

Heidrun Schneider / Sandra Vietgen

# DZHW-Studienberechtigten-panel 2015

Daten- und Methodenbericht zu den Erhebungen des Studienberechtigtenjahrgangs 2015 (1. und 2. Befragungswelle)

**Daten- und Methodenbericht**

April 2021

Autor(inn)en:  
Heidrun Schneider  
Sandra Vietgen

Herausgeber:  
Deutsches Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW)  
Lange Laube 12 | 30159 Hannover | [www.dzhw.eu](http://www.dzhw.eu)  
Tel.: +49 511 450670-0 | Fax: +49 511 450670-960 | [info@dzhw.eu](mailto:info@dzhw.eu)

Geschäftsführung:  
Prof. Dr. Monika Jungbauer-Gans  
Karen Schlüter

Vorsitzender des Aufsichtsrats:  
Ministerialdirigent Peter Greisler

Registergericht:  
Amtsgericht Hannover | B 210251

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/>



# Inhaltsverzeichnis

|     |   |    |
|-----|---|----|
| I   | Einleitung .....  | 1  |
| II  | Datennutzungshinweise .....                                       | 2  |
| 1   | Inhalt und Anlage der Studie.....                                 | 5  |
| 2   | Erhebungsinstrumente.....   | 9  |
| 2.1 | Inhalte der Erhebungsinstrumente .....                            | 9  |
| 2.2 | Pretest.....  | 11 |
| 3   | Grundgesamtheit und Stichprobenverfahren .....                    | 13 |
| 4   | Durchführung der Erhebungen .....                                 | 14 |
| 5   | Rücklauf.....   | 18 |
| 6   | Datenaufbereitung .....   | 20 |
| 6.1 | Datenübertragung.....   | 20 |
| 6.2 | Codierung offener Angaben .....                                   | 20 |
| 6.3 | Datenprüfung und Datenbereinigung.....                            | 21 |
| 6.4 | Generierung von Variablen .....                                   | 22 |
| 6.5 | Erstellung der Datensätze .....                                   | 22 |
| 6.6 | Vergabe von Variablennamen, Variablenlabels und Wertelabels ..... | 22 |
| 6.7 | Codierung fehlender Werte .....                                   | 23 |
| 7   | Gewichtung .....  | 24 |
| 7.1 | Vorgehen und Anwendungshinweise .....                             | 24 |
| 7.2 | Gewichtung des Datensatzes .....                                  | 25 |
| 8   | Anonymisierung .....  | 28 |
| 9   | Literaturverzeichnis.....   | 32 |

## Abbildungsverzeichnis

|              |   |    |
|--------------|---|----|
| Abbildung 1: | Datenzugangswege und Analysepotential .....   | 3  |
| Abbildung 2: | Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe .....   | 6  |
| Abbildung 3: | Entwicklung von Brutto- und Nettostichprobe im Zeitverlauf .....  | 18 |
| Abbildung 4: | Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential<br>der Daten des Studienberechtigtenpanels 2015 ..... | 29 |

## Tabellenverzeichnis

|            |  |    |
|------------|--|----|
| Tabelle 1: | Anlage und konzeptionelle Veränderungen der DZHW-Studienberechtigten-<br>befragungsreihe von 1976 bis 2015 .....                       | 6  |
| Tabelle 2: | Brutto-, Nettostichproben und Rücklaufquoten des DZHW-<br>Studienberechtigtenpanels 2015 (Welle 1 und 2) .....                         | 19 |
| Tabelle 3: | Systematik für fehlende Werte .....  | 23 |
| Tabelle 4: | Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2015 .....  | 25 |
| Tabelle 5: | Maßnahmen der statistischen Anonymisierung der Daten des DZHW-<br>Studienberechtigtenpanels 2015 (Welle 1 und 2) nach Zugangsweg ..... | 30 |

## I Einleitung

Die DZHW-Studienberechtigtenpanels stellen eine Untersuchungsreihe zu den Ausbildungs- und Studienentscheidungen beziehungsweise Ausbildungs- und Studienverläufen von Schulabsolvent(inn)en mit allgemeiner Hochschulreife (einschließlich fachgebundener Hochschulreife) oder Fachhochschulreife (einschließlich fach- und landesgebundener Fachhochschulreife) dar.<sup>1</sup> Sie werden durch das Deutsche Zentrum für Hochschul- und Wissenschaftsforschung GmbH (DZHW) durchgeführt, vom Bundesministerium für Bildung und Forschung (BMBF) finanziert und dienen – in Ergänzung zur amtlichen Hochschulstatistik – dem nationalen Bildungsmonitoring sowie der Beantwortung von Fragestellungen der Hochschul- und Wissenschaftsforschung. Seit 1976 werden alle ein bis drei Jahre neue Studienberechtigtenjahrgänge befragt.

Das Forschungsdatenzentrum am DZHW (FDZ-DZHW) bereitet die Daten einiger Studienberechtigtenjahrgänge nachträglich zum Zweck der Nachnutzung auf. Sie werden über verschiedene Zugangswege als Scientific Use Files (SUF) für die wissenschaftliche Sekundärnutzung zur Verfügung gestellt. Neben den Datensätzen der Erhebungen werden auch die Dokumentationsmaterialien zu den Datensätzen und zur Durchführung der Studien bereitgestellt.<sup>2</sup>

Der vorliegende Daten- und Methodenbericht ist Teil der Dokumentation der ersten beiden Befragungswellen des Studienberechtigtenjahrgangs 2015 (doi: [10.21249/DZHW:gsl2015:1.0.1](https://doi.org/10.21249/DZHW:gsl2015:1.0.1)). Weitere Dokumentationsmaterialien zur Studie (Datensatzreports, Fragebögen, Variablenfragebögen etc.) können frei im Metadaten suchsystem des FDZ-DZHW (<https://metadata.fdz.dzhw.eu>) heruntergeladen werden.

Abschnitt II dieses Berichts stellt die zentralen Informationen zur Nutzung der Daten dieser Studie im Überblick dar. Kapitel 1 stellt Inhalt und Anlage der DZHW-Studienberechtigtenbefragungsreihe bis zum Jahr 2015 im Allgemeinen und des Studienberechtigtenpanels 2015 im Speziellen vor. Die weitere Gliederung des Berichts orientiert sich im Wesentlichen am Ablauf des Forschungsprozesses. In Kapitel 2 werden die eingesetzten Erhebungsinstrumente und in den Kapiteln 3 bis 6 der Erhebungsprozess beschrieben (Stichprobenziehung, Erhebungsablauf, Rücklauf, Datenaufbereitung). In den Kapiteln 7 und 8 folgt die Darstellung der vorgenommenen Gewichtung und Anonymisierung.

---

<sup>1</sup> Aktuelle Informationen zum DZHW-Studienberechtigtenpanel können über die Website des Projektes (<https://bildungswege.dzhw.eu>) abgerufen werden.

<sup>2</sup> Informationen zu verfügbaren Datensätzen und Dokumentationen werden auf der Website des FDZ-DZHW (<https://fdz.dzhw.eu>) zur Verfügung gestellt.

## II Datennutzungshinweise

**[Voraussetzungen der Datennutzung]** Die Daten des Studienberechtigtenpanels 2015 werden durch das FDZ des DZHW entsprechend der europäischen Datenschutzgrundverordnung (EU-DSGVO) anonymisiert bereitgestellt und ausschließlich zur wissenschaftlichen Nutzung freigegeben.<sup>3</sup> Das FDZ bietet ein *Scientific Use File* (SUF) für die wissenschaftliche Sekundärnutzung an.

Voraussetzungen für die Nutzung des SUF sind die Anstellung der Datennutzerin/des Datennutzers an einer wissenschaftlichen Einrichtung und der Abschluss eines Datennutzungsvertrags mit dem FDZ. Studierende oder Promovierende ohne eine Anstellung an einer wissenschaftlichen Einrichtung müssen gemeinsam mit einer/einem betreuenden Mitarbeiter(in) einen Datennutzungsvertrag abschließen. Im Zuge des Vertragsabschlusses wird durch das FDZ auch das Vorliegen eines wissenschaftlichen Nutzungsinteresses geprüft.

**[Datenzugang]** Das SUF des Studienberechtigtenpanels 2015 wird über zwei Zugangswege angeboten, die hinsichtlich des Speicherortes, der Möglichkeit der eigenständigen Verknüpfung mit externen Daten und der Kontrollmöglichkeiten des FDZ unterschiedlich restriktiv sind.

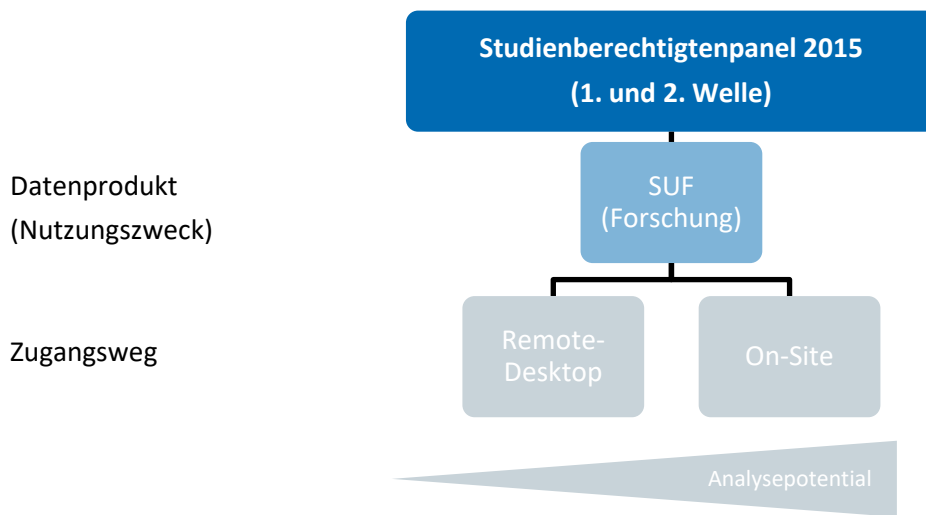
- **Remote-Desktop:** Die Daten werden auf einem Terminal-Server des FDZ bereitgestellt. Über eine besonders gesicherte Verbindung zwischen dem lokalen Computer der nutzenden Person und dem Terminal-Server des FDZ können die Daten mit der auf dem Terminal-Server vorhandenen Software analysiert werden. Das Übertragen der Daten auf den lokalen Computer ist nicht möglich. Analyseergebnisse werden erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.
- **On-Site:** Die Daten werden in den Räumlichkeiten des FDZ in einer kontrollierten Umgebung an einem speziell gesicherten Computer zur Analyse bereitgestellt. Wie beim Remote-Desktop-Zugang werden Analyseergebnisse erst nach einer Prüfung auf datenschutzrechtliche Unbedenklichkeit durch das FDZ freigegeben und zur Verfügung gestellt.

Die bereitgestellten Daten weisen je nach Zugangsweg einen unterschiedlich hohen Informationsgehalt auf und unterscheiden sich damit hinsichtlich ihres Analysepotentials (vgl. Abbildung 1). Dabei gilt: Je stärker der Datenzugriff der Nutzer(innen) durch technische und organisatorische Maßnahmen kontrolliert wird, desto mehr Informationen können für die Datennutzer(innen) bereitgestellt werden.<sup>4</sup> Mit diesem Vorgehen wird ein Höchstmaß an Nutzbarkeit und gleichzeitig ein bestmöglicher Schutz der bereitgestellten Daten sichergestellt.

<sup>3</sup> Das Datenschutzkonzept des FDZ ist angelehnt an den Portfolio-Ansatz von Lane, Heus und Mulcahy (2008, S. 6), an dem sich bereits das Leibniz-Institut für Bildungsverläufe (LIfBi) (vgl. Koberg, 2016, 699ff.) und das FDZ der Bundesagentur für Arbeit im Institut für Arbeitsmarkt- und Berufsforschung (vgl. Hochfellner, Müller, Schmucker und Roß (2012, S. 9) orientieren. Das FDZ des DZHW hat diesen Ansatz an die Anforderungen der eigenen Datenbestände angepasst und nutzt vier Kategorien von Maßnahmen zur Sicherstellung des Datenschutzes, die in unterschiedlicher Weise kombiniert werden können: Rechtlich-institutionelle Maßnahmen, informationelle Maßnahmen, technische Maßnahmen und statistische Maßnahmen.

<sup>4</sup> Zu den verschiedenen Anonymisierungsgraden und Analysepotentialen der verschiedenen SUF-Varianten vgl. Kapitel 8.

Abbildung 1: Datenzugangswege und Analysepotential



**[Datenprodukte]** Über den *Digital Object Identifier* (DOI) [10.21249/DZHW:gsl2015:1.0.1](https://doi.org/10.21249/DZHW:gsl2015:1.0.1) ist eine Website mit zentralen Informationen zur Studie, weiteren Dokumentationsmaterialien sowie einer Übersicht der zur Verfügung stehenden Datenprodukte zur Studie erreichbar.

Die bereitgestellten Daten des Studienberechtigtenpanels 2015 sind in zwei Datensätzen abgelegt. Es liegen ein Personendatensatz im wide-Format und ein Episodendatensatz im long-Format vor. Für das SUF werden für jeden im FDZ-DZHW angebotenen Zugangsweg beide Datensätze – jeweils mit zugangswegspezifischem Analysepotential (vgl. Abbildung 1) – bereitgestellt.

**[Gebühren der Datenbereitstellung]** SUF und CUF werden derzeit (Stand: April 2021) kostenfrei zur Verfügung gestellt. Änderungen bzw. die aktuelle Gebührenordnung können auf der Website des FDZ (<https://fdz.dzhw.eu>) eingesehen werden.

