetenzen mit dem Bildungserfolg von Personen aus zugewanderten Familien ist umstritten. Einerseits legen theoretische Ansätze und empirische Befunde nahe, dass Kompetenzen in der L1 den Erwerb der Zweitsprache und den Bildungserfolg im Aufnahmeland positiv beeinflussen können (z. B. Cummins, 1979; Edele et al., 2023; Edele & Stanat, 2016; Kempert et al., 2016). Andererseits werden negative oder nicht signifikante Zusammenhänge berichtet (z. B. Edele et al., 2023; Esser, 2006). Die empirische Befundlage ist unter anderem deshalb unzureichend, weil herkunftssprachliche Kompetenzen bislang nur selten systematisch mithilfe objektiver Indikatoren erfasst wurden (vgl. Edele, Seuring, et al., 2015; Kristen et al., 2011).

Um diese Annahmen im Rahmen des NEPS untersuchen zu können, wird die rezeptive Beherrschung der Herkunftssprachen Russisch und Türkisch erfasst. Die Tests richten sich an Schülerinnen und Schüler, die neben Deutsch entweder Russisch oder Türkisch sprechen und deren Familien aus der Türkei bzw. aus Staaten der ehemaligen Sowjetunion nach Deutschland zugewandert sind. Die in der Startkohorte 8 eingesetzten Hörverstehenstests stellen eine Weiterentwicklung der bereits in den Startkohorten 3 und 4 verwendeten L1-Tests dar (Tinkl et al., in Vorbereitung). Das Hörverstehen wurde als Testmodalität gewählt, da die Herkunftssprache häufig vor allem im familiären Kontext erworben wird und dieser Spracherwerb nicht notwendigerweise den Erwerb schriftsprachlicher Kompetenzen einschließt.

Die weiterentwickelten L1 Tests enthalten unabhängige und an die Jahrgangsstufe 6 angepasste Texteinheiten. Da sich die früheren Tests insbesondere für Schülerinnen und Schüler mit hohen L1-Kompetenzen als vergleichsweise leicht erwiesen und im oberen Kompetenzbereich nur eingeschränkt differenzierten (Edele et al., 2012; Edele, Schotte, et al., 2015), wurden zusätzliche anspruchsvollere Texteinheiten entwickelt. Um einen breiten Kompetenzbereich bei zugleich effizienter Testdauer abzudecken, wurden die L1 Tests in der SC8 als *branched-adaptive tests* umgesetzt. Im Unterschied zu vollständig adaptiven Tests, bei denen die Auswahl einzelner Aufgaben fortlaufend an die bisherigen Antworten angepasst wird, erfolgt beim *branched-adaptive Design* die Zuweisung zu einem Testpfad nach der Bearbeitung einer vorab festgelegten Gruppe von Aufgaben (Kubinger & Wild, 1989). Das zweistufige L1-Testdesign umfasste zunächst eine gemeinsame *routing stage* mit drei Texten mit insgesamt 14 Aufgaben. Bei mindestens neun richtigen Antworten erfolgte die Zuweisung zum *schweren Strang*, andernfalls zum *leichten Strang*. Anschließend wurden fünf weitere Texteinheiten bearbeitet, sodass jede Schülerin und jeder Schüler insgesamt acht Texteinheiten erhielt. Die Tests wurden als technologiebasierte Assessments (TBA) umgesetzt und von Schülerinnen und Schülern individuell am Computer mit Kopfhörern bearbeitet. Die Hörtexte wurden von professionellen Sprecher:innen mit Erstsprache Russisch bzw. Türkisch gesprochen und als standardisierte Audioaufnahmen wiedergegeben. Auch die zugehörigen Fragen und Antwortmöglichkeiten wurden auditiv präsentiert und zugleich auf dem Bildschirm angezeigt. Die Hörtexte selbst erschienen nicht in schriftlicher Form. Die Schülerinnen und Schüler beantworteten die Aufgaben im Multiple-Choice-Format direkt am Computer. Die Testkonzeption folgte dem Ziel, das Hörverstehen möglichst breit abzubilden. Deshalb wurden sowohl konzeptionell schriftsprachliche Texte in Form von Narrationen

² Der Begriff L1 (von „first language“) wird hier synonym mit der Sprache des Herkunftslands der Familie verwendet, unabhängig davon, ob die Sprache tatsächlich die erste von der Zielperson erworbene Sprache darstellt.

und Expositionen als auch konzeptionell mündliche Texte in Form von Dialogen eingesetzt. Die Hörtexte behandeln sowohl alltagsbezogene als auch schulbezogene Themen. Um eine Konfundierung der Testleistung mit dem inhaltlichen Vorwissen der Testpersonen möglichst zu vermeiden, wurden entweder Inhalte gewählt, mit denen alle Schülerinnen und Schüler in ähnlichem Maße vertraut sein sollten, oder Inhalte, die für alle gleichermaßen neuartig sein dürften. Die neu entwickelten Texteinheiten und Aufgaben wurden im Rahmen von Vorstudien erprobt.

