

Auswirkungen schulbezogener Itemformulierungen auf die gemessene Bias-Awareness von Lehrkräften

vorgelegt an der
Albert-Ludwigs-Universität Freiburg

von Lorenz Bung

18.12.2024

universität freiburg



MASTERARBEIT

zur Erlangung des akademischen Grades

Master of Education (M. Ed.)

an der

Albert-Ludwigs-Universität Freiburg

Wirtschafts- und Verhaltenswissenschaftliche Fakultät

Thema: Auswirkungen schulbezogener Itemformulierungen auf die gemessene Bias-Awareness von Lehrkräften

Kandidat: Lorenz Bung
 Schlossbergring 36
 79098 Freiburg

Prüfer: Prof. Dr. Meike Bonefeld

Ausgabedatum: 01.12.2024

Abgabedatum: 18.12.2024

Inhaltsverzeichnis

| | |
|--|-----------|
| Abbildungsverzeichnis | IV |
| Tabellenverzeichnis | V |
| 1 Einleitung | 1 |
| 2 Bisherige Forschung und Theorie | 3 |
| 2.1 Kognitive Verzerrungen | 3 |
| 2.2 Implizite Stereotypen | 8 |
| 2.3 Bias Awareness | 11 |
| 2.4 Implicit Bias und Bias Awareness im schulischen Kon- text | 13 |
| 2.5 Itemformulierung und Erfassung der Bias Awareness . | 15 |
| 3 Forschungsgegenstand | 17 |
| 3.1 Ziele und Leitfragen | 17 |
| 3.2 Hypothesen | 19 |
| 3.3 Methodik | 20 |
| 4 Auswertung und Ergebnisse | 23 |

| | | |
|----------|--|-----------|
| 4.1 | Datenaufbereitung | 23 |
| 4.2 | Deskriptive Statistik | 26 |
| 4.2.1 | Gesamte Stichprobe | 26 |
| 4.2.2 | Lehramtsstudierende | 27 |
| 4.2.3 | Lehrkräfte | 28 |
| 4.2.4 | Skalenkonsistenz | 28 |
| 4.3 | Inferenzstatistik | 29 |
| 4.3.1 | Überprüfung der Verteilungen | 30 |
| 4.3.2 | Ergebnisse zur gesamten Stichprobe | 31 |
| 4.3.3 | Ergebnisse zu Lehramtsstudierenden | 34 |
| 4.3.4 | Ergebnisse zu Lehrkräften mit Praxiserfahrung | 38 |
| 4.3.5 | Differenzen zwischen Lehramtsstudierenden und Lehrkräften | 42 |
| 5 | Diskussion | 43 |
| 5.1 | Aussagekraft der Ergebnisse | 43 |
| 5.1.1 | Unterschiede bei schulkontextspezifischer Er- fassung der Bias Awareness (H1) | 44 |
| 5.1.2 | Einfluss eigener Schülerinnen und Schüler auf die gemessene Bias Awareness (H2) | 46 |
| 5.1.3 | Differenzen bei Erfassung der Bias Awareness in Bezug auf stereotypisierte Gruppen (H3) . . | 47 |

| | | |
|----------|---|-------------|
| 5.1.4 | Kontraste zwischen Lehramtsstudierenden und Lehrkräften mit Praxiserfahrung (H4) | 48 |
| 5.2 | Limitationen | 49 |
| 5.3 | Zukünftige Arbeit | 52 |
| 6 | Fazit und Zusammenfassung | 54 |
| | Literaturverzeichnis | I |
| A | Verwendete Skalen zur Erhebung der Bias Awareness | XIII |
| B | R-Code zur statistischen Datenanalyse | XX |

Abbildungsverzeichnis

| | | |
|-----|---|----|
| 2.1 | Cognitive Bias Codex | 4 |
| 2.2 | Kreislauf der Integration von neuen Informationen . . . | 5 |
| 4.1 | Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala | 31 |
| 4.2 | Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala | 33 |
| 4.3 | Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende | 35 |
| 4.4 | Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende | 36 |
| 4.5 | Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung | 38 |
| 4.6 | Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung | 40 |

Tabellenverzeichnis

| | | |
|-----|--|----|
| 2.1 | Stereotype Content Model | 10 |
| 4.1 | Cronbach's Alpha (raw) der verschiedenen Bias Awareness Skalen bei unterschiedlichen Stichproben | 29 |
| 4.2 | Ergebnisse des Shapiro-Wilk-Tests der Mittelwerte der verschiedenen Bias Awareness Skalen bei unterschiedlichen Stichproben | 30 |
| 4.3 | Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung | 32 |
| 4.4 | Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala | 34 |
| 4.5 | Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung, eingeschränkt auf Lehramtsstudierende | 35 |
| 4.6 | Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende | 37 |

| | | |
|-----|---|----|
| 4.7 | Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung, eingeschränkt auf Lehrkräfte mit Praxiserfahrung | 39 |
| 4.8 | Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung | 41 |
| 4.9 | Ergebnisse der zweiseitigen Welch-t-Tests für unabhängige Stichproben beim Vergleich von Lehramtsstudierenden und Lehrkräften | 42 |

1. Einleitung

Die Tätigkeit als Lehrkraft ist ein sehr komplexes Handlungsfeld. Nicht nur innerhalb des Unterrichtsgeschehens, sondern auch außerhalb davon – beispielsweise bei der Leistungsbeurteilung von Schülerinnen und Schülern – werden hohe kognitive Anforderungen an Lehrkräfte gestellt, denen sie in häufig geringer Zeit gerecht werden müssen.

Nach der *cognitive load theory* ([Atkinson und Shiffrin, 1968](#)) handelt es sich beim menschlichen Arbeitgedächtnis um einen limitierten Speicher, der nur wenige Informationen gleichzeitig verarbeiten kann. Um trotz dieser Begrenzung nicht im Informationsüberfluss zu ertrinken, werden schnelle heuristische Methoden verwendet. Infolgedessen unterliegen Menschen vielen kognitiven Verzerrungen, welche ihre Wahrnehmung oder Urteilsfähigkeit beeinflussen können.

Eine häufig auftretende kognitive Verzerrung ist die Bildung von impliziten Stereotypen, also eine Reduzierung von Individuen auf bestimmte Merkmale oder Überzeugungen ([Greenwald und Banaji, 1995](#)). Es hat sich gezeigt, dass implizite Stereotypen häufig die Ursache offener Diskriminierung sind ([Agerström und Rooth, 2011](#); [Picho et al., 2013](#)). Auch für den schulischen Kontext konnte gezeigt werden, dass durch Lehrkräfte stereotypisierte Schülerinnen und Schüler unter negativen Folgen leiden ([Martiny und Froehlich, 2020](#)). Das Ausmaß und die negativen Auswirkungen impliziter Stereotypen lassen sich jedoch reduzieren, wenn sich die entsprechenden Personen über deren Existenz bewusst werden ([Monteith, 1993](#); [Perry et al.,](#)

2015). Der Grad dieses Bewusstseins wird Bias Awareness genannt.

Bei der Erfassung der Bias Awareness kommen häufig Fragebögen zum Einsatz, z.B. bei [Perry et al. \(2015\)](#). In verschiedenen Studien gab es Hinweise darauf, dass die Formulierung der Fragen einen großen Einfluss auf den gemessenen Wert haben kann ([Kaminski, 2017](#); [Bing et al., 2004](#); [Guyatt et al., 1999](#)).

Es stellt sich daher die Frage, welchen Einfluss die schulbezogene Formulierung der Items zur Erfassung der Bias Awareness auf den gemessenen Wert hat. In dieser Arbeit soll dieser Frage nachgegangen und analysiert werden, ob es diesbezüglich zusätzliche Differenzen zwischen Lehramtsstudierenden und Lehrkräften mit Praxiserfahrung gibt.

2. Bisherige Forschung und Theorie

Um die notwendige Basis für die vorliegende Arbeit zu schaffen, wird zunächst ein Blick auf die bisherige Literatur geworfen. Beginnend mit einem allgemeinen Blick auf kognitive Verzerrungen als Teil der psychologisch-kognitiven Forschung, ihren Ursachen sowie Beispielen der genannten Phänomene liegt der Fokus insbesondere auf impliziten Stereotypen. Als Möglichkeit, die Auswirkungen des implicit bias zu reduzieren wird anschließend eine Einführung zur Thematik der Bias Awareness geliefert sowie bisherige Literatur dazu überprüft. Anschließend wird diskutiert, inwiefern Biases im schulischen Kontext präsent sind und in welchen Aspekten sich ihre Auswirkungen auf das Lehren und Lernen äußern. Schließlich wird darauf eingegangen, welche Einflüsse die Itemformulierung bei Fragebögen auf die Erfassung der Bias Awareness hat.

2.1 Kognitive Verzerrungen

Unter kognitiven Verzerrungen (englisch: *cognitive biases* oder *cognitive illusions*) versteht man psychologische Effekte, welche zu einer Wahrnehmung oder Beurteilung führen, die von der realen Situation abweicht (Pohl, 2004). Die zahlreichen Manifestationen dieser Verzerrungen äußern sich auf verschiedene Art und Weise – einen Überblick liefert [Abbildung 2.1](#).

Wie in [Abbildung 2.2](#) dargestellt werden bei der Integration neuer Informationen vier Phasen durchlaufen. Aus der riesigen Menge der

2. Bisherige Forschung und Theorie

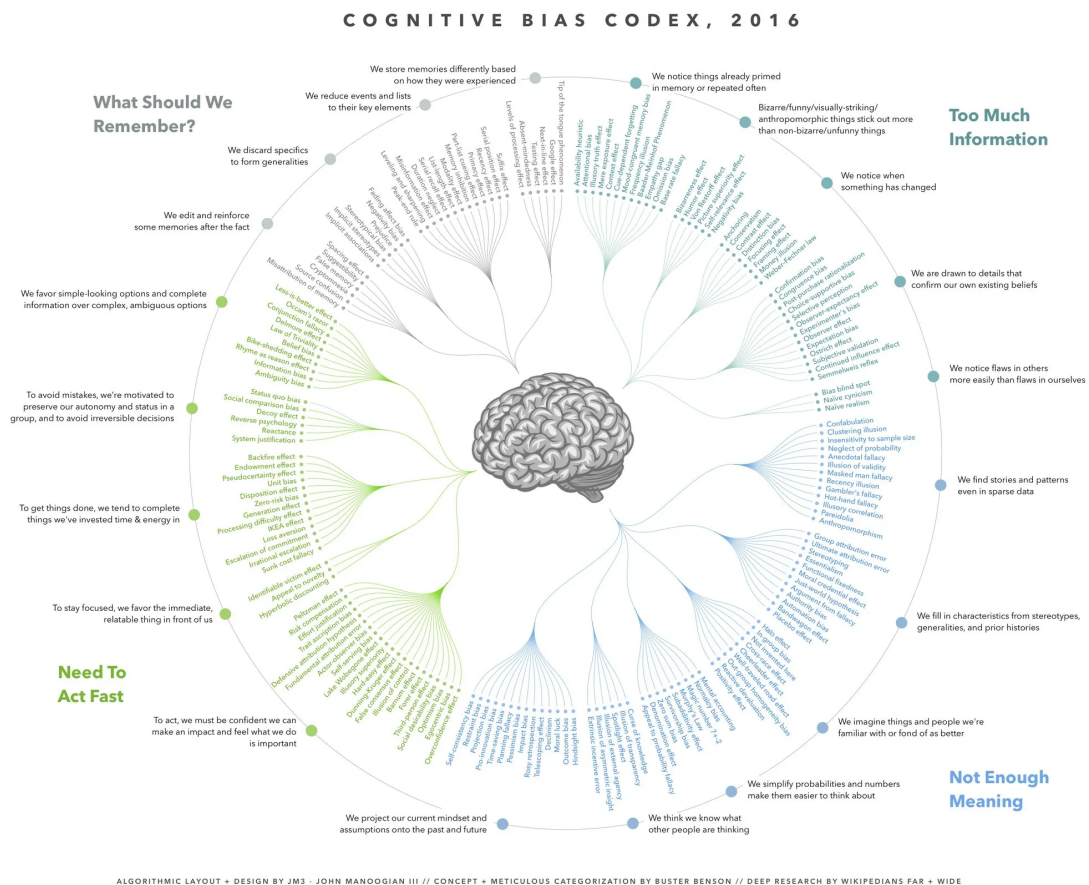


Abbildung 2.1: Cognitive Bias Codex – Beispiele verschiedener kognitiver Biases (Benson, 2016)

auf uns einströmenden Informationen werden zunächst einige wenige selektiert, aus welchen wir uns mithilfe unserer bisherigen Erinnerung einen Sinn machen können. Unser Vorwissen wird dann genutzt, um sie in einen Kontext einzubetten, welcher anschließend unser Handeln beeinflusst. Diese Handlungen fließen danach wiederum in unsere Erinnerung ein, auf die bei neuen Informationen zugegriffen wird.

Bei jedem Phasenübergang finden dabei Filtereffekte statt. Nach Benson (2016) kristallisieren sich dabei vier Problemstellungen heraus, mit denen das menschliche Gehirn tagtäglich konfrontiert wird. Jede der folgenden Kategorien repräsentiert dabei einen der dargestellten Übergänge zwischen den Phasen.

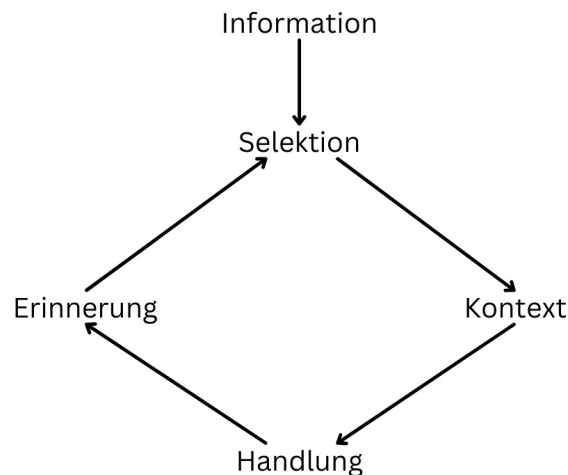


Abbildung 2.2: Kreislauf der Integration von neuen Informationen

1. **Informationsüberflutung:** Bereits seit Beginn der Existenz menschlicher Aufzeichnungen von Informationen existiert das Problem der Überflutung durch ebenjene (Bawden und Robinson, 2020). Aufgrund technischer Entwicklungen und der Verbreitung des Internets hat sich die Menge an Informationen, die auf uns einwirken, insbesondere in den letzten Jahrzehnten extrem vergrößert (Siegler und Schmidt, 2010). So gaben 22.5% der Deutschen im Jahr 2021 Informationsüberflutung als eine der Hauptursachen für Stress an (Meyer et al., 2021). Dies beeinflusst auch die Qualität professioneller Entscheidungen, beispielsweise im Gesundheitssystem (Hall und Walton, 2004), beim Management von Personal (Volnhals und Hirsch, 2008) und Marketing (Meyer, 1998) oder in der Wissenschaft (Landhuis, 2016). Nach der *cognitive load theory* (Atkinson und Shiffrin, 1968) ist das menschliche Arbeitsgedächtnis sehr begrenzt. Aufgrund der übermächtigen Menge an Informationen, die auf uns einwirken, sind unsere Gehirne daher dazu gezwungen, aggressiv zu filtern (Savolainen, 2007). Das Resultat können fehlerhafte Schlussfolgerungen sein, beispielsweise in Form eines *confirmation bias* (Goette et al., 2019).

2. *Mangel an Kontext*: Die Welt, in der wir leben, ist sehr komplex. Dennoch ist es notwendig, dass wir aus den wenigen uns gegebenen Informationen korrekte Vorhersagen treffen und ihnen eine angemessene Bedeutung zuweisen. Das Erkennen von Mustern trotz eines Mangels an Kontext ist in der Realität häufig unausweichlich. Doch gerade deshalb kann es vorkommen, dass wir eine Übergeneralisierung vornehmen und Beziehungen zwischen Informationen herstellen, obwohl diese nicht oder anders existieren.

Ein Beispiel dafür ist die sogenannte *Konfabulation* (Gilboa und Moscovitch, 2002): Eine Person meint, sich an ein Ereignis zu erinnern und kann konkrete Erinnerungen dazu nennen – jedoch stellen sich diese bei Überprüfung als falsch bzw. ausgedacht heraus. Bekannt ist diese Art der kognitiven Verzerrung z.B. durch den *Mandela-Effekt*, eine kollektive Konfabulation, nach der Nelson Mandela bereits in den 1980er-Jahren in Haft (und nicht erst 2013 in Freiheit) gestorben sei. Ein ähnlicher Effekt kann auch in Bezug auf visuelle Erinnerungen beobachtet werden (Prasad und Bainbridge, 2022), beispielsweise erinnern sich viele Menschen an einen Monopoly-Mann mit Monokel, obwohl dieser nie mit Monokel dargestellt worden ist.

Besonders relevant für diese Arbeit ist auch die Bildung von *Stereotypen*. Dabei handelt es sich um “weitverbreitete, jedoch feste und stark vereinfachte Bilder oder Vorstellungen zu einem bestimmten Typ von Personen oder Dingen” (Bordalo et al., 2016, Übers. d. Verf.). Insbesondere in Bezug auf die Zugehörigkeit zu ethnischen Gruppen (Brigham, 1971; Guichard und Connolly, 1977; Mastro, 2009) und Geschlecht (Hoffman und Hurst, 1990; Heilman, 2012; Haines et al., 2016; Ellemers, 2018) gab und gibt es immer noch viel aktuelle Forschung, die sich mit den (insbesondere negativen) Auswirkungen von Stereotypen befasst.

3. *Notwendigkeit, schnell zu handeln*: Trotz der Fülle an Informationen, die auf uns einströmen, dürfen wir uns nicht durch das Abwägen von Optionen aufhalten lassen, sondern müssen zuverlässig schnelle und gute Entscheidungen treffen. Aus diesem Grund liegt unser Fokus häufig auf den offensichtlichen und einfachen Dingen, anstatt sich in den komplizierten Details der Situation zu verlieren. Ein Beispiel hierfür ist der *identifiable victim effect*, nach dem Menschen eher bereit sind, Ressourcen für konkrete und identifizierbare als für abstrakte, statistische Opfer bereitzustellen (Jenni und Loewenstein, 1997). Weiterhin beschäftigen wir uns eher mit Trivialitäten als komplexen Problemen (Parkinson, 1958), wie der *bike shed effect* visualisiert: Je weniger komplex das Problem, desto mehr Zeit wird dafür aufgewendet, eine Entscheidung darüber zu treffen.

Damit wir trotz des Informationsmangels nicht zögern zu handeln, überschätzen wir unsere Fähigkeiten und Urteilkraft häufig (*overconfidence effect*, Moore und Healy (2008)) und legen höheren Wert auf unsere eigene Wahrnehmung (Ross und Sicoly, 1979). Diese Überschätzung findet insbesondere dann statt, wenn unsere tatsächliche Kompetenz relativ niedrig ist, wie der bekannte *Dunning-Kruger-Effekt* (Dunning, 2011) zeigt.

4. *Fehlbarkeit von Erinnerungen*: Nachdem wir der Informationsflut mit radikaler Filterung und Selektion begegnet sind, ihr vermeintlichen Kontext zugewiesen haben und auf Basis dieser Entscheidungen gehandelt haben, stehen wir vor einem weiteren Problem. Über die Nützlichkeit dieser Informationen in der Zukunft muss spekuliert werden, und so kommt es zu Verzerrungen der Realität. Liegt eine Liste von Informationen vor, spielt beispielsweise die Position innerhalb dieser eine große Rolle für die Erinnerung (Murdock Jr, 1962; Phillips et al., 1967).

Dass unsere Erfahrungen beeinflussen, wie wir uns an Dinge erinnern, zeigt sich unter Anderem beim *self-reference effect* (Rogers et al., 1977). Hierbei sind Erinnerungen an Informationen, die uns nicht persönlich betreffen, deutlich weniger zuverlässig als jene, die eine hohe persönliche Relevanz haben. Auch der sogenannte *google effect* (Sparrow et al., 2011) spiegelt wider, dass unsere Erinnerung abhängig von der Situation ist: Im Wissen, dass die betreffende Information durch eine kurze Internetrecherche schnell wieder abrufbar ist, sehen wir nur eine reduzierte Notwendigkeit, sich insbesondere an spezielle Details zu erinnern.

2.2 Implizite Stereotypen

Beim *impliziten Stereotyp* (oder auch *implicit bias*) handelt es sich um “eine Reihe von Überzeugungen über Merkmale, welche für die Mitglieder einer sozialen Kategorie charakteristisch sind” (Greenwald und Banaji, 1995, Übers. d. Verf.).

Wie schon durch den Namen angedeutet, handelt es sich hierbei um ein Teilgebiet der Stereotypen, welche schon in [Abschnitt 2.1](#) angesprochen wurden. Um die Bedeutung des Begriffs abzugrenzen ist es essenziell hinzuzufügen, dass es sich beim *implicit bias* um unbewusste Einflüsse auf das Denken handelt. Die ebenfalls gebräuchliche Bezeichnung des *explicit stereotype* stellt im Gegensatz dazu aktive und bewusste Einstellungen dar.

Eine weitere wichtige Unterscheidung muss zwischen Stereotypen und Vorurteilen gemacht werden. Während es sich beim sozialpsychologischen Begriff des Stereotyps um kognitive Zuordnungen zu

einer bestimmten sozialen Kategorie handelt, bezeichnet das Vorurteil die subjektive Bewertung und emotionale Reaktion (Degner, 2022). Bei beiden Begriffen handelt es sich jedoch um Konzepte, welche die Einstellungen zu Menschen bestimmter sozialer Gruppen unabhängig von deren individuellen Eigenschaften widerspiegeln.

