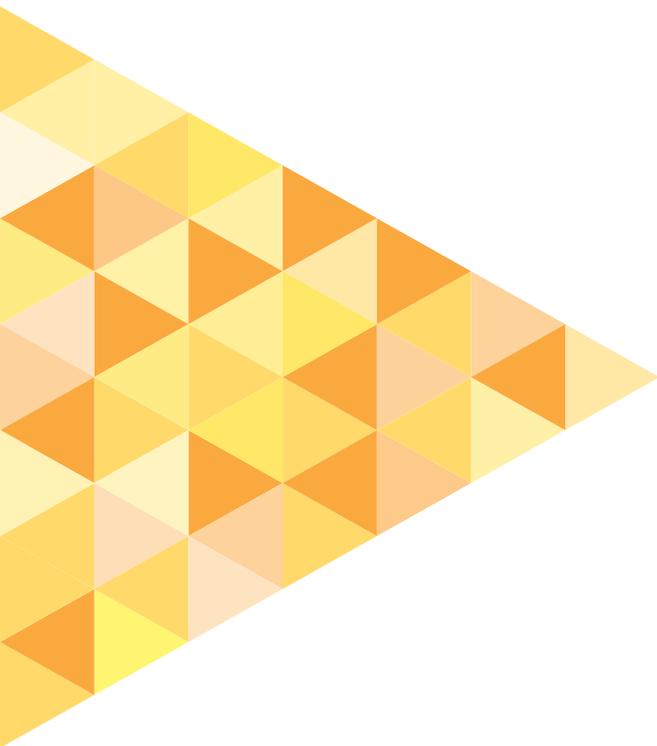


Rafael Schratz

Persönliche Identifikationscodes: Überlegungen im Rahmen einer Befragung von Schülerinnen und Schülern

Version vom 30.03.2020



BIBB-Preprint

Zitiervorschlag:

Schratz, Rafael : Persönliche Identifikationscodes :
Überlegungen im Rahmen einer Befragung von Schülerinnen
und Schülern. Version 1.0 Bonn, 2020



Version 1.0
April 2020

Herausgeber:

Bundesinstitut für Berufsbildung
Robert-Schuman-Platz 3
53175 Bonn
Internet: www.vet-repository.info
E-Mail: repository@bibb.de

CC Lizenz

Der Inhalt dieses Werkes steht unter Creative-Commons-Lizenz
(Lizenztyp: Namensnennung – Keine kommerzielle Nutzung –
Keine Bearbeitung – 4.0 International).
Weitere Informationen finden sie im Internet auf unserer
Creative-Commons-Infoseite www.bibb.de/cc-lizenz.

Bibliografische Information der Deutschen Nationalbibliothek

Diese Netzpublikation wurde bei der Deutschen
Nationalbibliothek angemeldet und archiviert:
urn:nbn:de:0035-vetrepository-776690-1

Persönliche Identifikationscodes: Überlegungen im Rahmen einer Befragung von Schülerinnen und Schülern

Inhalt

1. Ausgangslage	2
1.1. Ansätze zur anonymisierten Befragung bei Längsschnittstudien	2
1.2. Ausmaß des Datenverlusts	3
1.3. Länge des zu generierenden Codes	4
1.4. Kriterien und Empfehlungen zur Generierung von Codefragen	4
2. Selbstgenerierte Codes für die BIBB/TU DA-Berufsorientierungsstudie.....	5
2.1. Generierung der Codefragen (Befragung der Ausbildungsbotschafter/-innen)	5
2.2. Abgleich mit den Gütekriterien	6
2.3. Zu den einzelnen Fragen	7
2.4. Layout und Gestaltung	8
2.5. Zuordnung der Fragebögen und weiteres Vorgehen	8
2.6. Zusammenfassung.....	8
Literatur	9

