

Daten- und Methodenbericht
Dezember 2022

Andreas Woisch | Barbara Franke | Heiko Quast | Dennis Föste-Eggers | Hanna Menges | Ann-Christin Renneberg | Laura Schoger | Anna Meine | Daniel Klein | Anne Weber

DZHW- Studienberechtigten- panel 2018

Daten- und Methodenbericht zum DZHW-
Studienberechtigtenpanel 2018 (1. Befragungswelle)

Dieses Werk steht unter der Creative Commons Namensnennung – Nicht kommerziell – Weitergabe unter gleichen Bedingungen 3.0 Deutschland Lizenz (CC-BY-NC-SA)

<https://creativecommons.org/licenses/by-nc-sa/3.0/de/>



Projektleitung

Dr. Heike Spangenberg
Telefon +49 (0)511 450670-151
E-Mail: spangenberg@dzhw.eu

Projektmitarbeiter*innen

Andreas Woisch
Barbara Franke
Heiko Quast
Dennis Föste-Eggers
Hanna Mentges
Ann-Christin Renneberg
Laura Schoger
Anna Meine

Impressum

Herausgeber

Deutsches Zentrum für Hochschul- und
Wissenschaftsforschung GmbH (DZHW)
Lange Laube 12 | 30159 Hannover | www.dzhw.eu
Postfach 2920 | 30029 Hannover
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960

Geschäftsführung

Prof. Dr. Monika Jungbauer-Gans
Dr. habil. Thorsten Kowalke

Vorsitzender des Aufsichtsrats

Ministerialdirigent Peter Greisler

Registergericht

Amtsgericht Hannover | B 210251
Umsatzsteuer-Identifikationsnummer:
DE291239300

Dezember 2022

Inhaltsverzeichnis

| | |
|---|-----------|
| Tabellen-/Abbildungsverzeichnis | II |
| 1 Inhalt und Anlage der Studie | 3 |
| 2 Erhebungsinstrumente | 6 |
| 2.1 Inhalte der Erhebungsinstrumente | 6 |
| 2.2 Sekundär genutzte Bestandteile der Erhebungsinstrumente | 7 |
| 2.3 Pretest | 10 |
| 3 Grundgesamtheit und Stichprobenverfahren | 12 |
| 4 Durchführung der Erhebungen | 14 |
| 5 Rücklauf | 17 |
| 6 Datenaufbereitung | 19 |
| 6.1 Datenübertragung | 19 |
| 6.2 Codierung offener Angaben | 19 |
| 6.3 Datenprüfung und Datenbereinigung | 20 |
| 6.4 Codierung fehlender Werte | 21 |
| 6.5 Datensatzerstellung | 22 |
| 6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels | 22 |
| 7 Gewichtung | 24 |
| 7.1 Vorgehen und Anwendungshinweise | 24 |
| 7.2 Gewichtung des Datensatzes | 25 |
| 8 Anonymisierung | 27 |
| 9 Literatur | 31 |

Tabellen-/Abbildungsverzeichnis

| | |
|--|----|
| Abbildung 1: Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe..... | 4 |
| Tabelle 1: Quellennachweise zu eingesetzten Erhebungsinstrumenten (Schüler*innenbefragung) | 7 |
| Tabelle 2: Quellennachweise zu eingesetzten Erhebungsinstrumenten (Schulleitungsbefragung) | 9 |
| Tabelle 3: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Studienberechtigtenpanels 2018..... | 17 |
| Abbildung 2: Rücklauf des DZHW-Studienberechtigtenpanels 2018 im Zeitverlauf, erste Welle | 18 |
| Tabelle 4: Vercodete Merkmale und verwendete Codierlisten im Studienberechtigtenpanel 2018 (erste Welle) | 20 |
| Tabelle 5: Missingsystematik im DZHW-Studienberechtigtenpanel 2018..... | 21 |
| Tabelle 6: Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2018..... | 24 |
| Tabelle 7: Zugangswege im FDZ-DZHW..... | 27 |
| Tabelle 8: Maßnahmen der statistischen Anonymisierung der Daten | 29 |

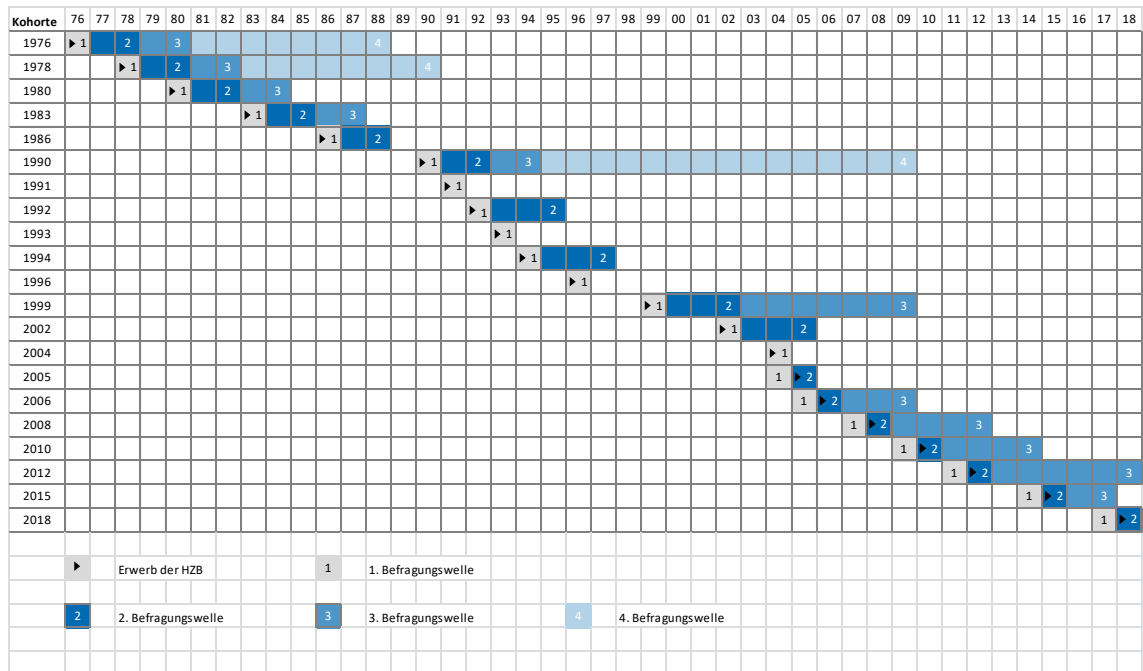
1 Inhalt und Anlage der Studie

[Studienreihe] Das DZHW-Studienberechtigtenpanel 2018 ist Teil der DZHW-Studienberechtigtenbefragungsreihe, in der anhand von standardisierten Mehrfachbefragungen Informationen zum Übergang von der Schule in Studium und Berufsausbildung sowie die nachschulischen Bildungs- und Erwerbsverläufe bei Schulabsolvent*innen mit Hochschulzugangsberechtigung in ihrer Gesamtheit erfasst werden. Das erste Studienberechtigtenpanel wurde 1976 durchgeführt, seitdem wurden 21 Studienberechtigtenjahrgänge (Kohorten) untersucht. Die Befragungen werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) gefördert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring sowie der Beantwortung von Fragestellungen der Hochschul- und Wissenschaftsforschung.

Die Grundgesamtheit der Kohorten umfasst jeweils die (zukünftig) studienberechtigten Schüler*innen, die im jeweiligen Schuljahr an allgemeinbildenden und beruflichen Schulen in Deutschland die allgemeine bzw. fachgebundene Hochschulreife, die volle Fachhochschulreife bzw. die fach- oder landesgebundene Fachhochschulreife anstreben.

In der Regel werden für jeden Studienberechtigtenjahrgang mehrere Befragungswellen zu unterschiedlichen Zeitpunkten vor und nach dem Erwerb der Hochschulzugangsberechtigung durchgeführt. Es handelt sich somit um ein kombiniertes Kohorten-Panel-Design. Die Befragungen der Studienberechtigtenkohorten umfassen ein bis vier Wellen. Bis 1986 verfolgte das DZHW das Ziel, jeden Studienberechtigtenjahrgang kurz nach sowie 2 ½, 4 ½ und 12 ½ Jahre nach Schulabgang zu befragen. Ab 1990 wurde die Befragung zu verspäteten oder korrigierten Ausbildungs- und Studienentscheidungen zweieinhalb Jahre nach Schulabgang gestrichen. Seit 2005 wird eine zusätzliche Befragung ein halbes Jahr vor dem Erwerb der Hochschulzugangsberechtigung durchgeführt. Nach dieser ersten Befragungswelle folgt die zweite Welle einer Kohorte, die ein halbes Jahr nach dem Erwerb der Hochschulzugangsberechtigung (HZB) erfolgt. Die dritte Befragungswelle folgt etwa zweieinhalb Jahre nach dem Studienabschluss. Über eine Spanne von über 40 Jahren entstand so eine Zeitreihe von 21 Jahrgängen mit insgesamt 52 Befragungswellen (vgl. Abbildung 1).

Abbildung 1: Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe



[Analysepotential] Um einen Zeitvergleich zu ermöglichen und unter den verschiedenen Kohorten vergleichen zu können, wird zum Teil in gleichen Befragungswellen ein identischer Kern an Informationen erhoben. Dadurch können auch mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends der nachschulischen Werdegänge in den Blick genommen werden. Zudem wird innerhalb einer Kohorte ein Teil der Fragen in den verschiedenen Befragungswellen wiederholt gestellt. Dies ermöglicht die Betrachtung intra-individueller Veränderungen zwischen den Wellen (z. B. kausale Panelanalysen). Darüber hinaus werden in Abhängigkeit von aktuellen Entwicklungen und Forschungsinteressen in einzelnen Kohorten bestimmte Aspekte vertiefend oder ergänzend abgefragt.

[Einordnung ins Forschungsfeld] Das Stichproben- und Erhebungsdesign sowie die damit verbundenen Analysemöglichkeiten unterscheiden das DZHW-Studienberechtigtenpanel von anderen in Deutschland durchgeführten Befragungen von Studienberechtigten. Keine andere Befragung ermöglicht bundesweite Analysen. Zusätzlich weisen andere Erhebungen in diesem Feld keine oder deutlich kürzere Zeitreihen auf. Andere Studienberechtigtenbefragungen sind beispielsweise die sächsische Abiturientenbefragung¹ (durchgeführt durch das Kompetenzzentrum für Bildungs- und Hochschulplanung an der TU Dresden), die TOSCA-Studie² (inzwischen durchgeführt vom Hector-Institut für Empirische Bildungsforschung an der Universität Tübingen) sowie das Berliner-Studienberechtigtenpanel Best Up³ (durchgeführt vom Deutschen Institut für Wirtschaftsforschung (DIW) und dem Wissenschaftszentrum Berlin (WZB)).