Die Bildung von Stereotypen und deren Anwendung im Alltag findet bereits im Kindesalter statt (Martiny und Froehlich, 2020). Schon bei Kleinkindern im Alter von wenigen Monaten findet eine Identifizierung charakteristischer sozialer Eigenschaften statt (Killen und Rutland, 2013). Dies geschieht im Rahmen sozialer Kategorisierung, bei welcher es sich um eine Vorstufe zur Stereotypenbildung handelt. Hierbei werden Menschen mit einer oder mehreren Gemeinsamkeiten zu einer sozialen Gruppe zusammengefasst. Die Einordnung von Objekten und Menschen in Kategorien ist ein Prozess, der regelmäßig und weitgehend automatisch stattfindet (Macrae und Bodenhausen, 2000). Diese Gruppierung hilft dabei, die Komplexität sozialer und so die damit einhergehende kognitive Belastung zu reduzieren sowie leichtere Vorhersagen zu treffen (Killen und Rutland, 2013). Beispielsweise ist die Erinnerung von Kindern an Informationen, welche ihrem kognitivem Schema entsprechen, stärker als wenn dies nicht der Fall ist (Bigler und Liben, 1993).

Beobachten lassen sich die Effekte dieser Kategorisierung ebenfalls bereits im frühen Kindesalter: So zeigen schon drei Monate alte Säuglinge Präferenzen gegenüber den Gesichtern von Mitgliedern der eigenen Ethnie gegenüber denen von Mitgliedern anderer Ethnien (Kelly et al., 2005). Das Vorhandensein impliziter Stereotype ließ sich ab einem Alter von 4–9 Jahren zeigen (Bigler und Liben, 1993), während explizite Vorurteile ebenfalls ab einem Alter von 4 Jahren gebildet werden (Aboud, 1993). Es wird davon ausgegan-

gen, dass viele Stereotype erlernt sind und stark durch das soziale Umfeld geprägt werden. Kinder nutzen dabei den Vergleich mit Beispielen (Martin und Halverson, 1981) und Gegenbeispielen (Slaby und Frey, 1975), um die eigene Gruppenzugehörigkeit abzugrenzen und erwartetes Verhalten vorherzusagen (Tajfel und Turner, 1979).

Nach dem *Stereotype Content Model* (Fiske et al., 2002) lassen sich Stereotype anhand der zwei Dimensionen *Wärme* und *Kompetenz* kategorisieren. Während die Dimension der Wärme beschreibt, ob die betreffende Gruppe als freundlich oder feindlich gesinnt wahrgenommen wird, stellt die Dimension der Kompetenz den zugewiesenen sozialen Status dar. Wie in Tabelle 2.1 dargestellt, werden je nach Wärme und Kompetenz unterschiedliche Emotionen bzw. Vorurteile mit den Stereotypen verbunden (Cuddy et al., 2008). So werden beispielsweise arme Menschen, welche selbst einen niedrigen sozio-ökonomischen Status haben, jedoch gleichzeitig als Bedrohung für den eigenen Status wahrgenommen werden, häufig mit der Emotion Verachtung in Verbindung gebracht.

| | | Kompetenz | |
|-------|---------|----------------------------------|------------------------------|
| | | niedrig | hoch |
| Wärme | hoch | Mitleid z.B. ältere Menschen | Bewunderung z.B. Freunde |
| | niedrig | Verachtung z.B. arme Menschen | Neid z.B. reiche Menschen |

Tabelle 2.1: Stereotype Content Model (Darstellung nach Fiske et al. (2002))

Auch wenn die Bildung von Stereotypen ein universell stattfindender Prozess ist, der grundsätzlich viele Vorteile mit sich bringt, lassen sich viele negative Effekte darauf zurückführen. Beispielsweise führen stereotypisierte Gruppen negatives Feedback eher auf Stereotype zurück, während positives Feedback eher als Mitleid interpretiert wird (Crocker et al., 1991). Weiterhin können implizite Ste-

reotype zu selbsterfüllenden Prophezeiungen führen (Latu et al., 2015). Zum Beispiel kann ein Arbeitgeber, der ein negatives Bild von Frauen in Führungspositionen hat, diese in Bewerbungsgesprächen durch seine Rückmeldungen negativ beeinflussen und somit zu einem schlechteren Ergebnis der Bewerberin führen. Zusätzlich können bekannte Stereotype das Selbstbild von Mitgliedern der betroffenen Personengruppe negativ beeinflussen (Cox et al., 2012; Ebert und Steffens, 2013). Dies kann zu einer Internalisierung der mit dem Stereotyp verbundenen Eigenschaften führen. Die Angst von Betroffenen, aufgrund von Stereotypen zu beurteilt zu werden oder existierende Vorurteile durch ihr Verhalten zu bestätigen, ist ein häufiges Phänomen, was unter dem Namen *stereotype threat* bekannt ist (Steele und Aronson, 1995; Keller und Dauenheimer, 2003). Der Effekt der selbsterfüllenden Prophezeiung akkumuliert sich außerdem (Madon et al., 2018), d.h. je mehr Menschen stereotypische Erwartungen haben, desto größer die Reaktion im Verhalten der betroffenen Person.

Stereotype und Vorurteile sind häufig die Ursache offener Diskriminierung. Beispielsweise lässt sich die Diskriminierung von bestimmten Gruppen bei Jobinterviews auf die Stereotypisierung zurückführen – ein Effekt, der sich bei übergewichtigen Menschen (Agerström und Rooth, 2011), Homosexuellen (Tilcsik, 2011) sowie weiblichen Bewerberinnen (Picho et al., 2013; Latu et al., 2015) beobachten lässt.

2.3 Bias Awareness

Wie bereits in Abschnitt 2.2 beschrieben handelt es sich beim implicit bias um kognitive Verzerrungen, über deren Existenz sich das

Individuum nicht bewusst ist. In Studien konnte jedoch gezeigt werden, dass Vorurteile nach einer Intervention, welche über deren Vorhandensein aufklärt, wesentlich geringer ausfallen (Penner, 1970; Grube et al., 1994). Auch neuere Studien konnten bestätigen, dass sich Menschen weniger diskriminierend verhalten oder die negativen Effekte ihrer Vorurteile kompensieren, wenn ihnen diese bewusst gemacht werden (Monteith, 1993; Monteith et al., 2010a,b; Pope et al., 2018). Rohner et al. konnten außerdem demonstrieren, dass der attractiveness bias bei Bewerbungsgesprächen reduziert werden kann, wenn die Versuchspersonen vor der Bewertung der Bewerbungen über die Existenz des Bias aufgeklärt wurden und das Ausmaß ihres eigenen Bias einschätzen mussten. Dieser Effekt trat nicht auf, wenn lediglich eine Aufklärung stattfand. Dies spricht dafür, dass erst die aktive Einschätzung der kognitiven Verzerrungen (eventuell durch eine damit verbundene Erhöhung der Bias Awareness) zur Reduzierung der negativen Auswirkungen führt.

Der Grad, zu welchem sich ein Individuum über die eigenen implicit biases bewusst ist, wird nach Perry et al. (2015) als *bias awareness* bezeichnet. Zusätzlich zu einer Reduktion der Auswirkungen des implicit bias konnten Perry et al. zeigen, dass eine höhere Bias Awareness dabei helfen kann, den Einfluss von Biases auf beobachtete Handlungen zu erkennen.

Doch warum reduzieren sich die Effekte der Vorurteile beim Bewusstwerden über deren Existenz? Nach der *Belief System Theory* (Rokeach, 1968; Grube et al., 1994) leben Menschen nach einer Menge an Werten, die ein zusammenhängendes System bilden. Ihre Handlungen und ihr Denken richten die Personen nach diesen Wertevorstellungen aus und nutzen diese, um ihr Handeln als gut oder schlecht zu bewerten. In manchen Situationen – beispielsweise nach der Aufklärung über Vorurteile – werden wir nun damit konfron-

tiert, dass unser Verhalten nicht im Einklang mit unserem Wertesystem steht. Diese Differenz führt dazu, dass wir zukünftig in ähnlichen Situationen stärker darauf achten, unsere Handlungen an unseren Werten auszurichten.

Nach [Perry et al. \(2015\)](#) gibt es trotz der generellen Tendenz, dass eine höhere Bias Awareness das Ausmaß von Vorurteilen sowie deren Effekte reduzieren kann, starke individuelle Unterschiede. So existieren beispielsweise Differenzen in der individuellen Motivation, sich frei von Vorurteilen zu verhalten ([Devine et al., 2002](#); [Plant und Devine, 1998](#)). Zusätzlich erkennen manche Personen auch nicht, dass ihr Verhalten ihren Werten widerspricht ([Monteith und Voils, 1998](#)) oder reagieren sich bei Konfrontation defensiv. Doch auch nach dem Erkennen der eigenen Vorurteile existieren Unterschiede im Ausmaß, zu dem die daraus resultierenden negativen Effekte ausgeglichen werden können ([Moskowitz und Li, 2011](#)).

2.4 Implicit Bias und Bias Awareness im schulischen Kontext

Es gibt zahlreiche Gründe dafür, dass auch heute noch viele Differenzen in Bezug auf schulischen Erfolg im deutschen Schulsystem bestehen. Im IQB-Bildungstrend 2021 konnten sowohl geschlechtsbezogene, soziale als auch zuwanderungsbezogene Disparitäten festgestellt werden ([Stanat et al., 2021](#)).

Die Erwartungen der Lehrkräfte haben einen signifikanten Einfluss auf die schulischen Erfolge sowie die Motivation der unterrichteten Kinder ([Gentrup et al., 2020](#)). Auch die impliziten Stereotype der Lehrkräfte spielen dabei eine Rolle für die Entwicklung von Unterschieden in Bezug auf Leistung und Motivation der Schülerinnen

und Schüler. Beispielsweise werden Kinder mit türkischem Migrationshintergrund und Jungen im Fach Deutsch bei gleichen Ergebnissen wie die Vergleichsgruppe schlechter bewertet, während Kinder mit osteuropäischem Migrationshintergrund oder mit höherem sozio-ökonomischen Status im Fach Mathematik bei gleichen Leistungen besser bewertet werden (Lorenz et al., 2016). Zahlreiche Studien konnten die Existenz und Auswirkungen sowohl geschlechtsbezogener (Keller und Dauenheimer, 2003; Hermann, 2020; Muntoni und Retelsdorf, 2020), ethnischer (Steele und Aronson, 1995; Glock und Karbach, 2015; Glock und Klapproth, 2017; Glock und Kleen, 2019; Costa et al., 2023) als auch sozioökonomischer Stereotype (Tobisch und Dresel, 2017; Kleen und Glock, 2020) im schulischen Kontext feststellen.

Martiny und Froehlich argumentieren, dass nicht nur die negative Bewertung durch biasbeeinflusste Lehrkräfte einen negativen Effekt auf den schulischen Erfolg der betroffenen Schülerinnen und Schüler hat, sondern auch das Gruppenzugehörigkeitsgefühl leidet (Martiny und Froehlich, 2020). Infolge des stereotype threats können Stereotype den betroffenen Personen das Gefühl geben, nur als Repräsentant ihrer Gruppe und nicht als Individuum gesehen zu werden und infolgedessen die wahrgenommene Eingebundenheit in der Klasse reduzieren (Walton und Carr, 2012). Dies wiederum kann sich negativ auf die Motivation und die Leistungen der Schülerinnen und Schüler auswirken.

Auch die Identifikation mit der Domäne, also beispielsweise dem akademischen Bereich an sich, leidet darunter, wenn die Zugehörigkeit zu einer bestimmten sozialen Gruppe stereotypisiert ist (Crocker und Major, 1989). Nach Osborne und Jones (2011) geht eine höhere Domänenidentifikation mit dem Bildungsbereich jedoch häufig mit besseren akademischen Leistungen einher, was weitere ne-

gative Effekte für stereotypisierte Personen nach sich zieht.

Zusätzlich zu den negativen Auswirkungen für Schülerinnen und Schüler durch implizite Stereotypen konnte auch gezeigt werden, dass implicit biases in Bezug auf ethnische Hintergründe zum Risiko für Burnout der Lehrkraft beitragen können ([Costa et al., 2023](#)).

Es gilt also, den implicit bias der Lehrkräfte so weit wie möglich zu reduzieren. Wie in [Abschnitt 2.3](#) bereits diskutiert, kann eine erhöhte Bias Awareness dazu beitragen, sowohl die Auswirkungen als auch das Ausmaß der impliziten Stereotype zu reduzieren.

Zu den Auswirkungen der Bias Awareness im schulischen Bereich gibt es bisher nur wenig Forschung. Es konnte jedoch im Rahmen einer Interventionsstudie gezeigt werden, dass eine Sensibilisierung für das Thema die Bias Awareness von Lehramtsstudierenden signifikant erhöhen kann ([Bonefeld, 2022](#)).

2.5 Itemformulierung und Erfassung der Bias Awareness

Eine große Rolle bei der Evaluierungen von Konstrukten wie der Bias Awareness spielen die Fragebogenitems, welche zur Erfassung des Konstrukts genutzt werden. Dabei spielt nicht nur die sprachliche Klarheit der Items eine Rolle, sondern auch die Reihenfolge, in welcher die verschiedenen Items angeordnet sind ([Baur und Blasius, 2014](#), S. 675ff.).

Einen entscheidenden Einfluss auf die Validität des Fragebogens hat auch die Formulierung der Fragebogenitems. So können gesellschaftliche Erwartungen die Antworten der getesteten Personen beeinflussen, sodass diese Angaben entsprechend der sozialen Er-

wünschtheit machen ([Kemper et al., 2012](#); [Kaminski, 2017](#)).

Weiterhin können kulturelle Unterschiede in der Formulierung der Fragebogenitems zu einer Verzerrung der Ergebnisse führen ([Harzing et al., 2013](#)). Es kann daher notwendig sein, Fragebögen an die Kultur der Zielgruppe entsprechend anzupassen ([Poortinga und van de Vijver, 2013](#)).

Doch nicht nur die sprachliche Klarheit und die kulturelle Komponente der Fragebogenitemformulierung haben einen Einfluss auf das Ergebnis der Befragung. Es spielt auch eine Rolle, ob die Fragen innerhalb der einzelnen Items allgemein oder kontextspezifisch formuliert sind. Beispielsweise konnten Bing et al. zeigen, dass eine kontextgebundene Itemformulierung Messfehler reduzieren und die Validität der Messung erhöhen kann ([Bing et al., 2004](#)).

Ein ähnlicher Effekt entsteht bei speziellem Framing der Fragen. Dabei wird durch die Formulierung bereits eine Assoziation bei den Versuchspersonen hervorgerufen, welche das Ergebnis beeinflussen kann ([Guyatt et al., 1999](#)).

Auch das Format der Items kann eine Rolle spielen: So reduziert eine nuancenreiche Auswahl an Antwortmöglichkeiten das Risiko für biasbehaftete Antworten ([Elson, 2017](#)). Eine zu große Anzahl an Möglichkeiten ist jedoch auch nicht erwünscht, da es so zur Überforderung der Befragten kommen kann ([Knutsen und Presser, 2010](#)). Wird eine Likert-Skala mit ungerader Anzahl an Auswahloptionen gewählt, kann es zusätzlich zu einem Mittelpunkt-Bias kommen, bei dem die mittlere Kategorie überdurchschnittlich häufig gewählt wird ([Kulas und Stachowski, 2009](#)).

3. Forschungsgegenstand

Nachdem im vorherigen Kapitel die theoretischen Grundlagen gelegt worden sind, geht es nun darum, das Thema der vorliegenden Arbeit genauer zu definieren. Dazu werden zunächst die Ziele der Arbeit formuliert sowie Leitfragen aufgestellt, welche erforscht werden sollen. Im Anschluss daran werden auf diesen Zielen bzw. Fragen basierende Hypothesen entwickelt, welche statistisch überprüft werden sollen. Nachdem diese Vorüberlegungen angestellt worden sind, werden der Entwurf sowie die Durchführung der Datenerhebung beschrieben und diskutiert.

3.1 Ziele und Leitfragen

Wie die bisherige Forschung gezeigt hat, unterliegen sehr viele Menschen dem Einfluss von implicit biases. Zusätzlich wurde gezeigt, dass die Bias Awareness einen maßgeblichen Einfluss auf den Grad der Auswirkungen dieser kognitiven Verzerrungen hat. Es wurde diskutiert, dass implicit biases und ihre Einflüsse auch in der Schule präsent sind und unter Anderem Auswirkungen auf die Bewertungen und den schulischen Erfolg der Schülerinnen und Schüler haben kann.

Wie bereits in [Abschnitt 2.5](#) diskutiert, kann die Formulierung von Fragebogenitems einen erheblichen Einfluss auf die damit gemessenen Werte haben. Ein Grund dafür ist das Framing der Fragen, welche die abgegebenen Antworten in eine bestimmte Richtung len-

ken können. Es stellt sich daher die Frage, ob die gemessene Bias Awareness von Lehrkräften geringer oder höher ist, wenn sie bei der Erfassung an Schülerinnen und Schülern denken. Gegebenenfalls verstärkt sich dieser Effekt auch, wenn es sich dabei um Schülerinnen und Schüler handelt, die sie zusätzlich selbst unterrichten.

Ebenfalls kann es sein, dass sich die gemessene Bias Awareness in Bezug auf eine stereotypisierte Gruppe gegenüber einer anderen stereotypisierten Gruppe unterscheidet. Zusätzlich könnten sich innerhalb einer Gruppe Unterschiede in der gemessenen Bias Awareness ergeben, wenn sie einerseits allgemein und andererseits in Bezug auf den schulischen Kontext erfasst wird.

Weiterhin stellt sich die Frage, ob es Unterschiede in Abhängigkeit von der Erfahrung der Lehrkräfte gibt. Es könnte beispielsweise sein, dass die Bias Awareness von Lehrkräften mit Berufserfahrung höher ist als die von Studierenden, welche bisher nur wenig oder keine Praxiserfahrung besitzen.

Basierend auf diesen Überlegungen lassen sich folgende Leitfragen formulieren:

- Unterscheidet sich die gemessene Bias Awareness bei allgemein formulierten Fragebogenitems gegenüber der gemessenen Bias Awareness bei Verwendung von schulspezifischen Formulierungen?
- Bestehen Unterschiede, wenn die Fragebogenitems so formuliert sind, dass Lehrkräfte Aussagen zu selbst unterrichteten Schülerinnen und Schülern treffen müssen?
- Spielt es bei der Erfassung der Bias Awareness eine Rolle, ob die Items schul- und stereotypenspezifisch (z.B. in Bezug auf Migrationshintergrund) formuliert sind im Vergleich zu allgemeiner, stereotypenspezifischer Formulierung?

- Gibt es Differenzen zwischen Lehrkräften und Lehramtsstudierenden?

Die vorliegende Arbeit soll einen Versuch darstellen, Antworten auf diese Leitfragen zu finden und aufzuzeigen, in welchen Bereichen zusätzliche Forschung zur Beantwortung der Fragen notwendig ist.

3.2 Hypothesen

Auf Basis der oben formulierten Ziele und Leitfragen für diese Arbeit lassen sich nun Hypothesen aufstellen, welche sich anschließend mit statistischen Methoden verifizieren oder falsifizieren lassen:

H1: *Die Bias Awareness von Lehrkräften gegenüber Schülerinnen und Schülern unterscheidet sich gegenüber Personen, die keine Schülerinnen oder Schüler sind.*

H2: *Die Bias Awareness von Lehrkräften gegenüber Schülerinnen und Schülern, die sie selbst unterrichten, unterscheidet sich im Vergleich zu ihrer Bias Awareness gegenüber Schülerinnen und Schülern allgemein.*

H3: *Die mit schul- und stereotypenspezifisch formulierten Items gemessene Bias Awareness unterscheidet sich von der mit allgemeinen, stereotypenspezifischen Items gemessenen Bias Awareness.*

H4: *Die Bias Awareness von Lehramtsstudierenden ist unterschiedlich im Vergleich zur Bias Awareness von Lehrkräften mit Praxiserfahrung.*

Mithilfe von H1 wird untersucht, ob und wenn ja, wie groß die Unter-

schiede der gemessenen Bias Awareness zwischen allgemein formulierten Items und schulspezifischen Items sind. Ob Unterschiede bei der Erfassung der Bias Awareness bezüglich eigener Schülerinnen und Schüler im Vergleich zu Schülerinnen und Schülern im Allgemeinen vorliegen, wird mit H2 untersucht. Da die Erfassung nicht nur allgemein erfolgt, sondern auch Skalen mit stereotypenspezifischer Formulierung getestet werden, sollen Differenzen in den Messungen mit H3 analysiert werden. Schlussendlich soll die Frage, ob die gemessene Bias Awareness sich zwischen Lehrkräften mit Berufserfahrung und Lehramtsstudierenden unterscheidet, mithilfe von H4 beantwortet werden.

3.3 Methodik

Zur Beantwortung der Forschungsfrage und zum damit verbundenen Überprüfen der aufgestellten Hypothesen wurde eine quantitative Datenerhebung durchgeführt. Diese fand im Rahmen eines Online-Fragebogens statt, dessen Bearbeitungsdauer sich auf etwa 10–15 Minuten belief. Hierfür wurde die Online-Plattform SoSci Survey ([Leiner, 2024](#)) verwendet, was es ermöglichte, den Fragebogen einfach zur Verfügung zu stellen. Die Zielgruppe des Fragebogens waren sowohl Lehramtsstudierende und Referendare als auch fertig ausgebildete Lehrkräfte mit Praxiserfahrung.

Der Fragebogen setzte sich dabei aus mehreren Seiten zusammen, in denen zunächst soziodemographische Informationen wie Geschlecht, Alter sowie Geburtsland abgefragt wurden. Darauf folgend wurden berufliche Angaben wie die Formale Bildung, aktuelle Beschäftigung und Schulart erhoben. Anschließend wurde die Bias Awareness zuerst allgemein, dann mit Bezug auf den schulischen Kontext erfasst.