1. Ausgangslage

Wie die Berufswahlforschung gezeigt hat, lassen sich junge Menschen bei ihrer Entscheidungsfindung für oder gegen Berufe von unterschiedlichen Motiven leiten. Ihre Berufswahl wird, so ein Ergebnis des BIBB-Forschungsprojektes „Bildungsorientierungen“, auch durch ihren oft unbewussten Wunsch nach sozialer Anerkennung beeinflusst (vgl. u. a. MATTHES, 2019: S. 167). Davon ausgehend untersuchen das Bundesinstitut für Berufsbildung (BIBB) und die Technische Universität Darmstadt in einer gemeinsamen Berufsorientierungsstudie, inwieweit Angebote der Berufsorientierung dazu beitragen können, das Berufswahlspektrum von Jugendlichen zu erweitern und sie in einer reflektierten und autonomen Berufswahl zu stärken. Es soll jeweils untersucht werden, wie die im Rahmen der Studie entwickelte Unterrichtseinheit „die Logiken der Berufswahl“, sowie der Einsatz von Ausbildungsbotschafterinnen und Ausbildungsbotschaftern (Abbos) in Schulen das Berufswahlspektrum und ausgewählte Aspekte der Berufswahlkompetenz von Jugendlichen beeinflussen. Um die Wirkung der Abbos zu prüfen, werden jeweils vor und nach ihrem Einsatz Schülerinnen und Schüler (SuS) mithilfe eines Fragebogens befragt. Daher ist es zwingend erforderlich, die Fragebögen einer Person zu den verschiedenen Befragungszeitpunkten einander zuordnen zu können. Gleichzeitig ist das Recht auf informationelle Selbstbestimmung der Schülerinnen und Schüler ein schützenswertes Gut, welches u. a. durch Datenschutzgesetze des Bundes und der jeweiligen Bundesländer geschützt wird. Somit ist es notwendig, eine Möglichkeit zu finden, die sowohl die Einhaltung der Rechte der Befragten, als auch die Zuordnung der Fragebögen ermöglicht.

Der folgende Beitrag fasst wesentliche Überlegungen zur Nutzung von persönlichen Identifikationscodes im Rahmen der o. g. Befragung zusammen und erörtert die Vor- und Nachteile unterschiedlicher Vorgehensweisen.

Unter folgenden Links können weitere Informationen zur BIBB/TU DA Berufsorientierungsstudie und zum BIBB-Forschungsprojekt „Bildungsorientierungen“ abgerufen werden:

<https://www.bibb.de/bibb-tuda-bo-studie>

<https://www.bibb.de/de/8475.php>

1.1. Ansätze zur anonymisierten Befragung bei Längsschnittstudien

Eine Literaturübersicht von AUDETTE et al. 2020 gibt einen aktuellen Überblick über vier Ansätze für die Verknüpfung von Fragebögen bei anonymisierten Längsschnittbefragungen und ihre jeweiligen Vor- und Nachteile (ebd.: S. 166f):

- a) Nicht-anonymisierte Erhebung, spätere Anonymisierung,
- b) bereits existierende eindeutige Identifikatoren,
 - i) bspw. Matrikelnummern / volles Geburtsdatum,
- c) elektronische Anonymisierungssysteme,
 - i) Applikationen, die die Befragungsdaten in anonymisierter Form übermitteln,
- d) selbst generierte Identifikationscodes (SGIC),
 - i) Code, den Befragte auf Basis der Antworten einer Kombination vordefinierter Fragen selbst generieren,
 - (1) (bspw. erster Buchstabe des eigenen Vornamens, Tag der Geburt, ...).

Konkret auf die BIBB/TU DA-Studie bezogen, sprechen gewichtige Gründe gegen die Ansätze a bis c:

- Zu a): Eine nicht anonymisierte Erhebung scheidet aus, da die Daten bereits anonymisiert erhoben werden sollen.
- Zu b): Geeignete, bereits existierende Identifikatoren sind für die Zielgruppe nicht bekannt. Das entsprechende Verfahren scheidet damit aus. Weiterhin erreichen auch diese Verfahren teilweise nur Zuordnungsraten von 72 Prozent (DIRENGA et al., 2016: S. 8.).
- Zu c): Ein elektronisches Anonymisierungssystem scheitert bereits an dem damit verbundenen zeitlichen und organisatorischen Aufwand. Der Aufwand bei der Implementierung kann mit bis zu 18 Monaten angesetzt werden (AUDETTE et al., 2020: S. 172). Dies ist mit dem Projektzeitplan nicht vereinbar.

Eine Reihe von Publikationen führen die Vorteile von SGICs als Methode zum Zuordnen von Fragebögen aus (vgl. DIRENGA et al., 2016; LIPPE et al., 2019; YUREK et al., 2008). Die beste Kombination von Vor- und Nachteilen sehen auch AUDETTE et al. bei diesem Ansatz gegeben (AUDETTE et al., 2020: S. 172f & 180):

Vorteile (AUDETTE et al., 2020):

- Faktische Anonymität,
- wahrgenommene Anonymität durch Befragte (Anmk.: und Schulpersonal) erhöht Teilnahmebereitschaft,
- Kombination persönlich relevanter Fragen ist besser erinnerbar als Identifikationscode,
- keine zufällige Re-Identifikation (auch in Bezug auf Bestätigungsfehler),
- keine Sammlung persönlicher Informationen → Einhaltung rechtlicher Normen erleichtert (Anmk.: Bspw. Datenschutz),
- für verschiedene Datenformate (Papier, elektronisch) anwendbar.