**[Pflichten der Datennutzer(innen)]** Die Datennutzer(innen) sind verpflichtet, folgende Regeln<sup>5</sup> einzuhalten:

- **Wissenschaftliche Nutzung:** Die Daten dürfen ausschließlich für wissenschaftliche Zwecke verwendet werden. Eine kommerzielle Nutzung ist untersagt.
- **De-Anonymisierungsverbot:** Jeder Versuch der Re-Identifikation von Analyseeinheiten (z. B. Personen, Haushalten, Institutionen) ist verboten.
- **Gebot zur Mitteilung von Sicherheitslücken:** Falls Datennutzer(innen) Kenntnis von Sicherheitslücken hinsichtlich Datenschutz bzw. Datensicherheit erlangen, müssen diese dem FDZ unverzüglich angezeigt werden.
- **Keine Weitergabe der Daten:** SUF dürfen nur durch die Person genutzt werden, die den Datennutzungsvertrag abgeschlossen hat.

<sup>5</sup> Der Datennutzungsvertrag regelt die Nutzungsbedingungen im Detail.

- **Löschungsgebot:** Download-SUF sind nach Ablauf der vereinbarten Nutzungsdauer (in der Regel 1,5 Jahre) von jeglichen Rechnern, Servern und Datenträgern zu löschen. Ebenso müssen alle Sicherungskopien, modifizierten Datensätze (z. B. Arbeits-, Auszugs- oder Hilfsdateien) sowie Ausdrücke vernichtet werden.
- **Bereitstellung/Meldung von Publikationen:** Jede Art von Publikation, die aus der Arbeit mit Daten des FDZ-DZHW hervorgeht, muss dem FDZ im Voraus gemeldet und nach Veröffentlichung unverzüglich und in elektronischer Form zur Verfügung gestellt werden.
- **Zitationspflicht:** Die verwendeten Daten müssen in Veröffentlichungen, anderen Arbeiten (z. B. Abschlussarbeiten) und Vorträgen laut der Vorgaben des FDZ zitiert werden (vgl. Zitationsanleitung unter [10.21249/DZHW:gsl2015:1.0.1](#)).



## 1 Inhalt und Anlage der Studie

**[Studienreihe]** Das DZHW-Studienberechtigtenpanel 2015 ist Teil der DZHW-Studienberechtigtenbefragungsreihe, in der anhand von standardisierten Mehrfachbefragungen Informationen zu den nachschulischen Werdegängen von Schulabsolvent(inn)en mit (schulischer) Hochschulzugangsberechtigung erfasst werden.<sup>6</sup> Die Untersuchungsreihe beginnt mit der Befragung des Studienberechtigtenjahrgangs 1976 in der damaligen BRD; seitdem wurde jeder zweite oder dritte Studienberechtigtenjahrgang (Kohorte) befragt. Die Schulabsolvent(inn)en in den neuen Bundesländern wurden erstmalig 1990 einbezogen. Die Grundgesamtheit aller DZHW-Studienberechtigtenbefragungen sind die Schulabsolvent(inn)en allgemeinbildender und berufsbildender Schulen eines Schuljahres mit bundesdeutscher Hochschulreife.

In der Regel<sup>7</sup> werden für jeden Studienberechtigtenjahrgang mehrere Befragungswellen zu unterschiedlichen Zeitpunkten vor und nach Erwerb der Hochschulzugangsberechtigung durchgeführt, das heißt es handelt sich um ein kombiniertes Kohorten-Panel-Design. Tabelle 1 hebt die konzeptionellen Änderungen hervor. Bis 1986 verfolgte das DZHW das Ziel jeden Studienberechtigtenjahrgang kurz nach sowie 2 ½, 4 ½ und 12 ½ Jahre nach Schulabschluss zu befragen. Ab 1990 wurde die Befragung zu verspäteten oder korrigierten Ausbildungs- und Studienentscheidungen zweieinhalb Jahre nach Schulabschluss gestrichen.<sup>8</sup> Mit dem Jahrgang 2015 wurde dieser Erhebungszeitpunkt wieder aufgenommen. Seit dem Studienberechtigtenjahrgang 2005 werden die Studienberechtigtenjahrgänge zusätzlich bereits ein halbes Jahr vor Schulabschluss befragt.

Die verschiedenen Erhebungen wurden als schriftlich-postalische Paper-and-Pencil-Befragung (PAPI) durchgeführt. Die Erhebungsinstrumente aller Kohorten enthalten Fragen zu Bildungsinformationen, nachschulischem Werdegang, zum Studium, zum Übergang in den Beruf und zu soziodemographischen und bildungsbiographischen Merkmalen. Der thematische Fokus einer Befragungswelle orientiert sich an der jeweils aktuellen Bildungs-, Berufs- und Lebensphase der Befragten zum Befragungszeitpunkt. Über eine Spanne von fast 40 Jahren bis zur vorliegenden Befragung entstand so eine Zeitreihe von 20 Jahrgängen mit insgesamt 47 Befragungswellen (vgl. Abbildung 2).

---

<sup>6</sup> Aktuelle Informationen zum Studienberechtigtenpanel sowie den bisher verfügbaren Scientific Use Files der Erhebungsreihe finden Sie unter <https://metadata.fdz.dzhw.eu>.

<sup>7</sup> Die Ausnahmen bilden die Studienberechtigtenjahrgänge 1991, 1993, 1996 und 2004, die nur einmal befragt wurden.

<sup>8</sup> Eine weitere Ausnahme stellt die, in Kooperation mit dem Bundesinstitut für Berufsbildung (BIBB) durchgeführte, zusätzliche Befragungswelle des Studienberechtigtenjahrgangs 1990 dar, die 5 ½ Jahre nach Schulabschluss erfolgte.

**Tabelle 1: Anlage und konzeptionelle Veränderungen der DZHW-Studienberechtigtenbefragungsreihe von 1976 bis 2015**

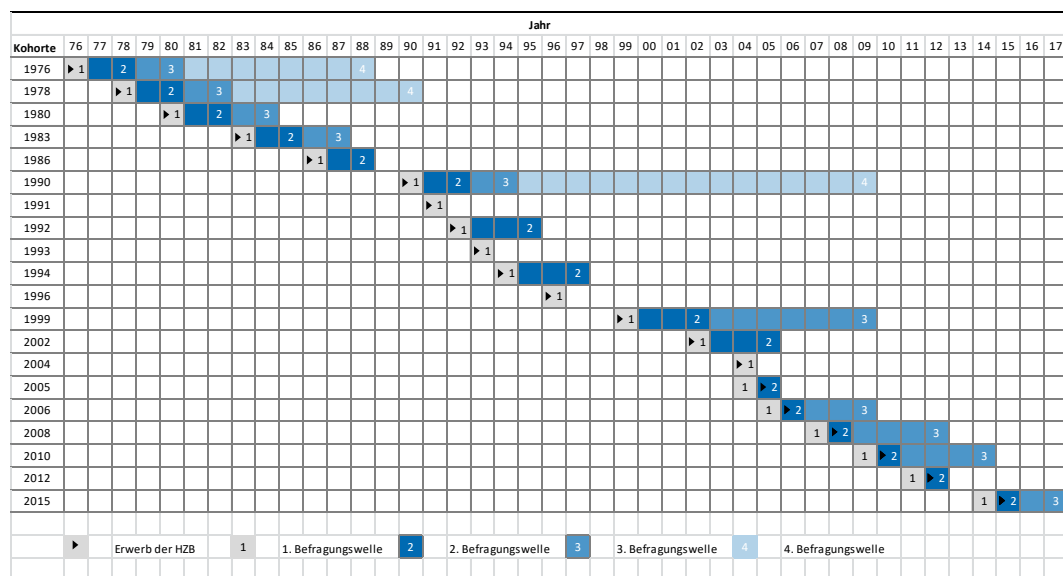
| Befragungszeitpunkt                     | Thematischer Fokus   | Studienberechtigtenjahrgänge <sup>a</sup> |               |                       |           |
|---|--|---|---------------|-----------------------|-----------|
|   |  | bis 1986 <sup>b</sup>                     | 1990 bis 2004 | 2005 bis 2012         | seit 2015 |
| ½ Jahr vor Schulabschluss               | Ausbildungs- und Studienabsichten                                  | -   | -             | 1. Welle              | 1. Welle  |
| ½ Jahr nach Schulabschluss              | Ausbildungs- und Studienentscheidungen                             | 1. Welle                                  | 1. Welle      | 2. Welle              | 2. Welle  |
| 2 ½ Jahre nach Schulabschluss           | verspätete oder korrigierte Ausbildungs- und Studienentscheidungen | 2. Welle                                  | -             | -                     | 3. Welle  |
| 3 ½ bzw. 4 ½ Jahre nach Schulabschluss  | Ausbildungs- und Studienverläufe                                   | 3. Welle                                  | 2. Welle      | 3. Welle <sup>c</sup> | 4. Welle  |
| 10 ½ bis 20 ½ Jahre nach Schulabschluss | Erwerbseintritt und -verläufe                                      | 4. Welle                                  | 3. Welle      | -                     | -         |

<sup>a</sup> Für einige Studienberechtigtenjahrgänge konnten nicht alle Wellen realisiert werden (siehe Abbildung 2).

<sup>b</sup> Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

<sup>c</sup> Für den Studienberechtigtenjahrgang 2012 konnte die 3. Welle erst 6 ½ Jahre nach Erwerb der Hochschulzugangsberechtigung realisiert werden.

**Abbildung 2: Kohorten-Panel-Design der DZHW-Studienberechtigtenbefragungsreihe**



Anmerkung: Die Befragungen der Jahrgänge 1976 bis 1986 richteten sich ausschließlich an Hochschulzugangsberechtigte aus den Bundesländern der damaligen Bundesrepublik Deutschland.

Quelle: Schneider, Franke, Woisch und Spangenberg (2017); eigene Überarbeitung

**[Analysepotential]** In allen Kohorten wird je Befragungswelle ein identischer Kern an Informationen erhoben. Auf dieser Grundlage können mit Zeitreihen- bzw. Kohortenvergleichen langfristige Trends im Übergang von der Schule in das weitere Ausbildungssystem und in den Arbeitsmarkt in den Blick genommen werden. Besonders hervorzuheben ist, dass in allen Befragungen von Studienberechtigtenjahrgängen, die mit mehr als einem halben Jahr Abstand nach Erwerb der Hochschulreife erfolgten, halbjahres- oder monatsgenaue kontinuierliche Verlaufsdaten zum individuellen Tätigkeitsverlauf seit dem Schulabschluss erfasst werden, die sich für Ereignisdaten- und Sequenzmusteranalysen eignen. Prinzipiell variiert innerhalb eines Befragungsjahrgangs der thematische Fokus von Welle zu Welle. Spezifische Fragen wurden jedoch wiederholt gestellt, sodass die Angaben derselben Befragten aus verschiedenen Wellen direkt miteinander verknüpft und intra-individuelle Veränderungen über den Zeitverlauf betrachtet werden können. Auf Basis des DZHW-Studienberechtigtenpanels können somit Querschnitts-, Trend-, Ereignisdaten-, Sequenzmuster- sowie eingeschränkt kausale Panelanalysen durchgeführt werden. Die Daten der Studienberechtigtenpanels werden in der Regel anhand der Merkmale Geschlecht, Bundesland, Art der Schule und der Hochschulreife gewichtet und an die Grundgesamtheit angeglichen. Zusätzlich werden paneltypische Ausfallprozesse in der Gewichtung der Daten berücksichtigt.

**[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 Abiturientenbefragung<sup>9</sup> (durchgeführt seit 1996 alle zwei Jahre durch das Kompetenzzentrum für Bildungs- und Hochschulplanung, welches im Jahr 2015 in das Zentrum für Qualitätsanalyse (ZQA) der TU Dresden integriert wurde), die TOSCA-Studie<sup>10</sup> (inzwischen durchgeführt vom Hector-Institut für Empirische Bildungsforschung an der Universität Tübingen) sowie das Berliner-Studienberechtigtenpanel *Best Up*<sup>11</sup> (durchgeführt vom Deutschen Institut für Wirtschaftsforschung (DIW) und dem Wissenschaftszentrum Berlin (WZB)).