[Spezifika des Studienberechtigtenpanels 2018] Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Studienberechtigtenkohorte 2018 folgende Spezifika auf. Thematisch wird ein Schwerpunkt auf Diversität mit einem besonderen Fokus auf Migrationshintergrund

¹ vgl. dazu: <https://tu-dresden.de/zqa/forschung/Forschungsprojekte/saechsische-abiturientenstudie> (abgerufen am 26.10.2022)

² vgl. dazu: <https://uni-tuebingen.de/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/fachbereich-sozialwissenschaften/hector-institut-fuer-empirische-bildungsforschung/forschung/aktuelle-studien/tosca/> (abgerufen am 26.10.2022)

³ vgl. dazu: <https://www.wzb.eu/de/forschung/dynamiken-sozialer-ungleichheiten/ausbildung-und-arbeitsmarkt/best-up> (abgerufen am 26.10.2022)

gelegt. Zudem werden erstmals innerhalb dieser Befragungsreihe die Schulleitungen zu verschiedenen Themen befragt.

Darauf hinzuweisen ist, dass zunächst nur die Daten der ersten Welle des DZHW-Studienberechtigtenpanels 2018 veröffentlicht werden (doi: [10.21249/DZHW:gsl2018:1.0.0](https://doi.org/10.21249/DZHW:gsl2018:1.0.0)) und sich der vorliegende Daten- und Methodenbericht entsprechend auf diese Welle fokussiert.

2 Erhebungsinstrumente

Die erste Befragungswelle im DZHW-Studienberechtigtenpanel 2018 wurde als schriftlich-postalische Paper-and-Pencil-Befragung (Paper and Pencil Interview; PAPI) durchgeführt, in der sowohl für die Schüler*innen als auch für die Schulleitungen als Erhebungsinstrument jeweils ein standardisierter Papierfragebogen zum Einsatz kam.

Kapitel 2.1 stellt die zentralen Inhalte der Erhebungsinstrumente vor. Kapitel 2.2 führt die sekundär genutzten Bestandteile der Erhebungsinstrumente auf. Kapitel 2.3 beschreibt die zur Prüfung und Verbesserung der Fragebögen durchgeführten Pretests.

2.1 Inhalte der Erhebungsinstrumente

Die Fragen der ersten Welle im Schüler*innenfragebogen beziehen sich auf fünf große Themenkomplexe: Fragen zur Schulzeit, zu der Zeit nach der Schule, zur Entscheidungsfindung, zu Persönlichkeit und Werten sowie Angaben zur Person. Der erste Themenkomplex beinhaltet Fragen zum besuchten Schulzweig, zur Art des angestrebten Abschlusses, zum Notendurchschnitt sowie zur Zufriedenheit mit der eigenen Leistung sowie mit dem Schulunterricht im Allgemeinen. Außerdem werden Fragen zu den persönlichen Stärken und Schwächen, zu Prüfungsangst und beruflichen Interessen gestellt. Die Berufs- und Bildungsabsichten sind Bestandteil des zweiten Themenkomplexes, welcher Fragen zum angestrebten Ausbildungsberuf, Studium, Hochschule und den Wünschen der Eltern beinhaltet. Die Entscheidungsfindung wird im dritten Teil durch Fragen zum aktuellen Informationsstand über Ausbildungs- und Studienmöglichkeiten, Probleme bei der Wahl des nachschulischen Werdegangs, Unterstützung bei der Entscheidungsfindung und die Berufsaussichten bei Ausbildung und Studium abgefragt. Um die Berufs- und Lebensziele, Risikobereitschaft, Religiosität und politische Partizipation geht es im vierten Themenkomplex, während im letzten Teil Fragen zum Geschlecht, Alter und Geburtsland der Befragten sowie Bücherbesitz und Geburtsland der Eltern sowie deren Ausbildungsabschlüsse gestellt werden.⁴ Hervorzuheben ist, dass in die erste Befragungswelle gezielt Befragungsinstrumente implementiert wurden, um Prozesse und Mechanismen abbilden bzw. erklären zu können, die später zu den mittels der späteren Wellen erfassten tatsächlichen Entscheidungen, Bildungsverläufen und Bildungserträgen (z. B. gesellschaftliches und politisches Engagement) führen. Die Operationalisierung von Konstrukten verschiedener theoretischer Ansätze zur Erklärung von verschiedenen Bildungs- und Berufsentscheidungen stellten daher einen Schwerpunkt in der ersten Befragungswelle dar. Da das DZHW-Studienberechtigtenpanel 2018 außerdem das Thema Diversität mit einem besonderen Fokus auf Migrationshintergrund als einen Hauptschwerpunkt betrachtet, wurde sowohl nach Indikatoren zur Bestimmung des Migrationshintergrundes als auch zu Einstellungen gegenüber der Situation der Migrant*innen in Deutschland gefragt.

Erstmals innerhalb der Befragungsreihe wurden im Rahmen der ersten Befragungswelle ergänzend auch die Schulleitungen anhand eines gesonderten Fragebogens zu verschiedenen Themen befragt, wie beispielsweise der Zusammensetzung der Schüler*innenschaft, dem Umgang der Schulen mit

⁴ Aufgrund von Datenschutzbestimmungen konnten für die Länder Hessen und Berlin in der ersten Welle keine Angaben zu den Eltern der Befragten erhoben werden (vgl. Kapitel 4). Die entsprechenden Fragen wurden jedoch in der zweiten Welle für Befragte aus Hessen und Berlin gestellt.

Diversität, räumlichen, personellen und finanziellen Ressourcen sowie zu Maßnahmen der Berufs- und Studienorientierung.

2.2 Sekundär genutzte Bestandteile der Erhebungsinstrumente

Teile der Erhebungsinstrumente wurden sekundär genutzt, die entsprechenden Quellennachweise sind in Tabelle 1 für die Schüler*innenbefragung und in Tabelle 2 für die Schulleitungsbefragung zusammengestellt.

Tabelle 1: Quellennachweise zu eingesetzten Erhebungsinstrumenten (Schüler*innenbefragung)

| Fragebogenblock | Fragennr. | Quelle und Zitation |
|--------------------------|--|---|
| 1 – Fragen zur Schulzeit | Frage 1.11 | Adaptiert nach Centers for Disease Control and Prevention (CDC) (1993). Behavioral Risk Factor Surveillance System Survey Questionnaire. Atlanta, Georgia: U.S. Department of Health and Human Services, Centers for Disease Control and Prevention, S. 3. https://www.cdc.gov/brfss/questionnaires/pdf-ques/93brfss.pdf (abgerufen am 10.11.2022). |
| | | FDZ-LifBi (2022). Codebook. NEPS Startkohorte 5 – Studierende. Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 16.0.0, S. 1554. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/16-0-0/SC5_16-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 1.12 | Adaptiert nach OECD (2017). PISA 2015 Technical Report, OECD Publishing, S. 315-317. https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf (abgerufen am 10.11.2022). |
| | | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 51-52. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| Frage 1.15 | Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2014). Allgemeine Selbstwirksamkeit Kurzsкала (ASKU). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis35. | |
| Frage 1.16 | Items 1, 2, 7: Adaptiert nach FDZ-LifBi (2022). Codebook. NEPS Startkohorte 5 – Studierende. Hochschulstudium und Übergang in den Beruf. Scientific Use File Version 16.0.0, S. 730, 729, 728. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/16-0-0/SC5_16-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). | |
| | Items 3, 4, 5, 6, 8, 9: Adaptiert nach Stocké, V. (2005). Einstellung zu Bildung. Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis136. | |
| 2 – Zeit nach der Schule | Frage 2.1 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 – Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 6971-6973. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| 3 - Entscheidungsfindung | Frage 3.8 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 – Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 6713-6720. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 3.10 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 – Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 3897, 3861. https://www.neps- |

| | | |
|------------------------------|------------|--|
| | | data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 3.11 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 — Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 6723-6724. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 3.12 | Item 3: Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 — Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 160. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 3.13 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 — Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 2941-2945. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| 4 – Persönlichkeit und Werte | Frage 4.1 | Adaptiert nach Beierlein, C., Kovaleva, A., Kemper, C. J., & Rammstedt, B. (2015). Kurzsкала zur Erfassung der Risikobereitschaft (R-1). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis236. |
| | Frage 4.2 | Adaptiert nach European Social Survey (2016). ESS Round 8 Source Questionnaire. London: ESS ERIC Headquarters c/o City University London, S. 21. https://stessrelpubprodwe.blob.core.windows.net/data/round8/fieldwork/source/ESS8_source_questionnaires.pdf (abgerufen am 10.11.2022). |
| | Frage 4.3 | Adaptiert nach Abele, A. E., Stief, M. & Krüsken, J. (2002). Persönliche Ziele von Mathematikern beim Berufseinstieg. Ein Vergleich offener und geschlossener Erhebungsmethoden. Zeitschrift für pädagogische Psychologie, 16(3/4), S. 193-205. https://doi.org/10.1024//1010-0652.16.34.193 (S. 196-197) (abgerufen am 16.12.2022). |
| | Frage 4.4 | GESIS – Leibniz-Institut für Sozialwissenschaften (o. J.). German Longitudinal Election Study. GLES 2013. Langfrist-Panel 2009-2013-2017. ZA5322, Version 1.1.0, Fragebogendokumentation, S. 22. https://access.gesis.org/dbk/60396 (abgerufen am 10.11.2022). |
| | Frage 4.5 | Adaptiert nach European Social Survey (2016). ESS Round 8 Source Questionnaire. London: ESS ERIC Headquarters c/o City University London, S. 10. https://stessrelpubprodwe.blob.core.windows.net/data/round8/fieldwork/source/ESS8_source_questionnaires.pdf (abgerufen am 10.11.2022). |
| | Frage 4.6 | Adaptiert nach Schupp, J. & Gerlitz, J.-Y. (2008). Big Five Inventory-SOEP (BFI-S). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis54. |
| | Frage 4.7 | Adaptiert nach Breen, R., van de Werfhorst, H. G. & Jæger, M. M. (2014). Deciding under Doubt: A Theory of Risk Aversion, Time Discounting Preferences, and Educational Decision-making. European Sociological Review, 30(2), 258–270. http://www.jstor.org/stable/24479880 (S. 263-264) (abgerufen am 10.11.2022). |
| 5 - Personenangaben | Frage 5.1 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 — Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 4161. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |
| | Frage 5.9 | Adaptiert nach Beierlein, C., Kovaleva, A., László, Z., Kemper, C. J. & Rammstedt, B. (2015). Kurzsкала zur Erfassung der Allgemeinen Lebenszufriedenheit (L-1). Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS). doi:10.6102/zis229. |