Nachteil (AUDETTE et al., 2020):

- Teilnehmende müssen sich genau an die abgefragten Informationen erinnern und diese wiedergeben können, daher ist mit Datenverlust zu rechnen, der durch sorgfältige Auswahl der Fragen reduziert werden kann.

1.2. Ausmaß des Datenverlusts

In vielen Fällen ist eine exakte Übereinstimmung der Codes einer Person zu verschiedenen Zeitpunkten nicht gegeben. Durch fehlerhafte Eingabe oder mangelnde Stabilität der abgefragten Informationen kommt es zu einzelnen Abweichungen. Eine Balance zwischen der Vermeidung von Datenverlust durch nicht vorgenommene Zuordnungen und falsch-positiver Zuordnung stellt die oft angewendete off-one bzw. off-two Regel dar. Bei dieser dürfen Codes auch bei einer bzw. bei zwei Abweichungen zugeordnet werden (SCHNELL et al., 2010: S. 393). Wird eine hundertprozentig exakte Passung vorausgesetzt, ist mit Datenverlusten von bis zu 50 Prozent, andernfalls von 20 bis 30 Prozent zu rechnen. Mithilfe sorgfältig ausgewählter Fragen lässt sich die Datenqualität verbessern, so dass der Datenverlust auf zehn Prozent reduziert werden kann (SCHNELL et al., 2010: 393f). In der Übersichtsstudie von AUDETTE et al. (2020) wurden Studien, die SGICs verwendeten, in Bezug auf die Anzahl der verwendeten Codeelemente und Übereinstimmungsraten verglichen. Die Befunde sind konsistent mit den von SCHNELL et al. (2010) berichteten Werten.

Abbildung 1: Passungsraten von Studien unter Verwendung von SGICs

	Minimum	Maximum	Ø	n
Codeelemente	3	9	5,85	(n=33)
Passung (perfekt)	42,00 %	94,41 %	65,27 %	(n=25)
Passung (one-off)	51,30 %	98,71 %	80,87 %	(n=23)
Abstand in Monaten	0,25	24	8,61	(n=33)

(AUDETTE et al., 2020: S. 176; eigene Darstellung)

Weitere wichtige Faktoren in Bezug auf die Passungsrate sind die Zielgruppe (SCHNELL 2019: S. 393) und mutmaßlich auch der zeitliche Abstand zwischen den Befragungen.

1.3. Länge des zu generierenden Codes

In den von AUDETTE et al. (2020) ausgewerteten Studien kamen drei bis neun Codeelemente zum Einsatz, wobei sieben der häufigste Wert war (AUDETTE et al., 2020: S. 176). Bezüglich der Länge des Codes ist eine Abwägung zwischen „accuracy“ und „identifying power“ vorzunehmen (DIRENGA et al., 2016: S. 3). Die „accuracy“ bezeichnet, wie konsistent eine Person zu verschiedenen Zeitpunkten den gleichen Code generiert. „identifying power“ besagt, ob der Code einer Person einzigartig ist, oder ob mehrere Personen den gleichen Code generieren. Ebenso zu berücksichtigen sind die Belastung der Befragten und die Gefahr, durch Erhebung zu vieler Informationen Probanden de-anonymisieren zu können. Eine Empfehlung ist, sieben bzw. acht Elemente zu verwenden, auch da dies von den Befragten noch als akzeptabel empfunden wird (DIRENGA et al., 2016: S. 3).

1.4. Kriterien und Empfehlungen zur Generierung von Codefragen

In der Literatur finden sich zahlreiche Kriterien und Empfehlungen zur Generierung der Codefragen, von denen einige im Folgenden dargestellt werden:

Kriterien nach (DIRENGA et al., 2016: S. 3)

Codefragen **müssen** ...

