

Auswirkungen verschiedener Aufsichtsbedingungen auf die Datenqualität bei standardisierten Befragungen im Klassenraum¹

Till Stefes

Beitrag zur Veranstaltung »Themen der empirischen Sozialforschung« der Sektion Methoden der empirischen Sozialforschung

Einleitung

Verschiedenste wissenschaftliche Disziplinen nutzen zur Befragung von jungen Menschen standardisierte, selbst-administrierte Befragungen in Klassenräumen, wobei Lehrkräfte oder externe Aufsichtspersonen den störungsfreien Ablauf sicherstellen (Hallfors et al. 2000; Lucia et al. 2007; Alibali und Nathan 2010; Bartlett et al. 2017; March et al. 2022), was aus mehreren Gründen praktisch ist: Befragungen, die in Schulen durchgeführt werden, gelten anderen Modi gegenüber als valider (Kann et al. 2002; Cops et al. 2016). Die Zusammenarbeit mit Schulen ist zeitsparend, da Forschende ganze Klassen auf einmal befragen können und weder Hausbesuche machen (Walser und Killias 2012) noch auf postalische Antworten warten müssen. Die Rekrutierung von Befragten über Schulen ist eine effiziente Strategie, um repräsentative Stichproben zu realisieren (Ellonen et al. 2023). Dabei kann die Schule als durchaus ambivalente Befragungssituation beschrieben werden.

Lehrkräfte können für den Erfolg einer Befragung von entscheidender Bedeutung sein: Eltern vertrauen ihnen, Schüler verlassen sich auf ihr Fachwissen und ihre Unterstützung und sie fungieren als Autoritätspersonen in der Befragungssituation. Sich jedoch allein auf sie zu verlassen, um eine wissenschaftliche Untersuchung durchzuführen, könnte Ergebnisse verzerren, da sie das Antwortverhalten der Schüler*innen beeinflussen können (Strange et al. 2003; Heath et al. 2013; Rasberry et al. 2020). Die Anwesenheit externer Aufsichtspersonen kann das Risiko dieser Verzerrungen verringern, wie etwa Bidonde et al. (2023) vermuten. Ihre Anwesenheit ist jedoch mit zusätzlichen und oft erheblichen Kosten verbunden (Walser und Killias 2012). Es stellen sich daher zwei Fragen: (1) Gefährdet eine unkontrollierte Aufsicht durch Lehrkräfte die Datenqualität und (2) ist die Investition in eine externe Aufsicht notwendig?

Dieser Beitrag stützt sich auf einen Datensatz, der aus einem quasi-experimentellen Studiendesign hervorgegangen ist. Die untersuchte Befragung wurde 2021 im Ruhrgebiet durchgeführt und testete verschiedene Formen der digitalen, hybriden und persönlichen Aufsicht. UWE („Umwelt, Wohlbefinden

¹ Auf Englisch erschienen als: Stefes, Till. 2024: Data Quality of Different Modes of Supervision in Classroom Surveys. Education Sciences 14 (3): 299. <https://doi.org/10.3390/educsci14030299>

und Entwicklung“) ist eine wiederholte Querschnittsstudie, die bisher nur in Klassenräumen durchgeführt wird. Sie dient mehreren Zwecken, darunter der soziologischen Erforschung des gelingenden Aufwachsens, aber vor allem soll sie steuerungsrelevante Daten für Akteur*innen in Schulen und Kommunen liefern. Das Ziel dieses Beitrages ist es zu untersuchen, wie sich Item-Nonresponse, Befragungsdauer und Antwortmuster zwischen den Gruppen je nach ihrer Aufsicht unterscheiden.

Bisherige Forschung

Eine wichtige Voraussetzung für erfolgreiche Forschungsarbeiten, an denen Schulen und Schulpersonal beteiligt sind, ist die Anpassung des Befragungsmodus an die Bedürfnisse der Schulen (Hatch et al. 2023). Die von Rasberry et al. (2020) vorgestellte Studie zeigt jedoch, wie schwierig Umfrageforschung in Schulen sein kann. Obwohl offenbar alle aus der Literatur bekannten Fallstricke bedacht wurden, hatten die Forschenden im Feld mit nachlässigen Lehrkräften und verschwundenen Fragebögen zu kämpfen.