Literatur

- Cummins, J. (1979). Linguistic interdependence and the educational development of bilingual children. *Review of Educational Research* 49(2), 222–251. <https://doi.org/https://doi.org/10.3102/00346543049002222>
- Edele, A., Schotte, K., Hecht, M., & Stanat, P. (2012). Listening comprehension tests of immigrant students' first language (L1) Russian and Turkish in grade 9: Scaling procedure and results. *NEPS Working Paper No. 13*. <https://doi.org/10.5157/NEPS:WP13:1.0>
- Edele, A., Schotte, K., & Stanat, P. (2015). Assessment of immigrant students' listening comprehension in their first languages (L1) Russian and Turkish in grade 9: Extended report of test construction and validation. *NEPS Working Paper No. 57*. <https://doi.org/10.5157/NEPS:WP57:1.0>
- Edele, A., Seuring, J., Kristen, C., & Stanat, P. (2015). Why bother with testing? The validity of immigrants' self-assessed language proficiency. *Social Science Research*, 52, 99–123. <https://doi.org/10.1016/j.ssresearch.2014.12.017>
- Edele, A., Seuring, J., Schotte, K., Kristen, C., & Stanat, P. (2023). Is the first language a resource, an obstacle, or irrelevant for language minority students' education? In S. Weinert, G. J. Blossfeld, & H.-P. Blossfeld (Eds.), *Education, Competence Development and Career Trajectories: Analysing Data of the National Educational Panel Study (NEPS)* (pp. 349–367). Springer. https://doi.org/10.1007/978-3-031-27007-9_16
- Edele, A., & Stanat, P. (2016). The role of first-language listening comprehension in second-language reading comprehension. *Journal of Educational Psychology*, 108(2), 163–180. <https://doi.org/10.1037/edu0000060>
- Esser, H. (2006). *Sprache und Integration: Die sozialen Bedingungen und Folgen des Spracherwerbs von Migranten* (1. Aufl. ed.). Campus Verlag.
- Kempert, S., Edele, A., Rauch, D., Wolf, K. M., Paetsch, J., Darsow, A., Maluch, J., & Stanat, P. (2016). Die Rolle der Sprache für zugewanderungsbezogene Ungleichheiten im Bildungserfolg [The role of language for immigration-related disparities in academic achievement]. In C. Diehl, C. Hunkler, & C. Kristen (Eds.), *Ethnische Ungleichheiten im Bildungsverlauf: Mechanismen, Befunde, Debatten* (pp. 157–241). Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-04322-3_5
- Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., & Will, G. (2011). The education of migrants and their children across the life course. *Zeitschrift für Erziehungswissenschaft*, 14(2), 121–137. <https://doi.org/10.1007/s11618-011-0194-3>
- Kubinger, K. D., & Wild, B. (1989). Die Optimierung der Meßgenauigkeit beim "branched"-adaptiven Testen. In K. D. Kubinger (Ed.), *Moderne Testtheorie* (pp. 187–218). Psychologie Verlags Union.
- Tinkl, L. A., Jindra, C., & Edele, A. (in preparation). *Assessment of immigrant students' listening comprehension in their first language (L1) Russian and Turkish in Grade 6 in Starting Cohort 8: further development, test construction and validation*. (NEPS Working Paper)

Metakognition

Unter Metakognition wird das Wissen über und die Kontrolle des eigenen kognitiven Systems verstanden. Gemäß Flavell (1979) und Brown (1987) werden deklarative und prozedurale Aspekte der Metakognition unterschieden, die beide im Nationalen Bildungspanel erfasst werden. In dieser Startkohorte wird nur die prozedurale Metakognition administriert.

Prozedurale Metakognition

Zur prozeduralen Metakognition gehört die Regulation des Lernprozesses durch Aktivitäten der Planung, Überwachung und Kontrolle. Der prozedurale Aspekt der Metakognition wird im Rahmen von NEPS in Kombination mit den Kompetenztests der einzelnen Domänen dabei nicht als direktes Maß derartiger Planungs-, Überwachungs- und Kontrollaktivitäten gemessen, sondern als metakognitives Urteil, das sich auf die Überwachung der Lernleistung während (bzw. kurz nach) der Lernphase bezieht (s.a. Nelson & Narens, 1990). Hierzu werden die Studienteilnehmerinnen und -teilnehmer nach Bearbeitung der jeweiligen Kompetenztests gebeten, ihre eigene Leistung in dem gerade bearbeiteten Test einzuschätzen. Erfragt wird die Anzahl der vermutlich richtig gelösten Aufgaben.

In der Startkohorte 8 erfolgt diese Selbsteinschätzung auf einer 11-stufigen Skala. Die Einschätzungen werden von 0 (vermutlich keine Aufgabe richtig gelöst) bis 10 (vermutlich alle Aufgaben richtig gelöst) kodiert.

Literatur

- 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 (Eds.), *The psychology of learning and motivation* (pp. 125-141). New York: Academic Press.