- a. ... für jede Person zutreffend sein (nicht z. B. Name eines Haustiers);
- b. ... den Befragungspersonen, aber nicht den Forschenden gut bekannt sein (nicht „Blutgruppe“);
- c. ... zeitlich unveränderlich sein (nicht die Anzahl der Geschwister) (Anmk.: Die Anzahl der älteren Geschwister weist dagegen eine hohe Stabilität auf).

Codefragen **sollten** weiterhin ...

- d. ... eindeutig angebbbar sein (nicht Haarfarbe);
- e. ... nicht beobachtbar sein (nicht Geschlecht);
- f. ... etwas sein, was eine Person preisgeben wird (Anmk.: Bzw. gefahrlos preisgeben kann) (nicht PIN-Codes);
- g. ... eine hohe Variation haben (also nicht etwas sein, auf das alle die gleiche Antwort geben);
- h. ... einfach verständlich sein (nicht: „Der fünfte Buchstabe des Mädchennamens der Mutter, oder der letzte, wenn der Name weniger als fünf Buchstaben hat.“).

Kriterien nach (AUDETTE et al., 2020: S. 178ff):

- a. Salienz,
- b. Konstanz,
- c. keine Sensitivität,
- d. Konsistenz des Formats (bspw. Name der Schule: Abkürzung oder ausgeschrieben),
- e. Schwierigkeit der Dekodierung.

Kriterien nach (YUREK et al., 2008: 439ff)

- a. Stabilität,
- b. Variabilität,
- c. Relevanz für Befragungsperson.

Zusammenfassend werden folgende Empfehlungen in der Literatur wiederholt genannt: Die Antworten auf die Fragen dürfen sich im Zeitverlauf nicht ändern (Stabilität) (u. a. YUREK et al., 2008). Codefragen müssen für alle Befragten verständlich sein, um zusätzlichen Datenverlust und eine Verzerrung der Befragungsdaten in Hinsicht auf das Sprachverständnis der Befragten zu verhindern (u. a. PÖGE, 2008: S. 69ff).

Weiterhin müssen die Antworten den Befragten bewusst sein, dies ist typischerweise bei Fragen der Fall, die sie selbst betreffen oder die für sie relevant sind (AUDETTE et al., 2020; YUREK et al., 2008). Andererseits gilt es, sensitive Fragen zu vermeiden (u. a. AUDETTE et al., 2020). Fragen müssen eine hinreichende Antwortvarianz ermöglichen (YUREK et al., 2008). Weiterhin sollen sie einerseits einen eindeutigen Code generieren (DIRENGA et al., 2016; PÖGE, 2008), andererseits soll es nicht möglich sein, darüber auf eine reale Person zu schließen (De-Anonymisierung). Die verschiedenen Kriterien stehen teilweise in einem Konfliktverhältnis zueinander oder zu anderen Sachzwängen, es ist daher in der Realität nicht möglich, allen Empfehlungen gleichermaßen Rechnung zu tragen.

2. Selbstgenerierte Codes für die BIBB/TU DA-Berufsorientierungsstudie

2.1. Generierung der Codefragen (Befragung der Ausbildungsbotschafter/-innen)

Die Befragungen der BIBB/TU DA Berufsorientierungsstudie finden im Rahmen des Besuchs von Abbos in Schulen bzw. einer Unterrichtseinheit statt. Geplant ist im Rahmen eines Cross Over-Designs zwei Berufe zu den verschiedenen Messzeitpunkten zu bewerten. Die Information, welcher Beruf zuerst bewertet wurde, ist ebenfalls bekannt und muss nicht erhoben werden.

Daher liegen über die Befragten durch die Beteiligung an der Erhebung, zum Zeitpunkt der Zuordnung, folgende Angaben vor:

- Schule,
- Klasse/Gruppe,
- Datum der Befragung,
- Reihenfolge der Berufsbewertungen.

Das bedeutet, dass falsche und fehlende Angaben seitens der SuS für diese vier Codeelemente ausgeschlossen werden können. Mithilfe der ersten drei Informationen können Teilgruppen von etwa 40 Personen zugeschnitten werden. Die vierte Information (Reihenfolge der Berufsbewertungen) ermöglicht es, diese Gruppe abermals zu halbieren.

Den Empfehlungen von DIRENGA et al. (2016) folgend fehlen noch vier Elemente, um die empfohlene Anzahl von acht Codeelementen zu erreichen. Eine geeignete Information, das Geschlecht, wird zum Zeitpunkt M1 bereits im Fragebogen erhoben und bietet sich daher besonders an. Zu den anderen Befragungszeitpunkten (M2, M3) wird diese Angabe im Codegenerierungsbogen erhoben.