Auf einer dritten Seite wurden die Fragen zusätzlich geschlechtsspezifisch formuliert.

Die Erfassung der Bias Awareness erfolgte mit einer übersetzten und angepassten Version der von [Perry et al. \(2015\)](#) vorgestellten Bias Awareness Skala. Die Übersetzung und Anpassung der Skala auf den deutschen Kontext erfolgte durch [Bonefeld \(2022\)](#). Hier wurden auch zusätzlich kontextspezifische Formulierungen verwendet; neben einer allgemeinen Erfassung (Beispielitem: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber manchen Menschen habe”) wurden beispielsweise auch die Bias Awareness gegenüber Frauen (Beispielitem: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Frauen habe”), Männern und sozioökonomischem Status (Beispielitem: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Menschen mit einem niedrigen sozio-ökonomischen Status habe”) erfasst.

Zusätzlich zu den allgemeinen Skalen zur Erfassung der Bias Awareness wurden eigenkonstruierte Skalen verwendet, in welchen jede der genannten Formulierungen auf den schulischen Kontext angepasst wurde. Dies geschah sowohl für die allgemeine Bias Awareness (Beispielitem der neuen Skala: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen habe”) als auch in Bezug auf Migrationshintergrund, sozioökonomischen Status (Beispielitem der neuen Skala: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen mit einem niedrigen sozio-ökonomischen Status habe”) sowie männliches und

weibliches Geschlecht. Die allgemein formulierte Skala zur Erfassung der Bias Awareness wurde zusätzlich so modifiziert, dass sich die Formulierung auf durch die Lehrkraft selbst unterrichtete Schülerinnen und Schüler bezog (Beispielitem: “Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen habe, welche ich aktuell unterrichte”).

Zur Erfassung wurde im Unterschied zu den in [Perry et al. \(2015\)](#) und [Bonefeld \(2022\)](#) verwendeten Skalen eine 6-stufige anstatt einer 7-stufigen Likert-Skala verwendet. Wie bereits in [Abschnitt 2.5](#) beschrieben, können Likert-Skalen mit einer ungeraden Anzahl an Antwortoptionen zu einer Verzerrung des Ergebnisses durch eine Tendenz zur Mitte führen. Den möglichen Auswirkungen dieses Effekts sollte hier durch die Wahl der 6-stufigen Likert-Skala vorgebeugt werden.

4. Auswertung und Ergebnisse

Auf Basis der theoretischen Überlegungen sowie den im vorherigen Kapitel formulierten Fragestellungen sollen nun die Hypothesen getestet werden. Zum Erreichen dieses Ziels werden zunächst die erhobenen Daten aufbereitet, um anschließend Aussagen über die vorliegende Stichprobe treffen zu können. Im Anschluss daran wird mithilfe statistischer Methoden analysiert, welche Informationen die gemessenen Werte liefern und welche Rückschlüsse daraus in Bezug auf die aufgestellten Hypothesen gezogen werden können.

Die Datenaufbereitung sowie deskriptive und Inferenzstatistik wurden mithilfe der Skriptsprache R durchgeführt. Der vollständige Code ist einsehbar in [Anhang B](#).

4.1 Datenaufbereitung

Bevor die erhobenen Daten für statistische Zwecke verwendet werden können, müssen sie zunächst in Hinblick auf ihre Qualität beurteilt und mögliche Verunreinigungen der Werte korrigiert werden. Weiterhin müssen die rohen Daten für ihre Verwendbarkeit zunächst einer Aufbereitung unterzogen werden.

Der Notwendigkeit der manuellen Bereinigung und Aufbereitung der Daten wurde bereits durch viele Einstellungen bei der Erstellung des Fragebogens entgegengewirkt. Beispielsweise wurden die Items der Bias-Awareness-Skala, die invertiert codiert sind, bereits im Frage-

bogen als solche gekennzeichnet. Dies hatte den Vorteil, dass in den exportierten Daten bereits korrekt gepolte Werte vorlagen und diese nicht erst umcodiert werden mussten.

Weiterhin wurde durch Eingabefilter verhindert, dass Werte fehlerhaft eingegeben wurden oder ganz fehlten. Es wurde zum Beispiel jede Frage als verpflichtend gekennzeichnet, so dass erst bei Abbruch des Fragebogens fehlende Werte auftreten konnten. In Fällen, bei denen es Sinn ergab, wurden außerdem nur korrekte Eingaben zugelassen. Beispielsweise konnte bei der Frage nach dem Alter lediglich eine Ganzzahl zwischen 1 und 99 eingegeben werden.

Durch eine Filterfrage wurde außerdem sichergestellt, nur Studierende Fragen beantworteten, welche an diese Zielgruppe gerichtet waren. Ebenso wurden Fragen, welche sich ausschließlich an fertig ausgebildete Lehrkräfte richteten, durch die Filterfrage nur an diese Gruppe gestellt. Durch dieses Vorgehen musste nicht erst identifiziert werden, welche Werte überhaupt von einer Person aus der zulässigen Gruppe stammten.

Im ersten Schritt der Datenbereinigung wurden fehlerhafte Bearbeitungen des Fragebogens entfernt sowie Interviews mit unzureichenden Informationen, beispielsweise durch einen frühen Abbruch des Nutzers, gelöscht. Zunächst wurden dabei sämtliche Interviews herausgefiltert, welche vor Beantwortung der Fragen zur Bias Awareness in Bezug auf schulischen Kontext abgebrochen wurden. In einem zweiten Schritt wurde das Interview eines Nutzers gelöscht, welcher als Beschreibung seiner aktuellen Tätigkeit lediglich das Wort "Hallo" angegeben hatte.

Im Anschluss daran wurden unklare Eingaben durch Studienteilnehmer korrigiert. Diese Korrektur geschah manuell, da es sich um individuelle Unterschiede bei den Eingaben handelte. Es wurden bei-

spielsweise alle Studienteilnehmer, welche sich in ihrer Tätigkeitsbeschreibung als Studierende ausgewiesen hatten, jedoch noch zusätzlich als Lehrkraft arbeiteten, als Lehramtsstudierende codiert. Auch Personen, welche zwar als Lehrkraft arbeiteten, sich jedoch im Moment in Elternzeit befanden oder aus anderen Gründen eine andere Angabe gemacht hatten, wurden korrekt als Lehrkräfte markiert.

Ebenso wurden Personen, welche sich im Referendariat befinden, als Lehramtsstudierende markiert. Bei der Klassifizierung von Referendaren und Referendarinnen gibt es sowohl Argumente, welche für und gegen eine solche Kategorisierung sprechen. In diesem Fall wurde die Wahl getroffen, da insbesondere der Einfluss langjähriger Berufserfahrung auf die Bias Awareness untersucht werden sollte. Im Fall von Berufsanfängern und -anfängerinnen liegt eine solche Erfahrung noch nicht vor, weswegen die Einteilung als Lehramtsstudierende erfolgte.

Zur Identifizierung von Interviews mit niedriger Qualität wurde berechnet, wie hoch der Anteil gleicher Antworten bei den Fragen zur Bias Awareness ist. Im Anschluss wurden alle Interviews gelöscht, welche über einem manuell festgelegten Grenzwert von 75% lagen. So konnte der Verfälschung der Statistik durch Eingaben von Nutzern, welche immer die selben Antworten gaben und sich nicht mit dem Inhalt der Fragen auseinander setzten, vorgebeugt werden.

Abschließend erfolgte eine Korrektur in den Datenstrukturen der Rohdaten. Hierbei wurden beispielsweise die Werte zum Geburtsland sowie die diesbezüglich eventuell getätigten manuellen Eingaben der Studienteilnehmer kombiniert. Außerdem wurden für jeden Teilnehmer die Mittelwerte der in den zur Bias Awareness gemachten Angaben gebildet und in einer neuen Variablen abgespeichert. Dies

geschah für jede der verwendeten Skalen. Die hier errechneten Werte bilden die Basis für die in [Abschnitt 4.3](#) diskutierten statistischen Methoden wie Hypothesentests.

4.2 Deskriptive Statistik

Bei der Bereinigung und Aufbereitung der Rohdaten sind bereits viele Studienteilnehmer gefiltert worden, deren Angaben fehlerhaft oder nicht verwendbar waren. Im Folgenden kann nun die bereinigte Gruppe beschrieben und statistische Angaben über die Stichprobe gemacht werden. Die Interviews der Studienteilnehmer wurden in Hinblick auf die Forschungsfragen und Hypothesen in zwei Gruppen geteilt. Diesen Gruppen wurden die Teilnehmer zugeordnet in Abhängigkeit davon, ob es sich um Lehramtsstudierende bzw. Personen im Referendariat handelte oder um fertig ausgebildete Lehrkräfte. Im Folgenden werden daher sowohl die gesamte Stichprobe sowie jede der beiden Gruppen beschrieben.

4.2.1 Gesamte Stichprobe

In der Grundgesamtheit der vorliegenden Daten befanden sich $N = 89$ Studienteilnehmer, von denen 15% männlich, 85% weiblich und 0% divers waren. Das mittlere Alter der Studienteilnehmer betrug 31.9 Jahre ($SD = 10.8$). 79% der Befragten gaben an, bereits Lehr Erfahrung zu besitzen, wobei die Erfahrung im Mittel bei 6.3 Jahren lag ($SD = 8.4$). Unter den Teilnehmern befanden sich 10%, die selbst einen Migrationshintergrund hatten. Als Migrationshintergrund wurde hierbei verstanden, dass entweder die Befragten selbst oder min-

destens ein Elternteil außerhalb von Deutschland geboren wurde.

Bei Betrachtung der erlangten Abschlüsse zeigte sich ein diverses Bild. 1% der Studienteilnehmer hatten einen Hauptschulabschluss gemacht, 12% einen Realschulabschluss, 75% Abitur, 28% einen Bachelor-Abschluss, 20% einen Master-Abschluss, 2% Diplom, 54% ein abgeschlossenes Staatsexamen sowie 2% eine abgeschlossene Promotion.

4.2.2 Lehramtsstudierende

Unter den Studienteilnehmern befanden sich $N = 35$ Lehramtsstudierende, welche zu 9% männlich, 91% weiblich und 0% divers waren. Das mittlere Alter der Studierenden betrug 23.5 Jahre ($SD = 2.7$). Von ihnen gaben 46% an, bereits Lehrerfahrung zu besitzen, wobei die Erfahrung im Mittel bei 0.8 Jahren lag ($SD = 1.2$). 94% der Teilnehmer gaben an, bereits das 3-wöchige Orientierungspraktikum oder das 12-wöchige Schulpraxissemester abgeschlossen zu haben. Unter den Studierenden befanden sich 11%, die selbst einen Migrationshintergrund hatten.

Die Studiendauer der befragten Personen lag bei 6.9 Semestern ($SD = 3.9$). 3% der Studierenden hatten einen Hauptschulabschluss gemacht, 20% einen Realschulabschluss, 77% Abitur, 31% einen Bachelor-Abschluss, 17% einen Master-Abschluss sowie 14% ein abgeschlossenes Staatsexamen.

4.2.3 Lehrkräfte

Von den Studienteilnehmern waren $N = 54$ fertig ausgebildete Lehrkräfte. 19% von diesen waren männlichen Geschlechts, 81% weiblich und 0% divers. Im Mittel waren die Lehrkräfte 37.3 Jahre alt ($SD = 10.7$) und die durchschnittliche Lehrererfahrung betrug 9.8 Jahre ($SD = 9.1$). 9% der befragten Personen hatte selbst einen Migrationshintergrund.

In Bezug auf die erlangten Abschlüsse ergaben sich folgende Informationen: 7% der Lehrkräfte hatten einen Hauptschulabschluss gemacht, 74% Abitur, 26% einen Bachelor-Abschluss, 22% einen Master-Abschluss, 4% Diplom, 80% ein abgeschlossenes Staatsexamen sowie 4% eine abgeschlossene Promotion.

Von den befragten Lehrkräften waren zum Befragungszeitpunkt 13% an einer Grundschule tätig, 39% in der Sekundarstufe 1, 35% am Gymnasium, 7% an einer sonderpädagogischen Schule sowie 6% an Gemeinschafts- und Gesamtschulen.

4.2.4 Skalenkonsistenz

Auf Basis der in der Studie gesammelten Daten wurde die interne Konsistenz der verwendeten Skalen überprüft. Hierzu wurde für jede Skala Cronbach's Alpha berechnet. Auch dabei wurden die Werte sowohl für die gesamte Stichprobe als auch in Bezug auf Lehrkräfte und Studierende berechnet. Die Ergebnisse der Berechnungen sind in [Tabelle 4.1](#) aufgestellt.

| Variable | α_{ges} | α_{STU} | α_{LK} |
|--|-----------------------|-----------------------|----------------------|
| BA01: Allgemein | 0.7 | 0.64 | 0.73 |
| BA02: Allgemein – Migrationshintergrund | 0.81 | 0.83 | 0.79 |
| BA03: Allgemein – sozioökonomischer Hintergrund | 0.79 | 0.87 | 0.7 |
| BA04: Allgemein – Frauen | 0.73 | 0.63 | 0.77 |
| BA05: Allgemein – Männer | 0.8 | 0.81 | 0.79 |
| BA06: Schülerinnen und Schüler | 0.77 | 0.7 | 0.77 |
| BA07: eigene Schülerinnen und Schüler | 0.73 | 0.79 | 0.69 |
| BA08: Schülerinnen und Schüler – Migrationshintergrund | 0.84 | 0.84 | 0.82 |
| BA09: Schülerinnen und Schüler – sozioökonomischer Hintergrund | 0.82 | 0.81 | 0.82 |
| BA10: Schülerinnen und Schüler – Frauen | 0.73 | 0.79 | 0.68 |
| BA11: Schülerinnen und Schüler – Männer | 0.8 | 0.8 | 0.8 |

Tabelle 4.1: Cronbach's Alpha (raw) der verschiedenen Bias Awareness Skalen bei unterschiedlichen Stichproben

4.3 Inferenzstatistik

Im Folgenden sollen mithilfe der erhobenen Daten Rückschlüsse von der Stichprobe auf die Allgemeinheit gezogen werden. Im Fokus steht hierbei die Überprüfung der in [Abschnitt 3.2](#) formulierten Hypothesen. Hierzu wurden zunächst die Mittelwerte der verschiedenen Skalen berechnet und anschließend miteinander verglichen. Um zu untersuchen, ob es sich bei den Differenzen um statistisch signifikante Abweichungen handelt, wurden abhängige t-Tests verwendet. Die Entscheidung für einen abhängigen Test beruhte dabei auf der Tatsache, dass es sich in allen getesteten Fällen um die Messwerte zwei verschiedener Skalen bei derselben Testperson handelt.

4.3.1 Überprüfung der Verteilungen

Eine notwendige Voraussetzung für das Anwenden eines t-Tests ist die Normalverteilung der Variable. Bevor die Tests durchgeführt werden können muss also zunächst überprüft werden, ob diese Bedingung erfüllt ist. Hierfür wurde für den Mittelwert jeder Skala ein Shapiro-Wilk-Test durchgeführt. Das Signifikanzniveau wurde dabei auf $\alpha = 0.05$ festgelegt. Der Shapiro-Wilk-Test wurde sowohl für die gesamte Stichprobe, als auch für die Teilgruppe der Lehrkräfte und der Lehramtsstudierenden angewendet. Die Ergebnisse des Tests sind in [Tabelle 4.2](#) dargestellt.

| Variable | P _{ges} | P _{STU} | PLK |
|--|------------------|------------------|--------|
| BA01: Allgemein | 0.41 | 0.19 | 0.33 |
| BA02: Allgemein – Migrationshintergrund | 0.18 | 0.02 | 0.62 |
| BA02: Allgemein – sozioökonomischer Hintergrund | 0.31 | 0.56 | 0.31 |
| BA03: Allgemein – Frauen | <0.001 | <0.001 | <0.001 |
| BA04: Allgemein – Männer | <0.001 | 0.10 | 0.01 |
| BA05: Schülerinnen und Schüler | 0.26 | 0.49 | 0.56 |
| BA06: eigene Schülerinnen und Schüler | 0.19 | 0.23 | 0.04 |
| BA07: Schülerinnen und Schüler – Migrationshintergrund | 0.10 | 0.12 | 0.45 |
| BA08: Schülerinnen und Schüler – sozioökonomischer Hintergrund | 0.20 | 0.11 | 0.21 |
| BA09: Schülerinnen und Schüler – Frauen | <0.001 | <0.001 | <0.001 |
| BA10: Schülerinnen und Schüler – Männer | <0.001 | 0.02 | 0.02 |

Tabelle 4.2: Ergebnisse des Shapiro-Wilk-Tests der Mittelwerte der verschiedenen Bias Awareness Skalen bei unterschiedlichen Stichproben

Es ergab sich, dass zahlreiche Werte unter dem festgelegten Signifikanzniveau von 0.05 liegen. Eine Untersuchung der Histogramme zeigte, dass es sich möglicherweise um Variablen mit einer verschobenen und zusätzlich abgeschnittenen Normalverteilung handeln könnte. Da der t-Test in gewissem Maße robust gegenüber

abgeschnittenen Normalverteilungen ist, wurde dieser nachfolgend trotzdem angewendet. Dennoch wurde zusätzlich bei den betroffenen Variablen auch ein Wilcoxon-Vorzeichen-Rang-Test für abhängige Stichproben verwendet, welcher auch bei nicht-normalverteilten Variablen genutzt werden kann.

4.3.2 Ergebnisse zur gesamten Stichprobe

Zur Überprüfung der Hypothesen H1 und H2 wurden die Differenzen zwischen den Mittelwerten der allgemeinen Bias Awareness Skala (BA01), der allgemeinen Skala mit schulspezifisch formulierten Items (BA06) sowie der Skala mit Bezug auf eigene Schülerinnen und Schüler (BA07) untersucht. Die Ergebnisse dieser Untersuchung sind in [Abbildung 4.1](#) dargestellt.

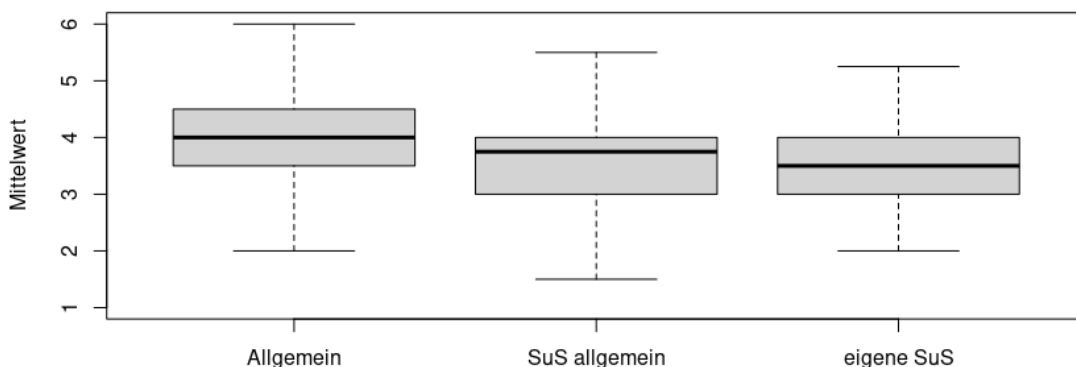


Abbildung 4.1: Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala

Um die statistische Relevanz der berechneten Differenzen in den Mittelwerten zu überprüfen, wurden jeweils zweiseitige abhängige t-Tests angewendet. Die Ergebnisse dieser Tests sind in [Tabelle 4.3](#) zusammengefasst.

4. Auswertung und Ergebnisse

| Variable | M | SD | df | t | p | Cohen's d |
|-------------|-------|-------|----|-------|--------|-----------|
| BA01 - BA06 | 0.463 | 0.896 | 87 | 4.849 | <0.001 | 0.517 |
| BA01 - BA07 | 0.546 | 0.881 | 86 | 5.778 | <0.001 | 0.619 |
| BA06 - BA07 | 0.075 | 0.541 | 86 | 1.289 | 0.201 | 0.138 |

Tabelle 4.3: Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung

Die mit der allgemeinen Skala gemessene Bias Awareness ($M = 4.03$, $SD = 0.85$) war signifikant höher als die mit der auf Schülerinnen und Schüler formulierten Skala erhobene Bias Awareness ($M = 3.56$, $SD = 0.88$); mit $t(88) = 4.849$; $p < 0.001$; $d = 0.517$. Nach [Cohen \(1992\)](#) handelt es sich somit um einen mittleren Effekt.

Bei Messung mit der allgemeinen Skala ($M = 4.03$, $SD = 0.85$) war die gemessene Bias Awareness ebenfalls signifikant höher als bei Messung mit in Bezug auf selbst unterrichtete Schülerinnen und Schüler formulierter Skala ($M = 3.49$, $SD = 0.81$); mit $t(88) = 5.778$; $p < 0.001$; $d = 0.619$. Auch hier handelt es sich um einen mittleren Effekt ([Cohen, 1992](#)).

Die Differenz zwischen der Messung mit der auf Schülerinnen und Schüler bezogenen Skala ($M = 3.56$, $SD = 0.88$) und der in Bezug auf selbst unterrichtete Schülerinnen und Schüler formulierten Skala ($M = 3.49$, $SD = 0.81$) war nicht signifikant; mit $t(88) = 1.289$; $p = 0.201$; $d = 0.138$.

Um Hypothese 3 zu testen wurden jeweils die Mittelwerte der Skalen mit stereotypenspezifischer Formulierung gebildet und mit den Mittelwerten der Skalen mit schul- und stereotypenspezifischer Formulierung verglichen. Eine Übersicht über die gemessenen Werte sind in [Abbildung 4.2](#) abgebildet.