Tabelle 2: Quellennachweise zu eingesetzten Erhebungsinstrumenten (Schulleitungsbefragung)

| Fragebogen-block | Frage-nr. | Quelle und Zitation |
|--|-----------|--|
| 1 – Hintergrundinformationen zur Schule und Person | Frage 1.4 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 266-268. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| | Frage 1.6 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 235. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| | Frage 1.8 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 284. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| 2 – Ausstattung der Schule | Frage 2.1 | Adaptiert nach Ditton, H. & Merz, D. (2013). Räumlichkeiten (Verfügbarkeit) - Schulleitung [Fragebogenskala: Version 1.0]. In: QualitätsSicherung in Schule und Unterricht - Fragebogenerhebung Erhebungszeitpunkt 1 (QuaSSU) [Skalenkollektion: Version 1.0]. Datenerhebung 2000. Frankfurt am Main: Forschungszentrum Bildung am DIPF. doi:10.7477/18:30:1. (vgl. auch https://www.fdz-bildung.de/skala.php?skala_id=4509&erhebung_id=30 (abgerufen am 10.11.2022)). |
| | Frage 2.2 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 237. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| | Frage 2.3 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 263-265. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| | Frage 2.4 | Adaptiert nach Mang, J., Ustjanzew, N., Leßke, I., Schiepe-Tiska, A. & Reiss, K. (2019). PISA 2015 Skalenhandbuch. Dokumentation der Erhebungsinstrumente. Münster: Waxmann. S. 262. https://www.pisa.tum.de/fileadmin/w00bgi/www/Berichtsbaende_und_Zusammenfassungen/Skalenhandbuch_PISA2015_openaccess.pdf (abgerufen am 10.11.2022). |
| | Frage 2.5 | Adaptiert nach Landkreis Rostock (Hrsg.) (2012). Anschlussperspektiven im Landkreis Rostock. Eine aktuelle Situationsbeschreibung am Übergang Schule-Beruf für den Zeitraum 2011/2012. PIW GmbH (2012). Bericht zum Auftrag des Projektes A-HA Anschluss Halten. „Durchführung einer Schulbefragung im Landkreis Rostock“. S. 24. http://www.piw.de/Piw/Dokumente/Bericht%20Schulbefragung%2011-3-2012.pdf (abgerufen 16.12.2022) |
| 3 – Unterstützung bei Berufs- und Ausbildungswahl | Frage 3.1 | Adaptiert nach Landkreis Rostock (Hrsg.) (2012). Anschlussperspektiven im Landkreis Rostock. Eine aktuelle Situationsbeschreibung am Übergang Schule-Beruf für den Zeitraum 2011/2012. PIW GmbH (2012). Bericht zum Auftrag des Projektes A-HA Anschluss Halten. „Durchführung einer Schulbefragung im Landkreis Rostock“. S. 24. http://www.piw.de/Piw/Dokumente/Bericht%20Schulbefragung%2011-3-2012.pdf (abgerufen 16.12.2022) |
| | Frage 3.4 | Adaptiert nach Landkreis Rostock (Hrsg.) (2012). Anschlussperspektiven im Landkreis Rostock. Eine aktuelle Situationsbeschreibung am Übergang Schule-Beruf für den Zeitraum 2011/2012. PIW GmbH (2012). Bericht zum Auftrag des Projektes A-HA Anschluss Halten. „Durchführung einer Schulbefragung im Landkreis |

| | | |
|-------------------------------------|-----------|---|
| | | Rostock". S. 24. http://www.piw.de/Piw/Dokumente/Bericht%20Schulbefragung%2021-3-2012.pdf (abgerufen 16.12.2022) |
| 4 – Umgang mit kultureller Vielfalt | Frage 4.1 | Adaptiert nach Fragebogen des PISA-Feldtests 2015 Deutschland, S. 96-97 (unveröffentlicht). ⁵ |
| | Frage 4.2 | Adaptiert nach FDZ-LifBi (2021). Codebook NEPS Startkohorte 4 – Klasse 9. Schule und Ausbildung. Scientific Use File Version 12.0.0, S. 1395-1398. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC4/12-0-0/SC4_12-0-0_Codebook_de.pdf (abgerufen am 10.11.2022). |

2.3 Pretest

[Ziel und Verfahren] Die Erhebungsinstrumente wurden im Vorfeld der Erhebung durch Pretests geprüft. Dabei sollte erstens für die bereits in vorherigen Kohorten eingesetzten Fragen und Antwortvorgaben geprüft werden, ob sie von dem Studienberechtigtenjahrgang 2018 gleich perzipiert werden wie von früheren Jahrgängen.⁶ Zweitens sollte für die neu eingesetzten Messinstrumente deren Verständlichkeit, Beantwortbarkeit, theoretische Aussagekraft, Reliabilität und Validität getestet werden. Drittens zog die Aufnahme der neuen Fragen grundsätzliche Veränderungen im Aufbau und Layout des Fragebogens sowie an der Befragungsdauer nach sich, die evaluiert werden sollten.

Eine erste Version der Erhebungsinstrumente wurde zunächst im Rahmen von Expertenbewertungen (vgl. Häder, 2015, S. 406–407) begutachtet. Darüber hinaus wurden ausgewählte (neu entwickelte oder angepasste) Instrumente im Rahmen von drei kognitiven Pretests geprüft. Neben diesen Bewertungen fanden zielgerichtete Entwicklungs- und Abschluss- Pretests⁷ statt. Diese wurden als sogenannter Pretest im Feld (vgl. Häder, 2015, S. 396), also unter möglichst ähnlichen Bedingungen wie in den tatsächlichen Befragungen, durchgeführt.

[Probanden] Die kognitiven Pretests wurden als interne Tests mit sowohl wissenschaftlichen Mitarbeiter*innen als auch studentischen Hilfskräften des DZHW durchgeführt. Durch die Befragung dieser Testpersonen wurden so gleichzeitig Expert*innen aus dem Bereich der Hochschulforschung einbezogen, welche die beiden Fragebögen entsprechend bewerten konnten.

Bei den Pretests im Feld handelt es sich um externe Pretests. In der ersten Welle wurden drei Schulen am Pretest des Schüler*innenfragebogens beteiligt. Die drei Schulen wiesen alle ein unterschiedliches Profil auf (ein allgemeinbildendes Gymnasium sowie zwei berufliche Schulen mit verschiedenen Schulzweigen), sodass 152 Personen aus unterschiedlichen Schulzweigen im Klassenverband befragt werden konnten. So wurde sichergestellt, dass die zum Teil schulzweigspezifischen Instrumente auch an den jeweiligen Schulzweigen getestet werden konnten. Der Schulleitungsfragebogen wurde von mehreren externen Lehrkräften getestet. Die externen Tester wurden durch persönliche Kontakte rekrutiert.

[Durchführung] Die externe Pretest-Phase des Schüler*innenfragebogens der ersten Welle fand zwischen dem 08.06.2017 und 14.06.2017, also etwa sechs Monate vor dem Feldstart statt. Dabei wurden an einer Schule Fragebögen mit einer Durchführungsanweisung postalisch versendet. Die

⁵ Ein ähnliches Instrument wurde außerdem verwendet in: Brown, M., Altrichter, H., Nayir, F., Nortvedt, G., Burns, D., Fellner, M., Gloppen, S. K., Helm, C., McNamara, G., O'Hara, J., Punzenberger, B., Skedsmo, G. & Wiese, E. F. (2019). Classroom Assessment that Recognises Cultural Difference - A European Perspective. Final Report (Leistungsbeurteilung unter Berücksichtigung kultureller Differenz - eine europäische Perspektive. Abschlussbericht), Dublin (M. Felner, Trans.). DOI: 10.13140/RG.2.2.14346.52164/2, S. 10-11 (Question number seven in Online Survey Questionnaire).

⁶ Die Erhebungsinstrumente des Studienberechtigtenpanels 2018 basieren im Wesentlichen auf dem für die Untersuchungsreihe etablierten Design und den in den zuvor befragten Kohorten der Studienberechtigten bewährten schriftlichen Fragebögen.

⁷ Abschluss-Pretests dienen im Gegensatz zu Entwicklungs-Pretests in der Regel kleineren Restkorrekturen der Frageinstrumente wie Kürzungen oder Umstellungen sowie der Überprüfung neuer Filterführungen oder Veränderungen des Druckbilds, vgl. Schnell, Hill und Esser (2005), S. 348.

beiden anderen Schulen wurden von Mitarbeiter*innen des Studienberechtigtenpanels aufgesucht, die die Testpersonen baten, den standardisierten Fragebogen zu bearbeiten und Verständnisprobleme, Kritik oder Anmerkungen zu notieren. Im Anschluss an die Bearbeitung des Fragebogens sollten die Befragten einen Feedbackfragebogen zu verschiedenen Aspekten der Befragung ausfüllen. Dabei wurden Informationen zur Ausfülldauer, zu Inhalt und Länge des Fragebogens, zu Aufbau und Layout, zur Verständlichkeit der Fragen und Ausfüllanweisungen sowie zur Vollständigkeit der Antwortmöglichkeiten erhoben. Auf Grundlage der Pretestergebnisse wurden die Formulierungen verschiedener Fragetexte präzisiert sowie die Reihenfolge von Items und Antwortkategorien in einzelnen Itembatterien und Mehrfachnennungen überarbeitet. Zudem wurde die Länge des Fragebogens verringert, indem Itembatterien gekürzt wurden, da die Ausfülldauer die ursprünglich geplanten 30 Minuten überstieg. Der grundsätzliche Aufbau hingegen blieb unverändert.

Der Pretest für den Schulleitungsfragebogen verlief dahingehend, dass drei interne Expert*innen den Fragebogenentwurf erhalten und dazu ein schriftliches Feedback gegeben haben.⁸ Das gleiche Verfahren wurde beim externen Pretest durch Lehrkräfte durchgeführt. Daraufhin wurden auch beim Schulleitungsfragebogen Formulierungen verschiedener Fragetexte präzisiert, Items und Antwortkategorien in einzelnen Itembatterien und Mehrfachnennungen überarbeitet.

⁸ Das Feedback wurde zwischen dem 17.07. und 18.07.2017 gegeben.

3 Grundgesamtheit und Stichprobenverfahren

[Grundgesamtheit und Inferenzpopulation] Die Grundgesamtheit des DZHW-Studienberechtigtenpanels 2018 umfasst alle (zukünftig) studienberechtigten Schüler*innen, die im Schuljahr 2017/2018 an allgemeinbildenden und beruflichen Schulen in Deutschland die allgemeine bzw. die fachgebundene Hochschulreife, die volle Fachhochschulreife bzw. die fach- oder landesgebundene Fachhochschulreife anstreben. Schüler*innen, die zum Befragungszeitraum bereits eine Hochschulreife vorwiesen, wurden ebenso aus der Befragung ausgeschlossen wie Schüler*innen, die im Schuljahr 2017/18 lediglich den schulischen Teil der Fachhochschulreife erwerben.