Die letzten drei gewählten Codefragen erheben den Geburtsmonat, den ersten Buchstaben des Vornamens und die Anzahl der älteren Geschwister. Alle Fragen wurden in der einschlägigen Literatur erfolgreich angewendet (u. a. YUREK et al., 2008 (Geburtsmonat); DILORIO et al., 2000; PÖGE, 2011 (Initial Vorname); DILORIO et al., 2000; YUREK et al. 2008 (ältere Geschwister) und erfüllen weitgehend die Empfehlungen des ersten Teils dieses Textes. Wegen der Problematik der Einhaltung des Datenschutzes bei Befragungen an Schulen (vgl. u. a. PÖGE, 2005: S. 59) und da die ersten fünf Angaben, also die vier bereits vorliegenden Informationen und das Geschlecht, bereits eine gute Eingrenzung ermöglichen, wird vorgeschlagen, die übrigen Codes in aggregierter Form zu erheben. Denn es besteht zumindest die theoretische Möglichkeit, dass die in der Klasse ausgefüllten Fragebögen durch Unbefugte mit hinreichendem Zusatzwissen de-anonymisiert werden könnten. Für den Geburtsmonat soll daher nur das Quartal, für das Initial nur die Buchstaben von A bis H, I bis P und Q bis Z, für die Anzahl der älteren Geschwister nur jeweils Angaben „0“, „1“, „2 und mehr“ erhoben werden. Abbildung 2 zeigt, wie die Angaben erfragt werden sollen:

Abbildung 2: Fragen zur Generierung des Identifikationscodes (Zeitpunkt M2)

14. Wir benötigen nun wieder einen Zuordnungsschlüssel (Code). Dafür sind die folgenden Fragen. Es sind dieselben wie vorhin.

Monat deiner Geburt:

<input type="checkbox"/> Januar, Februar oder März	<input type="checkbox"/> April, Mai oder Juni	<input type="checkbox"/> Juli, August oder September	<input type="checkbox"/> Oktober, November oder Dezember
---	--	---	---

erster Buchstabe deines ersten offiziellen Vornamens (deines ersten Vornamens wie in einem Ausweis):

<input type="checkbox"/> A, Ä, B, C, D, E, F, G oder H	<input type="checkbox"/> I, J, K, L, M, N, O, Ö, oder P	<input type="checkbox"/> Q, R, S, T, U, Ü, V, W, X, Y oder Z
---	--	---

Anzahl deiner Geschwister, die älter sind als du (einschließlich Halb- und Stiefgeschwister):

0 1 2 und
mehr

dein Geschlecht:

weiblich männlich

Quelle: Entwurf Fragebogen Abbo M2 (Stand: 31.01.2020) BIBB/TU DA Berufsorientierungsstudie 2020

2.2. Abgleich mit den Gütekriterien

Die richtigen Antworten auf die Codefragen sind zeitlich stabil. Es kann darauf vertraut werden, dass sich Geburtsmonat, Vorname und Geschlecht im Befragungszeitraum nicht verändern. In Bezug auf Geschwister wird explizit nach älteren Geschwistern gefragt um die Stabilität zu erhöhen. Der Einschluss von Stiefgeschwistern vermindert diesen Effekt zwar, es überwiegt jedoch die Überlegung, dass der Einschluss aller Geschwister eindeutiger und verständlicher ist, als der Ausschluss von bspw. nur den Stiefgeschwistern. Möglich wäre zudem ein Hinweis auf den Einschluss verstorbener Geschwister. Dies könnte die zeitliche Stabilität der Frage erhöhen. Hinsichtlich der Notwendigkeit auch eine Regelung über totgeborene Geschwister vorzunehmen, der steigenden Komplexität der Frage und der Belastung,

die Befragten mit dieser Thematik zu konfrontieren, sowie des vergleichsweise kurzen Befragungszeitraums wird davon jedoch abgesehen.

Um die Verständlichkeit der Code-Fragen zu gewährleisten, wurde auf komplizierte Anweisungen verzichtet. Es wird bspw. nach dem ersten Buchstaben des Vornamens gefragt, nicht nach dem dritten oder dem letzten Buchstaben, da für dieses Vorgehen Probleme erwartet werden (vgl. PÖGE, 2005: S. 59). Um eventuelle Unklarheiten zu beseitigen, wurden Erklärungen und Beispiele eingefügt, wie der Hinweis auf den offiziellen Namen in einem Ausweis, oder die Berücksichtigung von Halb- und Stiefgeschwistern.