Wenn Befragungen in Schulen durchgeführt werden, ist es durchaus üblich, mit Lehrkräften als „Assistenzaufsichtspersonen“ zusammenzuarbeiten, da diese bereits Autoritätspersonen in den Zielgruppen sind und in der Regel mit Gruppendynamik und Störungen umgehen können. Auch wenn ihre Fähigkeiten in dieser Hinsicht variieren können (March et al. 2022), sollte dies die Datenqualität in Bezug auf Validität und Rücklaufquote erhöhen (Bidonde et al. 2023). Sie können dazu beitragen, die Motivation der Befragten aufrechtzuerhalten und grundlegende Informationen, wie Postleitzahlen, zu erfassen. Retrospektive Daten könnten genauer sein, wenn eine Vertrauensperson dabei helfen kann, sich an Vergangenes zu erinnern. Lehrkräfte können außerdem Kompetenzen und Sozialverhalten der Befragten gut einschätzen. Darauf finden sich immer wieder Hinweise aus den Protokollen der UWE-Befragung, die hier analysiert werden. Zusammenfassend sprechen einige Argumente dafür, dass Lehrkräfte dem Befragungsprozess zuträglich sein können, woraus sich folgende Hypothese ergibt:

Hypothese 1a: Bei der Durchführung standardisierter Befragungen in Klassenräumen erhöht die Anwesenheit von Lehrkräften die Datenqualität.

Wenn Lehrkräfte während ihrer üblichen Unterrichtszeit eine Befragung durchführen oder dabei unterstützen sollen, müssen sie zwei Rollen miteinander in Einklang bringen. Sie fungieren als Aufsichtspersonen, sind aber gleichzeitig auch Dritte, ohne dass sie sich einer der beiden Rollen klar und eindeutig zuordnen ließen. In ihrer Rolle als Aufsichtspersonen verteilen sie Fragebögen oder Links zu Online-Umfragen und sind die Hauptinformationsquelle für Verständnisfragen. In vielen Fällen lesen sie auch eine Einführung vor. Aus Sicht der Forschenden gelten sie jedoch als Dritte. Vor allem, weil sie eine persönliche Beziehung zu den Befragten haben, was ihre Neutralität gefährdet. Vor der Durchführung der Befragung werden Lehrkräfte in den meisten Fällen nicht als Interviewer*innen geschult, obwohl Hatch et al. (2023) dies dringend empfehlen.

Lehrkräfte können Effekte sozialer Erwünschtheit hervorrufen oder verstärken und damit die Datenqualität gefährden. Soziale Erwünschtheit meint die Tendenz von Befragten, eher erwartbar sozial akzeptierte Antworten zu geben als ihre wahren Meinungen oder Verhaltensweisen zu äußern, was bei von Interviewer*innen geleiteten Befragungen häufiger auftritt (Atkeson et al. 2014). Jugendliche sind möglicherweise besonders anfällig für Effekte sozialer Erwünschtheit, da sie oft stark von Gleichaltrigen und sozialen Normen beeinflusst werden (Brown und Larson 2009). Die bloße Anwesenheit von Lehrkräften kann diesen Effekt erhöhen, da sie Autoritätspersonen sind (Möhring und Schlütz 2019, S. 49). Sie reproduzieren die gewohnte Klassenraumatmosphäre und somit finden sich die Befragten wahr-

scheinlich in ihrer sozialen Rolle als Schüler*innen wieder, die mit einem gewissen Grad von Gehorsamkeit verbunden ist.

Der Unterschied zwischen einer Befragung und einer Prüfung kann sowohl für Befragte als auch für Lehrkräfte schwer zu verinnerlichen sein. Eine häufige Beobachtung aus den Protokollen der UWE-Befragung ist, dass Lehrkräfte zur Eile neigen. Für eine erfolgreiche Befragung ist es entscheidend, dass sich die Befragten Zeit nehmen, Lehrkräfte dagegen sind an enge Zeitpläne gewöhnt und möchten die Dinge gerne rechtzeitig erledigen. Dies mag zum Teil auf die praktische Notwendigkeit zurückzuführen sein, dass Schulen ihre Personalressourcen effizient einsetzen müssen; was tendenziell zum Nachteil von wissenschaftlichen Befragungsprojekten in ihren Räumen geschieht (March et al. 2022). Darüber hinaus werden Umfragen von Lehrkräften oft als zusätzliche Arbeit empfunden, was durchaus zutreffend ist (March et al. 2022). Alibali und Nathan (2010) sowie Hatch et al. (2023) empfehlen daher dringend, diesem Umstand mit Geduld den Schulen und dem Lehrpersonal gegenüber zu begegnen, da deren Zeit begrenzt ist. Ausgehend von diesen Argumenten gegen die Anwesenheit von Lehrkräften würde eine Gegenhypothese zur ersten Hypothese lauten:

Hypothese 1b: Bei der Durchführung standardisierter Umfragen in Klassenräumen verringert die Anwesenheit von Lehrkräften die Datenqualität.