**[Spezifika des Studienberechtigtenpanels 2015]** Neben den allgemeinen Charakteristika der Studienreihe weist die hier betrachtete Befragung des Studienberechtigtenjahrgangs 2015 folgende Spezifika auf. So wurden – wie seit der Pilotstudie mit dem Studienberechtigtenjahrgang 2005 beziehungsweise der Befragung des Studienberechtigtenjahrgangs 2006 üblich – in einer ersten Welle bereits ein halbes Jahr vor Schulabschluss die Ausbildungs- und Studienabsichten in den Blick genommen. Erstmals wurde in einer zweiten Befragung ein Kalendarium zum Einsatz gebracht, mit dem monatsweise rückwirkend bis zum Jahr 2011 bis zu vier Tätigkeiten erfasst werden konnten. Wie vorangegangene Befragungen von Studienberechtigten fand auch die erste Befragung der Studienberechtigten des Jahres 2015, wenn möglich, im Klassenverbund und während des Unterrichts statt. Die Kohorte 2015 ist zum einen stark dadurch geprägt, dass es zwar zwischen 2007-2016 zu doppelten Abiturjahrgängen kam, allerdings der Jahrgang 2015 nicht direkt davon betroffen war. In vielen Bundesländern wurde eine Verkürzung der Schulzeit an allgemeinbildenden Gymnasien von insgesamt neun Jahren auf acht Jahre eingeführt, die ihren

---

<sup>9</sup> vgl. dazu: <https://tu-dresden.de/zqa/forschung/Forschungsprojekte/saechsische-abiturientenstudie> [Letzter Zugriff am 20.04.2021]

<sup>10</sup> vgl. dazu: <https://uni-tuebingen.de/fakultaeten/wirtschafts-und-sozialwissenschaftliche-fakultaet/faecher/fachbereich-sozialwissenschaften/hector-institut-fuer-empirische-bildungsforschung/forschung/aktuelle-studien/tosca/> [Letzter Zugriff am 20.04.2021]

<sup>11</sup> vgl. dazu: <http://www.best-up.eu/bestup.asp> [Letzter Zugriff am 20.04.2021]

Ausdruck beim Studienberechtigtenjahrgang darin findet, dass 88 Prozent der Studienberechtigten von allgemeinbildenden Gymnasien ihre Hochschulreife bereits nach acht Jahren erworben hatten (Schneider et al., 2017, S. 7). Zum anderen zeichnet sich die Kohorte 2015 dadurch aus, dass die im Rahmen des Bologna-Prozesses angestoßene Umstellung der traditionellen Studienabschlüsse (Diplom, Magister) auf die neue gestufte Studienstruktur mit den Abschlüssen Bachelor und Master, weitestgehend abgeschlossen ist. Im Wintersemester 2015/2016 führten 91 Prozent aller Studiengänge zu diesen neuen Studienabschlüssen (Hochschulrektorenkonferenz, 2015). Die Studienberechtigten hatten die Wahl zwischen 9.584 grundständigen Studiengängen (Hochschulrektorenkonferenz, 2015), wovon 53 Prozent zulassungsfrei zugänglich waren, und 327 anerkannten Ausbildungsberufen (Bundesinstitut für Berufsbildung, 2016). Auch das Angebot an dualen Studiengängen, in denen ein Hochschulstudium mit einer Berufsausbildung bzw. berufspraktischen Elementen kombiniert werden kann, ist mit 1.500 Studiengängen enorm gewachsen (Bundesinstitut für Berufsbildung, 2015). Dieser Zunahme an dualen Hochschulangeboten wurde mit dem Studienberechtigtenpanel 2015 erstmals Rechnung getragen, indem detailliert nach den Absichten diesbezüglich gefragt wurde und auch die Art des dualen Studiums erhoben wurde. Ein weiterer Schwerpunkt liegt auf der geographischen Mobilitätsbereitschaft und dem tatsächlichen Mobilitätsverhalten der (angehenden) studienberechtigten Schulabsolvent(inn)en von der Schule, in Ausbildung, Studium und Beruf. Diese Thematik hat in den letzten Jahren zunehmend an Bedeutung gewonnen, allerdings finden sich bisher nur wenige Studien, die wissenschaftlich abgesicherte Erkenntnisse über den Entstehungsprozess und die Wirkung unterschiedlicher Kontexte liefern. Schließlich lag neben diesen regionalen Kontexten ein weiterer Schwerpunkt auf der Erhebung von sozialen Kontexten bei der Studien- und Berufsausbildungswahl. Dem wurde durch die Integration neuer Befragungsinstrumente Rechnung getragen.

## 2 Erhebungsinstrumente

Für die Erhebung des Studienberechtigtenpanels 2015 wurde als Erhebungsinstrument jeweils ein standardisierter Papierfragebogen in deutscher Sprache eingesetzt.<sup>12</sup> Kapitel 2.1 stellt die zentralen Inhalte der zwei Erhebungsinstrumente vor. Kapitel 2.2 beschreibt die zur Prüfung und Verbesserung durchgeführten Pretests.

### 2.1 Inhalte der Erhebungsinstrumente

**[Charakteristika der Studienreihe]** Im Fokus des Studienberechtigtenpanels 2015 steht, wie bei den übrigen Kohorten der Studienreihe, die Beschreibung und Erklärung von Bildungsentscheidungen. Der Inhalt der Instrumente zur Erhebung des Bildungshandelns sowie dessen Erklärung orientiert sich – analog zu einer Klassifikation von Mertens (1976) – an dem zu erwartenden chronologischen Auftreten der unterschiedlichen Lebensphasen von der *nachschulischen Übergangsphase* (1) über die *Qualifizierungsphase* (2) bis zu der *Berufseintrittsphase* (3). Neben der Erfassung von Informationen zu Bildungsabsichten und -entscheidungen, werden in allen Erhebungsinstrumenten auch Einstellungen und Überzeugungen erfragt, um die Bildungsentscheidungen theoretisch fundiert erklären zu können.

Die Befragung der Studienberechtigten *vor* Schulabschluss<sup>13</sup> und Erwerb der Hochschulzugangsberechtigung (1. Welle) fokussiert auf die Ausbildungs- und Studienabsichten. Zu diesem Zweck erfolgt die Erfassung von Persönlichkeitseigenschaften, von Überzeugungen bezüglich Schulzeit, zum Informationsverhalten über Studien- und Ausbildungsmöglichkeiten, zu unterstützenden Personen bei der Studien- und Berufsausbildungswahl, zu potentiellen Zugängen zu Ressourcen verschiedener Dimensionen, zu geographischer Mobilitätsbereitschaft, zu Zukunftsplänen für die Zeit nach dem Schulabschluss und der diesbezüglichen Entscheidungsfindung.

In der ersten nachschulischen Befragung (2. Welle), ein halbes Jahr *nach* Erwerb der Hochschulzugangsberechtigung, rücken die Ausbildungs- und Studienentscheidungen in den Fokus. Dabei werden im Wesentlichen weitere Informationen zur Schulzeit (erstmalig im Studienberechtigtenpanel in Form eines monatsgenauen Kalendariums für die Jahre 2011-2015) sowie zur Person, die aktuell ausgeübte Tätigkeit, detaillierte Informationen zu den (ggf. beabsichtigten) Bildungsschritten, zu Motiven des nachschulischen Werdeganges, Gründe für eine verzögerte Aufnahme der nachschulischen Qualifizierung, zu potentiellen Zugängen zu Ressourcen verschiedener Dimensionen, zu Berufsaussichten sowie allgemein die Berufs- und Lebensziele erfasst.

**[Spezifika des Studienberechtigtenpanels 2015]** Im Vergleich zu vorangegangenen ersten Befragungen des Studienberechtigtenpanels wurde der Fragebogen der ersten Befragungswelle grundlegend überarbeitet und der Umfang um zwei DIN A4 Seiten (auf insgesamt acht Seiten) erweitert. Mit Blick auf die thematischen Schwerpunkte des Studienberechtigtenpanels 2015 wurden in der ersten Befragung einige bereits in vorherigen Studienberechtigtenbefragungen

---

<sup>12</sup> Die Fragebögen der beiden Wellen können von der Website des FDZ heruntergeladen werden. Ebenso stehen dort Flussdiagramme der Filterführung der zwei Fragebögen zur Verfügung.

<sup>13</sup> Eine Befragung *vor* Schulabschluss fand erstmalig in der Befragung des Studienberechtigtenjahrgangs 2005 statt und ist seitdem Standard des Studienberechtigtenpanels. Die vorherigen Kohorten (1976 bis 2004) wurden erst *nach* Schulabschluss kontaktiert und befragt (vgl. Kapitel 1).

eingesetzte Befragungsinstrumente aufgegriffen und darüber hinaus neue Messinstrumente integriert.

Eine schon vor dem Schulabschluss bestehende Absicht, ein *duales Studium* absolvieren zu wollen, wurde durch eine Ergänzung des Kerninstrumentes zur Studienabsicht (Frage 18 in Welle 1) erfasst. Die neu entwickelte Frage 20 (in Welle 1) spezifiziert ferner die Abfolge bzw. Parallelität der geplanten Doppelqualifizierung und ermöglicht damit genauere Aussagen zu bereits vor dem Schulabschluss geplanten Doppelqualifizierungen, die in vorherigen Studienberechtigtenbefragungen nicht eindeutig ausgewiesen werden konnten.

Außerdem wurde das bisher bereits in den DZHW-Studienanfängeruntersuchungen eingesetzte Instrumentarium zu Motiven bei der Wahl der Hochschule in modifizierter und erweiterter Form aufgenommen, um Aussagen darüber treffen zu können, welche Motive für die Wahl des künftigen Studien-/Ausbildungsortes bei angehenden Studienberechtigten eine Rolle spielen.

Um den *sozialen Kontext der Befragten bei der Studien- und Ausbildungsentscheidung* näher zu beleuchten, wurde den Befragten in dieser Kohorte zusätzlich die Möglichkeit gegeben, den Unterstützerkreis bei der Planung des nachschulischen Werdeganges und deren Beurteilung (Frage 29 in Welle 1) mit einer offenen Angabe von relevanten Personen zu ergänzen.

Neben diesen Schwerpunktthemen konnten in der ersten Befragung der Studienberechtigten des Jahres 2015 weitere Themenbereiche aufgenommen und vertieft werden. Von den angehenden Studienberechtigten mit Berufsausbildungsabsicht wurde erstmalig der angestrebte Ausbildungsberuf und das Vorliegen eines Ausbildungsvertrages erfasst (Frage 15 in Welle 1). Für angehende Studienberechtigte mit Studienabsicht wurde zusätzlich die Hochschule erhoben, an der sie studieren möchten (Frage 19 in Welle 1). Darüber hinaus wurde erstmals zur besseren Differenzierung der sozialen Herkunft der Studienberechtigten erfasst, ob ein oder beide Eltern teil(e) eine abgeschlossene Promotion vorweisen (Frage 45 in Welle 1).

Mit dem Messinstrument (Frage 32 in Welle 1), welches die Kurzskala zur Erfassung der allgemeinen Selbstwirksamkeit (ASKU [Item 1-3]; Beierlein, Kovaleva, Kemper & Rammstedt, 2012), ein Item zur Erfassung von nonkonformen Verhalten [Item 4] sowie die getestete Skala zur Erhebung von Impulsivem Verhalten (I-8 [Item 6-12]; (Kovaleva, Beierlein, Kemper & Rammstedt, 2012) kombiniert, wurden Persönlichkeitseigenschaften erhoben, die Einfluss auf die Bewältigung der Ausbildungs-/Studienentscheidung von Studienberechtigten haben, so dass die dahinter liegenden Mechanismen noch besser als bisher erklärt werden können. Des Weiteren wurde in der ersten Befragung zusätzlich zum bereits etablierten Befragungsinstrument, wie hoch die Chancen eingeschätzt werden, ein Hochschulstudium erfolgreich zu bewältigen (Frage 24 in Welle 1) auch danach gefragt, wie hoch die Chancen eingeschätzt werden, eine Berufsausbildung erfolgreich zu bewältigen (Frage 23 in Welle 1).

In der zweiten Befragungswelle des Studienberechtigtenpanels 2015 wurden im Unterschied zu vorangegangenen Studienberechtigtenbefragungen erstmals die bisherigen Bildungswege der Studienberechtigten von Januar 2011 bis Dezember 2015 mittels eines Kalendariums monatsweise genau erfragt, um ein detailliertes Bild der sehr unterschiedlichen Wege zur Hochschulreife geben zu können (Frage 11 in Welle 2). Damit wurde erhoben, welche verschiedenen Arten von Schulen von den Studienberechtigten besucht worden sind. Von besonderem Interesse sind dabei Studienberechtigte, die über einen eher untypischen Bildungsweg das (Fach-)Abitur erlangt haben (Realschulabschluss mit anschließendem Besuch eines Gymnasiums etc.). Außerdem können mit dem Instrumentarium u. a. Auslandsaufenthalte, Praktika, Jobs etc. während der letzten Schuljahre erhoben werden. Ferner wurde mit Hilfe dieses Instrumentariums auch herausgefunden

den, welche Tätigkeiten die Studienberechtigten zwischen Erwerb der Hochschulreife und der Aufnahme einer Berufsausbildung/eines Studium ausgeübt haben. Des Weiteren können die Angaben im Kalendarium mit dazu beitragen die Datenqualität insgesamt zu erhöhen, da ein umfassenderes Bild des bisherigen Bildungsweges (einschl. ausgewählter Tätigkeiten während der Schulzeit) von den Studienberechtigten vorliegt.

Darüber hinaus wurden zwei Instrumente integriert, die stärker als bisher Aufschluss über den genauen Entscheidungsprozess der Studien- und Berufsausbildungswahl geben. Dabei handelt es sich zum einen um ein Instrument, womit die Reihenfolge von zu treffenden Teilentscheidungen auf einem Zeitstrahl von den Studienberechtigten eingezeichnet werden sollen (Frage 12 in Welle 2). Zum anderen werden die Studienberechtigten gebeten, rückblickend die Entscheidungsfindung zu bewerten (Frage 1 in Welle 2).

Im Abschnitt mit den Fragen zum nachschulischen Werdegang wurde erstmals eine Frage zur Art des dualen Studiums ergänzt, um zwischen den verschiedenen Arten eines dualen Studiums differenzieren zu können (Frage 23 in Welle 2). Des Weiteren wurde die Möglichkeit gegeben, bei einem ausbildungsintegrierten dualen Studium die Berufsausbildung zu nennen. Außerdem wurde in die Frage zur Einschätzung von Berufsaussichten erstmals zusätzlich zur Einschätzung der Berufsaussichten für Absolventen eines Studiums allgemein, Einschätzung der Berufsaussichten für Absolventen eines beruflichen Ausbildungsweges ohne Studium allgemein sowie der Einschätzung der persönlichen Berufsaussichten, auch mit aufgenommen, wie die allgemeinen Berufsaussichten für Absolventen eines dualen Studiums eingeschätzt werden (Frage 34 in Welle 2).