Den Empfehlungen in der Literatur folgend betreffen die Fragen die Probanden selbst, da diese für sie leichter abrufbar sind. Es wird beispielsweise nach dem Namen der Befragten, nicht nach dem der Mutter gefragt, da letztere Frage schlechtere Passungswerte liefert (vgl. AUDETTE et al., 2020: S. 179f). Gleichzeitig wurden zu persönliche, sensitive Fragen möglichst vermieden, so wurden über die Aggregation keine Daten über sehr große Familien erfasst.

Es kann weiterhin von einer hinreichenden Antwortvarianz ausgegangen werden. Zwar schwanken Geburten im Jahresverlauf, jedoch konnte auf Basis von Daten der Genesis-Online Datenbank des Bundesamtes für Statistik (GENESIS-ONLINE DATENBANK (A): Abruf 30.01.2020; eigene Berechnungen) nachgewiesen werden, dass es sich zumindest für den relevanten Zeitraum von 2000 bis 2005 nur um eine leichte Schwankung handelt. Die Statistik der Lebendgeborenen im verfügbaren Jahr 2018 ergab Anteile von 46,5 Prozent Erstgeborenen, 35,3 Prozent Zweitgeborenen und 18,2 Prozent mit zwei oder mehr älteren Geschwistern (DESTATIS, 2019). Es wird erwartet, dass diese Ergebnisse näherungsweise auf frühere Jahrgänge übertragbar sind.

Eine Ungleichverteilung der Schüleranteile nach Geschlecht betrifft insbesondere Hauptschulen. Basierend auf Daten der Genesis-Online Datenbank, im Jahr 2017/2018 wurde der Anteil der männlichen Hauptschüler, für zwei der für die Befragung relevanten Bundesländer berechnet. Die Anteile wurden mit 60,3 Prozent (Hessen) und 58,6 Prozent (Nordrhein-Westfalen) als vertretbar angesehen (GENESIS-ONLINE DATENBANK (B): Abruf 31.01.2020; eigene Berechnungen). Eine eigene Auswertung der angegebenen Anfangsbuchstaben einer Schülerbefragung (POLLICH, 2010: S. 187; eigene Berechnungen) ergab zudem für den oben ausgeführten Zuschnitt der Buchstabengruppen in der BIBB-TU DA Berufsorientierungsstudie, eine Besetzung von 35 Prozent, 39 Prozent und 25 Prozent. Da nur wenige Nennungen mit Umlauten oder Sonderzeichen zu erwarten sind, werden diese aus datenschutzrechtlichen Überlegungen zusammen mit den anderen Kategorien abgefragt.

2.3. Zu den einzelnen Fragen

Die Frage nach dem Geburtsmonat statt der in der Literatur verbreiteteren Frage nach dem Geburtstag (u. a. POLLICH, 2010: S. 7) ermöglicht es, die Antwortmöglichkeiten auszuschreiben und somit Flüchtigkeitsfehler, durch eine Verwechslung von Ziffern beim Lesen, zu reduzieren. Es handelt sich um eine Frage mit persönlicher Relevanz für die Befragten, deren Antwort daher auch leicht abrufbar ist.

Auch der eigene Vorname ist persönlich relevant und erreichte in einer Untersuchung von PÖGE (2011: S. 120f) sehr gute Fehlerraten, die mit 0,5 Prozent niedriger, als die für den Vornamen der Mutter (1,1 %) und des Vaters (1,6 %) waren. Die Tendenz, dass eigene Namen eine bessere Zuordnung ermöglichen, als die tatsächlichen Namen von Familienangehörigen, zeigte sich auch in der Übersichtsstudie von AUDETTE et al. (2020: S. 179), auch wenn Daten nur für den in Deutschland nicht gebräuchlichen Mittelnamen vorliegen. In den Empfehlungen von LIPPE et al. (2019) findet sich ebenfalls eine aggregierte Abfrage des eigenen Vornamens, allerdings mit den Grenzen A bis M und N bis Z (LIPPE et al., 2019: S. 122).

Zwar schneiden die Fragen zu den Geschwistern in Bezug auf ihre Passungsrate schlechter ab, als die anderen verwendeten Fragen, nach sorgfältiger Abwägung aller Gesichtspunkte, auch des Risikos einer De-Anonymisierung und der erwarteten Akzeptanz in der Zielgruppe, handelt es sich jedoch um die bestmögliche Option für die letzte Codefrage.