Verglichen wurden jeweils die allgemeinen Skalen zu Migrationshintergrund (BA02), sozioökonomischem Status (BA03) sowie in Bezug

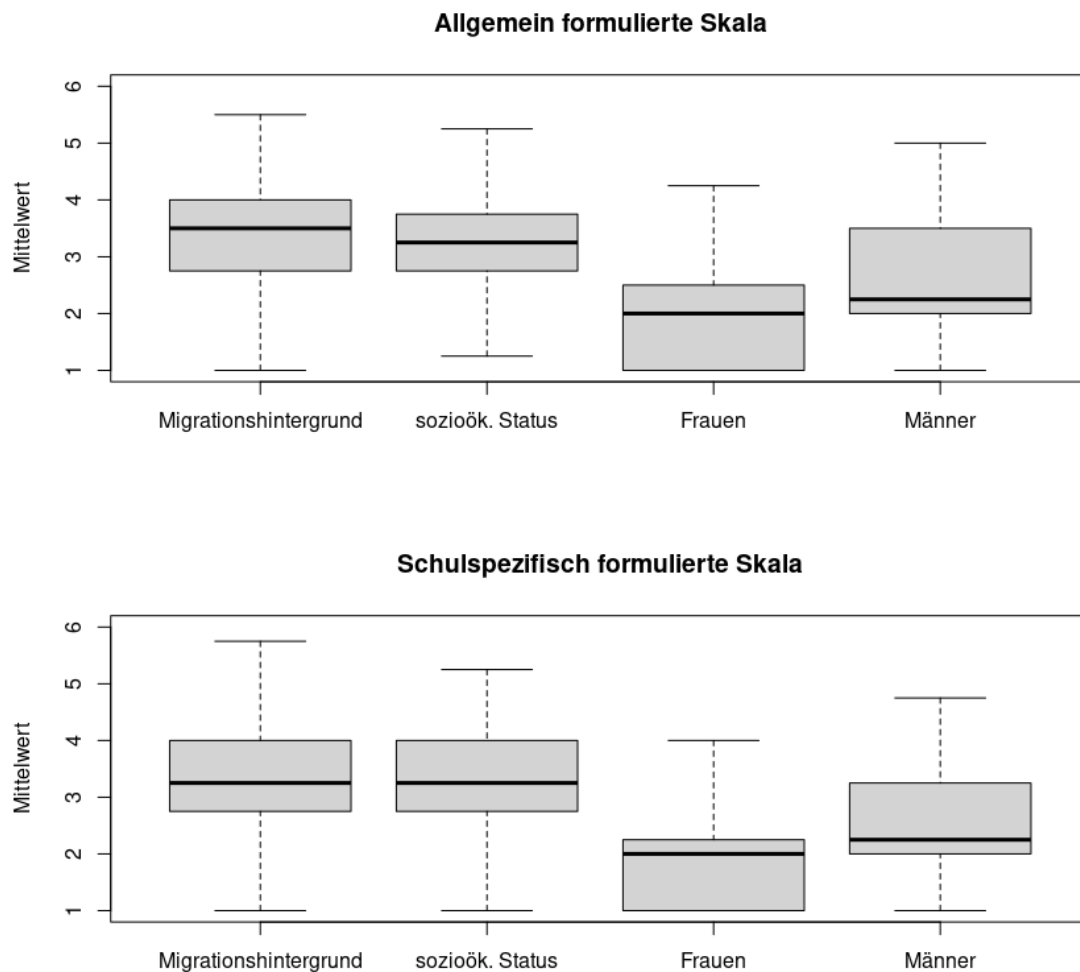


Abbildung 4.2: Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala

auf Frauen (BA04) und Männer (BA05) mit den schulspezifisch formulierten Skalen zu Migrationshintergrund (BA08), sozioökonomischem Status (BA09), Frauen (BA10) und Männern (BA11). Um zu überprüfen, ob es sich bei den Abweichungen um statistisch signifikante Differenzen handelt, wurden erneut zweiseitige abhängige t-Tests verwendet. Die Effektstärke wurde mithilfe von Cohen's d angegeben. Da keine der geschlechtsbezogenen Skalen den Shapiro-Wilk-Test bestanden hatte, wurde in diesen Fällen zusätzlich ein Wilcoxon-Vorzeichen-Rang-Test durchgeführt und die Wilcoxon-Effektstärke r berechnet. Die Ergebnisse sind in [Tabelle 4.4](#) dargestellt.

4. Auswertung und Ergebnisse

| Variable | M | SD | df | t | p | Cohen's d |
|-----------------------------|-------|-------|----|-------|-------|-----------|
| BA02 - BA08 | 0.107 | 0.613 | 83 | 1.603 | 0.113 | 0.175 |
| BA03 - BA09 | 0.158 | 0.577 | 83 | 2.507 | 0.014 | 0.274 |
| BA04 - BA10 (t-Test) | 0.077 | 0.618 | 80 | 1.123 | 0.265 | 0.125 |
| BA04 - BA10 (Wilcoxon-Test) | | | | | 0.254 | r = 0.23 |
| BA05 - BA11 (t-Test) | 0.133 | 0.775 | 80 | 1.542 | 0.127 | 0.171 |
| BA05 - BA11 (Wilcoxon-Test) | | | | | 0.369 | r = 0.15 |

Tabelle 4.4: Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala

Signifikante Effekte zeigten sich lediglich beim Vergleich der allgemeinen, in Bezug auf sozioökonomischen Status formulierten Skala ($M = 3.37$, $SD = 0.88$) und ihrer schulspezifischen Variante ($M = 3.23$, $SD = 0.9$); mit $t(83) = 2.507$; $p = 0.014$; $d = 0.274$. Hierbei handelt es sich allerdings nur um einen schwachen Effekt.

Bei allen anderen getesteten Variablen konnte keine statistische Signifikanz festgestellt werden. Auch die Verwendung des Wilcoxon-Vorzeichen-Rang-Tests in Fällen, in denen der Shapiro-Wilk-Test fehlgeschlagen war, lieferte keinen Hinweis auf statistisch signifikante Abweichungen.

4.3.3 Ergebnisse zu Lehramtsstudierenden

Nachdem in [Abschnitt 4.3.2](#) Aussagen zur gesamten Stichprobe getroffen wurden, werden im Folgenden nur die erhobenen Daten von Lehramtsstudierenden betrachtet. Dabei werden erneut zuerst die Differenzen zwischen den Skalen zur allgemeinen Bias Awareness (BA01) betrachtet, jeweils generell sowie in Bezug auf Schülerinnen und Schüler (BA06) sowie selbst unterrichtete Schülerinnen und

Schüler (BA07). Die Ergebnisse bei Einschränkung auf die Gruppe der Lehramtsstudierenden finden sich hierbei in [Abbildung 4.3](#).

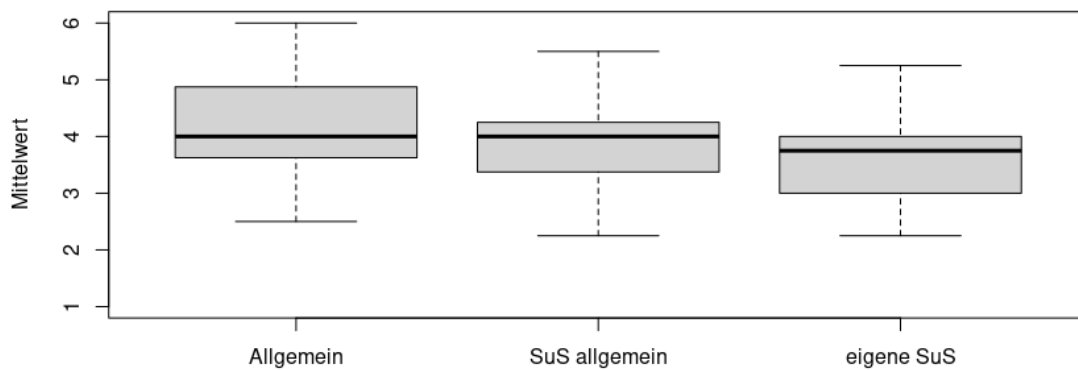


Abbildung 4.3: Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende

Die statistische Relevanz der Differenzen wurde mithilfe eines zweiseitigen abhängigen t-Tests geprüft. Aufgestellt sind die Ergebnisse dieser Tests in [Tabelle 4.5](#).

| Variable | M | SD | df | t | p | Cohen's d |
|-------------|-------|-------|----|-------|-------|-----------|
| BA01 - BA06 | 0.336 | 0.998 | 34 | 1.99 | 0.055 | 0.336 |
| BA01 - BA07 | 0.640 | 1.048 | 33 | 3.558 | 0.001 | 0.610 |
| BA06 - BA07 | 0.287 | 0.530 | 33 | 3.156 | 0.003 | 0.541 |

Tabelle 4.5: Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung, eingeschränkt auf Lehramtsstudierende

Für die allgemein formulierte Skala zur Messung der Bias Awareness ($M = 4.2$, $SD = 0.87$) im Vergleich mit der auf Schülerinnen und Schüler bezogenen Skala ($M = 3.86$, $SD = 0.81$) konnte keine statistische Signifikanz gemessen werden ($p = 0.055 > 0.05 = \alpha$).

Zwischen der allgemeinen Skala ($M = 4.2$, $SD = 0.87$) und der auf selbst unterrichtete Schülerinnen und Schüler bezogenen Skala ($M = 3.59$, $SD = 0.85$) bestand ein signifikanter Unterschied; mit $t(33) =$

3.558; $p = 0.001$; $d = 0.61$. Nach (Cohen, 1992) stellt die gemessene Differenz einen Effekt mittlerer Stärke dar.

Weiterhin konnte ein statistisch signifikanter Unterschied zwischen der allgemein schulbezogenen Skala ($M = 3.86$, $SD = 0.81$) und der auf eigene Schülerinnen und Schüler formulierten Skala ($M = 3.59$, $SD = 0.85$) festgestellt werden; mit $t(33) = 3.156$; $p = 0.003$; $d = 0.541$. Auch hier handelt es sich um einen mittleren Effekt.

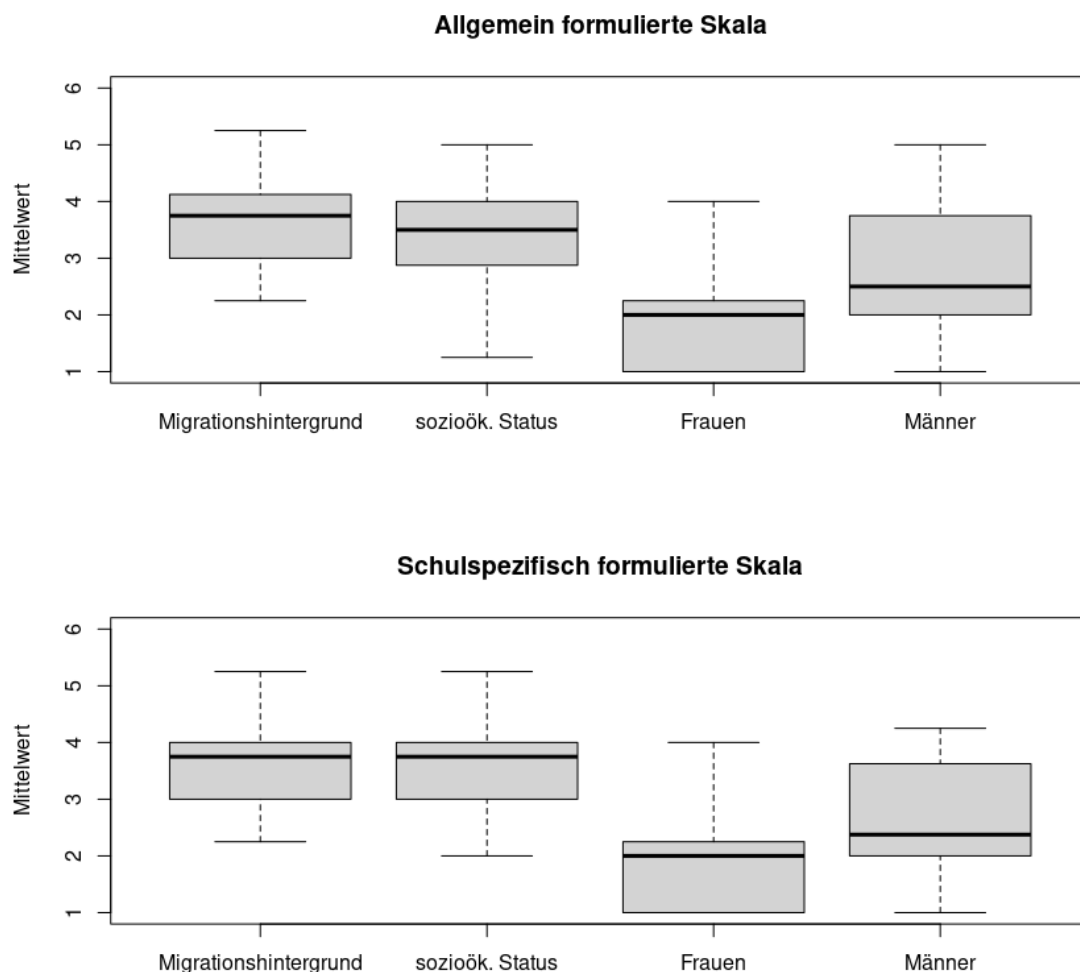


Abbildung 4.4: Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende

Ebenso wurden die Mittelwerte der stereotypenspezifisch formulierten Skala zur Messung der Bias Awareness mit der schul- und ste-

reotypenspezifisch formulierten Skala verglichen. Die Ergebnisse sind in [Abbildung 4.4](#) dargestellt.

Auch hier wurden jeweils die allgemeinen Skalen zu Migrationshintergrund (BA02), sozioökonomischem Status (BA03), in Bezug auf Frauen (BA04) und in Bezug auf Männer (BA05) mit den jeweiligen schulspezifisch formulierten Skalen zu Migrationshintergrund (BA08), sozioökonomischem Status (BA09), Frauen (BA10) und Männern (BA11) verglichen. Für die Überprüfung der statistischen Signifikanz wurden zweiseitige abhängige t-Tests verwendet. Die Effektstärken wurden mithilfe Cohen's d berechnet. In den Fällen, in welchen der Shapiro-Wilk-Test fehlgeschlagen war, wurde zusätzlich zum t-Test ein Wilcoxon-Vorzeichen-Rang-Test verwendet sowie die Wilcoxon-Effektstärke r berechnet. Die Ergebnisse der Tests sind in [Tabelle 4.6](#) aufgelistet.

| Variable | M | SD | df | t | p | Cohen's d |
|-----------------------------|--------|-------|----|--------|-------|-----------|
| BA02 - BA08 (t-Test) | -0.015 | 0.649 | 32 | -0.134 | 0.894 | 0.023 |
| BA02 - BA08 (Wilcoxon-Test) | | | | | 0.961 | r = -0.01 |
| BA03 - BA09 | 0.008 | 0.666 | 32 | 0.065 | 0.948 | 0.011 |
| BA04 - BA10 (t-Test) | -0.086 | 0.498 | 31 | -0.975 | 0.337 | 0.172 |
| BA04 - BA10 (Wilcoxon-Test) | | | | | 0.430 | r = -0.27 |
| BA05 - BA11 (t-Test) | 0.086 | 0.548 | 31 | 0.886 | 0.382 | 0.157 |
| BA05 - BA11 (Wilcoxon-Test) | | | | | 0.404 | r = 0.23 |

Tabelle 4.6: Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehramtsstudierende

Bei keiner der getesteten Variablen konnte ein statistisch signifikanter Unterschied festgestellt werden. Auch in den Fällen, in denen nicht-normalverteilte Variablen getestet wurden konnte der zusätzlich durchgeführte Wilcoxon-Vorzeichen-Rang-Test keine statistische Signifikanz feststellen.

4.3.4 Ergebnisse zu Lehrkräften mit Praxiserfahrung

Analog zum Vorgehen in [Abschnitt 4.3.3](#) wurde ebenfalls die Gruppe der Lehrkräfte in der vorliegenden Stichprobe einzeln betrachtet. Auch hier wurden zunächst die Differenzen zwischen der allgemeinen Skala zur Bias Awareness (BA01) und den Skalen mit Formulierungen mit Bezug auf Schülerinnen und Schüler (BA06) sowie eigene Schülerinnen und Schüler (BA07) untersucht. Die Ergebnisse des lehrkraftbezogenen Vergleichs sind in [Abbildung 4.5](#) dargestellt.

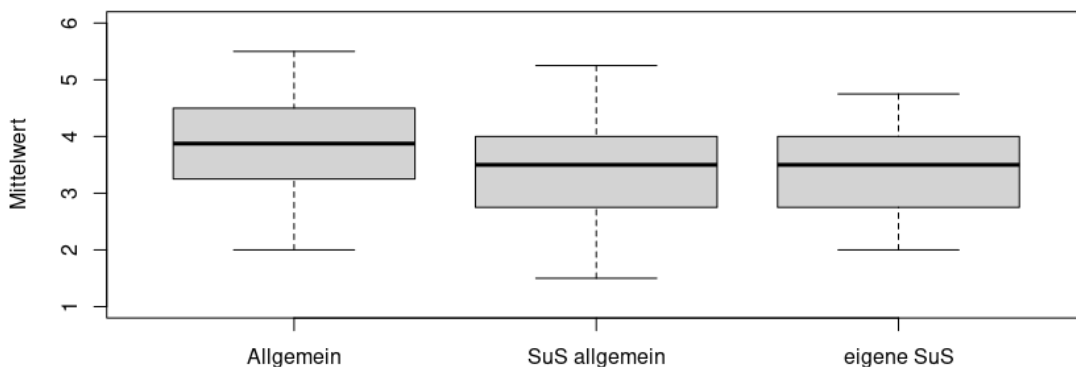


Abbildung 4.5: Unterschiede in den Mittelwerten bei Erfassung mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung

Zur Überprüfung der statistischen Relevanz wurden zweiseitige abhängige t-Tests angewandt. Die Effektstärke wurde durch Cohen's d berechnet. Da der Shapiro-Wilk-Test bei der auf selbst unterrichtete Schülerinnen und Schüler bezogenen Skala (BA07) fehlgeschlagen war, wurde hier zusätzlich ein Wilcoxon-Vorzeichen-Rang-Test durchgeführt sowie die Wilcoxon-Effektstärke r angegeben. Die Ergebnisse der durchgeführten Tests sind in [Tabelle 4.7](#) aufgestellt.

Die Differenz der Mittelwerte bei Messung der Bias Awareness mit allgemeiner Skala ($M = 3.92$, $SD = 0.82$) im Vergleich zur Messung

| Variable | M | SD | df | t | p | Cohen's d |
|-----------------------------|--------|-------|----|--------|--------|-----------|
| BA01 - BA06 | 0.547 | 0.820 | 52 | 4.855 | <0.001 | 0.667 |
| BA01 - BA07 (t-Test) | 0.486 | 0.760 | 52 | 4.653 | <0.001 | 0.639 |
| BA01 - BA07 (Wilcoxon-Test) | | | | | <0.001 | r = 0.69 |
| BA06 - BA07 (t-Test) | -0.061 | 0.507 | 52 | -0.881 | 0.383 | 0.123 |
| BA06 - BA07 (Wilcoxon-Test) | | | | | 0.481 | r = -0.13 |

Tabelle 4.7: Ergebnisse der zweiseitigen abhängigen t-Tests bei Vergleich allgemeiner und schulspezifischer Formulierung, eingeschränkt auf Lehrkräfte mit Praxiserfahrung

mit auf Schülerinnen und Schüler bezogener Skala ($M = 3.36$, $SD = 0.87$) stellte einen statistisch signifikanten Unterschied dar; mit $t(52) = 4.855$; $p < 0.001$; $d = 0.667$. Nach [Cohen \(1992\)](#) handelt es sich dabei um einen Effekt mittlerer Stärke.

Auch die Unterschiede bei Messung mit der allgemeinen Skala ($M = 3.92$, $SD = 0.82$) verglichen mit der auf selbst unterrichtete Schülerinnen und Schüler formulierten Skala ($M = 3.42$, $SD = 0.78$) waren statistisch signifikant; mit $t(52) = 4.653$; $p < 0.001$; $d = 0.639$. Der zusätzlich durchgeführte Wilcoxon-Test bestätigte dies; mit $p < 0.001$; $r = 0.69$. Auch hier handelt es sich um einen mittleren Effekt.

Bei Vergleich der in Bezug auf Schülerinnen und Schüler formulierten Skala ($M = 3.36$, $SD = 0.87$) und der in Bezug auf selbst unterrichtete Schülerinnen und Schüler formulierten Skala ($M = 3.42$, $SD = 0.78$) konnte weder mit dem t-Test ($t(52) = -0.881$; $p = 0.383$; $d = 0.123$) noch mit dem Wilcoxon-Test ($p = 0.481$; $r = -0.13$) statistische Signifikanz festgestellt werden.

Zusätzlich zum Vergleich der allgemeinen Skalen wurden auch die allgemein formulierten stereotypenspezifischen Skalen jeweils mit ihrer schulspezifisch formulierten Variante verglichen. Die Ergebnisse bei Betrachtung von Lehrkräften mit Praxiserfahrung sind in [Abbildung 4.6](#) dargestellt.