Die Befragung der Studienberechtigten erfolgte deutschlandweit und berücksichtigte dabei im Einzelnen folgende Schularten:

- Gymnasien
- Gesamtschulen/ Gemeinschaftsschulen/ Stadtteilschulen/ IGS/ Waldorfschulen mit gymnasialer Oberstufe/ Oberschule
- Abendgymnasien
- Berufliche Gymnasien/ Fachgymnasien /gymnasiale Oberstufe (u. a. im OSZ, BSZ)/ FOS 13/ BOS 13 (AHR)
- Kollegs
- Berufsoberschulen/ Fachoberschulen (FHR)
- (Höhere) Berufsfachschulen/ Höhere Handelsschulen/ Berufsschule(plus)/ Realschule plus (nur in Rheinland-Pfalz)/ Berufskollegs (nur in Baden-Württemberg)
- Fachschulen
- Fachakademien (nur in Bayern)

[Quotierte geschichtete Klumpenstichprobe] Aufgrund fehlender oder nicht zugänglicher Listen von Studienberechtigten auf Individualebene war es nicht möglich, eine einfache Zufallsstichprobe auf Ebene der Schüler*innen zu ziehen. Daher mussten die Individuen über die Schulen rekrutiert werden. Als Stichprobendesign wurde ein disproportional geschichtetes Klumpenstichprobenverfahren als Spezialfall einer mehrstufigen Zufallsauswahl (vgl. Häder, 2015, S. 169) gewählt. Die Schulen bzw. Schulzweige⁹ stellen dabei die „primary sampling units“ (Klumpen) und die Studienberechtigten des Jahrgangs 2018 die „secondary sampling units“ innerhalb dieser Klumpen dar.¹⁰ Die Auswahlgesamtheit bildete eine Liste mit allen relevanten Schulen bzw. Schulzweigen. Die Schichtung erfolgte anhand der Merkmale Bundesland und Schultyp. Innerhalb der Schichten wurden Ziehungsquoten für die Klumpen und Individuen festgelegt, die sich zum einen an der Verteilung des Studienberechtigtenjahrgangs 2014/2015 laut amtlicher Statistik orientierten und zum anderen auch die Rücklaufquoten der letzten Erhebungen berücksichtigten. Im Zuge dessen erhielten Studienberechtigte aus

⁹ Manche berufsbildenden Schulen bestehen aus verschiedenen Schulzweigen, von denen jedoch nur jene für die Stichprobe relevant waren, die zum Erwerb der (Fach)-Hochschulzugangsberechtigung führten. Daher wurden an bestimmten Schulen nur ausgewählte Schulzweige berücksichtigt.

¹⁰ Beispiel für den Personenkreis eines Klumpens: Alle Studienberechtigten des Prüfungsjahrgangs 2018 des allgemein-bildenden Gymnasiums XY in Bundesland YZ.

Bremen, Mecklenburg-Vorpommern, Sachsen-Anhalt und dem Saarland sowie Schüler*innen von Abendgymnasien und Kollegs überproportionale Auswahlwahrscheinlichkeiten, um in diesen Gruppen ausreichend große Fallzahlen zu generieren. Darüber hinaus wurden auch Schulzweige in Hessen überproportional häufig gezogen, da in diesem Bundesland eine neue Auflage im Genehmigungsverfahren¹¹ geringere Antwortquoten der Schulen vermuten ließen.

Im Rahmen der praktischen Durchführung des Ziehungsverfahrens erfolgte zunächst in jeder Schicht die Auswahl der Klumpen (Schulen) über eine einfache Zufallsauswahl. Die ausgewählten Schulen meldeten im Folgenden ihre Schulzweige und die Schüler*innenzahlen zurück, sodass ein zweiter Auswahlsschritt durchgeführt werden konnte, der die Erfüllung der Ziehungsquoten auf Individual-ebene zum Ziel hatte. In diesem zweiten Schritt wurden in jeder Schicht sukzessive so viele Schulen/Schulzweige zufällig gezogen, dass mindestens 97 Prozent der schichtspezifischen Ziehungsquote für die Studienberechtigten erreicht wurde.¹² Hiernach wurde an den gezogenen und zur Teilnahme bereiten Schulen bzw. Schulzweigen eine Vollerhebung unter allen Schüler*innen durchgeführt.

Im Unterschied zu den übrigen Bundesländern wurde in Hamburg eine Vollerhebung durchgeführt, sodass das oben geschilderte Stichprobenverfahren für Hamburg nicht zur Anwendung kam.

¹¹ In Hessen wurde erstmals gefordert, dass die Schulen, die an der Befragung teilnehmen wollen, vorab ihre Schul- bzw. Gesamtkonferenz um Erlaubnis bitten mussten. Diese Konferenzen tagen jedoch in der Regel nur einmal pro Halbjahr.

¹² Die Prüfung, ob die Ziehungsquote in einer Schicht erreicht wurde, erfolgte nach jedem Ziehungsschritt. Erfüllte beispielsweise die kumulierte Schüler*innenzahl von fünf Schulen die Ziehungsquote für die Schüler*innenzahl in der spezifischen Schicht, so wurde nach der Auswahl der fünften Schule keine weitere Schule gezogen.

4 Durchführung der Erhebungen

[Kontaktaufnahme und Adresspflege] Die gezogenen Schulen wurden vom DZHW angeschrieben und um ihre Teilnahme gebeten. Zudem teilte das DZHW den Schulen die Kriterien mit, anhand derer sie die Zielpersonen für das DZHW-Studienberechtigtenpanel 2018 identifizieren sollten (vgl. Kapitel 3).¹³ Da die Schulen die Kontaktdaten ihrer Schüler*innen aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Schüler*innenanzahl mit. Anschließend sendete das DZHW die passende Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch an die jeweiligen Ansprechpartner*innen der Schulzweige. Diese Kontaktpersonen wurden gebeten die Unterlagen an die Schüler*innen weiterzugeben und, wenn möglich, ihnen zu erlauben die Fragebogen im Klassenverband und während des Unterrichts ausfüllen zu lassen. Anschließend sollten die Fragebogen bestenfalls per Sammelpaket an das DZHW versendet werden. Es war jedoch auch möglich, den Fragebogen zu Hause auszufüllen und individuell an das DZHW zurückzusenden. In Mecklenburg-Vorpommern war das Ausfüllen während der Unterrichtszeit generell nicht erlaubt.

Um die weiterhin teilnahmebereiten Personen in der zweiten Befragungswelle direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Kontaktdaten (Postanschrift und/oder E-Mail-Adresse) erfasst. Beim Eingang eines Fragebogens im DZHW wurde sowohl auf den Fragebogen als auch auf den Adressabschnitt des Fragebogens per Paginierstempel eine eindeutige Identifikationsnummer gestempelt. Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt dann vom Fragebogen abgetrennt und nach der Erfassung getrennt von den Befragungsdaten auf einem geschützten Server gespeichert und verarbeitet.

[Vorabinformation] In einigen Bundesländern war für die Teilnahme von minderjährigen Schüler*innen an der Befragung die Einwilligung der Eltern notwendig. In diesen Bundesländern (Schleswig-Holstein, Brandenburg, Mecklenburg-Vorpommern, Niedersachsen, Hessen, Bayern, Saarland, Sachsen, Thüringen), bekamen die Schüler*innen zwei Wochen vor Versand der Erhebungsunterlagen eine Vorabinformation bestehend aus einem Anschreiben für die Schüler*innen, einem Elternanschreiben und einer Elterneinwilligungserklärung zugeschickt.¹⁴ Die Vorabinformationen unterschieden sich teilweise aufgrund von Datenschutzbestimmungen der jeweiligen Bundesländer voneinander. In Hamburg wurde ein Anschreiben an die Schüler*innen mit der Bitte um Teilnahme an der DZHW-Studienberechtigtenbefragung von der Behörde für Schule und Berufsbildung an allgemeinbildende Schulen bzw. dem Hamburger Institut für Berufliche Bildung an berufsbildende Schulen als Vorabinformation versendet.

[Erhebungsunterlagen] Die Erhebungsunterlagen bestanden pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen), einem Fragebogen, einer Studienfächerliste, einem

¹³ An dieser Stelle können durch die Schulen vereinzelt auch Personen als stichprobenzugehörig identifiziert worden sein, die nicht zur Grundgesamtheit gehörten (Overcoverage). Dies kam beispielsweise vor, wenn Schulabgänger*innen nur den schulischen Teil der Fachhochschulreife erworben haben ohne anschließend eine volle Fachhochschulreife anzustreben oder die Schulen den Fragebogen an Schüler*innen eines falschen Jahrgangs verteilt haben. Hierunter fallen sowohl Schüler*innen, die bereits in einem früheren Bildungsgang eine schulische Hochschulzugangsberechtigung erworben haben als auch solche, die nicht zum Abschlussjahrgang 2018 gehören, weil sie ihre Hochschulreife erst zu einem späteren Zeitpunkt erreichen werden.

¹⁴ Die Vorabinformationen wurden ab Mitte November 2017 an die jeweiligen Schulen verschickt.

Kugelschreiber mit DZHW-Logo und einem an das DZHW adressierten frankierten Umschlag, sodass auch wenn der Fragebogen zu Hause ausgefüllt wurde, eine kostenfreie Rücksendung für die Befragten möglich war.

Die Anschreiben an die Schüler*innen unterschieden sich teilweise je nach Datenschutzbestimmung der Bundesländer voneinander: Neben unterschiedlichen Anschreiben mussten in vier Bundesländern (Hamburg, Hessen, Saarland und Berlin) eigene Fragebogenvarianten verwendet werden.

- Im Fragebogen in Hamburg wurde Frage 4.7 aus der Standardvariante des Fragebogens (bevorzugtes Anfangsgehalt) aus Platzgründen als Frage 3.14 in Fragenblock 3 verschoben, damit die Befragten am Ende des Fragebogens um eine Einwilligung gebeten werden konnten, dass die Befragungsdaten für vertiefende Forschungsfragen durch die Vertrauensstelle der Behörde für Schule und Berufsbildung mit schulischen Daten verknüpft werden dürfen.¹⁵
- Im Saarland wurde auf dem Deckblatt des Fragebogens ein zusätzlicher Hinweis zur Einwilligung in die Teilnahme und Nutzung der Daten aufgenommen. Zudem musste bei den Hinweisen zur Verwendung der Adressdaten ein Satz geändert werden (anstatt „Datenauswertungen erfolgen immer anonym“ steht hier „Die Daten werden in einer Form veröffentlicht, die keine Rückschlüsse auf einzelne Personen ermöglicht“).
- Der Fragebogen in Hessen enthält keine Fragen über die Eltern der Befragten, da dies aus Datenschutzgründen nicht erlaubt war. Konkret fehlen die Fragen zur Anzahl der Bücher der Eltern, zu den Geburtsländern der Eltern und zu den beruflichen Ausbildungsabschlüssen der Eltern (im Standardfragebogen wurden diese Informationen in den Fragen 5.1, 5.6, 5.7 und 5.8 erfasst). Außerdem gibt es auf der letzten Seite in Block 6 ein zusätzliches Feld für die Einwilligung zur Verwendung der angegebenen Daten im Fragebogen. Darüber hinaus wurde bei den Hinweisen zur Verwendung der Adressdaten die gleiche Satzänderung wie im Saarland vorgenommen.
- In Berlin durften wie in Hessen ebenfalls keine Fragen über die Eltern gestellt werden. Zudem wurde vor der ersten Frage ein Feld für die Einwilligung zur Verwendung der angegebenen Daten im Fragebogen aufgenommen. Darüber hinaus musste auch in Berlin bei den Hinweisen zur Verwendung der Adressdaten die gleiche Satzänderung wie im Saarland und in Hessen vorgenommen werden. Ergänzend wurde bei der Angabe der Adresse ein Feld für die Einwilligung zur Verwendung der Kontaktdaten eingefügt.