2.4. Layout und Gestaltung

Ausgehend von den Erfahrungen von PÖGE (2005: S. 58) wurde zugunsten von Ankreuzfeldern weitgehend auf händische Angaben verzichtet, da hierbei die Gefahr unleserlicher Eingaben besteht. Dem steht zwar der Effekt entgegen, dass das vollständige, handschriftliche Ausfüllen Flüchtigkeitsfehler reduziert, gleichzeitig bieten die vorgegebenen Antwortoptionen aber auch eine Orientierung, was genau gefragt ist (DIRENGA, 2016: S. 4ff; LIPPE et al., 2019: S. 121).

2.5. Zuordnung der Fragebögen und weiteres Vorgehen

Das Standardzuordnungsverfahren, wie u. a. von SCHNELL et al. (2010) beschrieben, sieht folgendes Vorgehen vor: Im ersten Schritt werden perfekt übereinstimmende Fragebögen zugeordnet. Alle nur partiell übereinstimmenden Identifikationscodes werden im Anschluss zugeordnet, wobei zunächst eine, im nächsten Schritt gegebenenfalls zwei Abweichungen erlaubt sind. Es existieren zudem komplexere Zuordnungsverfahren, etwa auf Basis der Levenshtein-Distanz (SCHNELL et al., 2010: S. 393 & 399). Ob diese für die BIBB/TU DA Berufsorientierungsstudie zur Anwendung kommen, hängt auch von der mit dem Standardverfahren realisierbaren Zuordnungsrate ab.

Als zusätzliche Maßnahme zur Sicherstellung der Anonymität der Befragten werden die Angaben zu den Schulen nach der Zuordnung durch eine laufende Nummer ersetzt. Für die Schulen, in denen eine Messung nur zu den Zeitpunkten M1 und M2 stattfindet, werden die zu diesen beiden Zeitpunkten jeweils von der gleichen Person ausgefüllten Fragebögen, unmittelbar im Anschluss an die Erhebung einander zugeordnet. Gespeichert werden in der zur Auswertung bestimmten Datendatei lediglich eine laufende Nummer sowie der Arbeitsagenturbezirk, zu dem die Schule gehört. Werden zusätzlich Messungen zu den Zeitpunkten M3 und M4 vorgenommen, werden die Klartextangaben zur Schule in einer gesonderten Schnittstellendatei hinterlegt, die sowohl die laufende Nummer, als auch die Angaben zur Schule enthält. Die Angaben zur Schule werden nach der erfolgreichen Zuordnung der Fragebögen im Anschluss an M3 und M4 gelöscht. In die Auswertungsdatei gelangen auch in diesem Fall nur die laufende Nummer und der Arbeitsagenturbezirk der Schule.

2.6. Zusammenfassung

Durch Befragte selbst generierte Identifikationscodes sind eine effektive Methode zur Zuordnung von Fragebögen in Panelbefragungen bei gleichzeitiger Einhaltung des Datenschutzes. Hohe Verlustraten durch fehlerhafte oder anderweitig abweichende Angaben zu verschiedenen Zeitpunkten stellen jedoch eine Herausforderung dar, die durch die sorgfältige Auswahl von Fragen zur Codegenerierung bewältigt werden muss. Im Rahmen der BIBB/TU DA Berufsorientierungsstudie soll daher zur Generierung des Identifikationscodes zusätzlich auf bereits existierende Angaben (Schule, Klasse/Gruppe, Datum der Befragung und die Reihenfolge, in der Berufe bewertet wurden) zurückgegriffen werden. Auf diese Weise müssen lediglich vier der insgesamt acht Codeelemente bei den Befragten erhoben werden. Durch die Anwendung des beschriebenen Vorgehens wird eine gute Datenqualität in Bezug auf die Zuordnungsrate der Fragebögen erwartet.