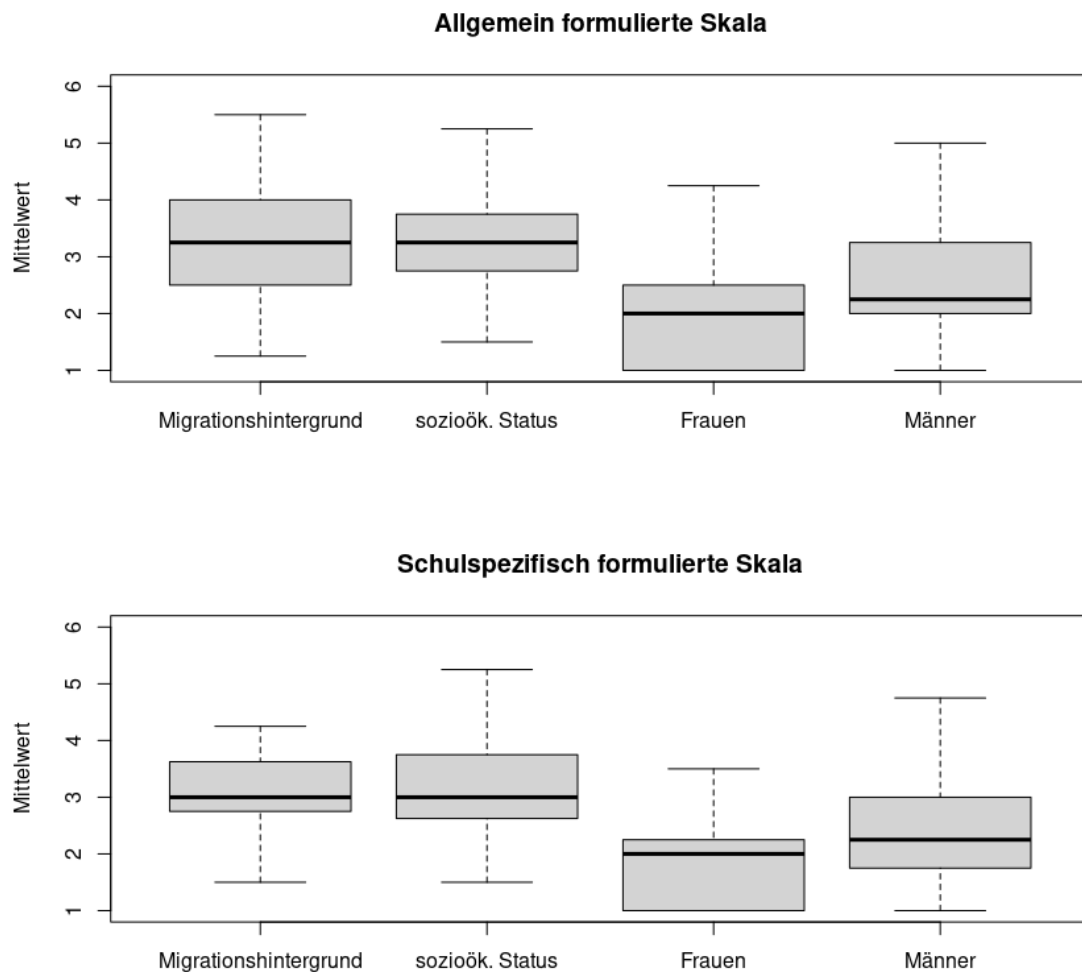


Abbildung 4.6: Unterschiede in den Mittelwerten bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung

Wie auch beim Vergleich der allgemeinen Skalen wurden zur Überprüfung der statistischen Signifikanz zweiseitige abhängige t-Tests durchgeführt und die Effektstärken in Form von Cohen's d berechnet. In den Fällen, in denen der Shapiro-Wilk-Test negativ ausgefallen war, wurde zusätzlich ein Wilcoxon-Vorzeichen-Rang-Test durchgeführt und die zugehörige Effektstärke r angegeben. Die Ergebnisse der Tests sind in [Tabelle 4.8](#) aufgelistet.

Es zeigte sich, dass ein statistisch signifikanter Unterschied zwi-

| Variable | M | SD | df | t | p | Cohen's d |
|-----------------------------|-------|-------|----|-------|--------|-----------|
| BA02 - BA08 | 0.186 | 0.581 | 50 | 2.291 | 0.026 | 0.321 |
| BA03 - BA09 | 0.255 | 0.762 | 50 | 3.687 | <0.001 | 0.516 |
| BA04 - BA10 (t-Test) | 0.184 | 0.669 | 48 | 1.922 | 0.061 | 0.275 |
| BA04 - BA10 (Wilcoxon-Test) | | | | | 0.046 | r = 0.50 |
| BA05 - BA11 (t-Test) | 0.163 | 0.896 | 48 | 1.275 | 0.209 | 0.182 |
| BA05 - BA11 (Wilcoxon-Test) | | | | | 0.617 | r = 0.11 |

Tabelle 4.8: Ergebnisse der zweiseitigen abhängigen t-Tests bzw. Wilcoxon-Vorzeichen-Rang-Tests bei Erfassung der stereotypenspezifischen Bias Awareness mit allgemeiner vs. schulspezifischer Skala, eingeschränkt auf Lehrkräfte mit Praxiserfahrung

schen der allgemein formulierten Skala zu Migrationshintergrund ($M = 3.29$, $SD = 0.93$) und ihrer schulbezogenen Variante ($M = 3.11$, $SD = 0.92$) vorlag; mit $t(50) = 2.291$; $p = 0.026$; $d = 0.321$. Es handelt sich hierbei um einen schwachen Effekt (Cohen, 1992).

Ebenso konnte statistische Signifikanz zwischen der allgemein formulierten Skala zum sozioökonomischen Status ($M = 3.34$, $SD = 0.77$) und ihrer schulbezogenen Variante ($M = 3.12$, $SD = 0.83$) festgestellt werden; mit $t(50) = 3.687$; $p < 0.001$; $d = 0.516$. Es handelt sich um einen Effekt mittlerer Stärke.

Die Differenz der Mittelwerte zwischen der allgemein formulierten Skala bezogen auf Frauen ($M = 2.02$, $SD = 1$) und ihrer schulbezogenen Variante ($M = 1.83$, $SD = 0.84$) war bei Verwendung des t-Tests nicht signifikant ($p = 0.061$). Der Wilcoxon-Vorzeichen-Rang-Test konnte allerdings mit $p = 0.046$ statistische Signifikanz bei einer Effektstärke von $r = 0.50$ zeigen.

Keine statistische Signifikanz konnte bei Vergleich der allgemein formulierten Skala mit Bezug auf Männer ($M = 2.53$, $SD = 1.07$) und ihre schulbezogenen Variante ($M = 2.36$, $SD = 0.99$) festgestellt werden. Sowohl der t-Test ($p = 0.209$) als auch der Wilcoxon-Test ($p = 0.617$) lieferten ein negatives Ergebnis.

4.3.5 Differenzen zwischen Lehramtsstudierenden und Lehrkräften

Nachdem die Ergebnisse zu den einzelnen Gruppen innerhalb der Stichprobe beschrieben wurden, kann nun untersucht werden, welche Differenzen zwischen Lehramtsstudierenden und Lehrkräften vorliegen. Hierzu wurden pro verwendeter Skala die erreichten Mittelwerte miteinander verglichen. Um die statistische Signifikanz der Unterschiede zu untersuchen, wurden zweiseitige Welch-t-Tests für unabhängige Stichproben verwendet. Die Ergebnisse sind in [Tabelle 4.9](#) dargestellt.

| Variable | M | SD (pooled) | df | t | p | Cohen's d |
|--|--------|-------------|--------|--------|-------|-----------|
| BA01 _{STU} – BA01 _{LK} | 0.283 | 0.841 | 70.156 | 1.537 | 0.129 | 0.337 |
| BA02 _{STU} – BA02 _{LK} | 0.201 | 1.002 | 64.169 | 0.893 | 0.375 | 0.201 |
| BA03 _{STU} – BA03 _{LK} | 0.070 | 0.883 | 58.472 | 0.340 | 0.735 | 0.078 |
| BA04 _{STU} – BA04 _{LK} | -0.195 | 0.925 | 76.187 | -0.975 | 0.333 | 0.211 |
| BA05 _{STU} – BA05 _{LK} | 0.178 | 1.085 | 65.05 | 0.716 | 0.477 | 0.164 |
| BA06 _{STU} – BA06 _{LK} | 0.501 | 0.847 | 76.994 | 2.761 | 0.007 | 0.591 |
| BA07 _{STU} – BA07 _{LK} | 0.164 | 0.810 | 66.049 | 0.902 | 0.370 | 0.202 |
| BA08 _{STU} – BA08 _{LK} | 0.438 | 1.004 | 59.164 | 1.874 | 0.066 | 0.436 |
| BA09 _{STU} – BA09 _{LK} | 0.291 | 0.894 | 59.808 | 1.405 | 0.165 | 0.326 |
| BA10 _{STU} – BA10 _{LK} | 0.075 | 0.873 | 62.119 | 0.369 | 0.713 | 0.086 |
| BA11 _{STU} – BA11 _{LK} | 0.255 | 1.027 | 61.801 | 1.070 | 0.289 | 0.248 |

Tabelle 4.9: Ergebnisse der zweiseitigen Welch-t-Tests für unabhängige Stichproben beim Vergleich von Lehramtsstudierenden und Lehrkräften

Eine statistisch signifikante Differenz konnte lediglich beim Vergleich der Mittelwerte bei der allgemein formulierten, auf Schülerinnen und Schüler bezogenen Skala von Lehramtsstudierenden ($M = 3.864$, $SD = 0.805$) und den Lehrkräften mit Praxiserfahrung ($M = 3.363$, $SD = 0.874$) festgestellt werden; mit $t(76.994) = 2.761$; $p = 0.007$; $d = 0.591$. Hierbei handelt es sich nach [Cohen \(1992\)](#) um einen mittleren Effekt.

5. Diskussion

Im Folgenden sollen die Ergebnisse der Befragung sowie ihre statistische Bedeutung genutzt werden, um Rückschlüsse auf die in [Abschnitt 3.2](#) aufgestellten Hypothesen zu ziehen. Weiterhin soll kritisch diskutiert werden, unter welchen Bedingungen die Ergebnisse entstanden sind und über welche Aussagekraft diese verfügen. Hierbei soll auch beleuchtet werden, welchen Limitationen die vorliegende Arbeit unterliegt. Zuletzt wird beleuchtet, welche zukünftige Forschung verfolgt werden und welche neuen wissenschaftlichen Erkenntnisse die Resultate dieser Arbeit vertiefen könnten.

5.1 Aussagekraft der Ergebnisse

Bevor die Ergebnisse im Folgenden in Hinblick auf die aufgestellten Hypothesen interpretiert und zur Beantwortung der Leitfragen genutzt werden, muss zunächst die interne Konsistenz der verwendeten Skalen diskutiert werden. Die Berechnung der Werte für Cronbach's Alpha in [Tabelle 4.1](#) eine Spannweite der Werte von 0.63 (α_{STU} bei der allgemeinen, in Bezug auf Frauen formulierten Skala) bis 0.84 (α_{ges} und α_{STU} bei der auf Schülerinnen und Schüler sowie Migrationshintergrund bezogenen Skala). Nach [Streiner \(2003\)](#) handelt es sich bei Werten von 0.6 – 0.7 um fragwürdige bis akzeptable Werte, im Bereich 0.7 – 0.8 um akzeptable bis gute Werte und im Bereich von 0.8 – 0.9 um sehr gute Werte. Mit Hinblick auf die von [Perry et al. \(2015\)](#) vorgestellten Werte in einem Bereich von

$\alpha = 0.78$ bis $\alpha = 0.87$ handelt es sich somit um realistische Ergebnisse mit teilweise akzeptabler, teilweiser guter interner Konsistenz.

Niedrige Werte für Cronbach's Alpha konnten insbesondere bei der allgemeinen, in Bezug auf Frauen formulierten Skala festgestellt werden. Auch bei Verwendung der schulspezifisch formulierten Variante lagen die Werte eher im akzeptablen Bereich. Die von Bonefeld (2022) präsentierten Werte für diese Skala lagen im Bereich von $\alpha = 0.56$ bis $\alpha = 0.73$. Es war also zu erwarten, dass hier niedrigere Werte erreicht werden würden. Die Ursache hierfür bei Verwendung der auf Frauen bezogenen Skala bleibt eine Frage, der in zukünftigen Arbeiten nachgegangen werden kann.

Weiterhin konnten niedrige Werte für Cronbach's Alpha bei der Skala zur Erfassung der generellen Bias Awareness festgestellt werden ($\alpha_{\text{ges}} = 0.7$, $\alpha_{\text{STU}} = 0.64$, $\alpha_{\text{LK}} = 0.73$). Der Grund hierfür kann nicht abschließend geklärt werden. Eine mögliche Erklärung ist, dass es sich hierbei um die erste abgefragte Skala im Fragebogen handelte. Es wäre möglich, dass Teilnehmer im Laufe der Befragung Antworten mit höherer Konsistenz gaben als noch zu Beginn des Interviews.

5.1.1 Unterschiede bei schulkontextspezifischer Erfassung der Bias Awareness (H1)

Bei Betrachtung der gesamten Stichprobe ergab sich, dass die Bias Awareness bei Erfassung mit der allgemeinen Skala signifikant höher lag als bei Erfassung mithilfe einer auf Schülerinnen und Schüler oder selbst unterrichtete Schülerinnen und Schüler formulierten Skala. In beiden Fällen konnte ein mittlerer Effekt festgestellt werden.

Bei Einschränkung der Stichprobe auf Lehrkräfte konnte das selbe Ergebnis beobachtet werden, wobei im Fall der auf eigene Schülerinnen und Schüler formulierten Skala zusätzlich ein Wilcoxon-Test verwendet wurde. Sowohl der t-Test als auch der Wilcoxon-Test lieferten jedoch mit $p < 0.001$ positive Ergebnisse. Bei Betrachtung der Lehramtsstudierenden konnte statistische Signifikanz bei Vergleich der allgemeinen und der auf selbst unterrichtete Schülerinnen und Schüler bezogenen Skala gezeigt werden.

Mit Hinblick auf die obenstehenden Ergebnisse und Bezug auf die erste Hypothese kann die Alternativhypothese somit angenommen werden. Der beobachtete Effekt kann aufgrund mehrerer Erklärungen aufgetreten sein.

Zunächst ist es möglich, dass generell eine geringere Bias Awareness gegenüber Schülerinnen und Schülern vorliegt. Ein Grund hierfür könnte eine größere Nähe zu den betroffenen Personen und damit verbundene Verzerrungen in der Selbstwahrnehmung sein. Es wäre jedoch auch möglich, dass die Differenzen bei den erfassten Werten aufgrund anderer Gründe zustande kommen. Ein Einflussfaktor ist beispielsweise die gesellschaftliche Erwartung, als Lehrkraft keine Biases gegenüber Schülerinnen und Schülern zu haben, welche dazu geführt haben könnte, dass entsprechende Antworten gegeben wurden. Zusätzlich kann es sein, dass Lehrkräfte tatsächlich ein geringeres Ausmaß an Biases gegenüber Schülerinnen und Schülern aufweisen und dies die gemessene Bias Awareness verzerrt.

5.1.2 Einfluss eigener Schülerinnen und Schüler auf die gemessene Bias Awareness (H2)

Zur Überprüfung der zweiten Hypothese wurde untersucht, wie die Differenz in der Bias Awareness bei Erfassung mit auf Schülerinnen und Schüler formulierter Skala im Vergleich mit der auf selbst unterrichtete Schülerinnen und Schüler formulierten Skala war.

Unter Betrachtung der gesamten Stichprobe ergab sich keine statistisch signifikante Abweichung ($p = 0.201$). Signifikanz konnte jedoch bei Einschränkung der Stichprobe auf Lehramtsstudierende gezeigt werden ($p = 0.003$). Die Effektstärke in Form von Cohen's d betrug hierbei $d = 0.541$, also ein mittlerer Effekt. Unter den Lehrkräften konnte weder mit dem t -Test ($p = 0.383$) noch dem Wilcoxon-Test ($p = 0.481$) ein statistisch signifikanter Effekt festgestellt werden.

Die Hypothese lässt sich somit nicht eindeutig akzeptieren oder verwerfen. Betrachtet man lediglich Lehramtsstudierende, kann die Nullhypothese verworfen werden. In Hinblick darauf, dass Lehrkräfte mit längerer Berufserfahrung jedoch selbst unterrichtete Schülerinnen und Schüler meist über einen deutlich längeren Zeitraum unterrichten und eventuell auch auf Beziehungsebene eine engere persönliche Bindung zu ihnen aufweisen, muss bei Beurteilung in Hinblick auf die Hypothese besonderer Wert auf sie gelegt werden. Da weder in der Gesamtstichprobe noch bei Einschränkung auf Lehrkräfte ein statistisch signifikanter Effekt gemessen werden konnte, kann daher die Alternativhypothese verworfen werden.

5.1.3 Differenzen bei Erfassung der Bias Awareness in Bezug auf stereotypisierte Gruppen (H3)

Wird die gesamte Stichprobe betrachtet, konnten ausschließlich bei Verwendung der in Bezug auf sozioökonomischen Status formulierten Skala statistisch signifikante Unterschiede zwischen der allgemeinen und schulspezifischen Variante festgestellt werden ($p = 0.014$). Bei Berechnung der Effektstärke mittels Cohen's d zeigte sich mit $d = 0.274$ jedoch lediglich ein schwacher Effekt.

Schränkt man die betrachtete Stichprobe auf die Lehramtsstudierenden ein, zeigte sich sogar in keinem Fall ein statistisch signifikanter Effekt. Auch der Wilcoxon-Test konnte in den Fällen, in denen er zusätzlich verwendet wurde, keine statistische Signifikanz zeigen.

Werden jedoch nur die Lehrkräfte mit Praxiserfahrung berücksichtigt, so wurde bei der in Bezug auf Migrationshintergrund ($p = 0.026$), sozioökonomischen Status ($p < 0.001$) sowie Frauen (t-Test: $p = 0.061$, Wilcoxon-Test: $p = 0.046$) formulierten Skala eine signifikante Differenz festgestellt. Es handelte sich hierbei um Effekte im Bereich von $d = 0.275$ bis $d = 0.516$, also um Effekte schwacher bis mittlerer Stärke.

Mit Hinblick auf die Hypothese zeigt sich somit ein differenziertes Bild. Unter Einschränkung auf die Lehramtsstudierenden muss die Nullhypothese angenommen werden. Bei Betrachtung der Lehrkräfte mit Praxiserfahrungen kann in Bezug auf Migrationshintergrund, sozioökonomischen Status und Frauen von einem schwachen bis mittleren Effekt ausgegangen werden und somit die Nullhypothese verworfen werden. Aufgrund der fehlenden Effekte in Bezug auf Männer lässt sich dieses Ergebnis jedoch nicht auf die allgemein formulierte Hypothese generalisieren. Hier ist weitere Forschung nötig,

um die Gründe für den fehlenden Effekt in Bezug auf Männer zu erforschen sowie weitere mögliche Einschränkungen aufzudecken.

Bei Erfassung der stereotypenspezifischen Bias Awareness kam es in einigen Fällen zu nicht-normalverteilten Variablen. Hier wäre es interessant, die Fragestellung erneut mithilfe anderer Daten zu untersuchen, um so möglicherweise eine Normalverteilung zu erreichen und einen aussagekräftigeren t-Test durchführen zu können. Dies könnte die Notwendigkeit einer Modifizierung der Skala bedeuten, da es möglich ist, dass die erfassten Daten einer abgeschnittenen Normalverteilung entsprechen. In diesem Fall müsste die Skala angepasst werden, um diesem Effekt vorzubeugen.

5.1.4 Kontraste zwischen Lehramtsstudierenden und Lehrkräften mit Praxiserfahrung (H4)

Bei Vergleich der Ergebnisse der Lehramtsstudierenden mit denen der Lehrkräfte mit Praxiserfahrung in [Abschnitt 4.3.5](#) konnte ausschließlich bei Betrachtung der allgemeinen, auf Schülerinnen und Schüler formulierten Skala eine signifikante Differenz festgestellt werden ($p = 0.007$). Hierbei handelte es sich mit $d = 0.591$ um einen Effekt mittlerer Stärke. Die Differenzen bei allen anderen Skalen wiesen keine statistische Signifikanz auf (d.h. $p > \alpha = 0.05$).

In Hinblick auf die Hypothese lässt sich somit nicht generalisiert schließen, dass die Bias Awareness von Lehramtsstudierenden signifikant von der Bias Awareness von Lehrkräften mit Praxiserfahrung abweicht. Die Nullhypothese muss daher akzeptiert werden.

Für das Fehlen des erwarteten Effekts gibt es mehrere mögliche Erklärungen. Einerseits kann es sein, dass tatsächlich keine signifikan-

ten Unterschiede beim latenten Konstrukt der Bias Awareness zwischen Lehramtsstudierenden und Lehrkräften vorliegt. Möglich wäre, dass sich die Bias Awareness im Laufe der beruflichen Tätigkeit nicht oder nur sehr wenig verändert, beispielsweise durch das Fehlen von entsprechenden Aufklärungen in Fortbildungen. Dies würde dazu führen, dass sich keine Unterschiede zwischen den beiden Gruppen feststellen lassen. Diesbezüglich wäre es von Interesse, mögliche erklärende Theorien für das Fehlen dieser Differenzen zu entwickeln und diese durch experimentelle Studien zu belegen.

Eine weitere mögliche Erklärung ist die Zusammensetzung der Stichprobe der Lehramtsstudierenden. Die befragten Personen studierten im Mittel bereits 6.9 Semester lang und hatten zu 94% bereits das Praxissemester oder Orientierungspraktikum absolviert. Ebenso lag der Anteil an Studierenden, welche bereits einen universitären Abschluss hatten, relativ hoch. Insgesamt hatten die befragten Personen aus der Stichprobe der Studierenden bereits also verhältnismäßig viel Lehrerfahrung. Es stellt sich die Frage, ob sich die hier präsentierten Ergebnisse bei Studienanfängern reproduzieren lassen.