Neu gestellt wurde außerdem eine Frage, mit der gemessen werden soll, inwieweit die Studierenden/Auszubildenden in ihr(e) Berufsausbildung/Studium integriert sind (Frage 32 in Welle 2).

Darüber hinaus wurde das Kerninstrument zur Erfassung der aufgenommenen nachschulischen Qualifizierung (Frage 19 in Welle 1) erweitert: Im Block a) zum gewählten Studium wurde die Frage ergänzt, ob es sich bei der angegebenen Hochschule um die Wunschhochschule handelt, im Block b) zur gewählten Berufsausbildung wurde zusätzlich nach der PLZ der beruflichen Schule und ggf. der PLZ des Ausbildungsbetriebes gefragt und im Block c) bei der gewählten Berufstätigkeit wurde ebenfalls nach der PLZ des Arbeitsortes gefragt. Ferner wurde erstmalig von den Studienberechtigten erhoben, welche Rolle Zulassungsbeschränkungen bei der Wahl der Hochschule gespielt haben (Frage 24 in Welle 2).

### 2.2 Pretest

**[Ziel und Verfahren]** Die Erhebungsinstrumente der zwei Befragungswellen 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 2015 gleich perzipiert werden würden wie von früheren Jahrgängen.<sup>14</sup> Zweitens sollte für die in beiden Wellen des Studienberechtigtenpanels 2015 neu eingesetzten Messinstrumente deren Verständlichkeit, Beantwortbarkeit, theoretische Aussagekraft, Reliabilität und Validität

---

<sup>14</sup> Die Erhebungsinstrumente des Studienberechtigtenpanels 2015 basieren im Wesentlichen auf dem für die Untersuchungsreihe etablierten Design und den in den Befragungen der Studienberechtigten 2005, 2006, 2008, 2010 und 2012 bewährten schriftlichen Fragebögen.

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 in beiden Wellen zunächst im Rahmen von *Expertenbewertungen* (vgl. Häder, 2015, S. 406–407) begutachtet. Neben diesen Bewertungen fanden in den einzelnen Wellen zielgerichtete „Entwicklungs- und Abschluss-Pretests“ (Schnell, Hill & Esser, 2011, S. 341)<sup>15</sup> statt. Zum einen wurden jeweils in frühen Phasen der Fragebogenentwicklung qualitative Interviews mit einem Rohentwurf des Fragebogens bzw. mit einzelnen Fragen durchgeführt. Dabei wurden auch bewährte qualitative Techniken, wie bspw. „frame of reference probing“, „paraphrasing“ und „think aloud“, eingesetzt (vgl. Schnell et al., 2011, S. 342). Zum anderen wurden jeweils gegen Ende der Fragebogenentwicklung sogenannte *Pretests im Feld* (vgl. Häder, 2015, S. 396), also unter möglichst ähnlichen Bedingungen wie in den tatsächlichen Befragungen, durchgeführt (vgl. zu diesem Vorgehen auch das „Zwei-Phasen-Pretesting“ von Prüfer & Rexroth, 2000).

**[Proband(inn)en]** In beiden Befragungswellen wurden zum einen etwa zehn Mitarbeiter(innen) des DZHW als Expert(inn)en aus dem Bereich der Hochschulforschung einbezogen sowie weitere drei Expert(inn)en außerhalb des DZHW. Zum anderen wurden externe Testpersonen rekrutiert, die im gleichen Alter beziehungsweise in einer ähnlichen Lebenssituation wie die Befragten waren.

**[Durchführung]** Der Standard-Pretest der ersten Befragungswelle fand zum Ende des Schuljahres 2013/2014, also ein halbes Jahr vor dem jeweiligen Feldstart, statt. Er wurde an einer berufsbildenden Schule mit verschiedenen Schulformen durchgeführt, um die Auswahl der Befragten strukturell möglichst nah an der endgültigen Stichprobe zu halten und um insbesondere Personen einzubeziehen, die eher ungewöhnliche Bildungswege haben. Damit sollte gewährleistet werden, dass auch diese Befragten die Fragebogen ansprechend finden. Am Pretest nahmen insgesamt 52 Schüler(innen) aus drei Klassen teil. Der Fragebogen wurde während des Unterrichts im Beisein von DZHW-Mitarbeiterinnen ausgefüllt. Im Anschluss erörterten diese mit den Schüler(inne)n das Verständnis der Fragen und Antwortvorgaben sowie Schwierigkeiten mit dem Erhebungsinstrument.

Für die zweite Welle wurde im Sommer 2015 ein „Zwei-Phasen-Pretesting“ (vgl. Prüfer & Rexroth, 2000) durchgeführt. Zuerst wurden einzelne Fragen des vorläufigen Erhebungsinstrumentes mit neun Personen kognitiv gepretestet (vgl. Lenzner, Neuert & Otto, 2015). Im Zentrum standen vor allem die für die Befragung neu konzipierten Instrumente (u. a. Frage 11, Frage 12 & Frage 23). Nicht alle getesteten Instrumente konnten später im endgültigen Erhebungsinstrument verwendet werden. Im Anschluss daran wurde ein Standard-Pretest durchgeführt, indem Studienberechtigten des Jahrgangs 2015<sup>16</sup> das vorläufige Erhebungsinstrument postalisch zugesandt wurde. Das Ausfüllen der Fragebögen war mit der Bitte verbunden, Ausfüllprobleme, Filterfehler, Unverständlichkeiten, Ausfülldauer sowie unvollständige Antwortvorgaben zu notieren und (portofrei) an das DZHW zurückzusenden.

<sup>15</sup> „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.

<sup>16</sup> Diese Pretestteilnehmer(innen) waren nicht Teil der Stichprobe, hatten aber an der ersten Befragungswelle teilgenommen und damit ihre Adresse angegeben (vgl. Kap.6.3).



### 3 Grundgesamtheit und Stichprobenverfahren

**[Grundgesamtheit]** Die Grundgesamtheit der Befragung des Studienberechtigtenjahrgangs 2015 umfasst alle Schülerinnen und Schüler der Bundesrepublik Deutschland, die im Schuljahr 2014/2015 an allgemeinbildenden oder berufsbildenden Schulen die allgemeine oder fachgebundene Hochschulreife, die Fachhochschulreife, die fach- oder landesgebundene Fachhochschulreife erwarben.

**[Stichprobenverfahren]** Aufgrund fehlender oder nicht zugänglicher Listen von Studienberechtigten auf Individualebene war es nicht möglich, eine einfache Zufallsstichprobe zu ziehen. So 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. Schulzweige<sup>17</sup> stellen dabei die „primary sampling units“ (Klumpen), die Studienberechtigten des Jahrgangs 2015 die „secondary sampling units“ innerhalb dieser Klumpen dar.<sup>18</sup> 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 2011/2012 laut amtlicher Statistik<sup>19</sup> orientierten und zum anderen die Rücklaufquoten der letzten Erhebung, dem Studienberechtigtenpanel 2012, berücksichtigten. Im Zuge dessen erhielten Studienberechtigte aller Schulzweige aus Bayern, Bremen, Mecklenburg-Vorpommern, Saarland, Sachsen-Anhalt und Thüringen überproportionale Auswahlwahrscheinlichkeiten, um in diesen Gruppen ausreichend große Fallzahlen zu generieren. Darüber hinaus wurde aus diesem Grund jeweils für die Schulformen Abendgymnasium, Abendkolleg, (Höhere) Berufsfachschulen und Fachschulen pro Bundesland (sofern diese Schulform vorhanden ist) eine Mindestanzahl von 50 Schüler(innen) festgelegt.

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 Schülerzahlen zurück, sodass ein zweiter Auswahlschritt durchgeführt werden konnte, der die Erfüllung der Ziehungsquoten auf Individualebene zum Ziel hatte. In diesem zweiten Schritt wurden in jeder Schicht *sukzessive* so viele Schulen/Schulzweige zufällig gezogen, dass mindestens 97 der schichtspezifischen Ziehungsquote für die Studienberechtigten erreicht wurde.<sup>20</sup> Hiernach wurden die gezogenen Schulen (oder Schulzweige an diesen) komplett erhoben.

<sup>17</sup> Manche berufsbildenden Schulen bestehen aus verschiedenen Schulzweigen, von denen jedoch nur jene für die Stichprobe relevant waren, die zum Erwerb der Hochschulzugangsberechtigung führten. Daher wurden in manchen Fällen nur bestimmte Schulzweige berücksichtigt.

<sup>18</sup> Beispiel für den Personenkreis eines Klumpens: Alle Studienberechtigten des Prüfungsjahrgangs 2015 des allgemeinbildenden Gymnasiums XY in Bundesland YZ.

<sup>19</sup> Es handelt sich um die zum Zeitpunkt der Stichprobenziehung aktuellsten veröffentlichten Daten des Statistischen Bundesamtes (2014a); (2014b). Genutzt werden konnten Daten der amtlichen Statistik über die Anzahl der Absolvent(innen) allgemeinbildender und berufsbildender Schulen: Allgemeinbildende Schulen - Schuljahr 2012/13 Fachserie 11 Reihe 1; S. 287-295 und Berufliche Schulen, Schuljahr 2012/13 Fachserie 11 Reihe 2; S. 457-488. Die Daten der Schulabsolvent(innen) beider Publikationen beziehen sich jeweils auf das Schuljahr 2011/12. Außerdem musste berücksichtigt werden, dass im Schuljahr 2011/2012 Baden-Württemberg, Berlin, Brandenburg und Bremen doppelte Abiturient(innen)jahrgänge hatte, was auf das Schuljahr 2014/2015 nicht zutraf.

<sup>20</sup> Die Prüfung, ob die Ziehungsquote in einer Schicht erreicht wurde, erfolgte nach jedem Ziehungsschritt. Erfüllte beispielsweise die kumulierte Schülerzahl von fünf Schulen die Ziehungsquote für die Schülerzahl in der spezifischen Schicht, so wurde nach der Auswahl der fünften Schule keine weitere Schule gezogen.

## 4 Durchführung der Erhebungen

**[Kontaktaufnahme und Adresspflege]** Im Vorfeld der ersten Befragungswelle trat das DZHW, wie schon bei vorangegangenen Studienberechtigtenbefragungen, zunächst mit der Bitte an die Kommission für Statistik der Kultusministerkonferenz (KMK) heran, die geplante Untersuchung im Vorfeld zu unterstützen, in dem den 16 Kultusministerien mit einem Schreiben die Bedeutung der DZHW-Studienberechtigtenbefragung für Bund und Länder verdeutlicht wurde.

Im Anschluss daran wurden die zuständigen Behörden und Ministerien der 16 Bundesländer kontaktiert und um die benötigten Genehmigungs- bzw. Zustimmungsschreiben<sup>21</sup> gebeten sowie um ein entsprechendes Empfehlungsschreiben<sup>22</sup>. Es wurde jede Landesbehörde gesondert gebeten *erstens*, die in die Stichprobe einzubeziehenden Schulen anschreiben und um Mitarbeit bitten zu dürfen und, *zweitens*, die DZHW-Studienberechtigtenbefragung mit einem entsprechenden Empfehlungsschreiben zu unterstützen. Hierbei sollten die Länderbehörden erneut auf die Bedeutung der DZHW-Studienberechtigtenbefragung hinweisen, die Schulen um ihre Unterstützung bitten und soweit möglich das Ausfüllen der Befragungsunterlagen im Rahmen einer Schulstunde befürworten.

---

<sup>21</sup> In Nordrhein-Westfalen sowie in Hessen wurde keine Genehmigung zur Durchführung der Befragung benötigt. Es liegt allein in der Verantwortung der Schule, ob diese mit ihren Schüler(inne)n teilnimmt. Für alle anderen Bundesländer existiert ein Genehmigungsschreiben.

<sup>22</sup> Ein explizites Empfehlungsschreiben an die Schulen steht für neun der insgesamt 16 Bundesländer zur Verfügung.

Als Nächstes wurden die Schulleiter(innen) der per Zufallsauswahl gezogenen Schulen vom DZHW angeschrieben und um ihre Teilnahme gebeten.<sup>23 24</sup> Zudem teilte das DZHW den Schulen die Kriterien mit, anhand derer sie die Zielpersonen für das Studienberechtigtenpanel 2015 identifizieren sollten (vgl. Kapitel 3).<sup>25</sup> Da die Schulen die Kontaktdaten ihrer Schüler(innen) aus Datenschutzgründen nicht herausgeben durften, teilten sie dem DZHW lediglich die jeweilige Schülerzahl (ggf. für einzelne Schulzweige) sowie eine(n) Ansprechpartner(in) auf der beigelegten Antwortpostkarte mit. Einige Wochen nach dem ersten Anschreiben, erhielten alle Schulleitungen, von denen bis zu diesem Zeitpunkt noch keine Reaktion vorlag, ein Erinnerungsschreiben, in dem sie ein weiteres Mal um die Mitwirkung an der Studienberechtigtenbefragung gebeten wurden.<sup>26</sup> Darüber hinaus mussten in einigen Bundesländern Schulen bzw. Schulzweige zusätzlich auch telefonisch kontaktiert werden, da auch auf die schriftliche Erinnerung keine Reaktion erfolgt war. An die auf der Antwortpostkarte angegebenen Ansprechpersonen wurden im Dezember 2014 die Pakete mit der passenden Anzahl an Erhebungsunterlagen für die erste Befragungswelle postalisch gesendet.<sup>27</sup> Aufgrund des direkten Feldzugangs erfolgte die Befragung über die Schulen in den Klassenverbänden und, wenn möglich, während des Unterrichts.<sup>28</sup>

Um die weiterhin teilnahmebereiten Personen in den Folgewellen direkt durch das DZHW kontaktieren zu können, wurden im Fragebogen der ersten Welle deren Adressdaten erfragt. Beim Eingang eines Fragebogens im DZHW wurde dieser neben der eingedruckten Identifikationsnummer mit einer laufenden Nummer versehen. Außerdem wurde zu diesem Zeitpunkt der Abschnitt mit der Adressangabe vom Fragebogen abgetrennt und mit derselben laufenden Nummer wie der Fragebogen versehen. Im Weiteren wurde eine Referenzliste mit Identifikations- und laufender Nummer sowie den zugehörigen Adressen erstellt.<sup>30</sup> In der zweiten Welle wurden alle Personen kontaktiert, die ihre Adresse in der ersten Welle angegeben hatten. Dafür wurden die Identifikationsnummer und die Adresse in die Fragebögen gedruckt, sodass die Befragungsunterlagen direkt an die Postadressen der Teilnehmer(innen) versandt werden konnten.<sup>31</sup> In dieser Erhebungswelle erfolgte vor dem Versand der Erhebungsunterlagen ein Adressabgleich mit der Deutschen Post, um auch Personen zu berücksichtigen, die in der Zwischenzeit umgezogen wa-

---

<sup>23</sup> Im Rahmen dieses Erstkontaktes mit den Schulen wurde erstmalig auch ein Projekt-Flyer übermittelt, der wichtige Informationen zum Forschungsprojekt und dessen Ablauf enthielt.