Literatur

AUDETTE, Lillian M.; HAMMOND, Marie S.; ROCHESTER, Natalie K.: Methodological Issues With Coding Participants in Anonymous Psychological Longitudinal Studies. In: Educational and Psychological Measurement, 80, (2020) 1, S. 163-185. – DOI: 10.1177/0013164419843576 (Stand: 20.01.2020)

DESTATIS (BUNDESAMT FÜR STATISTIK): Lebendgeborene nach der Geburtenfolge 2018. Wiesbaden 2019. – URL: <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bevoelkerung/Geburten/Tabellen/lebendgeborene-insgesamt.html> (Stand: 31.01.2020)

DILORIO, Colleen; SOET, Johanna E.; VAN MARTER, Deborah; WOODRING, Tammy M.; DUDLEY, William N.: Focus on Research Methods: An Evaluation of a Self-Generated Identification Code. In: Research in Nursing & Health, 23, (2000) 2, S. 167–174. – DOI: 10.1002/(SICI)1098-240X(200004)23:2<167::AID-NUR9>3.0.CO;2-K (Stand: 13.03.2020 15:54)

DIRENGA, Julie; TIMMERMANN, Dion; LUND, Jorrid; KAUTZ, Christian: Design and Application of Self-Generated Identification Codes (SGICs) for Matching Longitudinal Data. 44. SEFI Conference. Tampere (Finnland), 12.09. - 15.09.2016. – URL: http://sefibenvwh.cluster023.hosting.ovh.net/wp-content/uploads/2017/09/direnga-design-and-application-of-self-generated-identification-codes-172_a.pdf (Stand: 24.01.2020)

GENESIS-ONLINE DATENBANK (A): Lebendgeborene: Deutschland, Monate, Geschlecht (12612-0002). Wiesbaden. – URL: <https://www-genesis.destatis.de> (Stand: 31.01.2020)

GENESIS-ONLINE DATENBANK (B): Bundesländer, Schuljahr, Geschlecht, Schulart (21111-0003). Wiesbaden. – URL: <https://www-genesis.destatis.de> (Stand: 31.01.2020)

LIPPE, Megan; JOHNSON, Bailey; CARTER, Patricia: Protecting Student Anonymity in Research Using a Subject-Generated Identification Code. In: Journal of Professional Nursing, 35, (2019) 2, S. 120-123

MATTHES, Stephanie: Warum werden Berufe nicht gewählt? Die Relevanz von Attraktions- und Aversionsfaktoren in der Berufsfindung. Bonn 2019. – URL: <https://www.bibb.de/veroeffentlichungen/de/publication/show/9795> (Stand: 25.03.2020)

PÖGE, Andreas: Persönliche Codes bei Längsschnittstudien: Ein Erfahrungsbericht. In: ZA-Information / Zentralarchiv für Empirische Sozialforschung, 56, (2005), S. 50-69. – URL: <https://www.ssoar.info/ssoar/handle/document/19853> (Stand: 24.01.2020)

PÖGE, Andreas: Persönliche Codes 'reloaded'. In: Methoden - Daten - Analysen, 2, (2008) 1, S. 59-70. – URL: <https://www.ssoar.info/ssoar/handle/document/12654> (Stand: 24.01.2020)

PÖGE, Andreas: Fehlertolerante Zuordnung unverschlüsselter und verschlüsselter selbstgenerierter Codes im empirischen Test. In: Methoden - Daten - Analysen, 5, (2011) 1, S. 109-134. – URL: <https://www.ssoar.info/ssoar/handle/document/25504> (Stand: 24.01.2020)

POLLICH, Daniela: Methodendokumentation der kriminologischen Schülerbefragung in Duisburg 2002 und 2007 - Sechs-Wellen-Befragung. In: Schriftenreihe: Jugendkriminalität in der modernen Stadt – Methoden Nr. 16/2010. Münster, Bielefeld 2010, S. 1-207. – URL: https://www.uni-bielefeld.de/soz/krimstadt/pdf/Methodenbericht_Duisburg_Sechs-Wellen-Panel.pdf (Stand: 24.01.2020)

SCHNELL, Rainer: Selbst generierte Codes für Panelstudien. In: SCHNELL, Rainer (Hrsg.): Survey-Interviews: Methoden standardisierter Befragungen. 2. Aufl. Wiesbaden 2019, S. 391-394

SCHNELL, Rainer; BACHTALER, Tobias; REIHER, Jörg: Improving the Use of Self-Generated Identification Codes. In: Evaluation Review, 34, (2010) 5, S. 391-418. – DOI: 10.1177/0193841X10387576 (Stand: 20.01.2020)

YUREK, Leo A.; VASEY, Joseph; HAVENS, Donna Sullivan: The Use of Self-Generated Identification Codes in Longitudinal Research. In: Evaluation Review, 32, (2008) 5, S. 435-452. – DOI: 10.1177/0193841X08316676 (Stand: 20.01.2020)