5.2 Limitationen

Wie bereits in [Abschnitt 4.1](#) genauer beschrieben wurden im Rahmen der Datenaufbereitung Maßnahmen ergriffen, um die Qualität der vorliegenden Daten zu steigern. Ein Nebeneffekt der Auswahl von verpflichtenden Fragen beim Fragebogen war es, dass in weniger Interviews Fragen am Ende des Bogens beantwortet waren als noch am Anfang. Weiterhin war es möglich, dass Teilnehmer gegen Ende kein Interesse mehr an einer seriösen Teilnahme hatten und

somit keine ehrlichen Antworten mehr gaben. Durch die so festgestellten Effekte kann es zu Verzerrungen in den Daten gekommen sein, welche nicht durch deren Aufbereitung und Bereinigung ausgeglichen werden konnten.

Die Werte für die erlangten Abschlüsse der Studienteilnehmer sind dahingehend überraschend, dass sämtliche Befragten entweder Lehramtsstudierende oder bereits als Lehrkräfte tätig waren und eine Abiturquote von 25% somit sehr gering ist. Eine mögliche Erklärung ist, dass Teilnehmer die Frage falsch verstanden haben können. Gefragt war nach allen erlangten Bildungsabschlüssen, dies könnte aus verschiedenen Gründen jedoch als Frage nach dem höchsten erlangten Bildungsabschluss missverstanden worden sein.

Ein weiterer Punkt ist die Zusammensetzung der vorliegenden Stichprobe. Zunächst war der Anteil der Lehrkräfte mit 61% höher als der Anteil an Lehramtsstudierenden mit 39%. Dies kann insbesondere bei Vergleich der beiden Gruppen miteinander die Aussagekraft der Vergleiche einschränken und das durch die Gesamtstichprobe dargestellte Bild verzerren. Weiterhin lag innerhalb der Lehramtsstudierenden die Quote der Personen mit abgeschlossenem Masterstudium oder Staatsexamen mit 17% bzw. 14% relativ hoch. Auch die durchschnittliche Studiendauer war mit einem Mittelwert von 6.9 Semestern relativ hoch, wobei es hier eine große Varianz gab ($SD = 3.9$). Dies kann die Aussagekraft zu dieser Gruppe einschränken, da zusätzlich auch sehr viele der Befragten schon Praxiserfahrung mitbrachten. Der Anteil an weiblichen befragten Personen war mit 85% (bzw. 81% der Lehrkräfte und sogar 91% der Lehramtsstudierenden) sehr hoch. Dies kann einen Einfluss auf die Richtung und das Ausmaß der vorhandenen geschlechtsbezogenen Biases sowie die geschlechtsbezogene Bias Awareness gehabt haben.

Die Skalen zur Bias Awareness dienen der Messung des Bewusstseins über das Vorhandensein eigener Biases. Was nicht durch sie erfasst werden kann, ist das Ausmaß der tatsächlichen Existenz solcher Biases. So kann es beispielsweise sein, dass eine Lehrkraft auf der Skala zur Erfassung der schulbezogenen Bias Awareness in Bezug auf den sozioökonomischen Status einen niedrigen Wert erlangt. Der Grund dafür muss jedoch nicht unbedingt sein, dass die Person sich ihrer Biases nicht bewusst ist, sondern kann auch aus der einfachen Tatsache resultieren, dass das Ausmaß ihrer Biases gegenüber dieser stereotypisierten Gruppe sehr gering ist.

Die Bias Awareness wurde in der vorliegenden Arbeit ausschließlich auf Basis von Selbsteinschätzungen mithilfe der Skala nach [Perry et al. \(2015\)](#) erhoben. Es wäre möglich, dass die Erhebung mithilfe anderer Skalen ein anderes Bild der Bias Awareness der befragten Personen zeichnet. Hier wäre unter Anderem auch die zusätzliche Verwendung von Skalen zur Messung des tatsächlichen Vorhandenseins von Biases nützlich, um die Ursache niedrigen Abschneidens auf der Bias Awareness Skala festlegen zu können.

Zusätzlich fand keine Kontrolle der befragten Personen statt. Es wäre durchaus möglich, dass sich Lehramtsstudierende als Lehrkräfte ausgegeben haben oder umgekehrt oder sogar Personen ohne Hintergrund im Bildungsbereich an der Befragung teilgenommen haben.

Weiterhin ist der Einfluss sozialer Erwartungen auf die Antworten der befragten Personen ein Punkt, der bei der Einordnung der Ergebnisse beachtet werden muss. Beispielsweise ist es möglich, dass Personen, die in einer rassistisch geprägten Gesellschaft leben, auf Fragen in Bezug auf Rassismus eher positiv antworten, auch wenn sie diese Werte persönlich nur wenig vertreten. Der gesellschaftli-

che Einfluss kann ebenfalls einen Einfluss auf die Bias Awareness haben und so die hier erhobenen Daten beeinflussen.

5.3 Zukünftige Arbeit

Es wäre interessant zu untersuchen, wie das Ausmaß der tatsächlich vorhandenen Biases beeinflussen kann, welches Ergebnis befragte Personen auf der Skala zur Erfassung der Bias Awareness erreichen. Hier wäre es beispielsweise möglich, beides abzufragen und bei einem hohen Ausmaß der Beurteilung durch die Bias Awareness Skala eine höhere Bedeutung zuzumessen.

Da in der im Rahmen dieser Arbeit durchgeführten Befragung hauptsächlich weibliche Lehrkräfte befragt wurden, könnte eine interessante Frage sein, inwiefern sich das Geschlecht der befragten Personen auf die Bias Awareness oder das Vorhandensein von Biases auswirkt. Insbesondere im Hinblick auf die Erfassung der Bias Awareness mit der geschlechtsspezifisch formulierten Skala wäre es möglich, dass sich Unterschiede zwischen der auf Frauen und auf Männer formulierten Skala ergeben.

Eine weitere interessante Frage ist die nach dem Zusammenhang zwischen der Berufserfahrung von Lehrkräften und ihrer Bias Awareness. Es könnte zum Beispiel festgestellt werden, dass es in den ersten Berufsjahren im Vergleich zum Studium zu einem Anstieg der Bias Awareness kommt, diese sich jedoch über das weitere Berufsleben hinweg nicht weiter verändert. Weiterhin könnten Langzeitstudien durchgeführt werden, welche die Bias Awareness von Lehramtsstudierenden untersucht und dieselben Testpersonen nach mehrjähriger Berufserfahrung erneut befragt.

Zusätzlich könnte man untersuchen, welchen Zusammenhang es zwischen der Bias Awareness und der Schulart gibt, an der die befragten Lehrkräfte tätig sind. Es wäre möglich, dass es hier zu Differenzen je nach Tätigkeitsbereich kommt.

6. Fazit und Zusammenfassung

Es wurde untersucht, welcher Zusammenhang zwischen der Formulierung von Fragebogenitems zur Erfassung der Bias Awareness und den dabei gemessenen Werten besteht. Zusätzlich wurde untersucht, ob es diesbezüglich Differenzen zwischen Lehramtsstudierenden und Lehrkräften gibt.

Es konnte gezeigt werden, dass die mit einer allgemeinen Skala erfasste Bias Awareness signifikant höher lag gegenüber einer Erfassung mithilfe einer auf den schulischen Kontext angepassten Skala. Dieser Effekt zeigte sich sowohl in der gesamten Stichprobe auch bei getrennter Betrachtung der Lehramtsstudierenden und Lehrkräfte.

Ein signifikanter Unterschied zwischen allgemeiner, schulbezogener Erfassung und in Bezug auf selbst unterrichtete Schülerinnen und Schüler formulierter Skala konnte lediglich bei den Lehramtsstudierenden festgestellt werden.

Bei Analyse der Unterschiede zwischen allgemeiner stereotypenspezifischer Erfassung und ihrer schulbezogenen Variante zeigte sich ein gemischtes Bild, bei dem unter den Lehrkräften in Bezug auf Migrationshintergrund, sozioökonomischen Status und Frauen ein schwacher bis mittlerer Effekt festgestellt werden konnte. Es ist weitere Forschung nötig, um diesen Effekt zu bestätigen und möglicherweise zu generalisieren.

Die Existenz statistisch signifikanter Differenzen zwischen den Lehramtsstudierenden und Lehrkräften mit Praxiserfahrung konnten nicht

festgestellt werden.

Zusammenfassend bleibt festzuhalten, dass weitere wissenschaftliche Erkenntnisse im Bereich der schulbezogenen Bias Awareness nötig sind, um die bestehenden Forschungslücken zu schließen und eine solide wissenschaftliche Basis zu liefern.

Literaturverzeichnis

Frances E Aboud. The developmental psychology of racial prejudice. *Transcultural Psychiatric Research Review*, 30(3):229–242, 1993. 9

Jens Agerström und Dan-Olof Rooth. The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology*, 96(4):790, 2011. 1, 11

Richard Chatham Atkinson und Richard Martin Shiffrin. Human memory: A proposed system and its control processes. *The Psychology of Learning and Motivation: Advances in Research and Theory*, 2:89–195, 1968. 1, 5

Nina Baur und Jörg Blasius. *Handbuch Methoden der empirischen Sozialforschung*, volume 13. Springer VS, Wiesbaden, 2014. ISBN 978-3-531-18939-0. doi: 10.1007/978-3-531-18939-0. 15

David Bawden und Lyn Robinson. Information overload: An overview. *Oxford Encyclopedia of Political Decision Making*, 2020. doi: 10.1093/acrefore/9780190228637.013.1360. 5

Buster Benson. Cognitive bias cheat sheet: An organized list of cognitive biases because thinking is hard, 2016. URL <https://betterhumans.pub/cognitive-bias-cheat-sheet-55a472476b18>. Accessed 10.09.2024. 4

Rebecca S Bigler und Lynn S Liben. A cognitive-developmental approach to racial stereotyping and reconstructive memory in euro-american children. *Child development*, 64(5):1507–1518, 1993. 9

- Mark N Bing, James C Whanger, H Kristl Davison, und Jayson B VanHook. Incremental validity of the frame-of-reference effect in personality scale scores: a replication and extension. *Journal of Applied Psychology*, 89(1):150, 2004. [2](#), [16](#)
- Meike Bonefeld. Reflexion eigener Stereotype als Motor zur nachhaltigen Stereotypreduktion bei angehenden Lehrkräften. In *Stereotype in der Schule II: Ursachen und Möglichkeiten der Intervention*, pages 341–377. Springer, 2022. [15](#), [21](#), [22](#), [44](#), [XIII](#)
- Pedro Bordalo, Katherine Coffman, Nicola Gennaioli, und Andrei Shleifer. Stereotypes. *The Quarterly Journal of Economics*, 131(4):1753–1794, 2016. [6](#)
- John C Brigham. Ethnic stereotypes. *Psychological Bulletin*, 76(1): 15, 1971. [6](#)
- Jacob Cohen. A power primer. *Psychological Bulletin*, 112(1):155–159, 1992. [32](#), [36](#), [39](#), [41](#), [42](#)
- Sara Costa, Sabine Pirchio, Anna Shevchuk, und Sabine Glock. Does teachers’ ethnic bias stress them out? The role of teachers’ implicit attitudes toward and expectations of ethnic minority students in teachers’ burnout. *International Journal of Intercultural Relations*, 93:101757, 2023. [14](#), [15](#)
- William TL Cox, Lyn Y Abramson, Patricia G Devine, und Steven D Hollon. Stereotypes, prejudice, and depression: The integrated perspective. *Perspectives on Psychological Science*, 7(5):427–449, 2012. [11](#)
- Jennifer Crocker und Brenda Major. Social stigma and self-esteem: The self-protective properties of stigma. *Psychological review*, 96(4):608–630, 1989. [14](#)

- Jennifer Crocker, Kristin Voelkl, Maria Testa, und Brenda Major. Social stigma: The affective consequences of attributional ambiguity. *Journal of personality and social psychology*, 60(2):218, 1991. [10](#)
- Amy JC Cuddy, Susan T Fiske, und Peter Glick. Warmth and competence as universal dimensions of social perception: The stereotype content model and the bias map. *Advances in experimental social psychology*, 40:61–149, 2008. [10](#)
- Juliane Degner. *Vorurteile: Haben immer nur die anderen*. Springer, Berlin/Heidelberg, 2022. ISBN 978-3-662-60571-4. doi: 10.1007/978-3-662-60572-1. [9](#)
- Patricia G Devine, E Ashby Plant, David M Amodio, Eddie Harmon-Jones, und Stephanie L Vance. The regulation of explicit and implicit race bias: the role of motivations to respond without prejudice. *Journal of personality and social psychology*, 82(5):835, 2002. [13](#)
- David Dunning. The Dunning–Kruger effect: On being ignorant of one’s own ignorance. In *Advances in experimental social psychology*, volume 44, pages 247–296. Elsevier, 2011. [7](#)
- Irena D Ebert und Melanie C Steffens. Positionsartikel zum Forschungsprogramm Explizite und implizite geschlechterbezogene Kognitionen heute. *GENDER–Zeitschrift für Geschlecht, Kultur und Gesellschaft*, 5(3):9–10, 2013. [11](#)
- Naomi Ellemers. Gender stereotypes. *Annual review of psychology*, 69(1):275–298, 2018. [6](#)
- Malte Elson. Question wording and item formulation. In Jörg Matthes, Robert F Potter, und Christine S Davis, editors, *International Encyclopedia of Communication Research Methods*. Wiley-Blackwell, 2017. [16](#)

Susan T Fiske, Amy J. C. Cuddy, Peter Glick, und Jun Xu. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, 82(6):878–902, 2002. [10](#)

Sarah Gentrup, Georg Lorenz, Cornelia Kristen, und Irena Kogan. Self-fulfilling prophecies in the classroom: Teacher expectations, teacher feedback and student achievement. *Learning and Instruction*, 66:101296, 2020. [13](#)

Asaf Gilboa und Morris Moscovitch. The cognitive neuroscience of confabulation: A review and a model. *Handbook of memory disorders*, 2:315–342, 2002. [6](#)

Sabine Glock und Julia Karbach. Preservice teachers' implicit attitudes toward racial minority students: Evidence from three implicit measures. *Studies in Educational Evaluation*, 45:55–61, 2015. ISSN 0191-491X. doi: <https://doi.org/10.1016/j.stueduc.2015.03.006>. [14](#)

Sabine Glock und Florian Klapproth. Bad boys, good girls? implicit and explicit attitudes toward ethnic minority students among elementary and secondary school teachers. *Studies in Educational Evaluation*, 53:77–86, 2017. ISSN 0191-491X. doi: <https://doi.org/10.1016/j.stueduc.2017.04.002>. [14](#)

Sabine Glock und Hannah Kleen. Attitudes toward students from ethnic minority groups: The roles of preservice teachers' own ethnic backgrounds and teacher efficacy activation. *Studies in Educational Evaluation*, 62:82–91, 2019. ISSN 0191-491X. doi: <https://doi.org/10.1016/j.stueduc.2019.04.010>. [14](#)

Lorenz Goette, Hua-Jing Han, und Benson Tsz Kin Leung. Information overload and confirmation bias. *Cambridge Working Papers in Economics*, 2019. [5](#)

Anthony G Greenwald und Mahzarin R Banaji. Implicit social cognition: attitudes, self-esteem, and stereotypes. *Psychological review*, 102(1):4, 1995. 1, 8

Joel W. Grube, Daniel M. Mayton II, und Sandra J. Ball-Rokeach. Inducing change in values, attitudes, and behaviors: Belief system theory and the method of value self-confrontation. *Journal of Social Issues*, 50(4):153–173, 1994. doi: <https://doi.org/10.1111/j.1540-4560.1994.tb01202.x>. 12

Charles P Guichard und Margaret A Connolly. Ethnic group stereotypes: A new look at an old problem. *The Journal of Negro Education*, 46(3):344–357, 1977. 6

Gordon H Guyatt, Deborah J Cook, Derek King, Geoffrey R Norman, SL Kane, und C Van Ineveld. Effect of the framing of questionnaire items regarding satisfaction with training on residents' responses. *Academic Medicine*, 74(2):192–194, 1999. 2, 16

Elizabeth L Haines, Kay Deaux, und Nicole Lofaro. The times they are a-changing... or are they not? A comparison of gender stereotypes, 1983–2014. *Psychology of Women Quarterly*, 40(3):353–363, 2016. 6

Amanda Hall und Graham Walton. Information overload within the health care system: a literature review. *Health Information & Libraries Journal*, 21(2):102–108, 2004. 5

Anne-Wil Harzing, B Sebastian Reiche, und Markus Pudelko. Challenges in international survey research: A review with illustrations and suggested solutions for best practice. *European Journal of International Management*, 7(1):112–134, 2013. 16

Madeline E Heilman. Gender stereotypes and workplace bias. *Research in organizational Behavior*, 32:113–135, 2012. 6

Johanna M Hermann. Warum Mädchen schlechter rechnen und Jungen schlechter lesen—Wenn Geschlechtsstereotype zur Bedrohung für das eigene Leistungsvermögen in der Schule werden. *Stereotype in der Schule*, pages 33–70, 2020. [14](#)

Curt Hoffman und Nancy Hurst. Gender stereotypes: Perception or rationalization? *Journal of personality and social psychology*, 58 (2):197, 1990. [6](#)

Karen Jenni und George Loewenstein. Explaining the identifiable victim effect. *Journal of Risk and Uncertainty*, 14:235–257, 1997. doi: 10.1023/A:1007740225484. [7](#)

Katarina Karoline Kaminski. *Situational Judgment Tests und Kontextspezifität: Überprüfung des Einflusses von Sozialer Erwünschtheit und Überprüfung der Konstruktvalidität eines stark kontextspezifischen Situational Judgment Tests anhand des TEFEM (Test zur Erfassung von Führungs-und Entscheidungsverhalten im Militär)*. PhD thesis, Universität der Bundeswehr Hamburg, Hamburg, 2017. [2](#), [16](#)

Johannes Keller und Dirk Dauenheimer. Stereotype threat in the classroom: Dejection mediates the disrupting threat effect on women’s math performance. *Personality and Social Psychology Bulletin*, 29(3):371–381, 2003. [11](#), [14](#)

David J Kelly, Paul C Quinn, Alan M Slater, Kang Lee, Alan Gibson, Michael Smith, Liezhong Ge, und Olivier Pascalis. Three-month-olds, but not newborns, prefer own-race faces. *Developmental science*, 8(6):F31–F36, 2005. [9](#)

Christoph J Kemper, Constanze Beierlein, Doreen Bensch, Anastasiya Kovaleva, und Beatrice Rammstedt. Eine Kurzsкала zur Erfassung des Gamma-Faktors sozial erwünschten Antwortverhaltens:

- Die Kurzskala Soziale Erwünschtheit-Gamma (KSE-G). *GESIS – Working Papers*, 2012. 16
- Melanie Killen und Adam Rutland. Children and social exclusion: Morality, prejudice, and group identity. *Journal of Moral Education*, 42(2):258–260, 2013. doi: 10.1080/03057240.2013.774854. 9
- Hannah Kleen und Sabine Glock. Sag’ mir, wie du heißt, dann sage ich dir, wie du bist: Eine Untersuchung von Vornamen. *Stereotype in der Schule*, pages 99–131, 2020. 14
- Jon A. Knutsen und Stanley Presser. Question and questionnaire design. *Handbook of survey research*, pages 263–313, 2010. 16
- John T Kulas und Alicia A Stachowski. Middle category endorsement in odd-numbered likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, 43(3):489–493, 2009. 16
- Esther Landhuis. Scientific literature: Information overload. *Nature*, 535(7612):457–458, 2016. 5
- Ioana M Latu, Marianne Schmid Mast, und Tracie L Stewart. Gender biases in (inter) action: The role of interviewers’ and applicants’ implicit and explicit stereotypes in predicting women’s job interview outcomes. *Psychology of Women Quarterly*, 39(4):539–552, 2015. 11
- D. J. Leiner. Sosci survey (version 3.5.02), 2024. URL <https://www.soscisurvey.de>. Accessed 12.12.2024. 20
- Georg Lorenz, Sarah Gentrup, Cornelia Kristen, Petra Stanat, und Irena Kogan. Stereotype bei Lehrkräften? Eine Untersuchung systematisch verzerrter Lehrererwartungen. *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 68(1):89–111, 2016. 14

C Neil Macrae und Galen V Bodenhausen. Social cognition: Thinking categorically about others. *Annual review of psychology*, 51(1):93–120, 2000. ISSN 1545-2085. doi: 10.1146/annurev.psych.51.1.93. [9](#)

Stephanie Madon, Lee Jussim, Max Gyll, Heather Nofziger, Elizabeth R Salib, Jennifer Willard, und Kyle C Scherr. The accumulation of stereotype-based self-fulfilling prophecies. *Journal of Personality and Social Psychology*, 115(5):825, 2018. [11](#)

Carol Lynn Martin und Charles F Halverson. A schematic processing model of sex typing and stereotyping in children. *Child development*, pages 1119–1134, 1981. [10](#)

Sarah E Martiny und Laura Froehlich. Ein theoretischer und empirischer Überblick über die Entwicklung von Stereotypen und ihre Konsequenzen im Schulkontext. *Stereotype in der Schule*, pages 1–32, 2020. [1](#), [9](#), [14](#)

Dana Mastro. Effects of racial and ethnic stereotyping. In *Media effects*, pages 341–357. Routledge, 2009. [6](#)

Bertolt Meyer, Alexander Zill, und Dominik Dilba. Entspann dich, Deutschland! TK-Stressstudie 2021. 2021. [5](#)

Jörn-Axel Meyer. Information overload in marketing management. *Marketing Intelligence & Planning*, 16(3):200–209, 1998. [5](#)

Margo J Monteith. Self-regulation of prejudiced responses: Implications for progress in prejudice-reduction efforts. *Journal of personality and social psychology*, 65(3):469–485, 1993. [1](#), [12](#)