Die naheliegende Lösung wäre, Lehrkräfte in Umfragemethodik zu schulen, wie es Hatch et al. (2023) getan haben. Das ist teuer und zeitaufwändig, daher oft keine Option. Stattdessen können geschulte Interviewer*innen eingreifen, wenn Lehrkräfte unbeabsichtigt Antworten verzerren, und bei Verständnisfragen helfen. Insbesondere wenn diese Personen selbst aus dem Forschungszusammenhang kommen, können sie anwesenden Lehrkräften und Befragten helfen, den Zweck der Studie zu verstehen, was zu einer höheren Motivation aller Beteiligten führt (March et al. 2022). Daher lautet eine mögliche Hypothese zum Einfluss von externer Aufsicht auf die Datenqualität wie folgt:

Hypothese 2a: Bei der Durchführung standardisierter Umfragen in Klassenräumen erhöht die Anwesenheit externer Aufsicht die Datenqualität.

Laut Epstein (1998) können Erwachsene, die in einem schulischen Umfeld arbeiten, sowohl von Schüler*innen als auch Lehrkräften einfach als zusätzliche Lehrkräfte wahrgenommen werden. Dies kann sich negativ auf die Objektivität der beaufsichtigenden Personen auswirken und auch die Antworten der Jugendlichen beeinflussen. Wenn dies der Fall ist, lohnt sich der Einsatz von Aufsichtspersonen möglicherweise nicht einmal. Strange et al. (2003) fanden keinen signifikanten Unterschied in der Antwortrate, wenn die Umfragen entweder von einer Lehrkraft oder einem Forschenden durchgeführt wurden. Walser und Killias (2012) sowie Kivivuori et al. (2013) befassten sich mit hochsensiblen Fragebögen und stellten ebenfalls nur geringe signifikante Unterschiede im Antwortverhalten von Jugendlichen fest, die von Lehrkräften bzw. externen Aufsichtspersonen beaufsichtigt wurden. Die Gegenhypothese muss daher lauten:

Hypothese 2b: Bei der Durchführung standardisierter Befragungen in Klassenräumen erhöht die Anwesenheit externer Aufsichtspersonen die Datenqualität nicht.

Daten und Methoden

Die Hypothesen werden im Folgenden mit verschiedenen Regressionsanalysen auf der Grundlage von Daten aus der oben kurz beschriebenen Befragung UWE 2021 (verfügbar über Stefes 2023) getestet. Indikatoren der Datenqualität werden getrennt nach Aufsichtsmodus unter Berücksichtigung von Merkmalen der Befragten und der Befragungssituation geschätzt.

UWE befragt Kinder und Jugendliche mit standardisierten Fragebögen zu ihrem Wohlbefinden, ihrem Alltag und ihren sozialen Ressourcen (siehe Stefes et al. 2023). Die Analysen in dieser Studie verwenden Daten aus der Befragung in einer Stadt, die im Frühjahr 2021 trotz der COVID-19-Pandemie eine Zusammenarbeit mit den städtischen weiterführenden Schulen ermöglichte. Im Rahmen dieser Welle hat eine Kohorte von 923 Schüler*innen der siebten und neunten Klasse, im Alter von 12 bis 17 Jahren, Fragebögen ausgefüllt. Während dieses Zeitraums waren die Schulen je nach Inzidenz der COVID-19-Fälle geschlossen oder nur für eine begrenzte Anzahl von Schüler*innen geöffnet. Aufgrund dieser besonderen Umstände wurden in denselben Schulen unterschiedliche Aufsichtsformen angewendet, und sogar die zeitweise aufgeteilten Klassen wurden in unterschiedlichen Konstellationen befragt.

Die Studie sah drei Varianten der Aufsicht vor, darunter (A) nur Lehrkräfte anwesend, (B) Lehrkräfte und externe Aufsichtspersonen, die per Videokonferenztechnologie (VCT) anwesend waren, oder (C) externe Aufsichtspersonen, die per VCT ohne Lehrkräfte Aufsicht geführt haben. Die Fragebögen wurden selbstständig mit Leihgeräten der Schule, eigenen Geräten oder Papierfragebögen ausgefüllt. Die Zuordnung erfolgte ausschließlich auf Basis der wöchentlich wechselnden Kontaktbeschränkungen sowie den Möglichkeiten der Schulen, was die Arbeit mit Videokonferenz-Software oder Onlinesurveys angeht.

Indikatoren und Analysen zur Datenqualität

Anhand der Datensätze von UWE lassen sich vier Faktoren der Datenqualität zuverlässig messen: Item-Nonresponse, Dauer der Befragung und Straightlining.