Zudem wurde erstmals im Studienberechtigtenpanel ein Fragebogen für die Schulleitungen der ausgewählten Schulen mitgeschickt. In Berlin musste ein eigener Fragebogen für die Schulleitungen verwendet werden, da vor der ersten Frage ein Feld für die Einwilligung zur Verwendung der angegebenen Daten sowie am Ende des Fragebogens ein Zusatzhinweis, dass keine Angaben über identifizierbare Personen erfolgen durften, eingefügt werden mussten. Der im Saarland eingesetzte Fragebogen enthielt auf dem Deckblatt einen zusätzlichen Hinweis zur Einwilligung in die Teilnahme und Nutzung der Daten.

[Feldphase] Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich von Dezember 2017¹⁶ bis Oktober 2018.¹⁷ Aufgrund des angewendeten Kontaktverfahrens über die Schulen konnte das DZHW keinen direkten Einfluss auf den genauen Befragungszeitpunkt der Schüler*innen nehmen. Das Erinnerungsschreiben wurde an die Ansprechpartner*innen in den Schulen etwa vier bis acht Wochen nach Feldstart verschickt. Mitte März 2018 wurden alle Schulen, die eine Rücklaufquote von unter 25 Prozent aufwiesen, ein zweites Mal per E-Mail erinnert.

¹⁵ Zusätzlich lag den Befragungsunterlagen in Hamburg ein Schreiben mit Hinweisen zur beabsichtigten Datenverknüpfung bei. Eine Verknüpfung von Befragungs- und schulischen Kontextdaten war im Projekt Studienberechtigtenpanel 2018 zwar nicht vorgesehen, soll aber im Rahmen eines Zusatzprojektes realisiert werden.

¹⁶ Da die Erhebungsunterlagen in der ersten Welle nicht direkt an die Zielpersonen, sondern an die Ansprechpersonen der Schulen verschickt und durch diese ausgeteilt wurden, lässt sich das genaue Datum des Feldstarts nicht rekonstruieren. Von der Versandfirma wurden die ersten Pakete für die Schulen am 5.12.2017 verschickt (Niedersachsen, Sachsen, Thüringen, Mecklenburg-Vorpommern, Brandenburg, Rheinland-Pfalz). Im Selbstversand wurden (aufgrund von Verzögerungen durch die Versandfirma) bereits im November erste Pakete versandt. Der erste Fragebogenrücklauf kam am 1.12. (Schülerfragebogen) und am 8.12. (Schulleiterfragebogen) am DZHW an.

¹⁷ Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt. Dies war nicht zuletzt notwendig, da aufgrund gestiegener Datenschutzabstimmungen und -auflagen die Befragung in einzelnen Ländern erst verzögert starten konnte. Der letzte Fragebogen ging im Oktober 2018 ein.

[Rücklaufsteigernde Maßnahmen] Die rücklaufsteigernden Maßnahmen zielten zum einen auf Anreize für die Schulen, die Befragung organisatorisch zu unterstützen und zum anderen auf die individuelle Motivation der Befragten.

Für den Erstkontakt bestand die organisatorische Unterstützung der Schulen (und der verantwortlichen Ansprechpartner*innen) darin, Unterrichtszeit zum Ausfüllen bereitzustellen (außer in Mecklenburg-Vorpommern), die Fragebögen am Ende der Unterrichtsstunden einzusammeln und die Rücksendung zu organisieren. Als Gegenleistung wurden den Schulen bei ausreichender Fallzahl schulspezifische Auswertungen angeboten. Auch wurden für eine hohe Teilnahmebereitschaft Preise für die Schulen (zehn Mal 300 Euro) ausgelobt. Die Auslobung von Incentives an die Schulen war in Niedersachsen, Hessen, Bayern, Saarland, Sachsen und Thüringen nicht erlaubt. Als nützlich erwiesen sich zudem ein Hinweis auf die gesamtgesellschaftliche Bedeutung der Erhebung in den Anschreiben, ministerielle Aufforderungen die Befragung zu unterstützen sowie Anschreiben und die Kontaktierung der Schulen, von denen noch keine oder nur wenige Fragebögen eingegangen waren. Als kleines Dankeschön für die Ansprechpartner*innen in den Schulen wurde außerdem ein Notizheft mit Kugelschreiber mitgesandt.

Die auf die individuelle Motivation der Befragten zielende rücklaufsteigernde Maßnahme bestand in einem Hinweis auf den gesamtgesellschaftlichen und altersgruppenspezifischen Nutzen im Anschreiben an die Befragten. Als materieller individueller Anreiz wurden unter allen teilnehmenden Schüler*innen ein Notebook im Wert von 1.300 Euro, ein E-Book-Reader im Wert von 200 Euro und 20 Amazon-Gutscheine im Wert von jeweils 25 Euro verlost. In Niedersachsen, Hessen, Bayern, Saarland, Berlin, Sachsen und Thüringen war die Verlosung von Incentives an die Schüler*innen nicht erlaubt.

5 Rücklauf

Auf der Grundlage des Stichproben- und Erhebungsdesigns (siehe Kapitel 3 und 4) wurden für die Befragung der ersten Welle 2.901 Schulen angeschrieben. Von diesen Schulen erklärten sich 1.185 (41 Prozent) bereit zur Teilnahme und meldeten die Anzahl ihrer zu erwartenden Studienberechtigten zurück. Dies entsprach einer Zahl von 83.988 Schüler*innen. Die daraus gezogene bereinigte Bruttostichprobe der ersten Befragungswelle umfasste 82.413 Schüler*innen, die kontaktiert wurden.¹⁸

Die Rücklaufquote betrug 48,2 Prozent, sodass eine Nettostichprobe von 39.714 auswertbaren Fragebögen zu verzeichnen war.¹⁹ 63 Prozent der Teilnehmer*innen der ersten Welle, erklärten sich durch Angabe einer Adresse (postalisch und/oder E-Mail) mit einer Kontaktierung für weitere Befragungen einverstanden.²⁰ Die grundsätzliche Verweigerung der Teilnahme an Folgebefragungen (keine Angabe von Adressangaben für die Kontaktierung in der zweiten Welle) ist als paneltypischer Ausfallprozess zu nennen.

Tabelle 3: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Studienberechtigtenpanels 2018

| | Welle 1 |
|--------------------------------------|---------|
| Bruttostichprobe | 83.988 |
| Bereinigte Bruttostichprobe | 82.413 |
| Rücklauf Nettostichprobe (bereinigt) | 39.714 |
| Rücklaufquote | 48,2 % |

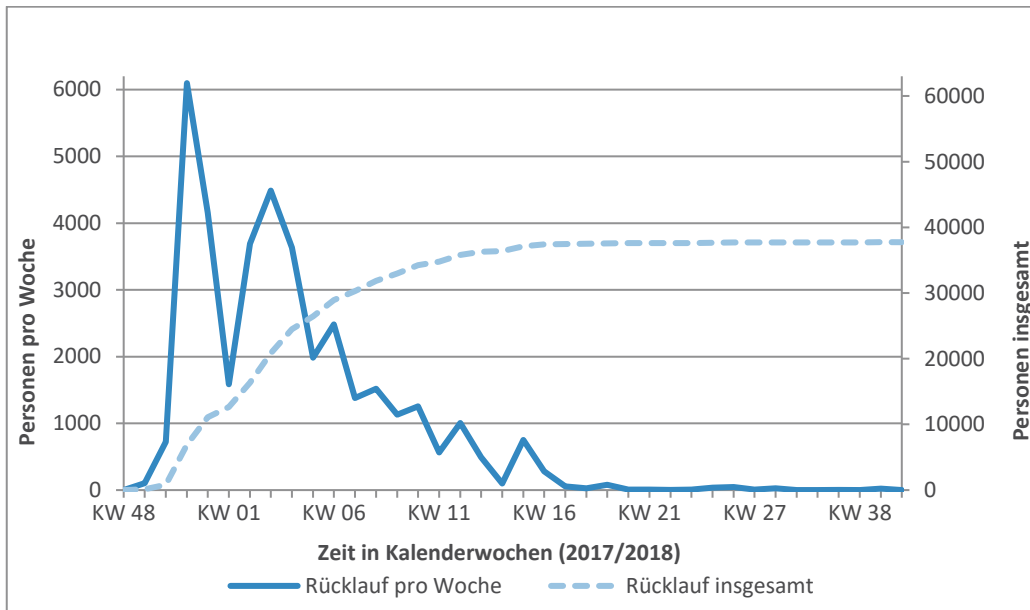
Wie Abbildung 2 zeigt, ist der Großteil (87 Prozent) der eingegangenen Fragebogen zwischen Dezember 2017 (KW 48) und Februar 2018 (KW 9) zurückgesendet worden. Aufgrund des teilweise verzögerten Versands von Einladungen war jedoch eine längere Feldphase nötig. Zum Ende der 17. KW lagen schließlich mehr als 99 Prozent aller gültigen Fragebogen vor, danach gingen nur noch vereinzelt Fragebogen ein.

¹⁸ Im Laufe der Rückmeldung von den Schulen wurden auch Personen im Nachhinein als nicht stichprobenzugehörig eingestuft, weil sie im relevanten Prüfungsjahr 2018 nur den schulischen Teil der Fachhochschulreife absolvierten. Diese Art von *Overcoverage* trat auf, da manche Schulen mehrere Schulzweige aufweisen, die nicht alle Teil der Stichprobe waren, jedoch von den Schulen zurückgemeldet wurden. Sowohl die Brutto- als auch die Nettostichprobe sind bereits um diese Zahlen bereinigt.

¹⁹ Insgesamt acht Befragte (Stand 05.07.2019) haben sich im Laufe der Feldphase zur zweiten Befragung der Studienberechtigten 2018 beim DZHW gemeldet und um nachträgliche Löschung ihrer Befragungsdaten gebeten. Die entsprechenden Fälle werden hier bereits nicht mehr als Rücklauf gezählt, da sie im Datensatz nicht mehr enthalten sind. Zudem wurde die Rücklaufzahl um Fälle bereinigt, die angaben, die Hochschulreife in einem anderen Jahr erworben zu haben bzw. dass sie an einer Berufsfachschule (voraussichtlich) den schulischen Teil der Hochschulreife erlangen und entsprechend nicht zur Grundgesamtheit gehören. Personen, die gemäß ihren Angaben in Welle 1 den schulischen Teil der Fachhochschulreife an einer anderen Schulform anstreben (n=28), sind weiterhin im Datensatz enthalten. Vorsätzlich falsch ausgefüllte und nicht verwertbare Fragebogen wurden ebenso aus dem Datensatz entfernt und hier im Rücklauf nicht berücksichtigt (siehe Kapitel 6.3).

²⁰ Hiervon haben 42 Befragte aus Berlin zwar eine Adresse angegeben, die dazugehörigen Befragungsdaten konnten aber nicht verwendet werden, da die entsprechende Einwilligung fehlte (Einwilligung in die Verwendung von Befragungs- bzw. Adressdaten erfolgte in Berlin separat).

Abbildung 2: Rücklauf des DZHW-Studienberechtigtenpanels 2018 im Zeitverlauf, erste Welle



An die Schulleitungen wurden insgesamt 1.104 Fragebogen verschickt. Dabei können Schulen auch mehrere Fragebogen für die Schulleitungen erhalten haben, da an manchen Schulen mehrere Schulzweige befragt wurden. Die Rücklaufquote betrug 59,1 Prozent, sodass eine Nettostichprobe von 652 auswertbaren Fragebogen der Schulleitungen zu verzeichnen war.