<sup>24</sup> In Rheinland-Pfalz und Bremen mussten zusätzliche Anschreiben an die Elternbeiräte angefertigt und versendet werden. In Rheinland-Pfalz war es außerdem notwendig, dazu auch die Schülervertretungen über die Durchführung der Befragung zu informieren.

<sup>25</sup> 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 Schulabsolvent(inn)en 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.

<sup>26</sup> Im Saarland erfolgte auf Grund der bereits weit fortgeschrittenen Feldphase in den anderen Bundesländern keine schriftliche Erinnerung, sondern es wurde Ende März 2015 direkt telefonisch erinnert.

<sup>27</sup> Dem Schulanschreiben waren eine Anleitung zur Durchführung der Erhebung, eine frankierte Antwortpostkarte zur Anforderung einer schulspezifischen Auswertung, der Hinweis auf Teilnahmeprämien (Ausnahmen: In den Bundesländern Bayern, Saarland, Sachsen und Thüringen waren Verlosungen nicht gestattet.) sowie für die Befragungsteilnehmer(innen) Kugelschreiber und für die Ansprechpersonen Kunststoffbecher (beides jeweils mit Projektaufdruck) beigelegt.

<sup>28</sup> In den Bundesländern Mecklenburg-Vorpommern und Saarland wurde das Verwenden von Unterrichtszeit durch die jeweilige Kultusbehörde nicht gestattet.

<sup>29</sup> Dass komplette Schulen bzw. Schulzweige trotz vorheriger Zusage nicht teilnahmen, stellt eine Ausnahme dar.

<sup>30</sup> Zur Gewährleistung des Datenschutzes wurde der Adressabschnitt vom Fragebogen abgetrennt und die Referenzliste getrennt von den Befragungsdaten auf einem geschützten Server gespeichert.

<sup>31</sup> In den Fragebögen konnten Teilnehmer(innen) in einem Adressabschnitt ihre Adressdaten korrigieren.

ren. Zudem wurden durch das DZHW Adressen nachrecherchiert, wenn Erhebungsunterlagen unzustellbar waren.<sup>32</sup>

**[Erhebungsunterlagen]** Die Erhebungsunterlagen bestanden in beiden Befragungswellen pro zu befragender Person aus einem Anschreiben (inkl. Datenschutzinformationen)<sup>33</sup>, dem Fragebogen sowie einem Incentive (1. Welle: Kugelschreiber passend zum Fragebogen gestaltet mit Aufdruck „Schule! Und dann? DZHW-Studienberechtigtenpanel 2015“; 2. Welle: Kleine Gummibärentüte mit Aufschrift „Danke“). Eine portofreie Rücksendung des ausgefüllten Fragebogens wurde über die gesammelte Rücksendung in kleinen Paketen durch die Schulen (1. Welle) bzw. über ein an das DZHW adressiertes Antwortkuvert gewährleistet.

**[Feldphase]** Der Erhebungszeitraum der ersten Befragungswelle erstreckte sich hauptsächlich von Dezember 2014 bis Ende April 2015.<sup>34</sup> Aufgrund des angewendeten Kontaktverfahrens über die Schulen konnte das DZHW keinen direkten Einfluss auf den genauen Befragungszeitpunkt nehmen. In der ersten Befragungswelle begann der Versand der Pakete mit den Befragungsunterlagen an die Schulen Anfang Dezember 2014.<sup>35</sup> Alle Schulen bzw. Schulzweige, von denen bis Mitte März 2015 keinerlei Rücklauf vorlag, wurden postalisch kontaktiert, um in Erfahrung zu bringen, ob die Befragungsunterlagen verteilt worden sind.

Der Erhebungszeitraum der zweiten Welle erstreckte sich hauptsächlich von Dezember 2015 bis April 2016.<sup>36</sup> Durch den nun im DZHW vorliegenden Adressbestand konnten konkrete Versandzeitpunkte für die Erhebungsunterlagen festgelegt werden, die jeweils als Reaktion auf ein Abflachen des Fragebogenrücklaufes eingesetzt wurden. In der zweiten Befragungswelle fanden insgesamt zwei Erinnerungsaktionen statt. Die erste Erinnerung wurde Anfang Februar 2016 durchgeführt, die zweite Erinnerung Mitte März 2016. Der zweiten Erinnerung wurde nochmals der Fragebogen beigelegt, um die Hürde zur Teilnahme, so gering wie möglich zu halten.

Die Erinnerungsschreiben wurden in der zweiten Befragungswelle gezielt nur an diejenigen Personen verschickt, die sich bis zum Zeitpunkt der jeweiligen Erinnerung noch nicht an der Erhebung beteiligt hatten.

**[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 verantwortlicher Ansprechpartner(innen)) darin, Unterrichtszeit zum Ausfüllen bereitzustellen, die Fragebögen am Ende der Unterrichtsstunden einzusammeln sowie gegebenenfalls bei einzelnen Klassen die Rückgabe der Fragebögen zu unterstützen. Als Gegenleistung wurden den Schulen schulspezifische Auswertungen angeboten. Auch wurden für eine hohe Teilnahmebereitschaft Preise für die Schulen (drei Mal 250 Euro) ausgelobt<sup>37</sup> und für die Ansprechpersonen an den Schulen ein Kunststoffbecher den Paketen beigelegt. Als nützlich

<sup>32</sup> Neben dem Adressaktualisierungsservice der Deutschen Post wurden auch Melderegisterauskünfte eingeholt.

<sup>33</sup> In den Bundesländern Bayern, Brandenburg, Nordrhein-Westfalen (nur Gymnasien), Saarland und Thüringen waren den Befragungsunterlagen der ersten Welle darüber hinaus Elternanschriften und Elterneinwilligungserklärungen beigelegt. In Sachsen waren diese bereits in den Unterlagen zur Vorabinformation enthalten.

<sup>34</sup> Die Feldzeit wurde so lange wie möglich – parallel zu den weiteren Projektschritten – ausgedehnt und jeder noch eintreffende Fragebogen bis September 2015 berücksichtigt.

<sup>35</sup> Die Genehmigung für die Befragung saarländischer Schulen lag erst Ende Februar 2015 vor, so dass die Rekrutierung der Schulen erst Anfang März 2015 und der Versand der Befragungsunterlagen in diesem Bundesland erst ab Mitte März 2015 erfolgen konnten.

<sup>36</sup> Bei der zweiten Befragung wurden nach Juli 2016 nur noch vereinzelt Fragebögen an das DZHW zurückgeschickt. Der letzte im Datensatz enthaltene Fragebogen ging im November 2016 ein.

<sup>37</sup> Ausnahmen bildeten die Bundesländer Bayern, Saarland, Sachsen und Thüringen, in denen eine Verlosung nicht gestattet worden war.

erwiesen sich zudem ein Hinweis auf die gesamtgesellschaftliche Bedeutung der Erhebung in den Anschreiben, ministerielle Aufforderungen die Befragung zu unterstützen sowie Anschreiben und die telefonische Kontaktierung bei Schulen, von denen noch keine Fragebögen eingegangen waren. Die auf die individuelle Motivation der Befragten zielenden rücklaufsteigernden Maßnahmen bestanden in einem Hinweis auf den gesamtgesellschaftlichen und altersgruppenspezifischen Nutzen im Anschreiben an die Befragten, einem zusammen mit der ersten Erinnerung an diese verschickten Flyer (in Welle 2), einer zweiten Erinnerung, der nochmals der Fragebogen beigelegt wurde, und einer Projektwebsite<sup>38</sup> (mit Informationen zum Projekt und daraus entstandenen Publikationen sowie der Möglichkeit der Kontaktaufnahme mittels Kontaktformular). Als materieller individueller Anreiz bestand in der ersten Welle für die Befragten die Möglichkeit zum Gewinn von einem *Notebook nach Wahl* im Wert von ca. 900 Euro sowie 20 *Einkaufsgutscheinen* im Wert von jeweils 50 Euro.<sup>39</sup> In der zweiten Welle wurden 25 *Amazon-Gutscheine* im Wert von jeweils 50 Euro, zwei *IKEA-Gutscheine* im Wert von jeweils 250 Euro und ein Apple iPhone 6S im Wert von 739 Euro verlost.

---

<sup>38</sup> Die Projektwebsite (<https://www.panel2015.de/>) wurde im Laufe der Erhebung für jede Welle aktualisiert.

<sup>39</sup> Ausnahmen bildeten die Bundesländer Bayern, Saarland, Sachsen und Thüringen, in denen eine Verlosung nicht gestattet worden war.

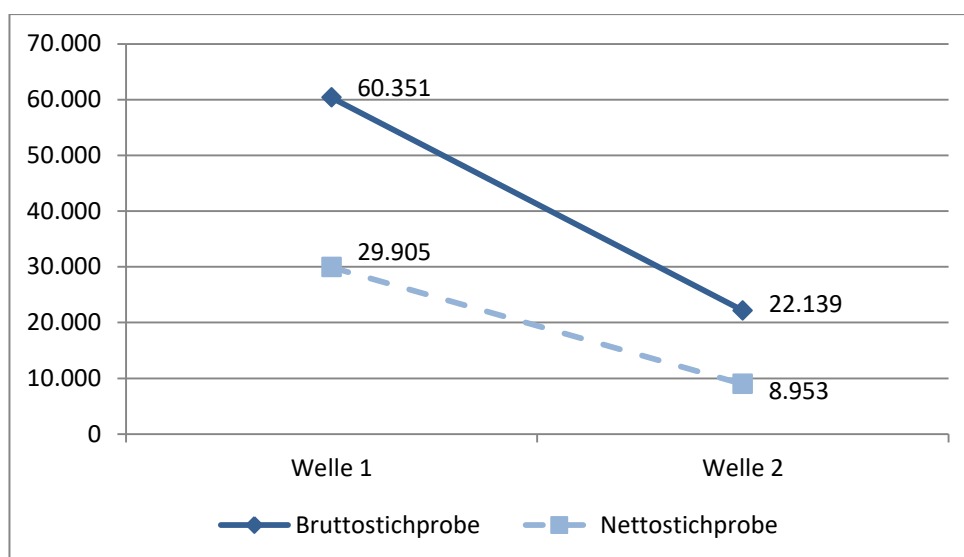
## 5 Rücklauf

**[Rücklauf]** Auf der Grundlage des Stichproben- und Erhebungsdesigns (siehe Kapitel 3) wurden für die Befragung der ersten Welle 1.660 Schulen angeschrieben. Von diesen Schulen meldeten 946 (57 Prozent) die Anzahl ihrer zu erwartenden Studienberechtigten zurück. Dies entsprach einer Zahl von 70.975 Schülerinnen und Schülern. Die daraus gezogene bereinigte Bruttostichprobe der ersten Befragungswelle umfasste letztlich 60.351 Studienberechtigte aus 789 Schulen bzw. Schulzweigen, die kontaktiert wurden (14 % der Grundgesamtheit vgl. Schneider et al., 2017, S. 9).<sup>40</sup> Die Rücklaufquote betrug 50 Prozent, sodass eine Nettostichprobe von 29.905 auswertbaren Fragebögen zu verzeichnen war. 22.139 Personen, also 74 Prozent der 29.905 Teilnehmer(innen) der ersten Welle, erklärten sich mit einer Kontaktierung für weitere Befragungen einverstanden (Schneider et al., 2017, S. 9).

In der zweiten Erhebung, welche auf postalischem Wege durchgeführt wurde, konnten folglich 22.139 Personen in die Bruttostichprobe einbezogen werden. Die Rücklaufquote der zweiten Befragung war im Vergleich zur ersten Welle mit 40 Prozent deutlich geringer, was zu einer Nettostichprobe von 8.953 Personen führte.

**[Panelausfälle]** Das Studienberechtigtenpanel 2015 ist von paneltypischen Ausfallprozessen betroffen.<sup>41</sup> Diese äußern sich in der grundsätzlichen Verweigerung zur Teilnahme an Folgebefragungen (keine Angabe von Adressangaben für die Kontaktierung in der zweiten Welle) oder in Form von Nichtteilnahmen nach einer (versuchten) Kontaktaufnahme in der zweiten Erhebungswelle. In Abbildung 3 und Tabelle 2 ist die beschriebene Entwicklung der Brutto- und Nettostichprobe im Zeitverlauf dargestellt.