Margo J Monteith und Corrine I Voils. Proneness to prejudiced responses: Toward understanding the authenticity of self-reported discrepancies. *Journal of personality and social psychology*, 75(4):901–916, 1998. [13](#)

- Margo J Monteith, Steven A Arthur, und Sara M Flynn. Self-regulation and bias. *Handbook of prejudice, stereotyping, and discrimination*, pages 493–507, 2010a. [12](#)
- Margo J Monteith, Aimee Y Mark, und Leslie Ashburn-Nardo. The self-regulation of prejudice: Toward understanding its lived character. *Group Processes & Intergroup Relations*, 13(2):183–200, 2010b. [12](#)
- Don A Moore und Paul J Healy. The trouble with overconfidence. *Psychological review*, 115(2):502, 2008. [7](#)
- Gordon B Moskowitz und Peizhong Li. Egalitarian goals trigger stereotype inhibition: A proactive form of stereotype control. *Journal of Experimental Social Psychology*, 47(1):103–116, 2011. [13](#)
- Francesca Muntoni und Jan Retelsdorf. Geschlechterstereotype in der Schule. *Stereotype in der Schule*, pages 71–97, 2020. [14](#)
- Bennet B Murdock Jr. The serial position effect of free recall. *Journal of experimental psychology*, 64(5):482, 1962. [7](#)
- Jason W Osborne und Brett D Jones. Identification with academics and motivation to achieve in school: How the structure of the self influences academic outcomes. *Educational Psychology Review*, 23:131–158, 2011. [14](#)
- C Northcote Parkinson. *Parkinson's Law or the Pursuit of Progress*. Murray, 1958. ISBN 9780719510496. [7](#)
- Louis A Penner. Interpersonal attraction toward a black person as a function of value importance. *Personality: An International Journal*, 2:175–187, 1970. [12](#)
- Sylvia P. Perry, Mary C. Murphy, und John F. Dovidio. Modern prejudice: Subtle, but unconscious? The role of Bias Awareness in

- Whites' perceptions of personal and others' biases. *Journal of Experimental Social Psychology*, 61:64–78, 2015. ISSN 0022-1031. doi: <https://doi.org/10.1016/j.jesp.2015.06.007>. 1, 2, 12, 13, 21, 22, 43, 51
- James L Phillips, Richard M Shiffrin, und Richard C Atkinson. Effects of list length on short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 6(3):303–311, 1967. 7
- Katherine Picho, Ariel Rodriguez, und Lauren Finnie. Exploring the moderating role of context on the mathematics performance of females under stereotype threat: A meta-analysis. *The Journal of social psychology*, 153(3):299–333, 2013. 1, 11
- E Ashby Plant und Patricia G Devine. Internal and external motivation to respond without prejudice. *Journal of personality and social psychology*, 75(3):811, 1998. 13
- Rüdiger Pohl. *Cognitive illusions: A handbook on fallacies and biases in thinking, judgement and memory*. Psychology press, East Sussex, 2004. ISBN 1-84169-351-0. 3
- Ype H Poortinga und Fons JR van de Vijver. Der Umgang mit methodischen Stolperfallen in der kulturvergleichenden Stressforschung. *Handbuch Stress und Kultur: Interkulturelle und kulturvergleichende Perspektiven*, pages 153–171, 2013. doi: <https://doi.org/10.1007/978-3-531-93449-5>. 16
- Devin G. Pope, Joseph Price, und Justin Wolfers. Awareness reduces racial bias. *Management Science*, 64(11):4988–4995, 2018. doi: 10.1287/mnsc.2017.2901. 12
- Deepasri Prasad und Wilma A Bainbridge. The visual mandela effect as evidence for shared and specific false memories across people. *Psychological Science*, 33(12):1971–1988, 2022. 6

- Timothy B Rogers, Nicholas A Kuiper, und William S Kirker. Self-reference and the encoding of personal information. *Journal of personality and social psychology*, 35(9):677, 1977. 8
- Jean-Christophe Rohner, Annika Waldén, Hanna Blomberg, und Pontus Carlsson. Reducing physical-attractiveness bias in hiring decisions : An experimental investigation. volume 13. Lund University, 2013.
- Milton Rokeach. A theory of organization and change within value-attitude systems. *Journal of social issues*, 24(1), 1968. 12
- Michael Ross und Fiore Sicoly. Egocentric biases in availability and attribution. *Journal of personality and social psychology*, 37(3): 322, 1979. 7
- Reijo Savolainen. Filtering and withdrawing: strategies for coping with information overload in everyday contexts. *Journal of information Science*, 33(5):611–621, 2007. 5
- M. G. Siegler und Eric Schmidt. Every 2 days we create as much information as we did up to 2003, 2010. URL <https://techcrunch.com/2010/08/04/schmidt-data>. Accessed 25.10.2024. 5
- Ronald G Slaby und Karin S Frey. Development of gender constancy and selective attention to same-sex models. *Child development*, pages 849–856, 1975. 10
- Betsy Sparrow, Jenny Liu, und Daniel M Wegner. Google effects on memory: Cognitive consequences of having information at our fingertips. *Science*, 333(6043):776–778, 2011. 8
- Petra Stanat, Stefan Schipolowski, Rebecca Schneider, Karoline A. Sachse, Sebastian Weirich, und Sofie Henschel. *IQB-Bildungstrend 2021: Kompetenzen in den Fächern Deutsch und*

- Mathematik am Ende der 4. Jahrgangsstufe im dritten Ländervergleich*. Waxmann Verlag GmbH, Münster, 2021. ISBN 978-3-8309-9606-4. doi: <https://doi.org/10.31244/9783830996064>. 13
- Claude M Steele und Joshua Aronson. Stereotype threat and the intellectual test performance of african americans. *Journal of personality and social psychology*, 69(5):797–811, 1995. 11, 14
- David L Streiner. Starting at the beginning: an introduction to coefficient alpha and internal consistency. *Journal of personality assessment*, 80(1):99–103, 2003. 43
- Henri Tajfel und John Turner. An integrative theory of intergroup conflict. *The social psychology of intergroup relations*, pages 33–47, 1979. 10
- András Tilcsik. Pride and prejudice: Employment discrimination against openly gay men in the united states. *American Journal of sociology*, 117(2):586–626, 2011. 11
- Anita Tobisch und Markus Dresel. Negatively or positively biased? dependencies of teachers' judgments and expectations based on students' ethnic and social backgrounds. *Social Psychology of Education*, 20:731–752, 2017. 14
- Martina Volnhals und Bernhard Hirsch. Information Overload und Controlling. *Controlling & Management*, 52(Suppl 1):50–57, 2008. 5
- Gregory M Walton und Priyanka B Carr. Social belonging and the motivation and intellectual achievement of negatively stereotyped students. In M Inzlicht und T Schmader, editors, *Stereotype threat: Theory, process and application*, pages 89–106. Oxford University Press, 2012. 14

A. Verwendete Skalen zur Erhebung der Bias Awareness

Die Skalen mit allgemeinem Kontext sind direkt aus (Bonefeld, 2022) entnommen. Bei den Skalen mit schulischem Kontext handelt es sich um angepasste Varianten dieser Skala.

Allgemeine Skala, allgemeiner Kontext

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber manchen Menschen habe.
2. Ich mache mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Ich mache mir Sorgen, dass ich unbewusste Vorurteile habe.
4. Ich mache mir nie Sorgen, dass ich mich auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Migrationshintergrund, allgemeiner Kontext

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Menschen mit Migrationshintergrund habe.

2. Wenn ich mit Menschen mit Migrationshintergrund spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Menschen mit Migrationshintergrund mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Menschen mit Migrationshintergrund habe.
4. Ich mache mir nie Sorgen, dass ich mich Menschen mit Migrationshintergrund gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf sozioökonomischen Status, allgemeiner Kontext

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Menschen mit einem niedrigen sozio-ökonomischen Status habe.
2. Wenn ich mit Menschen mit einem niedrigen sozio-ökonomischen Status spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Menschen mit einem niedrigen sozio-ökonomischen Status mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Menschen mit einem niedrigen sozio-ökonomischen Status habe.
4. Ich mache mir nie Sorgen, dass ich mich Menschen mit einem niedrigen sozio-ökonomischen Status gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Frauen, allgemeiner Kontext

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Frauen habe.
2. Wenn ich mit Frauen spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Frauen mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Frauen im Allgemeinen habe.
4. Ich mache mir nie Sorgen, dass ich mich Frauen gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Männer, allgemeiner Kontext

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Männern habe.
2. Wenn ich mit Männern spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Männer mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Männern im Allgemeinen habe.
4. Ich mache mir nie Sorgen, dass ich mich Männern gegenüber auf subtile Weise voreingenommen verhalten könnte.

Allgemeine Skala, Kontext Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen habe.
2. Wenn ich mit Schüler*innen spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Schüler*innen mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Schüler*innen habe.
4. Ich mache mir nie Sorgen, dass ich mich Schüler*innen gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Migrationshintergrund, Kontext Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen mit Migrationshintergrund habe.
2. Wenn ich mit Schüler*innen mit Migrationshintergrund spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Schüler*innen mit Migrationshintergrund mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Schüler*innen mit Migrationshintergrund habe.
4. Ich mache mir nie Sorgen, dass ich mich Schüler*innen mit Migrationshintergrund gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf sozioökonomischen Status, Kontext Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen mit einem niedrigen sozio-ökonomischen Status habe.
2. Wenn ich mit Schüler*innen mit einem niedrigen sozio-ökonomischen Status spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich Schüler*innen mit einem niedrigen sozio-ökonomischen Status mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Schüler*innen mit einem niedrigen sozio-ökonomischen Status habe.
4. Ich mache mir nie Sorgen, dass ich mich Schüler*innen mit einem niedrigen sozio-ökonomischen Status gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Frauen, Kontext Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber weiblichen Schülerinnen habe.
2. Wenn ich mit weiblichen Schülerinnen spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich weibliche Schülerinnen mag, mache ich mir Sor-

gen, dass ich unbewusste Vorurteile gegenüber weiblichen Schülerinnen im Allgemeinen habe.

4. Ich mache mir nie Sorgen, dass ich mich weiblichen Schülerinnen gegenüber auf subtile Weise voreingenommen verhalten könnte.

Skala in Bezug auf Männer, Kontext Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber männlichen Schülern habe.
2. Wenn ich mit männlichen Schülern spreche, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich männliche Schüler mag, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber männlichen Schülern im Allgemeinen habe.
4. Ich mache mir nie Sorgen, dass ich mich männlichen Schülern gegenüber auf subtile Weise voreingenommen verhalten könnte.

Allgemeine Skala, Kontext selbst unterrichtete Schülerinnen und Schüler

1. Obwohl ich weiß, dass es nicht angebracht ist, habe ich manchmal das Gefühl, dass ich unbewusst eine negative Einstellung gegenüber Schüler*innen habe, welche ich aktuell unterrichte.

2. Wenn ich mit Schüler*innen spreche, die ich aktuell unterrichte, mache ich mir manchmal Sorgen, dass ich mich ungewollt voreingenommen verhalte.
3. Auch wenn ich die Schüler*innen mag, die ich aktuell unterrichte, mache ich mir Sorgen, dass ich unbewusste Vorurteile gegenüber Schüler*innen habe, die ich aktuell unterrichte.
4. Ich mache mir nie Sorgen, dass ich mich Schüler*innen, die ich aktuell unterrichte, gegenüber auf subtile Weise voreingenommen verhalten könnte.

B. R-Code zur statistischen Datenanalyse

```
1 data <- data.frame(read.csv("data_awareness2024_2024-12-08_13-47.csv",
2                             header=TRUE, sep=";", dec=",", fileEncoding="UTF-16LE"))
3
4 ## DATA PREPARATION
5 ## CLEANUP
6 # Delete unused columns
7 data[,c('CASE', 'SERIAL', 'REF', 'QUESTNNR', 'MODE', 'MAILENT', 'Q_VIEWER', '
8         LASTDATA')] <- list(NULL)
9 # Delete users that quit before first bias awareness question
10 data <- data[!is.na(data$BA01_01),]
11 # Delete single user with gave unclear job description
12 data <- data[data$BE01_04 != "Hallo",]
13 # Replace -9 with NA
14 data[] <- lapply(data, function(x) replace(x, x == -9, NA))
15 # Discard users that didn't answer at least to page 6 with BA questions about
16     SuS
17 data <- data[data$LASTPAGE > 5,]
18
19 ## REFACTORING
20 # Combine columns for origin country
21 data$SD19 <- ifelse(data$SD19 == 1, "Deutschland", data$SD19_02)
22 data$SD19_02 <- NULL
23 # Combine columns for paternal origin country
24 data$SD20 <- ifelse(data$SD20 == 1, "Deutschland", data$SD20_02)
25 data$SD20_02 <- NULL
26 # Combine columns for maternal origin country
27 data$SD21 <- ifelse(data$SD21 == 1, "Deutschland", data$SD21_02)
28 data$SD21_02 <- NULL
29 # Combine columns for teaching experience
30 data$BE07_01 <- ifelse(data$BE07 == 2, 0, ifelse(data$BE07 == -9, NA, data$BE07
31         _01))
32 data$BE07 <- data$BE07_01
33 data$BE07_01 <- NULL
34
35 # Correct job descriptions
36 # Set correct job for students who gave unclear job description
37 data$BE01[grepl("stud", data$BE01_04, ignore.case=TRUE)] <- 1
38 # Now the other users which gave an additional job description are teachers
39 data$BE01[data$BE01_04 != "" & !grepl("stud", data$BE01_04, ignore.case=TRUE)]
40     <- 3
```

```
37 # Set job for users in Referendariat to students
38 data$BE01[data$BE01 == 2] <- 1
39 # Additional job descriptions can be deleted now
40 data[,c("BE01_04")] <- list(NULL)
41
42 percent_equal <- function(x) {
43   x <- na.omit(x)
44   if (length(x) == 0) {
45     return(NA)
46   }
47   hw <- max(table(x))
48   hw / length(x) * 100
49 }
50
51 # Users that gave the same answer more than 75% of the time can be discarded
52 data$ProzentGleich <- apply(data[,c(2:45)], 1, percent_equal)
53 data <- data[data$ProzentGleich < 75,]
54
55 # Combine 4-pairs into a single value representing the construct
56 data$BA01 <- rowMeans(data[,c("BA01_01", "BA01_02", "BA01_03", "BA01_04")], na.rm=TRUE)
57 data$BA02 <- rowMeans(data[,c("BA02_01", "BA02_02", "BA02_03", "BA02_04")], na.rm=TRUE)
58 data$BA03 <- rowMeans(data[,c("BA03_01", "BA03_02", "BA03_03", "BA03_04")], na.rm=TRUE)
59 data$BA04 <- rowMeans(data[,c("BA04_01", "BA04_02", "BA04_03", "BA04_04")], na.rm=TRUE)
60 data$BA05 <- rowMeans(data[,c("BA05_01", "BA05_02", "BA05_03", "BA05_04")], na.rm=TRUE)
61 data$BA06 <- rowMeans(data[,c("BA06_01", "BA06_02", "BA06_03", "BA06_04")], na.rm=TRUE)
62 data$BA07 <- rowMeans(data[,c("BA07_01", "BA07_02", "BA07_03", "BA07_04")], na.rm=TRUE)
63 data$BA08 <- rowMeans(data[,c("BA08_01", "BA08_02", "BA08_03", "BA08_04")], na.rm=TRUE)
64 data$BA09 <- rowMeans(data[,c("BA09_01", "BA09_02", "BA09_03", "BA09_04")], na.rm=TRUE)
65 data$BA10 <- rowMeans(data[,c("BA10_01", "BA10_02", "BA10_03", "BA10_04")], na.rm=TRUE)
66 data$BA11 <- rowMeans(data[,c("BA11_01", "BA11_02", "BA11_03", "BA11_04")], na.rm=TRUE)
67
68 ## DESCRIPTIVE STATISTICS
69 # Split data into student and teacher groups
70 data_STU <- data[data$BE01 == 1,]
71 data_LK <- data[data$BE01 == 3,]
72
73 # total amount of study participants
74 total_n <- nrow(data)
75 # Number of students
```

```
76 stu_n <- nrow(data_STU)
77 # Number of teachers
78 lk_n <- nrow(data_LK)
79
80 ## Analyse all study participants
81 # Calculate amount of female/male/diverse participants
82 total_female <- nrow(data[data$SD01 == 1,])
83 total_male <- nrow(data[data$SD01 == 2,])
84 total_diverse <- nrow(data[data$SD01 == 3,])
85 # Calculate average age
86 total_avg_age <- psych::describe(data$SD02_01, na.rm=TRUE)
87 # Calculate amount of participants with migration background
88 total_migr <- nrow(data[data$SD19 != "Deutschland"
89                        | data$SD20 != "Deutschland"
90                        | data$SD21 != "Deutschland",])
91 # Calculate amount of participants with teaching experience
92 total_exp <- nrow(data[data$BE07 > 0,])
93 # Calculate average teaching experience (in years)
94 total_avg_exp <- psych::describe(data$BE07, na.rm=TRUE)
95 # Calculate amount of participants with Hauptschulabschluss
96 total_fb_hs <- nrow(data[data$BE02_01 == 2,])
97 # Calculate amount of participants with Realschulabschluss
98 total_fb_rs <- nrow(data[data$BE02_02 == 2,])
99 # Calculate amount of participants with Fachhochschulreife
100 total_fb_fh <- nrow(data[data$BE02_03 == 2,])
101 # Calculate amount of participants with Abitur
102 total_fb_abi <- nrow(data[data$BE02_04 == 2,])
103 # Calculate amount of participants with bachelor's degree
104 total_fb_ba <- nrow(data[data$BE02_05 == 2,])
105 # Calculate amount of participants with master's degree
106 total_fb_ma <- nrow(data[data$BE02_06 == 2,])
107 # Calculate amount of participants with Diplom
108 total_fb_dipl <- nrow(data[data$BE02_07 == 2,])
109 # Calculate amount of participants with Staatsexamen
110 total_fb_stex <- nrow(data[data$BE02_08 == 2,])
111 # Calculate amount of participants with doctorate's degree
112 total_fb_doc <- nrow(data[data$BE02_09 == 2,])
113
114 ## Analyse students
115 # Calculate amount of female/male/diverse students
116 stu_female <- nrow(data_STU[data_STU$SD01 == 1,])
117 stu_male <- nrow(data_STU[data_STU$SD01 == 2,])
118 stu_diverse <- nrow(data_STU[data_STU$SD01 == 3,])
119 # Calculate average age
120 stu_avg_age <- psych::describe(data_STU$SD02_01, na.rm=TRUE)
121 # Calculate amount of students with migration background
122 stu_migr <- nrow(data_STU[data_STU$SD19 != "Deutschland"
123                        | data_STU$SD20 != "Deutschland"
124                        | data_STU$SD21 != "Deutschland",])
125 # Calculate amount of students with teaching experience
```

```
126 stu_exp <- nrow(data_STU[data_STU$BE07 > 0,])
127 # Calculate average teaching experience (in years)
128 stu_avg_exp <- psych::describe(data_STU$BE07, na.rm=TRUE)
129 # Calculate average semester
130 stu_avg_sem <- psych::describe(data_STU$BE03_01, na.rm=TRUE)
131 # Calculate amount that passed OSP or SPS
132 stu_osp_sps <- nrow(data_STU[data_STU$BE05 == 1
133                        | data_STU$BE06 == 1,])
134 # Calculate amount of students with Hauptschulabschluss
135 stu_fb_hs <- nrow(data_STU[data_STU$BE02_01 == 2,])
136 # Calculate amount of students with Realschulabschluss
137 stu_fb_rs <- nrow(data_STU[data_STU$BE02_02 == 2,])
138 # Calculate amount of students with Fachhochschulreife
139 stu_fb_fh <- nrow(data_STU[data_STU$BE02_03 == 2,])
140 # Calculate amount of students with Abitur
141 stu_fb_abi <- nrow(data_STU[data_STU$BE02_04 == 2,])
142 # Calculate amount of students with bachelor's degree
143 stu_fb_ba <- nrow(data_STU[data_STU$BE02_05 == 2,])
144 # Calculate amount of students with master's degree
145 stu_fb_ma <- nrow(data_STU[data_STU$BE02_06 == 2,])
146 # Calculate amount of students with Diplom
147 stu_fb_dipl <- nrow(data_STU[data_STU$BE02_07 == 2,])
148 # Calculate amount of students with Staatsexamen
149 stu_fb_stex <- nrow(data_STU[data_STU$BE02_08 == 2,])
150 # Calculate amount of students with doctorate's degree
151 stu_fb_doc <- nrow(data_STU[data_STU$BE02_09 == 2,])
152 # Calculate amount of students with goal Grundschule
153 stu_sa_gru <- nrow(data_STU[data_STU$BE08 == 1 & !is.na(data_STU$BE08),])
154 # Calculate amount of students with goal Sekundarstufe 1
155 stu_sa_ <- nrow(data_STU[data_STU$BE08 == 2 & !is.na(data_STU$BE08),])
156 # Calculate amount of students with goal Gymnasium
157 stu_sa_ <- nrow(data_STU[data_STU$BE08 == 3 & !is.na(data_STU$BE08),])
158 # Calculate amount of students with goal Berufliche Schule
159 stu_sa_ <- nrow(data_STU[data_STU$BE08 == 4 & !is.na(data_STU$BE08),])
160 # Calculate amount of students with goal Sonderpädagogik
161 stu_sa_ <- nrow(data_STU[data_STU$BE08 == 5 & !is.na(data_STU$BE08),])
162
163 ## Analyse teachers
164 # Calculate amount of female/male/diverse teachers
165 lk_female <- nrow(data_LK[data_LK$SD01 == 1,])
166 lk_male <- nrow(data_LK[data_LK$SD01 == 2,])
167 lk_diverse <- nrow(data_LK[data_LK$SD01 == 3,])
168 # Calculate average age
169 lk_avg_age <- psych::describe(data_LK$SD02_01, na.rm=TRUE)
170 # Calculate amount of teachers with migration background
171 lk_migr <- nrow(data_LK[data_LK$SD19 != "Deutschland"
172                        | data_LK$SD20 != "Deutschland"
173                        | data_LK$SD21 != "Deutschland",])
174 # Calculate average teaching experience (in years)
175 lk_avg_exp <- psych::describe(data_LK$BE07, na.rm=TRUE)
```

```

176 # Calculate amount of teachers with Hauptschulabschluss
177 lk_fb_hs <- nrow(data_LK[data_LK$BE02_01 == 2,])
178 # Calculate amount of teachers with Realschulabschluss
179 lk_fb_rs <- nrow(data_LK[data_LK$BE02_02 == 2,])
180 # Calculate amount of teachers with Fachhochschulreife
181 lk_fb_fh <- nrow(data_LK[data_LK$BE02_03 == 2,])
182 # Calculate amount of teachers with Abitur
183 lk_fb_abi <- nrow(data_LK[data_LK$BE02_04 == 2,])
184 # Calculate amount of teachers with bachelor's degree
185 lk_fb_ba <- nrow(data_LK[data_LK$BE02_05 == 2,])
186 # Calculate amount of teachers with master's degree
187 lk_fb_ma <- nrow(data_LK[data_LK$BE02_06 == 2,])
188 # Calculate amount of teachers with Diplom
189 lk_fb_dipl <- nrow(data_LK[data_LK$BE02_07 == 2,])
190 # Calculate amount of teachers with Staatsexamen
191 lk_fb_stex <- nrow(data_LK[data_LK$BE02_08 == 2,])
192 # Calculate amount of teachers with doctorate's degree
193 lk_fb_doc <- nrow(data_LK[data_LK$BE02_09 == 2,])
194 # Calculate amount of teachers at Grundschule
195 lk_sa_gru <- nrow(data_LK[data_LK$BE04 == 1 & !is.na(data_LK$BE04),])
196 # Calculate amount of teachers at Sekundarstufe 1
197 lk_sa_sek1 <- nrow(data_LK[data_LK$BE04 == 2 & !is.na(data_LK$BE04),])
198 # Calculate amount of teachers at Gymnasium
199 lk_sa_gym <- nrow(data_LK[data_LK$BE04 == 3 & !is.na(data_LK$BE04),])
200 # Calculate amount of teachers at Berufliche Schule
201 lk_sa_bs <- nrow(data_LK[data_LK$BE04 == 4 & !is.na(data_LK$BE04),])
202 # Calculate amount of teachers at Sonderpädagogik teacher
203 lk_sa_snd <- nrow(data_LK[data_LK$BE04 == 5 & !is.na(data_LK$BE04),])
204 # The rest
205 lk_sa_oth <- nrow(data_LK[data_LK$BE04 == 6 & !is.na(data_LK$BE04),])
206
207 # Run normalcy tests for combined variables, all participants
208 shapiro.test(data$BA01)
209 shapiro.test(data$BA02)
210 shapiro.test(data$BA03)
211 shapiro.test(data$BA04)
212 shapiro.test(data$BA05)
213 shapiro.test(data$BA06)
214 shapiro.test(data$BA07)
215 shapiro.test(data$BA08)
216 shapiro.test(data$BA09)
217 shapiro.test(data$BA10)
218 shapiro.test(data$BA11)
219
220 # Calculate cronbach alphas for all study participants
221 psych::alpha(data[,c("BA01_01", "BA01_02", "BA01_03", "BA01_04")])
222 psych::alpha(data[,c("BA02_01", "BA02_02", "BA02_03", "BA02_04")])
223 psych::alpha(data[,c("BA03_01", "BA03_02", "BA03_03", "BA03_04")])
224 psych::alpha(data[,c("BA04_01", "BA04_02", "BA04_03", "BA04_04")])
225 psych::alpha(data[,c("BA05_01", "BA05_02", "BA05_03", "BA05_04")])