6 Datenaufbereitung

Im Folgenden werden die verschiedenen Schritte der Datenaufbereitung beschrieben. Die Aufbereitungsprozesse der Gewichtung und Anonymisierung werden in den beiden folgenden Kapiteln 7 und 8 gesondert erläutert.

6.1 Datenübertragung

[Erstellung eines Codeplans] Zur weiteren Verarbeitung wurden die Angaben der Befragten aus den Papierfragebogen auf Basis eines Codeplans in ein computerlesbares Format übertragen. Der Codeplan wurde auf Basis des Fragebogens der Befragung erstellt. Dabei wurde vermerkt, welcher Frage bzw. Teilfrage eine Variable zugeordnet ist, welchen Namen diese Variable trägt und welche numerischen Codierungen für die standardisierten Antworten der Befragten verwendet werden. Um die Erfassungsreihenfolge festzulegen, wurden die Variablen zusätzlich nummeriert.

[Datenerfassung] Für die Datenübertragung wurden der Codeplan, weitere Anweisungen zur Datenerfassung sowie die vorbereiteten Papierfragebogen an einen externen Dienstleister übergeben. Die Erfassung der Angaben erfolgte dort manuell durch Schreibkräfte.

6.2 Codierung offener Angaben

Vor der Datenübertragung (vgl. Kapitel 6.1) erfolgte auf den Papierfragebogen eine Codierung der (halb-)offenen Angaben. Dabei wurden diesen anhand von Codierlisten numerische Codierungen zugeordnet. Je nach Variable wurde eine unterschiedliche Codierliste verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.) oder um bereits in anderen Studien eingesetzte projekteigene Schlüssel. Für einige Variablen wurden neue projekteigene Codierlisten auf Basis der in den Daten des Studienberechtigtenjahrgangs 2018 vorkommenden Nennungen entwickelt. Für eine halboffene Frage (Frage 1.1: Abfrage der Schulart) wurde keine neuen Variable mit numerischen Codierungen erstellt, sondern die Nennungen nur den vorhandenen (geschlossenen) Antwortkategorien zugeordnet.

In Tabelle 4 sind die codierten Merkmale sowie die jeweils verwendete Codierliste dargestellt. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport dokumentiert. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen selbst sind nicht im Datensatz enthalten (vgl. Kapitel 8).

Tabelle 4: Vercodete Merkmale und verwendete Codierlisten im Studienberechtigtenpanel 2018 (erste Welle)²¹

| Merkmal | Codierliste |
|--|--|
| Bezeichnung (Ausbildungs-/Wunsch)Beruf | Klassifikation der Berufe (KldB) 2010 (erste Fassung) (5-Steller) ^a |
| Hochschule | Deutsche Hochschulen: Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2017/18) ^b (ergänzt um eigene Sondercodes) Ausländische Hochschulen: Projekteigene Codierung nach Land der Hochschule |
| Geburtsland und Staatsbürgerschaft | Projekteigene Codierung |
| Sonstige offene Abfragen | Projekteigene Codierung oder Zuordnung zu vorgegebenen Kategorien |

^a <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-erste-Fassung/Generische-Publikationen/Systematiken/KldB2010-ersteFassung-Systematisches-Verzeichnis.xlsx> (abgerufen am 18.10.2022)

^b https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studenten_2_Pr%C3%BCfungen/Schluessselverzeichnis_WS_2017_Stand_2018_01_04.xlsx (abgerufen am 10.11.2022)

6.3 Datenprüfung und Datenbereinigung

[Manuelle Vorkorrektur] Bereits vor der Übertragung der Daten (vgl. Kapitel 6.1) wurden auf den Papierfragebogen eine manuelle Prüfung und gegebenenfalls eine Anpassung von Angaben der Befragten durchgeführt.²² Dies sollte vor allem die Erfassbarkeit der Daten erleichtern. Dafür wurde in erster Linie die Form der bestehenden Angaben verändert. Beispielsweise wurden schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen oder verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0). Für Mehrfachnennungen bei Fragen, die eigentlich nur eine Einfachauswahl vorsahen, wurden teils zusätzliche Codes ergänzt.

[Softwaregestützte Korrektur] Im Anschluss an die Datenübertragung erfolgte eine umfassende Prüfung und Korrektur der Daten mit Hilfe einer DZHW-eigenen Software. Dabei sollten zum einen mögliche Fehler bei der vorherigen manuellen Vorkorrektur und Datenübertragung, zum anderen inkonsistente Angaben der Befragten identifiziert werden.

Zu diesem Zweck wurden die erfassten Fragebogen-Daten in eine Datenbank eingelesen. Anschließend wurden anhand formaler Regeln gültige Wertebereiche und Antwortkombinationen definiert und geprüft. Folgende Typen von Prüfungen wurden vorgenommen:

- *Prüfung von Wertebereichen:* Es wurde geprüft, ob die erfasste Ausprägung einer Variablen in dem für diese Variable definierten Wertebereich lag. Beispielsweise durften bei Variablen mit 5er-Skalen nur Werte von 1 bis 5 auftreten; im Zuge der Codierung der offenen Berufsangaben mussten vergebene Codes 5-stellig sein.
- *Prüfung der Einhaltung der Filterführung:* Auf Grundlage der definierten Filterführung des Fragebogens wurde zum einen geprüft, ob für die jeweilige befragte Person Angaben zu erwarten gewesen wären, die aber nicht vorhanden waren (Vollständigkeitsprüfung), und zum anderen,

²¹ Darauf hingewiesen sei außerdem, dass für Studienfächer, die nicht in der dem Fragebogen angehängten Studienfachliste enthalten waren, zusätzliche Codes ergänzt wurden.

²² Die Zahl der vorgenommenen Korrekturen wurde nicht zentral, sondern nur auf den Papierfragebogen dokumentiert und ist daher nicht mehr systematisch rekonstruierbar.

ob für die jeweilige Person, Angaben vorhanden waren, die nicht hätten erfolgen dürfen (Filterverstöße).

- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben innerhalb des Fragebogens geprüft.

Insgesamt wurden mehrere hundert Konsistenzregeln definiert und getestet. Bei fehlenden, fehlerhaften oder unplausiblen Werten wurde zunächst mit Hilfe des Papierfragebogens geprüft, ob der entsprechende Wert falsch (bzw. nicht) übertragen worden war. War dies nicht der Fall, wurden auf Basis für die einzelnen Fragen spezifisch festgelegter Regeln sowie anderer Angaben im Fragebogen Umsetzungen oder Korrekturen vorgenommen. Die entsprechenden Regeln bzw. Codieranweisungen bei Inkonsistenzen wurden in einer Anleitung festgehalten. Fehlerkorrekturen wurden auf den Fragebogen dokumentiert²³ und in der Regel von mindestens einer weiteren Person geprüft.

[Löschung von Fällen] Vorsätzlich von den Befragten falsch ausgefüllte Fragebögen der ersten Welle wurden ebenso aus dem Datensatz entfernt wie Fälle, die nicht zur Grundgesamtheit gehörten (bspw. falscher Jahrgang; siehe Kapitel 3). Insgesamt wurden nach den genannten Kriterien in der ersten Befragungswelle 110 Fälle gelöscht.

6.4 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde für das DZHW-Studienberechtigtenpanel 2018 eine den projektspezifischen Anforderungen orientierte Systematik erstellt. Tabelle 5 stellt die verwendete Missingsystematik dar.

Tabelle 5: Missingsystematik im DZHW-Studienberechtigtenpanel 2018

| Werte | Bezeichnung | Erläuterung |
|-------|-------------------------|---|
| -999 | weiß nicht | „weiß nicht“, wenn dies explizit hingeschrieben wurde. |
| -998 | keine Angabe | Befragte Person hat keine Angabe gemacht, also nichts angekreuzt Wird auch vergeben, wenn eine Person eine Itematterie (die eigentlich mit 0 „nicht genannt“ und 1 „genannt“ vercodet wird) gänzlich nicht beantwortet hat |
| -996 | Abbruch | Wird für alle weiteren Items/Fragen vergeben, wenn ab einem bestimmten Punkt keine Angaben mehr vorliegen. |
| -995 | keine Teilnahme (Panel) | In dem integrierten Gesamtdatensatz (alle Fälle der ersten und zweiten Befragung des Jahrgangs 2018) werden auf den Variablen der zweiten Welle alle Fälle, die nur an der |

²³ Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nicht systematisch rekonstruierbar.

| | | |
|------|---|---|
| | | ersten Befragung teilgenommen haben auf -995 gesetzt. |
| -989 | Filter | filterbedingt fehlend (siehe Flowchart) |
| -988 | trifft nicht zu | Dieser Code wird nicht bei Filterführung vergeben, sondern wenn explizit eine Antwortoption „trifft nicht zu“ vorgesehen ist oder eine Antwortkategorie durch erfolgte Verwendung anderer Antwortkategorien bereits ausgeschlossen wurde. |
| -987 | designbedingt fehlend (Fragebogensplit) | In Berlin und Hessen durften in der ersten Welle keine Fragen zu den Eltern gestellt werden, weshalb sie in diesen Variablen den Missingcode „designbedingt fehlend“ enthalten. |
| -968 | unplausibler Wert | Je nach Frage unterschiedlich definiert (z. B. Wertebereich, unplausible Mehrfachnennung) |
| -966 | nicht bestimmbar | Wird bei offenen Angaben oder Anmerkungen vergeben, die nicht vercodet bzw. plausibilisiert werden konnten. |
| -949 | nicht valide | nicht valide (z. B. unkenntlich gemacht oder gestrichen) |

6.5 Datensatzerstellung

Die Daten der Schüler*innen- und der Schulleitungsbefragung wurden in einem Datensatz abgelegt und um die zusätzlich generierten Variablen (vgl. Kapitel 6.2, 7 und 8) ergänzt. Pro Schüler*in existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

Der Datensatz wird standardmäßig im Stata- und SPSS-Format bereitgestellt.

6.6 Vergabe von Variablennamen, Variablenlabels und Wertelabels

[Variablen- und Wertelabelvergabe] Für Variablen- und Wertelabels wurden Formulierungen des Fragebogens übernommen oder prägnante Kurzformen dieser Formulierungen gewählt. Dabei basieren die Variablenlabels in der Regel auf dem entsprechenden Fragetext. Grundlage für die Wertelabels sind je nach Fragetyp die Texte der Antwortoptionen bzw. eine Kombination der Texte von Frage und Antwortoption.

[Variablenbenennung] Die Benennung der Variablen erfolgt anhand von inhaltlichen Kriterien, orientiert sich also am Frageinhalt. Für Indikatoren, die in mehreren Befragungswellen verwendet wer-

den, wurden die Namen der zugehörigen Variablen durch die Vergabe eines identischen Stammes harmonisiert. Bei Fragebatterien wurde ein einheitlicher Variablenstamm gewählt und die einzelnen Items mit einer Zahl versehen (z. B. unterricht01-unterricht10).