**Abbildung 3: Entwicklung von Brutto- und Nettostichprobe im Zeitverlauf**



Anmerkung zur Brutto- und Nettostichprobe: Nur Personen mit Stichprobenberechtigung

<sup>40</sup> Im Laufe der Rückmeldung von den Schulen wurden auch Personen im Nachhinein als nicht stichprobenzugehörig eingestuft, weil sie im relevanten Prüfungsjahr 2015 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.

<sup>41</sup> Zu paneltypischen Ausfallprozessen vgl. Schnell et al. (2005), S. 233-234.

Die deutliche Diskrepanz zwischen den Rücklaufquoten der ersten Welle der Befragung von 2015 und der zweiten Welle kann teilweise durch die Form der Erhebung in der ersten Welle erklärt werden: Durch die Befragung im Klassenverbund wurden vermutlich in der ersten Welle auch Personen zur Teilnahme motiviert, die andernfalls nicht teilgenommen hätten. Diese fehlende grundsätzliche Teilnahmebereitschaft trug möglicherweise zu einer höheren Verweigerungsquote in der – auf postalischem Weg zugestellten – zweiten Welle bei und äußerte sich ebenfalls in der recht hohen Anzahl an Bögen ohne Adressrückmeldungen.

**Tabelle 2: Brutto-, Nettostichproben und Rücklaufquoten des DZHW-Studienberechtigtenpanels 2015 (Welle 1 und 2)**

|   | <b>Welle 1</b> | <b>Welle 2</b> |
|---|----------------|----------------|
| Bruttostichprobe  | 60.351         | 22.139         |
| Nettostichprobe   | 29.905         | 8.953          |
| Rücklaufquote   | 50 %           | 40 %           |
| Anteil der Nettostichprobe an der vorangegangenen Nettostichprobe   | --             | 30 %           |
| Anteil Nettostichprobe an der Bruttostichprobe t1                   | 50 %           | 15 %           |
| Anteil der Bruttostichprobe an der vorangegangenen Bruttostichprobe | --             | 37 %           |
| Anteil der Bruttostichprobe an der Bruttostichprobe t1              | --             | 37 %           |

## 6 Datenaufbereitung

Im Folgenden werden die verschiedenen Arbeitsschritte der Datenaufbereitung beschrieben. Diese erfolgten in den zwei Befragungswellen analog. Die in den Kapiteln 6.1 bis 6.3 beschriebenen Tätigkeiten wurden bereits durch das Primärforschungsprojekt durchgeführt. Die Generierung von Variablen (Kapitel 6.4) wurde sowohl durch das Primärforschungsprojekt als auch im Rahmen der Datenedition durch das FDZ-DZHW vorgenommen. Die in den Kapiteln 6.5 bis 6.7 dargestellten Tätigkeiten wurden durch das FDZ durchgeführt, teilweise aufbauend auf Vorarbeiten des Primärforschungsprojektes. Die im Rahmen der Datenedition vorgenommenen Aufbereitungsprozesse der Gewichtung und Anonymisierung werden in den beiden folgenden Kapiteln 7 und 8 gesondert erläutert.

### 6.1 Datenübertragung

Zur weiteren Verarbeitung wurden die Angaben der Befragten aus den Papierfragebögen auf Basis eines Codeplans in ein computerlesbares Format übertragen. Zuvor wurden noch auf den Papierfragebögen numerische Codierungen für einen Teil der offenen Angaben vermerkt (vgl. Kapitel 6.2) sowie manuelle Vorarbeiten zur Erleichterung der Datenübertragung vorgenommen (vgl. Kapitel 6.3).

**[Erstellung eines Codeplans]** Auf Basis des Fragebogens der Befragung wurde ein Codeplan 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.<sup>42</sup>

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

### 6.2 Codierung offener Angaben

Vor der Datenübertragung erfolgte eine Codierung für einen Teil der (halb-)offenen Angaben. Dabei wurden diesen anhand einer Codierliste numerische Codierungen zugeordnet. Je nach Variable wurden unterschiedliche Codierlisten verwendet. Es handelt sich um Klassifikationsschlüssel der amtlichen Statistik (z. B. Klassifikation der Berufe, Schlüsselverzeichnis der Studenten- und Prüfungsstatistik etc.) oder um bereits in vorherigen Studien eingesetzte Schlüssel. Für einige Variablen wurden neue Codierlisten auf Basis der in den Daten der Befragung des Studienberechtigtenjahrgangs 2015 vorkommenden Nennungen entwickelt. Für einige halboffene Fragen wurden keine neuen Variablen mit numerischen Codierungen erstellt, sondern die Nennungen nur – sofern möglich – den vorhandenen (geschlossenen) Antwortkategorien zugeordnet. Der Datensatz beinhaltet ausschließlich die codierten numerischen Variablen, die offenen Nennungen

<sup>42</sup> Die Daten wurden in einem einfachen spaltenorientierten Textformat ohne eine die Variablennamen enthaltene Kopfzeile erfasst. Der Codeplan musste daher festlegen, in welcher Reihenfolge die Daten zu erfassen waren, damit die zu einer Variable zugehörigen Daten in der richtigen Spalte eingetragen werden konnten.



selbst sind nicht im Datensatz enthalten. Die Ausprägungen der einzelnen Variablen sind im Datensatzreport sowie im Metadatensuchsystem<sup>43</sup> dokumentiert.

### 6.3 Datenprüfung und Datenbereinigung

**[Manuelle Vorarbeiten]** Bereits vor der Übertragung der Daten wurden auf den Papierfragebögen, um die Erfassbarkeit der Daten zu erleichtern, schwer lesbare Angaben oder Streichungen der Befragten verdeutlicht, Zahlenangaben rechtsbündig in die dafür vorgesehenen Kästchen eingetragen, verbale Angaben von Noten in Ziffern übersetzt (z. B. „gut“ = 2,0).

**[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 Fehler bei der Datenübertragung, zum anderen inkonsistente Angaben der Befragten identifiziert werden.

Zu diesem Zweck wurden die erfassten Fragebogen-Daten (sowie in der zweiten Befragungswelle auch Daten der ersten Befragungswelle) 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 Variable in dem für diese Variable definierten Wertebereich lag.
- *Prüfung der Einhaltung der Filterführung:* Es wurde auf der einen Seite geprüft, ob aufgrund der im Fragebogen vorgesehenen Filterführung für die jeweilige befragte Person Angaben zu erwarten gewesen wären (Vollständigkeitsprüfung), und auf der anderen Seite, ob für die jeweilige Person keine Angaben hätten erfolgen dürfen (Filterverstöße).
- *Prüfung von Merkmalskombinationen:* Es wurde die Konsistenz der Angaben geprüft. Dabei konnten auch komplexere Merkmalskombinationen innerhalb der einzelnen Befragungswellen überprüft werden. Zudem wurde hier auch die Konsistenz von Angaben zwischen den Befragungswellen 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. Ansonsten wurde versucht, den korrekten Wert anhand anderer Fragebogen-Angaben zu erschließen. Im Zweifelsfall wurde ein spezifischer Missingcode vergeben (vgl. Kapitel 6.7). Fehlerkorrekturen wurden dokumentiert<sup>44</sup> und von mindestens einer weiteren Person geprüft.

**[Löschung von Fällen]** In beiden Wellen wurden Fälle aus dem Datensatz entfernt, die nicht zur Grundgesamtheit bzw. zur regulären Stichprobe gehörten.<sup>45</sup>

<sup>43</sup> <https://metadata.fdz.dzhw.eu/>

<sup>44</sup> Die Dokumentation der Fehlerkorrekturen erfolgte handschriftlich auf den Papierfragebögen und ist daher nur teilweise anhand der Rohdatensätze rekonstruierbar.

<sup>45</sup> Dies kam beispielsweise vor, wenn Schulabsolvent(inn)en 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. Darüber hinaus wurde auch Schulen die Teilnahme ermöglicht, die nicht zur regulären Stichprobe gehörten. Diese wurden im Datensatz gesondert kenntlich gemacht, bei Datenanalysen ausgeschlossen und den Datennutzer(inne)n nicht zur Verfügung gestellt.

## 6.4 Generierung von Variablen

Neben den Variablen, die die codierten Antworten der Befragten enthalten, beinhaltet der Datensatz zum Studienberechtigtenjahrgang 2015 auch generierte Variablen. Dabei handelt es sich zum einen um Variablen mit numerischen Codierungen von ursprünglich offenen Nennungen (vgl. Kapitel 6.2). Zum anderen wurden Variablen des Datenschutzes wegen verändert (vgl. Kapitel 8). Die überwiegende Anzahl der generierten Variablen wurde bereits durch das Primärforschungsprojekt erstellt. Der Variablenname einer durch das FDZ-DZHW generierten Variable ist durch das Suffix „\_g#“ gekennzeichnet. Eine Übersicht aller für das Studienberechtigtenpanel 2015 generierten Variablen sowie eine detaillierte Dokumentation der einzelnen Variablen mit Angabe ihrer jeweiligen Ausprägungen und Berechnungsvorschriften findet sich im Datensatzreport sowie im Metadatensuchsystem.<sup>46</sup>

## 6.5 Erstellung der Datensätze

**[Zusammenführung der Wellen]** Die Daten der beiden Befragungswellen wurden bereits vom Primärforschungsprojekt zusammengeführt. Die Zuordnung der Daten erfolgte über die im Rahmen der Feldphase vergebene Identifikationsnummer der Befragten (vgl. Kapitel 4).

**[Erstellung von Personen- und Episodendatensatz]** Die so zusammengeführten Daten wurden in zwei getrennten Datensätzen abgelegt. Der Personendatensatz enthält den Großteil der Befragungsdaten sowie die zusätzlichen generierten Variablen. Pro befragte Person existiert eine Datenzeile (wide-Format). Die Reihenfolge der Variablen orientiert sich an der Reihenfolge der zugehörigen Fragen im Fragebogen.

Der Episodendatensatz enthält die Antworten aus dem Kalendarium der zweiten Welle (Frage 11). Für jede befragte Person werden eine oder mehrere Episoden gespeichert. Dabei ist eine Episode definiert als ein Zeitraum, in dem eine bestimmte Tätigkeitsart (z. B. Berufsausbildung, Duales Studium) ausgeübt wird bzw. ein konkreter Status (z. B. Arbeitslosigkeit) besteht. Für jede Episode einer Person existiert jeweils eine Datenzeile (long-Format), welche monatsgenau den Anfang- und Endzeitpunkt der Tätigkeit und die Episodennummer angibt. Die Episoden wurden fallweise sortiert, das heißt alle Episoden einer Person folgen direkt aufeinander.

Das Zusammenführen von Personendatensatz und Episodendatensatz wird über die Identifikationsnummer der Person (Variable: pid) ermöglicht.

**[Datenformat]** Der Datensatz wird standardmäßig im Stata-Format bereitgestellt. Möglichkeiten zur Nutzung weiterer Formate bzw. Analysesoftware sind der Website zu entnehmen (vgl. auch Abschnitt II).

## 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. Bei generierten Variablen, denen bestimmte Klassifikationen

<sup>46</sup> <https://metadata.fdz.dzhw.eu/>

zugrunde liegen, wurden für die Wertelabels die Bezeichnungen der Schlüssel der Klassifikation wortgetreu übernommen. Die Variablen- und Wertelabels sind als zweisprachige Labels (Deutsch und Englisch) im gleichen Datensatz hinterlegt.

**[Variablenbenennung]** Die Variablenbenennung wurde überwiegend vom Primärforschungsprojekt vorgenommen. Variablen, die in genau der gleichen Art und Weise auch in vorherigen Studienberechtigtenkohorten erhoben worden sind, wurden mit gleichem Variablennamen auch im Studienberechtigtenpanel 2015 benannt (z. B. Frage 1 in Welle 1 „schulneu“, Frage 2 in Welle 1 „reifeart“ etc.). Häufig wurde bei der Variablenbenennung erst ein einheitlicher Stamm (möglichst inhaltlich sprechend) für alle Variablen einer Frage erstellt (z. B. Frage 7 in Welle 1 „schwunt“, Frage 11 in Welle 1 „begab“) und dann die dazugehörigen Variablen durchnummeriert. Anstelle einer Nummerierung kommt es auch vor, dass neben einem einheitlichen Stamm (z. B. Frage 22 in Welle 1 „ort“) sprechende Bezeichnungen vergeben worden sind (z. B. ebd. „ortbilig“, „ortelvt“, „orthomna“, etc.).

## 6.7 Codierung fehlender Werte

Zur Codierung fehlender Werte wurde folgende Systematik durch das Primärforschungsprojekt erstellt und während der Datenaufbereitung durch das FDZ um weitere Werte ergänzt:

**Tabelle 3: Systematik für fehlende Werte**

| Code | Wertelabel                              |
|------|---|
| 0    | Keine Angabe                            |
| -1   | Filter                                  |
| -2   | (ganze) Frage nicht beantwortet         |
| -3   | ganze Spalte nicht beantwortet          |
| -995 | keine Teilnahme (Panel)                 |
| -969 | unbekannter fehlender Wert <sup>a</sup> |
| -967 | anonymisiert                            |

<sup>a</sup> Dieser Wert wird vergeben, wenn keinerlei Ursache rekonstruiert werden kann.