```

```
226 psych::alpha(data[,c("BA06_01", "BA06_02", "BA06_03", "BA06_04")])
227 psych::alpha(data[,c("BA07_01", "BA07_02", "BA07_03", "BA07_04")])
228 psych::alpha(data[,c("BA08_01", "BA08_02", "BA08_03", "BA08_04")])
229 psych::alpha(data[,c("BA09_01", "BA09_02", "BA09_03", "BA09_04")])
230 psych::alpha(data[,c("BA10_01", "BA10_02", "BA10_03", "BA10_04")])
231 psych::alpha(data[,c("BA11_01", "BA11_02", "BA11_03", "BA11_04")])
232
233 # Run normalcy tests for combined variables, students only
234 shapiro.test(data_STU$BA01)
235 shapiro.test(data_STU$BA02)
236 shapiro.test(data_STU$BA03)
237 shapiro.test(data_STU$BA04)
238 shapiro.test(data_STU$BA05)
239 shapiro.test(data_STU$BA06)
240 shapiro.test(data_STU$BA07)
241 shapiro.test(data_STU$BA08)
242 shapiro.test(data_STU$BA09)
243 shapiro.test(data_STU$BA10)
244 shapiro.test(data_STU$BA11)
245
246 # Calculate cronbach alphas for students only
247 psych::alpha(data_STU[,c("BA01_01", "BA01_02", "BA01_03", "BA01_04")])
248 psych::alpha(data_STU[,c("BA02_01", "BA02_02", "BA02_03", "BA02_04")])
249 psych::alpha(data_STU[,c("BA03_01", "BA03_02", "BA03_03", "BA03_04")])
250 psych::alpha(data_STU[,c("BA04_01", "BA04_02", "BA04_03", "BA04_04")])
251 psych::alpha(data_STU[,c("BA05_01", "BA05_02", "BA05_03", "BA05_04")])
252 psych::alpha(data_STU[,c("BA06_01", "BA06_02", "BA06_03", "BA06_04")])
253 psych::alpha(data_STU[,c("BA07_01", "BA07_02", "BA07_03", "BA07_04")])
254 psych::alpha(data_STU[,c("BA08_01", "BA08_02", "BA08_03", "BA08_04")])
255 psych::alpha(data_STU[,c("BA09_01", "BA09_02", "BA09_03", "BA09_04")])
256 psych::alpha(data_STU[,c("BA10_01", "BA10_02", "BA10_03", "BA10_04")])
257 psych::alpha(data_STU[,c("BA11_01", "BA11_02", "BA11_03", "BA11_04")])
258
259 # Run normalcy tests for combined variables, teachers only
260 shapiro.test(data_LK$BA01)
261 shapiro.test(data_LK$BA02)
262 shapiro.test(data_LK$BA03)
263 shapiro.test(data_LK$BA04)
264 shapiro.test(data_LK$BA05)
265 shapiro.test(data_LK$BA06)
266 shapiro.test(data_LK$BA07)
267 shapiro.test(data_LK$BA08)
268 shapiro.test(data_LK$BA09)
269 shapiro.test(data_LK$BA10)
270 shapiro.test(data_LK$BA11)
271
272 # Calculate cronbach alphas for teachers only
273 psych::alpha(data_LK[,c("BA01_01", "BA01_02", "BA01_03", "BA01_04")])
274 psych::alpha(data_LK[,c("BA02_01", "BA02_02", "BA02_03", "BA02_04")])
275 psych::alpha(data_LK[,c("BA03_01", "BA03_02", "BA03_03", "BA03_04")])
```



```

276 psych::alpha(data_LK[,c("BA04_01", "BA04_02", "BA04_03", "BA04_04")])
277 psych::alpha(data_LK[,c("BA05_01", "BA05_02", "BA05_03", "BA05_04")])
278 psych::alpha(data_LK[,c("BA06_01", "BA06_02", "BA06_03", "BA06_04")])
279 psych::alpha(data_LK[,c("BA07_01", "BA07_02", "BA07_03", "BA07_04")])
280 psych::alpha(data_LK[,c("BA08_01", "BA08_02", "BA08_03", "BA08_04")])
281 psych::alpha(data_LK[,c("BA09_01", "BA09_02", "BA09_03", "BA09_04")])
282 psych::alpha(data_LK[,c("BA10_01", "BA10_02", "BA10_03", "BA10_04")])
283 psych::alpha(data_LK[,c("BA11_01", "BA11_02", "BA11_03", "BA11_04")])
284
285 ## INFERENCE STATISTICS
286 # TESTS WITH ALL DATA
287 # Test BA with allg items
288 # BA allg / allg <--> BA SuS / allg
289 t.test(data$BA01, data$BA06, paired=TRUE)
290 sd(data$BA01 - data$BA06, na.rm=TRUE)
291 print(lsr::cohensD(data$BA01, data$BA06, method="paired"))
292 # BA allg / allg <--> BA eigene SuS
293 t.test(data$BA01, data$BA07, paired=TRUE)
294 sd(data$BA01 - data$BA07, na.rm=TRUE)
295 print(lsr::cohensD(data$BA01, data$BA07, method="paired"))
296 # BA SuS / allg <--> BA eigene SuS
297 t.test(data$BA06, data$BA07, paired=TRUE)
298 sd(data$BA06 - data$BA07, na.rm=TRUE)
299 print(lsr::cohensD(data$BA06, data$BA07, method="paired"))
300
301 # Test BA with domain-specific items
302 # BA allg / migr <--> BA SuS / migr
303 t.test(data$BA02, data$BA08, paired=TRUE)
304 sd(data$BA02 - data$BA08, na.rm=TRUE)
305 print(lsr::cohensD(data$BA02, data$BA08, method="paired"))
306 # BA allg / sozök <--> BA SuS / sozök
307 t.test(data$BA03, data$BA09, paired=TRUE)
308 sd(data$BA03 - data$BA09, na.rm=TRUE)
309 print(lsr::cohensD(data$BA03, data$BA09, method="paired"))
310 # BA allg / frauen <--> BA SuS / frauen
311 t.test(data$BA04, data$BA10, paired=TRUE)
312 sd(data$BA04 - data$BA10, na.rm=TRUE)
313 print(lsr::cohensD(data$BA04, data$BA10, method="paired"))
314 # Additional wilcoxon test because shapiro-wilk failed
315 wilcox.test(data$BA04, data$BA10, exact=FALSE, paired=TRUE)
316 effectsize::rank_biserial(data$BA04, data$BA10, paired=TRUE)
317 # BA allg / männer <--> BA SuS / männer
318 t.test(data$BA05, data$BA11, paired=TRUE)
319 sd(data$BA05 - data$BA11, na.rm=TRUE)
320 print(lsr::cohensD(data$BA05, data$BA11, method="paired"))
321 # Additional wilcoxon test because shapiro-wilk failed
322 wilcox.test(data$BA05, data$BA11, exact=FALSE, paired=TRUE)
323 effectsize::rank_biserial(data$BA05, data$BA11, paired=TRUE)
324
325 # TESTS WITH STUDENTS ONLY

```

```

326 # Test BA with allg items
327 # BA allg / allg <--> BA SuS / allg
328 t.test(data_STU$BA01, data_STU$BA06, paired=TRUE)
329 sd(data_STU$BA01 - data_STU$BA06, na.rm=TRUE)
330 print(lsr::cohensD(data_STU$BA01, data_STU$BA06, method="paired"))
331 # BA allg / allg <--> BA eigene SuS
332 t.test(data_STU$BA01, data_STU$BA07, paired=TRUE)
333 sd(data_STU$BA01 - data_STU$BA07, na.rm=TRUE)
334 print(lsr::cohensD(data_STU$BA01, data_STU$BA07, method="paired"))
335 # BA SuS / allg <--> BA eigene SuS
336 t.test(data_STU$BA06, data_STU$BA07, paired=TRUE)
337 sd(data_STU$BA06 - data_STU$BA07, na.rm=TRUE)
338 print(lsr::cohensD(data_STU$BA06, data_STU$BA07, method="paired"))
339
340 # Test BA with domain-specific items
341 # BA allg / migr <--> BA SuS / migr
342 t.test(data_STU$BA02, data_STU$BA08, paired=TRUE)
343 sd(data_STU$BA02 - data_STU$BA08, na.rm=TRUE)
344 print(lsr::cohensD(data_STU$BA02, data_STU$BA08, method="paired"))
345 # Additional wilcoxon test because shapiro-wilk failed
346 wilcox.test(data_STU$BA02, data_STU$BA08, exact=FALSE, paired=TRUE)
347 effectsize::rank_biserial(data_STU$BA02, data_STU$BA08, paired=TRUE)
348 # BA allg / sozök <--> BA SuS / sozök
349 t.test(data_STU$BA03, data_STU$BA09, paired=TRUE)
350 sd(data_STU$BA03 - data_STU$BA09, na.rm=TRUE)
351 print(lsr::cohensD(data_STU$BA03, data_STU$BA09, method="paired"))
352 # BA allg / frauen <--> BA SuS / frauen
353 t.test(data_STU$BA04, data_STU$BA10, paired=TRUE)
354 sd(data_STU$BA04 - data_STU$BA10, na.rm=TRUE)
355 print(lsr::cohensD(data_STU$BA04, data_STU$BA10, method="paired"))
356 # Additional wilcoxon test because shapiro-wilk failed
357 wilcox.test(data_STU$BA04, data_STU$BA10, exact=FALSE, paired=TRUE)
358 effectsize::rank_biserial(data_STU$BA04, data_STU$BA10, paired=TRUE)
359 # BA allg / männer <--> BA SuS / männer
360 t.test(data_STU$BA05, data_STU$BA11, paired=TRUE)
361 sd(data_STU$BA05 - data_STU$BA11, na.rm=TRUE)
362 print(lsr::cohensD(data_STU$BA05, data_STU$BA11, method="paired"))
363 # Additional wilcoxon test because shapiro-wilk failed
364 wilcox.test(data_STU$BA05, data_STU$BA11, exact=FALSE, paired=TRUE)
365 effectsize::rank_biserial(data_STU$BA05, data_STU$BA11, paired=TRUE)
366
367 # TESTS WITH TEACHERS ONLY
368 # Test BA with allg items
369 # BA allg / allg <--> BA SuS / allg
370 t.test(data_LK$BA01, data_LK$BA06, paired=TRUE)
371 sd(data_LK$BA01 - data_LK$BA06, na.rm=TRUE)
372 print(lsr::cohensD(data_LK$BA01, data_LK$BA06, method="paired"))
373 # BA allg / allg <--> BA eigene SuS
374 t.test(data_LK$BA01, data_LK$BA07, paired=TRUE)
375 sd(data_LK$BA01 - data_LK$BA07, na.rm=TRUE)

```

```

376 print(lsr::cohensD(data_LK$BA01, data_LK$BA07, method="paired"))
377 # Additional wilcoxon test because shapiro-wilk failed
378 wilcox.test(data_LK$BA01, data_LK$BA07, exact=FALSE, paired=TRUE)
379 effectsize::rank_biserial(data_LK$BA01, data_LK$BA07, paired=TRUE)
380 # BA SuS/allg <--> BA eigene SuS
381 t.test(data_LK$BA06, data_LK$BA07, paired=TRUE)
382 sd(data_LK$BA06 - data_LK$BA07, na.rm=TRUE)
383 print(lsr::cohensD(data_LK$BA06, data_LK$BA07, method="paired"))
384 # Additional wilcoxon test because shapiro-wilk failed
385 wilcox.test(data_LK$BA06, data_LK$BA07, exact=FALSE, paired=TRUE)
386 effectsize::rank_biserial(data_LK$BA06, data_LK$BA07, paired=TRUE)
387
388 # Test BA with domain-specific items
389 # BA allg / migr <--> BA SuS / migr
390 t.test(data_LK$BA02, data_LK$BA08, paired=TRUE)
391 sd(data_LK$BA02 - data_LK$BA08, na.rm=TRUE)
392 print(lsr::cohensD(data_LK$BA02, data_LK$BA08, method="paired"))
393 # BA allg / sozök <--> BA SuS / sozök
394 t.test(data_LK$BA03, data_LK$BA09, paired=TRUE)
395 sd(data_LK$BA03 - data_LK$BA08, na.rm=TRUE)
396 print(lsr::cohensD(data_LK$BA03, data_LK$BA09, method="paired"))
397 # BA allg / frauen <--> BA SuS / frauen
398 t.test(data_LK$BA04, data_LK$BA10, paired=TRUE)
399 sd(data_LK$BA04 - data_LK$BA10, na.rm=TRUE)
400 print(lsr::cohensD(data_LK$BA04, data_LK$BA10, method="paired"))
401 # Additional wilcoxon test because shapiro-wilk failed
402 wilcox.test(data_LK$BA04, data_LK$BA10, exact=FALSE, paired=TRUE)
403 effectsize::rank_biserial(data_LK$BA04, data_LK$BA10, paired=TRUE)
404 # BA allg / männer <--> BA SuS / männer
405 t.test(data_LK$BA05, data_LK$BA11, paired=TRUE)
406 sd(data_LK$BA05 - data_LK$BA11, na.rm=TRUE)
407 print(lsr::cohensD(data_LK$BA05, data_LK$BA11, method="paired"))
408 # Additional wilcoxon test because shapiro-wilk failed
409 wilcox.test(data_LK$BA05, data_LK$BA11, exact=FALSE, paired=TRUE)
410 effectsize::rank_biserial(data_LK$BA05, data_LK$BA11, paired=TRUE)
411
412 # COMPARISON OF STUDENTS AND TEACHERS
413 # BA allg / allg
414 t.test(data_STU$BA01, data_LK$BA01, paired=FALSE)
415 mean(data_STU$BA01, na.rm=TRUE) - mean(data_LK$BA01, na.rm=TRUE)
416 effectsize::sd_pooled(data_STU$BA01, data_LK$BA01)
417 print(lsr::cohensD(data_STU$BA01, data_LK$BA01))
418 # BA allg / migr
419 t.test(data_STU$BA02, data_LK$BA02, paired=FALSE)
420 mean(data_STU$BA02, na.rm=TRUE) - mean(data_LK$BA02, na.rm=TRUE)
421 effectsize::sd_pooled(data_STU$BA02, data_LK$BA02)
422 print(lsr::cohensD(data_STU$BA02, data_LK$BA02))
423 # BA allg / sozök
424 t.test(data_STU$BA03, data_LK$BA03, paired=FALSE)
425 mean(data_STU$BA03, na.rm=TRUE) - mean(data_LK$BA03, na.rm=TRUE)

```

```
426 effectsize::sd_pooled(data_STU$BA03, data_LK$BA03)
427 print(lsr::cohensD(data_STU$BA03, data_LK$BA03))
428 # BA allg/frauen
429 t.test(data_STU$BA04, data_LK$BA04, paired=FALSE)
430 mean(data_STU$BA04, na.rm=TRUE) - mean(data_LK$BA04, na.rm=TRUE)
431 effectsize::sd_pooled(data_STU$BA04, data_LK$BA04)
432 print(lsr::cohensD(data_STU$BA04, data_LK$BA04))
433 # BA allg/männer
434 t.test(data_STU$BA05, data_LK$BA05, paired=FALSE)
435 mean(data_STU$BA05, na.rm=TRUE) - mean(data_LK$BA05, na.rm=TRUE)
436 effectsize::sd_pooled(data_STU$BA05, data_LK$BA05)
437 print(lsr::cohensD(data_STU$BA05, data_LK$BA05))
438 # BA SuS/allg
439 t.test(data_STU$BA06, data_LK$BA06, paired=FALSE)
440 mean(data_STU$BA06, na.rm=TRUE) - mean(data_LK$BA06, na.rm=TRUE)
441 effectsize::sd_pooled(data_STU$BA06, data_LK$BA06)
442 print(lsr::cohensD(data_STU$BA06, data_LK$BA06))
443 # BA own SuS/allg
444 t.test(data_STU$BA07, data_LK$BA07, paired=FALSE)
445 mean(data_STU$BA07, na.rm=TRUE) - mean(data_LK$BA07, na.rm=TRUE)
446 effectsize::sd_pooled(data_STU$BA07, data_LK$BA07)
447 print(lsr::cohensD(data_STU$BA07, data_LK$BA07))
448 # BA SuS/migr
449 t.test(data_STU$BA08, data_LK$BA08, paired=FALSE)
450 mean(data_STU$BA08, na.rm=TRUE) - mean(data_LK$BA08, na.rm=TRUE)
451 effectsize::sd_pooled(data_STU$BA08, data_LK$BA08)
452 print(lsr::cohensD(data_STU$BA08, data_LK$BA08))
453 # BA SuS/sozök
454 t.test(data_STU$BA09, data_LK$BA09, paired=FALSE)
455 mean(data_STU$BA09, na.rm=TRUE) - mean(data_LK$BA09, na.rm=TRUE)
456 effectsize::sd_pooled(data_STU$BA09, data_LK$BA09)
457 print(lsr::cohensD(data_STU$BA09, data_LK$BA09))
458 # BA SuS/frauen
459 t.test(data_STU$BA10, data_LK$BA10, paired=FALSE)
460 mean(data_STU$BA10, na.rm=TRUE) - mean(data_LK$BA10, na.rm=TRUE)
461 effectsize::sd_pooled(data_STU$BA10, data_LK$BA10)
462 print(lsr::cohensD(data_STU$BA10, data_LK$BA10))
463 # BA SuS/männer
464 t.test(data_STU$BA11, data_LK$BA11, paired=FALSE)
465 mean(data_STU$BA11, na.rm=TRUE) - mean(data_LK$BA11, na.rm=TRUE)
466 effectsize::sd_pooled(data_STU$BA11, data_LK$BA11)
467 print(lsr::cohensD(data_STU$BA11, data_LK$BA11))
```