7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe im Vergleich zur definierten Grundgesamtheit. Das Designgewicht ist zudem notwendig, um die Disproportionalität der Stichprobe auszugleichen. Es folgt zunächst eine allgemeinere Einführung in die Vorgehensweise und die Darstellung der erstellten Gewichte. Im Anschluss wird die Gewichtungsprozedur im Detail beschrieben.

7.1 Vorgehen und Anwendungshinweise

[Ursachen für die Verzerrungen der Stichprobe] Maßgeblich für die Verzerrungen der Stichprobe sind zwei Prozesse:

- **Designbedingte Verzerrung:** Disproportionalitäten werden bewusst erzeugt, um in bestimmten relevanten Subgruppen die Fallzahlen zu erhöhen (vgl. Kapitel 3).
- **Verzerrung durch Nonresponse:** Ausfallprozesse (z. B. Nichtteilnahmen, fehlende Erreichbarkeit, Verlust auf dem Postweg) führen zu einem verringerten Rücklauf und somit zu einer Differenz zwischen Brutto- und Nettostichprobe (vgl. Kapitel 5). Wenn diese Ausfallprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.²⁴ Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Missing not at Random), der einer Modellierung bedarf.²⁵

[Konzeptuelles Vorgehen] Es werden für die Auswertungen der ersten Welle sowohl ein Designgewicht als auch zwei Querschnittsgewichte bereitgestellt. Das Designgewicht ergibt sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Um die Querschnittsgewichte zu erzeugen, wurde, wie in vorangegangenen Studienberechtigtenbefragungen, auf eine Zellgewichtung zurückgegriffen, mit dem die Daten an ausgewählte Merkmale der Grundgesamtheit angepasst werden. In Tabelle 6 sind die im Datensatz bereitgestellten Gewichte dargestellt.

Tabelle 6: Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2018

| Variablenname | Beschreibung |
|---------------|--|
| desgew | Designgewicht Welle 1 |
| gew1 | Querschnittsgewicht Welle 1 (für Länderauswertungen) |
| gew2 | Querschnittsgewicht Welle 1 (für Bundesauswertungen) |

[Hinweise zur Anwendung der Gewichte] Bei den erstellten Gewichten der ersten Welle handelt es sich um probability weights, die in Stata mit Hilfe spezifischer Optionen berücksichtigt werden können.²⁶ Grundsätzlich ist zu beachten, dass Gewichte nur für jene Analysen sinnvolle Korrekturgrößen

²⁴ Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

²⁵ Siehe Rubin (1976) zu den unterschiedlichen Formen von Ausfallprozessen.

²⁶ Siehe hierzu die Stata-Hilfe (Befehl: help weights).

darstellen, in denen die zur Gewichtung verwendeten Variablen in dem Analysemodell enthalten sind oder mit diesen in einem Zusammenhang stehen. Aus diesem Grund sollten Gewichte immer mit Bedacht und Fokus auf die analysierte Fragestellung verwendet werden. Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für den Datensatz näher dargestellt.

7.2 Gewichtung des Datensatzes

[Designgewichtung] Die Berechnung des Designgewichts ergibt sich im vorliegenden Studiendesign direkt aus dem Stichprobenplan. Für jede Schicht wird ein spezifisches Designgewicht bestimmt, wobei die Schichten sich anhand der Merkmale Bundesland und Schulart²⁷ ergeben (vgl. Kapitel 3). Benötigt werden für jede Schicht die folgenden Informationen aus dem Stichprobenplan:

- Anzahl der Schulen bzw. Schulzweige in der Grundgesamtheit
- Anzahl der gezogenen Schulen bzw. Schulzweige je Schicht

Anhand dieser Informationen wird für jede Schicht die spezifische Auswahlwahrscheinlichkeit π bestimmt. Diese ergibt sich aus der Formel:

$$\pi = \frac{n(\text{Stichprobe})}{N(\text{Grundgesamtheit})}$$

Das Designgewicht ergibt sich schließlich als Kehrwert der Auswahlwahrscheinlichkeit:

$$\frac{1}{\pi}$$

[Querschnittsgewichtung] Für die erste Welle wurden mit einer Zellgewichtung Querschnittsgewichte erstellt. Hierbei wurden die Daten anhand der Merkmale Geschlecht, Schulgruppen²⁸ und Bundesland an die Grundgesamtheit angepasst, um mit Blick auf diese Merkmale Verzerrungen in der realisierten Stichprobe gegenüber der Verteilung in der Grundgesamtheit auszugleichen.²⁹ Diese Verzerrungen können sowohl aus Unit- oder Item-Nonresponse resultieren als auch aus dem disproportionalen Stichprobendesign. Im Unterschied zum Designgewicht gehen in die Querschnittsgewichte der ersten Welle (gew1 und gew2) nicht alle Befragten der Vollerhebung aus Hamburg ein. Stattdessen wurde in Hamburg ein Subsample an Schulen/Schulzweigen entsprechend der Grundgesamtheit zufallsbasiert gezogen und in den Querschnittsgewichten berücksichtigt.

Für jede Zelle der Gewichtungsmatrix wird anhand der Sollzahlen zunächst der Anteil bestimmt, den diese Zelle in der Grundgesamtheit repräsentiert (pN). Für die realisierte Stichprobe wird ebenfalls eine solche Quote gebildet ($pn = \text{Anteil realisierte Fälle in der jeweiligen Zelle} / \text{Gesamtfallzahl in der Stichprobe}$). Der Gewichtungsfaktor ergibt sich, indem aus beiden Anteilswerten ein Quotient gebildet wird (pN/pn).

Dieses Vorgehen wird zweimal durchgeführt. Beim ersten Gewicht gew1 werden die Gewichtungsfaktoren für jedes Land gesondert berechnet. Bezugsgruppe für die Bestimmung der Anteilswerte pN und pn ist dann die Fallzahl in Grundgesamtheit bzw. Stichprobe im jeweiligen Bundesland. Das

²⁷ Hierbei handelt es sich um die neun Schularten, die auch maßgeblich für die Stichprobenziehung waren (siehe Kapitel 3).

²⁸ Dies waren die folgenden Gruppen: (1) Gymnasium/Gesamtschule/Waldorfschule, (2) Abendgymnasium/Kolleg, (3) Fachgymnasium/andere berufsbildende Schulen mit Erwerb des Abiturs, (4) Fach- und Berufsoberschulen (mit Erwerb der FHR), (5) sonstige berufsbildende Schulen (mit Erwerb der FHR). Die Zuordnung zu den Schulgruppen erfolgt anhand der Angabe der Befragten zur besuchten Schule bzw. dem besuchten Schulzweig.

²⁹ Datengrundlage bildeten die zum Zeitpunkt der Gewichtung (Dezember 2018) aktuellsten vorliegenden Zahlen zu Schulabsolvent*innen mit Studienberechtigung des Schuljahres 2017/18. Diese wurden der Fachserie 11 (Reihe 1 für allgemeinbildende Schulen sowie Reihe 2 für berufsbildende Schulen) entnommen. Diese Fachserie enthält allerdings nur die Absolvent*innenzahlen des Jahres 2017.

Gewicht gew_2 berücksichtigt zusätzlich die Relationen der Länder untereinander, da die Anteilswerte hier bezogen auf die bundesweiten Fallzahlen bestimmt werden.

[Normierung auf die Fallzahl der Stichprobe] Das errechnete Designgewicht wurde, wie in der sozialwissenschaftlichen Forschungspraxis üblich, auf die Fallzahl der Stichprobe normiert. Da die Berechnung der Querschnittsgewichte auf relativen Häufigkeiten basiert, ist eine Normierung auf die Stichprobengröße bereits erfolgt.

[Trimming der Gewichte] Für das Designgewicht erfolgte keine Trimmung. Bei der Zellgewichtung wurden teilweise Zellzusammenfassungen vorgenommen. Zwingend notwendig waren Zellzusammenfassungen immer dann, wenn eine Zelle Fälle laut amtlicher Statistik aufwies, jedoch keine entsprechenden Fälle in der Stichprobe vorlagen. Solche „Leierzellen“ wurden dann nach inhaltlichen Kriterien mit einer anderen Zelle zusammengefasst.

Des Weiteren wurden extreme Gewichtungsfaktoren geprüft und in Einzelfällen ebenfalls mit anderen Zellen verschmolzen. Geprüft wurden in diesem Zusammenhang alle Gewichte mit Werten kleiner als 0,3 und größer als 6. Eine Zellverschmelzung wurde aber nur bei deutlichen Abweichungen von den Grenzwerten vorgenommen und wenn eine inhaltlich sinnvolle Zusammenführung mit einer anderen Zelle möglich war. In jedem Fall wurden lediglich Zellverschmelzungen zwischen Schularten vorgenommen. Eine Zusammenfassung von Zellen über Bundesländer hinweg sowie zwischen Männern und Frauen wurde aus inhaltlichen Gründen nicht angewendet.

8 Anonymisierung

[Datenschutzrechtlicher Rahmen] Für personenbezogene Daten³⁰, die in freiwilligen Befragungen durch das DZHW erhoben werden, gelten die EU-Datenschutz-Grundverordnung (EU-DSGVO) und das Bundesdatenschutzgesetz in seiner Neufassung vom 30. Juni 2017.³¹ Danach sind personenbezogene Daten für die Weitergabe zur wissenschaftlichen Sekundärnutzung (ohne Vorliegen einer Einverständniserklärung zur Sekundärnutzung der personenbezogenen Daten) in der Regel derart aufzubereiten, dass „die personenbezogenen Daten ohne Hinzuziehung zusätzlicher Informationen nicht mehr einer spezifischen betroffenen Person zugeordnet werden können, sofern diese zusätzlichen Informationen gesondert aufbewahrt werden und technischen und organisatorischen Maßnahmen unterliegen, die gewährleisten, dass die personenbezogenen Daten nicht einer identifizierten oder identifizierbaren natürlichen Person zugewiesen werden können“ (Art. 4 Abs. 5 DSGVO; s. auch Art. 89 DSGVO sowie Erwägungsgrund 26 DSGVO).

[Sicherstellung des Datenschutzes im FDZ-DZHW] Im FDZ-DZHW wird der Datenschutz der Befragten über rechtlich-organisatorische Maßnahmen, technische Maßnahmen und statistische Maßnahmen sichergestellt³², die in unterschiedlicher Weise kombiniert werden können. Aus den Kombinationen ergeben sich vier potenzielle Zugangswege, die in Tabelle 7 im Überblick dargestellt sind.³³ Je stärker der Datenzugang rechtlich-organisatorisch und technisch kontrolliert und beschränkt wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen anonymisiert, also um Informationen reduziert, werden, d. h. desto größer bleibt ihr Analysepotenzial.