## 7 Gewichtung

Die Gewichtung der Daten dient dem Ausgleich von Verzerrungen der Stichprobe im Vergleich zur definierten Grundgesamtheit. 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 Netto-Stichprobe (vgl. Kapitel 5). Wenn diese Ausfallprozesse unsystematisch sind (Missing Completely at Random), können sie ignoriert werden.<sup>47</sup> Jedoch unterliegen sie zumeist einem systematischen Ausfallprozess (Missing at Random, Not Missing at Random), der einer Modellierung bedarf.<sup>48</sup>

**[Konzeptuelles Vorgehen]** Im Zuge einer Gewichtungsprozedur sollten idealerweise zunächst designbedingte Disproportionalitäten ausgeglichen werden. Die hierfür benötigten *Designgewichte* ergeben sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Im Anschluss sollte eine Adjustierung der Designgewichte mit Hilfe von *Nonresponsegewichten* im Quer- und Längsschnitt erfolgen, die auf der Grundlage von Informationen über Teilnehmer(innen) und Nichtteilnehmer(innen) auf Individualebene erzeugt werden. In einem letzten Schritt können die nonresponse-adjustierten Designgewichte anhand von Merkmalsverteilungen aus der Grundgesamtheit kalibriert werden (Kalibrierung).

Vor dem Hintergrund der Erläuterungen zur Grundgesamtheit und zum Stichprobenverfahren in Kapitel 3 wird ersichtlich, dass ein derartiges idealtypisches Vorgehen für die Daten des Studienberechtigtenpanels 2015 nicht durchführbar ist. Das Stichprobendesign ermöglicht keine Ableitung von exakten Auswahlwahrscheinlichkeiten. Da in der ersten Befragungswelle zudem keine Informationen über Nichtteilnehmer(innen) vorliegen, können hier auch keine individuellen Ausfallgewichte gebildet werden.

<sup>47</sup> Insofern die Einbußen an statistischer Teststärke durch die Verringerung der Stichprobe als irrelevant erachtet werden.

<sup>48</sup> Siehe grundlegend zu den unterschiedlichen Formen von Ausfallprozessen Rubin (1976).

Als Gewichte für die erste Welle werden ein geschätztes Designgewicht sowie Querschnittsgewichte berechnet. Das *Designgewicht* ergibt sich bei zufallsgesteuerten Auswahlverfahren direkt aus dem Stichprobendesign. Die *Querschnittsgewichte* basieren auf einer Zellgewichtung mit der die Daten an ausgewählte Merkmale der Grundgesamtheit<sup>49</sup> (Geschlecht, Bundesland, Art der Schule, Art der Hochschulreife) angepasst wurden. Diejenigen Zellen, welche (vorab definierte) Ausreißerwerte in den Gewichten ( $<0,3$  und  $>7$ ) enthalten, werden mittels Zellverschmelzungen nivelliert. Aus der Merkmalsverteilung der Grundgesamtheit werden zugleich Informationen zu Nichtteilnehmer(inne)n gewonnen, sodass hiermit ebenfalls eine Form der Nonresponse-Adjustierung erfolgt. In Tabelle 4

sind die im Datensatz bereitgestellten Gewichte dargestellt.

**Tabelle 4: Bereitgestellte Gewichte zum DZHW-Studienberechtigtenpanel 2015**

| Variablenname | Beschreibung   |
|---------------|--|
| <i>gew1</i>   | Querschnittsgewicht Welle 1 (jedes Bundesland gesondert)                       |
| <i>gew2</i>   | Querschnittsgewicht Welle 1 (bundesweites Gewicht mit Länderrelationen)        |
| <i>gewao</i>  | Längsschnittgewicht 2-Wellen-Panel (jedes Bundesland gesondert)                |
| <i>gewbo</i>  | Längsschnittgewicht 2-Wellen-Panel (bundesweites Gewicht mit Länderrelationen) |

**[Hinweise zur Anwendung der Gewichte]** Bei den erstellten Gewichten handelt es sich um *probability weights*, die in Stata mit Hilfe ado-spezifischer Optionen berücksichtigt werden können.<sup>50</sup> Die Gewichte *gew1* und *gew2* sind für Auswertungen der ersten Welle vorgesehen, *gew1* für Auswertungen einzelner Bundesländer und *gew2* für Bundesauswertungen. Die Gewichte *gewao* und *gewbo* sind für Auswertungen des 2-Wellen-Panels konzipiert, *gewao* für Auswertungen einzelner Bundesländer und *gewbo* für Bundesauswertungen.

Grundlegend ist zu beachten, dass Gewichte nur für jene Analysen sinnvolle Korrekturgrößen 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.

## 7.2 Gewichtung des Datensatzes

Im Folgenden wird die Vorgehensweise bei der Erstellung der Gewichte für den Datensatz näher dargestellt.

**[Designgewichtung]** Die Berechnung der Designgewichte 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 Schulart<sup>51</sup> ergeben (vgl. Kapitel 3). Benötigt werden für jede Schicht die folgenden Informationen aus dem Stichprobenplan:

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

<sup>49</sup> Genutzt wurden die zum Zeitpunkt der Stichprobenziehung aktuellsten veröffentlichten Daten des Statistischen Bundesamtes (2015a) (2015b) über die Anzahl der Absolvent(inn)en allgemeinbildender und berufsbildender Schulen: Allgemeinbildende Schulen - Schuljahr 2014/15 Fachserie 11 Reihe 1 und Berufliche Schulen - Schuljahr 2014/15 Fachserie 11 Reihe 2. Die Daten der Schulabsolvent(inn)en beider Publikationen beziehen sich jeweils auf das Schuljahr 2013/14.

<sup>50</sup> Siehe hierzu die Stata-Hilfe (Befehl: *help weights*).

<sup>51</sup> Hierbei handelt es sich um die neun Schularten, die auch maßgeblich für die Stichprobenziehung waren (vgl. Kapitel 3).

Anhand dieser Informationen wird für jede Schicht die spezifische *Auswahlwahrscheinlichkeit*  $\pi$  bestimmt. Diese ergibt sich aus der Formel:

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

Das Designgewicht ergibt sich schließlich als Kehrwert der Auswahlwahrscheinlichkeit:

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

Die Designgewichte werden abschließend noch auf die Stichprobengröße normiert, indem die jeweiligen Gewichte durch den Mittelwert aller Gewichte geteilt werden.

**[Querschnittsgewichtung]** Für die erste Welle wurde mit einer Zellgewichtung Querschnittsgewichte erstellt. Hierbei wurden die Daten anhand der Merkmale Geschlecht, Schulgruppen<sup>52</sup> und Bundesland an die Grundgesamtheit angepasst, um mit Blick auf diese Merkmale Verzerrungen in der realisierten Stichprobe gegenüber der Verteilung in der Grundgesamtheit auszugleichen.<sup>53</sup> Diese Verzerrungen können sowohl aus (Unit- oder Item-Nonresponse) resultieren als auch aus dem disproportionalen Stichprobendesign.

Für jede Zelle der Gewichtungsmatrix wurde anhand der Sollzahlen zunächst der Anteil bestimmt, den diese Zelle in der Grundgesamtheit repräsentiert ( $p_N$ ). Für die realisierte Stichprobe wurde ebenfalls eine solche Quote gebildet ( $p_n$ =Anteil realisierte Fälle in der jeweiligen Zelle geteilt durch Gesamtfallzahl in der Stichprobe). Der Gewichtungsfaktor ergibt sich, indem aus beiden Anteilswerten ein Quotient gebildet wird ( $p_N/p_n$ ). Dieses Vorgehen wurde zweimal durchgeführt. Beim Gewicht *gew1* wurden die Gewichtungsfaktoren für jedes Land gesondert berechnet. Bezugsgruppe für die Bestimmung der Anteilswerte  $p_N$  und  $p_n$  war dann die Fallzahl in Grundgesamtheit bzw. Stichprobe im jeweiligen Bundesland. Das Gewicht *gew2* berücksichtigt zusätzlich die Relationen der Länder untereinander, da die Anteilswerte hier bezogen auf die bundesweiten Fallzahlen bestimmt wurden.

Für die zweite Erhebungswelle wurde kein Querschnittsgewicht erstellt, da keine neu hinzugekommenen Individuen befragt wurden (Auffrischungsstichprobe).

**[Längsschnittgewichtung]** Die Gewichte *gewao* und *gewbo* des Zwei-Wellen-Panels wurden ebenfalls mit einer Zellgewichtung erstellt. Hierbei wurden die Daten anhand der Merkmale Geschlecht, Schulgruppen<sup>54</sup> und Bundesland an die Grundgesamtheit angepasst, um mit Blick auf diese Merkmale Verzerrungen in der realisierten Stichprobe gegenüber der Verteilung in der

<sup>52</sup> Folgende Schulgruppen wurden unterschieden: (1) Gymnasium/Gesamtschule/Waldorfschule, (2) Abendgymnasium/Kolleg, (3) Fachgymnasium/andere berufsbildende Schulen mit Erwerb der allgemeinen Hochschulreife, (4) Fach- und Berufsoberschulen (mit Erwerb der Fachhochschulreife), (5) sonstige berufsbildende Schulen (mit Erwerb der Fachhochschulreife).

<sup>53</sup> Grundlage bildeten die Daten des Statistischen Bundesamtes (2015a); (2015b) zu Schulabsolvent(inn)en mit Studienberechtigung des Schuljahres 2013/14. Diese wurden der entsprechenden Fachserie 11 des Schuljahres 2014/15 (Reihe 1 für allgemeinbildende Schulen sowie Fachserie 11 Reihe 2 für berufsbildende Schulen) entnommen.

<sup>54</sup> Folgende Schulgruppen wurden unterschieden: (1) Gymnasium/Gesamtschule/Waldorfschule, (2) Abendgymnasium/Kolleg, (3) Fachgymnasium/andere berufsbildende Schulen mit Erwerb der allgemeinen Hochschulreife, (4) Fach- und Berufsoberschulen (mit Erwerb der Fachhochschulreife), (5) sonstige berufsbildende Schulen (mit Erwerb der Fachhochschulreife).

Grundgesamtheit auszugleichen.<sup>55</sup> Diese Verzerrungen können sowohl aus (Unit- oder Item-Nonresponse) resultieren als auch aus dem disproportionalen Stichprobendesign.

Für jede Zelle der Gewichtungsmatrix wurde anhand der Sollzahlen zunächst der Anteil bestimmt, den diese Zelle in der Grundgesamtheit repräsentiert ( $p_N$ ). Für die realisierte Stichprobe wurde ebenfalls eine solche Quote gebildet ( $p_n$ =Anteil realisierte Fälle in der jeweiligen Zelle geteilt durch Gesamtfallzahl in der Stichprobe). Der Gewichtungsfaktor ergibt sich, indem aus beiden Anteilswerten ein Quotient gebildet wird ( $p_N/p_n$ ). Dieses Vorgehen wurde zweimal durchgeführt. Beim Gewicht *gewao* wurden die Gewichtungsfaktoren für jedes Land gesondert berechnet. Bezugsgruppe für die Bestimmung der Anteilswerte  $p_N$  und  $p_n$  war dann die Fallzahl in Grundgesamtheit bzw. Stichprobe im jeweiligen Bundesland. Das Gewicht *gewbo* berücksichtigt zusätzlich die Relationen der Länder untereinander, da die Anteilswerte hier bezogen auf die bundesweiten Fallzahlen bestimmt wurden.

---

<sup>55</sup> Grundlage bildeten die Daten des Statistischen Bundesamtes (2016b); (2016a) zu Schulabsolvent(inn)en mit Studienberechtigung des Schuljahres 2014/15. Diese wurden der entsprechenden Fachserie 11 des Schuljahres 2015/16 (Reihe 1 für allgemeinbildende Schulen sowie Fachserie 11 Reihe 2 für berufsbildende Schulen) entnommen.

## 8 Anonymisierung

**[Datenschutzrechtlicher Rahmen]** Für personenbezogene Daten<sup>56</sup>, 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.<sup>57</sup> 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). Das heißt, für die Weitergabe von Daten aus wissenschaftlichen Forschungsprojekten an Dritte sind die Daten derart zu anonymisieren, dass kein Bezug zur Person mehr hergestellt werden kann.

**[Datenzugang, Anonymisierungsgrad und Analysepotential]** Das FDZ des DZHW stellt für das Studienberechtigtenpanel 2015 ein SUF für die wissenschaftliche Sekundärnutzung zur Verfügung. Die Anonymität der Befragten wird dabei über eine Kombination aus statistischen Maßnahmen und technischen Zugriffsbeschränkungen sichergestellt. Je stärker der Datenzugang technisch kontrolliert wird, desto geringer ist das Risiko einer De-Anonymisierung der Daten, desto weniger müssen die Daten mittels statistischer Maßnahmen um Informationen reduziert werden und desto größer bleibt ihr Analysepotential.

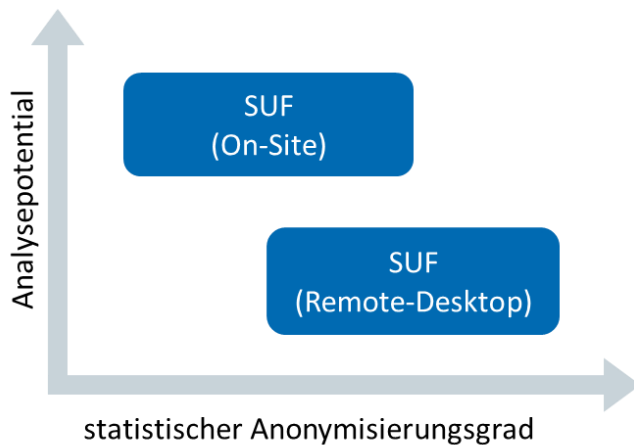
Das SUF wird über zwei verschiedene Zugangswege angeboten: Remote-Desktop und On-Site (für weiterführende Informationen s. Kapitel II). Für jeden Zugangsweg wird eine andere SUF-Variante bereitgestellt, die unterschiedlich stark anonymisiert worden ist und entsprechend weniger oder mehr Informationen umfasst. Abbildung 4 gibt einen Überblick über den jeweiligen Grad der statistischen Anonymisierung und dem damit verbundenen Analysepotential. Im Folgenden werden die durchgeführten statistischen Anonymisierungsmaßnahmen in Abhängigkeit vom Zugangsweg erläutert.