Tabelle 7: Zugangswege im FDZ-DZHW

| | Campus Use File (CUF): Download | Scientific Use File (SUF): Download | Scientific Use File (SUF): Remote-Desktop | Scientific Use File (SUF): On-Site |
|--------------------------------------|------------------------------------|--|--|---------------------------------------|
| rechtlich-organisatorische Maßnahmen | gering | moderat | moderat | hoch |
| technische Maßnahmen | sehr gering | gering | moderat | hoch |
| statistische Maßnahmen | sehr hoch | hoch | moderat | gering |

³⁰ „Personenbezogene Daten (sind) alle Informationen, die sich auf eine identifizierte oder identifizierbare natürliche Person (im Folgenden „betroffene Person“) beziehen; als identifizierbar wird eine natürliche Person angesehen, die direkt oder indirekt, insbesondere mittels Zuordnung zu einer Kennung wie einem Namen, zu einer Kennnummer, zu Standortdaten, zu einer Online-Kennung oder zu einem oder mehreren besonderen Merkmalen identifiziert werden kann, die Ausdruck der physischen, physiologischen, genetischen, psychischen, wirtschaftlichen, kulturellen oder sozialen Identität dieser natürlichen Person sind“ (Art. 4 DSGVO, S. 1).

³¹ Die DSGVO gilt grundsätzlich innerhalb der EU und somit ebenfalls für das DZHW. Das BDSG in seiner Neufassung vom 30. Juni 2017 (Gesetz zur Anpassung des Datenschutzrechts an die Verordnung (EU) 2016/679 und zur Umsetzung der Richtlinie (EU) 2016/680 (Datenschutz-Anpassungs- und Umsetzungsgesetz EU DSAnpUG-EU)) kommt teils zusätzlich zur Anwendung, da die DZHW GmbH juristisch als öffentliche Stelle des Bundes betrachtet wird (§ 2 Abs. 3 BDSG). Der Bund hält die absolute Mehrheit der Anteile der DZHW GmbH und das Institut erfüllt Aufgaben der öffentlichen Verwaltung des Bundes im weitesten Sinn.

³² Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, 6ff.), an dem sich bereits das Leibniz-Institut für Bildungswissenschaften (IfBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker & Roß, 2012, 9f.) orientieren.

³³ Für genauere Informationen vgl. <https://www.fdz.dzhw.eu/de/datennutzung>

Für die Daten des DZHW-Studienberechtigtenpanels 2018 gibt es nur die beiden Zugangswege „Scientific Use File (SUF): Remote-Desktop“ und „Scientific Use File (SUF): On-Site“. Im Folgenden werden für beide Zugangswege die durchgeführten statistischen Anonymisierungsmaßnahmen genauer erläutert.

[Statistische Anonymisierungsmaßnahmen] Im Rahmen der statistischen Anonymisierungsmaßnahmen wurden zunächst alle Informationen darauf geprüft, ob sich über diese Personen direkt identifizieren lassen. Diese sogenannten *direkten Identifikatoren*, wie Namen, Adressen und E-Mail Adressen, wurden bereits während der Feldphase in einem separaten Datensatz erfasst und sind somit nicht im Datensatz enthalten. Die Original-Identifikationsnummer wurde entfernt und durch eine neue zufällig vergebene Identifikationsnummer ersetzt.

Anschließend wurden die *Quasi-Identifikatoren* bestimmt, also Informationen, die in Kombination oder durch die Anspielung externer Informationen geeignet sind, eine Person indirekt zu identifizieren. Um eine eindeutige Zuordnung der Daten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Zugangsweg – aggregiert oder gar nicht freigegeben. Im Zuge der Aggregationen neu generierte Variablen wurden im Datensatz durch das Suffix „_g#“ im Variablennamen gekennzeichnet und hinter der jeweiligen Ausgangsvariablen positioniert.

Darüber hinaus empfehlen Ebel und Meyermann, offene Angaben zu löschen „selbst wenn die jeweiligen Fragestellungen an sich unproblematisch sind. Denn es besteht die Gefahr, dass Studienteilnehmer/-innen bei eigentlich unbedenklichen Fragen mit offener Antwortmöglichkeit kritische Informationen preisgegeben haben, die zu einer Identifikation führen könnten“ (Ebel & Meyermann, 2015, S. 5). Entsprechend werden keine offenen Angaben freigegeben.

Zuletzt wurde geprüft, ob in den Daten sensible Informationen, z. B. zur Gesundheit, sexuellen Orientierung und zu politischen Einstellungen, enthalten waren, für die bei den Befragten kein zusätzliches Einverständnis für die Sekundärnutzung eingeholt wurde. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (vgl. Koberg, 2016, S. 694) und sind daher besonders schützenswert (Art. 9 DSGVO, Erwägungsgrund 51 DSGVO).

Die nachfolgende Tabelle 8 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Zugangsweg dar.

Tabelle 8: Maßnahmen der statistischen Anonymisierung der Daten

| Merkmal ^d | Scientific Use File (SUF): On-Site | Scientific Use File (SUF): Remote-Desktop |
|---|---|---|
| Direkte Identifikatoren | Vergabe von zufälligen IDs für - Schüler*innen - Schule - Schulleitung | Vergabe von zufälligen IDs für - Schüler*innen - Schule - Schulleitung |
| Welle 1, Schulleitungsbefragung | | |
| Funktion an der Schule (Frage 1.1) | keine Freigabe | keine Freigabe |
| Herkunftsort/-stadtteil der Schüler*innen (Frage 1.3) | keine Freigabe | keine Freigabe |
| sonstige Räume (Frage 2.1 (offene Angabe)) | Vercodung ^e | Vercodung ^e |
| andere Geldquellen (Frage 2.4 (offene Angabe)) | Vercodung ^e | Vercodung ^e |
| sonstige Aktivitäten zur Stärkung der Praxiseinblicke der Lehrkräfte (Frage 3.4 (offene Angabe)) | Vercodung ^e | Vercodung ^e |
| andere Maßnahmen zur Unterstützung der Arbeit mit Schüler*innen mit Migrationshintergrund oder deren Eltern (Frage 4.2 (offene Angabe)) | Vercodung ^e | Vercodung ^e |
| Welle 1, Schüler*innenbefragung | | |
| Schulart (Frage 1.1) | Aggregation der Kategorien - „Gesamtschule/Gemeinschaftsschule mit gymnasialer Oberstufe“ und „freie Waldorfschule“ - „Abendgymnasium“ und „Kolleg (nicht Berufskolleg)“ - „Fachschule“ und „Fachakademie“ | Aggregation zu den Kategorien - „allgemeinbildende Schule“ - „berufliche Schule“ |
| (voraussichtlicher) Ausbildungs-/Wunschberuf (Frage 1.10, Frage 2.3 und Frage 2.13 (offene Angabe)) | Vercodung ^e Aggregation zu KldB-4-Steller ^a | Aggregation zu den Kategorien - „allgemeinbildende Schule“ - „berufliche Schule“ Aggregation zu KldB-4-Steller ^a |
| Gesundheitszustand (Frage 1.11) | keine Freigabe | keine Freigabe |
| angestrebte Hochschule (Frage 2.9 (offene Angabe)) | Vercodung ^e und anschließend zusätzlich: Aggregation zu Bundesland der Hochschule, wenn deutsche Hochschule ^c , und Kategorie „Ausland“, wenn ausländische Hochschule Aggregation zu zwei Hochschulart-Kategorien „Universität/Pädagogische Hochschule/Kirchl.-Theol. Hochschu- | Vercodung ^e und anschließend zusätzlich: Aggregation zu Bundesland der Hochschule, wenn deutsche Hochschule ^c , und Kategorie „Ausland“, wenn ausländische Hochschule Aggregation zu zwei Hochschulart-Kategorien „Universität/Pädagogische Hochschule/Kirchl.-Theol. Hochschu- |

| | | |
|--|---|--|
| | le/Kunsthochschule“ und „Fachhochschule/Berufs-, Verwaltungs-, Wirtschaftsakademie“, wenn deutsche Hochschule ^c | le/Kunsthochschule“ und „Fachhochschule/Berufs-, Verwaltungs-, Wirtschaftsakademie“, wenn deutsche Hochschule ^c |
| sonstige Schwierigkeiten bei Wahl des nachschulischen Werdegangs (Frage 3.4 (offene Angabe)) | Vercodung ^e und anschließend zusätzlich Aggregation der Kategorien „Gesundheitsprobleme“ und „keine Angabe“ zu „nicht bestimmbar“ | Vercodung ^e und anschließend zusätzlich Aggregation der Kategorien „Gesundheitsprobleme“ und „keine Angabe“ zu „nicht bestimmbar“ |
| Staatsangehörigkeit/ Geburtsland (Eltern) (wenn nicht Deutschland/deutsch) (Frage 5.4, Frage 5.5, Frage 5.6 und 5.7 (offene Angabe)) | Vercodung ^e und anschließend zusätzlich: Aggregation der Kategorien „Kurdistan“ und „keine Angabe“ zu „nicht bestimmbar“ Aggregation zu NEPS-Regionen ^b (dabei Aggregation der Kategorien „Kurdistan“ und „keine Angabe“ zu „nicht bestimmbar“) | Vercodung ^e und anschließend zusätzlich: Aggregation zu NEPS-Regionen ^b (dabei Aggregation der Kategorien „Kurdistan“ und „keine Angabe“ zu „nicht bestimmbar“) |

^a gemäß Klassifikation der Berufe (KldB) 2010 (erste Fassung), vgl. <https://statistik.arbeitsagentur.de/DE/Statischer-Content/Grundlagen/Klassifikationen/Klassifikation-der-Berufe/KldB2010-erste-Fassung/Generische-Publikationen/Systematiken/KldB2010-ersteFassung-Systematisches-Verzeichnis.xlsx> (abgerufen am 18.10.2022)

^b gemäß NEPS-Liste, vgl. https://www.neps-data.de/Portals/0/NEPS/Datenzentrum/Forschungsdaten/SC5/16-0-0/SC5_16-0-0_Anonymization.pdf, Appendix A.2 (abgerufen am 01.11.2022)

^c gemäß Destatis-Schlüsselverzeichnis für die Studenten- und Prüfungsstatistik (WiSe 2017/18), vgl. https://www.statistik-nord.de/fileadmin/Dokumente/Erhebungen/Hochschulstatistiken/1_Studenten_2_Pr%C3%BCfungen/Schlüsselverzeichnis_WS_2017_Stand_2018_01_04.xlsx (abgerufen am 10.11.2022) (ergänzt um eigene Sondercodes)

^d Die Fragennummerierung bezieht sich jeweils auf die Standardvariante des Erhebungsinstruments.

^e vgl. Kapitel 6.2

9 Literatur

- Ebel, T. & Meyermann, A. (2015). Hinweise zur Anonymisierung von quantitativen Daten. Forschungsdaten Bildung informiert. Bd.3. Frankfurt am Main: Deutsches Institut für Internationale Pädagogische Forschung.
- Häder, M. (2015). Empirische Sozialforschung. Eine Einführung (3. Aufl.). Wiesbaden: Springer Fachmedien Wiesbaden. doi:10.1007/978-3-531-19675-6.
- Hochfellner, D., Müller, D., Schmucker, A. & Roß, E. (2012). FDZ-Methodenreport. Datenschutz am Forschungsdatenzentrum (Nr. 06). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. v. Maurice, M. Bayer & J. Skopek (Hrsg.), *Methodological Issues of Longitudinal Surveys. The example of the National Educational Panel Study* (S. 691–708). Wiesbaden: Springer VS. doi:10.1007/978-3-658-11994-2.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), 2–16.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.