<sup>56</sup> „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).

<sup>57</sup> 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.



**Abbildung 4: Datenzugangsweg, statistischer Anonymisierungsgrad und Analysepotential der Daten des Studienberechtigtenpanels 2015**



**[Statistische Anonymisierungsmaßnahmen]** Im Rahmen der Anonymisierung sind zunächst alle Informationen, mit denen sich Personen oder Institutionen direkt identifizieren lassen, zu löschen. Diese sogenannten *direkten Identifikatoren*, wie Namen, Adressen oder E-Mail-Adressen, wurden im Rahmen des Studienberechtigtenpanels 2015 bereits während der Feldphase separat erfasst und nur genutzt, um eine Kontaktaufnahme zu ermöglichen. Diese Angaben sind somit nicht in den verschiedenen SUF-Varianten enthalten. Des Weiteren wurde, um einen Rückbezug auf die Originaldaten zu verhindern, die Original-Identifikationsnummer aus dem Datensatz 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.<sup>58</sup> Für das Studienberechtigtenpanel 2015 wurden beispielsweise folgende Quasi-Identifikatoren identifiziert: regionale Informationen (Geburtsland, Hochschule oder Arbeitsort), Staatsangehörigkeit, Sprache im Elternhaus, Schulart, schulische Prüfungsfächer, Hochschule, Studienfach, Abschlussart, Berufsangaben. Um eine eindeutige Zuordnung der Studienberechtigten zu unterbinden, wurden diese Schlüsselmerkmale – je nach Zugangsweg – aggregiert oder gelöscht (vgl. Tabelle 5). Beispielsweise ist das Merkmal Studienfach im SUF für die On-Site-Nutzung uneingeschränkt verfügbar. Im Remote-Desktop-SUF hingegen wird das Merkmal zu Studienbereichen aggregiert.

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). Die offenen Angaben wurden größtenteils bereits im Rahmen der Datenaufbereitung durch das Primärforschungsprojekt vercodet und werden daher in allen SUF-Varianten zur Verfügung gestellt. Teilweise wurden jedoch – in Abhängigkeit von der Sensibilität der enthaltenen Informationen und vom Zugangsweg – die vom Primärforschungsprojekt vorge-

<sup>58</sup> Dabei ist darauf hinzuweisen, dass die Identifikation einer Person bereits durch die Stichprobenauswahl erschwert wird, da eine Ungewissheit darüber besteht, ob eine befragte Person eine einzigartige Merkmalskombination in der Population aufweist.

nommenen Codierungen zusätzlich aggregiert. Nicht codierte offene Angaben wurden in allen SUF-Varianten gelöscht.

Zuletzt wurde geprüft, ob in den Daten *sensible Informationen*, z. B. zur Gesundheit, sexuellen Orientierung oder zu politischen Einstellungen, enthalten waren. Diese eignen sich zwar nicht notwendig zur Re-Identifikation von Individuen oder Institutionen, jedoch können die Informationen im Falle einer De-Anonymisierung nutzbringend sein (Koberg, 2016, S. 694) und sind daher besonders schützenswert (vgl. Art. 9 DSGVO, Erwägungsgrund 51 DSGVO). Die nachfolgende Tabelle 5 stellt in Kurzform die durchgeführten statistischen Anonymisierungsmaßnahmen je nach Zugangsweg dar.

**Tabelle 5: Maßnahmen der statistischen Anonymisierung der Daten des DZHW-Studienberechtigtenpanels 2015 (Welle 1 und 2) nach Zugangsweg<sup>59</sup>**

| Merkmal   | On-Site-SUF  | Remote-Desktop-SUF  |
|---|--|---|
| Direkte Identifikatoren                               | Löschung und Vergabe einer zufälligen ID   | Löschung und Vergabe einer zufälligen ID  |
| Schulart  | Aggregation von geringbesetzten Schularten (Codierliste cl-dzhw-20)                  | Aggregation zu allgemeinbildenden und beruflichen Schulen (Codierliste cl-dzhw-20)  |
| Prüfungsfächer/Schwerpunktfach (Schule, Berufsschule) | Freigabe   | Aggregation zu (Berufs-)Schulfächergruppen (Codierliste cl-dzhw-7)                  |
| Studienfach   | Freigabe   | Aggregation zu Studienbereichen <sup>60</sup>                                       |
| Angestrebter Studienabschluss                         | Aggregation von gering besetzten Arten   | Aggregation von gering besetzten Arten  |
| Hochschulwunsch: Hochschulort                         | Aggregation zu Bundesland und Ausland (ja/nein)                                      | Aggregation zu Bundesland und Ausland (ja/nein)                                     |
| Hochschule  | Aggregation zur Hochschulart   | Aggregation zur Hochschulart (Uni/FH)   |
| Hochschulort, Aufenthaltsort im Dezember 2015         | Aggregation zum Bundesland und Staat einer ausl. Hochschule                          | Aggregation zum Bundesland und Staat einer ausl. Hochschule <sup>61</sup>           |
| Ausbildungsberuf, berufliche Tätigkeit                | Freigabe   | Aggregation zu Berufsordnungen <sup>62</sup>  |
| Arbeitsort (Postleitzahl)                             | Reduktion auf 3 Stellen  | Reduktion auf 2 Stellen   |
| Staatsangehörigkeit                                   | 30 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen) | 7 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen) |
| Geburtsland   | 30 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen) | 7 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen) |
| Sprache im Elternhaus                                 | Erste Sprache:   | Erste Sprache:  |

<sup>59</sup> Detaillierte Informationen zu den anonymisierten Variablen sind dem Datensatzreport sowie dem MetadatenSuchsystem (<https://metadata.fdz.dzhw.eu>) zu entnehmen.

<sup>60</sup> Nach Schlüsselverzeichnis der Studenten- und Prüfungsstatistik (cl-destatis-studienfach-2016)

<sup>61</sup> Nur Staaten im Ausland mit mehr als einer Hochschule.

<sup>62</sup> Nach Klassifikation der Berufe von 1992 von Destatis (KlB-3-Steller).

| Merkmal  | On-Site-SUF   | Remote-Desktop-SUF  |
|--|---|---|
|  | Freigabe von 20 Sprachen, weitere Sprachen zu „Sonstige“ aggregiert<br>Zweite Sprache:<br>Freigabe von 3 Sprachen, weitere Sprachen zu „Sonstige“ aggregiert<br>Dritte Sprache:<br>Aggregation zu genannt/nicht genannt | Freigabe von 10 Sprachen, weitere Sprachen zu „Sonstige“ aggregiert<br>Zweite Sprache:<br>Freigabe von 3 Sprachen, weitere Sprachen zu „Sonstige“ aggregiert<br>Dritte Sprache:<br>Aggregation zu genannt/nicht genannt |
| Beruf der Eltern                                 | Aggregation zu Berufsordnungen <sup>63</sup>  | Aggregation zu Berufsordnungen <sup>64</sup>  |
| Geburtsland der Eltern                           | 30 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen)  | 7 Staaten einzeln ausgewiesen, ansonsten Aggregation (Staatengruppen, Weltregionen)   |
| Anzahl Brüder/Schwestern                         | Top-Codierung „3 und mehr Geschwister“  | Löschung  |
| Probleme bei Wahl des nachschulischen Werdegangs | Offene Angaben wurden vercodet und teilweise den vorhandenen Kategorien zugeordnet. Die Kategorien "Gesundheitsprobleme" und "soziale oder ethnische Herkunft, Geschlecht" werden zu „sonstige Gründe“ zusammengefasst  | Offene Angaben wurden vercodet und teilweise den vorhandenen Kategorien zugeordnet. Die Kategorien "Gesundheitsprobleme" und "soziale oder ethnische Herkunft, Geschlecht" werden zu „sonstige Gründe“ zusammengefasst  |
| Tätigkeit im Dez.'15: „Krankheit“                | Zu „Sonstige Gründe“ aggregiert   | Zu „Sonstige Gründe“ aggregiert   |

<sup>63</sup> Nach Klassifikation der Berufe von 1992 von Destatis (KIDB-3-Steller).

<sup>64</sup> Nach Klassifikation der Berufe von 1992 von Destatis (KIDB-3-Steller).

## 9 Literaturverzeichnis

- Beierlein, C., Kovaleva, A., Kemper, C. J. & Rammstedt, B. (2012). *Ein Messinstrument zur Erfassung subjektiver Kompetenzerwartungen: Allgemeine Selbstwirksamkeit Kurzskala (ASKU)* (GESIS Working Papers 2012/17). GESIS – Leibniz-Institut für Sozialwissenschaften. Verfügbar unter <https://nbn-resolving.org/urn:nbn:de:0168-ss0ar-292351>
- Bundesinstitut für Berufsbildung (Hrsg.). (2015). *AusbildungPlus. Duales Studium in Zahlen 2014*. Trends und Analysen. Bonn.
- Bundesinstitut für Berufsbildung (Hrsg.). (2016). *Datenreport zum Berufsbildungsbericht 2016. Informationen und Analysen zur Entwicklung der beruflichen Bildung*. Bielefeld: W. Bertelsmann Verlag.
- Ebel, T. & Meyermann, A. (2015). *Hinweise zur Anonymisierung von quantitativen Daten* (forschungsdaten bildung informiert Nr. 3). Verbund Forschungsdaten Bildung. Verfügbar unter [https://www.forschungsdaten-bildung.de/get\\_files.php?action=get\\_file&file=fdb-informiert\\_nr-7.pdf](https://www.forschungsdaten-bildung.de/get_files.php?action=get_file&file=fdb-informiert_nr-7.pdf)
- 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). *Datenschutz am Forschungsdatenzentrum* (FDZ-Methodenreport Nr. 6). Nürnberg: Institut für Arbeitsmarkt- und Berufsforschung (IAB).
- Hochschulrektorenkonferenz (Hrsg.). (2015). *Statistische Daten zu Studienangeboten an Hochschulen in Deutschland. Studiengänge, Studierende, Absolventinnen und Absolventen. Wintersemester 2015/16* (Statistiken zur Hochschulpolitik, 1/2015). Bonn: HRK.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. von 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
- Kovaleva, A., Beierlein, C., Kemper, C. J. & Rammstedt, B. (2012). *Eine Kurzskala zur Messung von Impulsivität nach dem UPPS-Ansatz: Die Skala Impulsives-Verhalten-8 (1-8)* (GESIS - Leibniz-Institut für Sozialwissenschaften, Hrsg.) (GESIS Working Papers 2012|20). Köln.
- Lane, J., Heus, P. & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy*, 1(1), pp. 2–16.
- Lenzner, T., Neuert, C. & Otto, W. (2015). *Kognitives Pretesting* (GESIS Survey Guidelines). Mannheim: GESIS – Leibniz-Institut für Sozialwissenschaften. doi:10.15465/gesis-sg\_010
- Mertens, D. (1976). Beziehungen zwischen Qualifikation und Arbeitsmarkt. In W. Schlaffke (Hrsg.), *Jugendarbeitslosigkeit. Unlösbare Aufgabe für das Bildungs- und Beschäftigungssystem?* (S. 68–117). Köln: Dt. Inst.-Verl.
- Prüfer, P. & Rexroth, M. (2000). *Zwei-Phasen-Pretesting* (ZUMA Arbeitsbericht Nr. 8). Zugriff am 25.10.2016.
- Rubin, D. B. (1976). Inference and missing data. *Biometrika*, 63(2), 581–592.
- Schneider, H., Franke, B., Woisch, A. & Spangenberg, H. (2017). *Erwerb der Hochschulreife und nachschulische Übergänge von Studienberechtigten. Studienberechtigte 2015 ein halbes Jahr vor und ein halbes Jahr nach Schulabschluss*. (Forum Hochschule Nr. 4). Hannover: Deutsches Zentrum für Hochschul- und Wissenschaftsforschung (DZHW).
- Schnell, R., Hill, P. B. & Esser, E. (2005). *Methoden der empirischen Sozialforschung* (7. Aufl.). München: Oldenbourg.
- Schnell, R., Hill, P. B. & Esser, E. (2011). *Methoden der empirischen Sozialforschung* (9. Aufl.). München: Oldenbourg.
- Statistisches Bundesamt. (2014a). *Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2013/2014* (Fachserie 11 Reihe 1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2014b). *Bildung und Kultur. Berufliche Schulen. Schuljahr 2012/2013* (Fachserie 11 Reihe 2). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2015a). *Bildung und Kultur. Allgemeinbildende Schulen. Schuljahr 2014/2015* (Fachserie 11 Reihe 1). Wiesbaden: Statistisches Bundesamt.
- Statistisches Bundesamt. (2015b). *Bildung und Kultur. Berufliche Schulen. Schuljahr 2014/2015* (Fachserie 11 Reihe 2). Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2016a). *Bildung und Kultur. Allgemeinbildende Schulen*. Schuljahr 2015/2016 (Fachserie 11 Reihe 1). Wiesbaden: Statistisches Bundesamt.

Statistisches Bundesamt. (2016b). *Bildung und Kultur. Berufliche Schulen*. Schuljahr 2015/2016 (Fachserie 11 Reihe 2). Wiesbaden: Statistisches Bundesamt.