Allen Befragten wurden 210 Fragen vorgelegt. Die große Anzahl von Fragen ist einer der Gründe, warum eine Aufsicht bisher als obligatorisch angesehen wurde. Die meisten Befragten beantworteten die meisten Fragen, aber es gibt auch viele fehlende Werte im Datensatz. Item-Nonresponse ist ein gängiger Indikator für die Qualität von Umfragedaten (z.B. Gummer et al. 2021; Leiner 2019). Für die Analysen in dieser Studie wird die Item-Nonresponse durch Zählen der Fragen mit fehlenden Werten im Datensatz gemessen. Einige Fragen wurden gefiltert und konnten nicht von allen Befragten beantwortet werden, beispielsweise Folgefragen zum Migrationshintergrund. Diese wurden aus der Zählung ausgeschlossen. Frühzeitig abgebrochene Beobachtungen wurden aus dieser Analyse ausgeschlossen, da sie zum großen Teil extreme Ausreißer sind. Die Entscheidung gegen einen Anteil oder fehlende Informationen anstelle der Rohzählung erfolgt daher, dass eine Zählung leichter zu interpretieren ist als ein Anteil und die Umwandlung in einen Prozentsatz keine zusätzlichen Informationen liefert. Da die resultierende Variable als Zähldaten beschrieben werden kann, ist eine Variante der Poisson-Regression angemessen (Little 1988; Little 1992). Eine Zero-Inflation kann ausgeschlossen werden, da jede Null im Datensatz für einen vollständigen Fragebogen steht. Daher wird ein negatives Binomialmodell zur Schätzung der Anzahl unbeantworteter Fragen angewendet.

Die Dauer der mit digitalen Fragebögen durchgeführten Befragungen wurde aufgezeichnet, sodass auf Anzeichen von Eile geprüft werden kann. Die Interviewdauer ist ein mehrdeutiger Indikator. Eine längere Dauer ist nicht gleichzusetzen mit höherer Qualität, genauso wenig wie weniger Ausfälle oder Nichtantworten Indikatoren für eine validere Messung sind. Fragen können zu langsam oder zu schnell beantwortet werden. Ersteres könnte bedeuten, dass die Befragten Schwierigkeiten haben, die Fragen zu verstehen, dass sie müde oder abgelenkt sind. Letzteres könnte darauf hindeuten, dass sie den Fragebogen zu schnell ausfüllen, alle Fragen gleich oder willkürlich beantworten und somit keine wahrheitsgemäßen Angaben machen. Ein sehr schnelles Ausfüllen des Fragebogens kann also ein Hinweis auf eine unachtsame oder gefälschte Antwort sein. Aus diesem Grund schlägt Leiner (2019) vor, die Ausfüllzeiten zu verwenden, um bedeutungslose Daten zu identifizieren. In Bezug auf die Antwortzeit für einzelne Fragen erwähnen Revilla und Ochoa (2015), dass hochqualifizierte Befragte sehr schnell

antworten könnten, obwohl extrem kurze Antwortzeiten die Möglichkeit ausschließen, dass die Befragten die Frage überhaupt gelesen haben. Eine kurze Antwortzeit im Allgemeinen hänge üblicherweise mit einer geringeren Qualität der Antworten zusammen. Ihre Ergebnisse stützen diese Behauptung überzeugend und stehen im Einklang mit den Erkenntnissen von Gummer und Roßmann (2015), die bei motivierteren Befragten eine längere Dauer feststellten.

Die Analyse schließt Abbrüche und Befragte aus, die Papierfragebögen ausgefüllt haben. Erstere, weil ihre Dauer für diese Frage irrelevant ist, Letztere, weil die erforderlichen Daten für diese Gruppe nicht erhoben werden konnten. Da Tourangeau et al. (2013) die Cox-Regression zur Modellierung der Interviewdauer verwendet haben, folgt die Studie hier ihrer Empfehlung. Um mögliche Ungenauigkeiten aufgrund von Multikollinearität auszuschließen, wurden Varianzinflationsfaktoren berechnet, die für alle Kovariaten zufriedenstellend sind.

Ob eine Antwort wahrheitsgemäß ist oder nicht, lässt sich nur schwer beurteilen. Ein sogenanntes Satisficing-Antwortverhalten kann dazu führen, dass Befragte ihre Antworten an der ersten Option ausrichten, die ihnen ausreichend passend erscheint – was den Aufwand zur Beendigung der Befragung deutlich verringert (Gummer et al. 2021). Wenn sie dann ihre nachfolgenden Antworten in einer Batterie an der ersten Wahl ausrichten, sprechen wir von Straightlining. Das Fehlen von Straightlining erhöht die Reliabilität der Daten, da es auf ein echtes Engagement der Befragten und eine wahrheitsgemäße Berichterstattung hindeutet (Leiner 2019; Gummer et al. 2021). UWE verwendet mehrere Item-Batterien oder Rasterfragen, die sich für eine Untersuchung möglicher Straightlining-Muster eignen. Viele Rasterfragen enthalten nur drei oder vier Items, die wiederholt mit derselben Option wahrheitsgemäß beantwortet werden können. Daher verwendet die in dieser Arbeit vorgestellte Analyse nur Item-Batterien, die mindestens fünf Items enthalten. Der Fragebogen enthält sechs Raster, die wahrscheinlich nicht wahrheitsgemäß beantwortet werden können, indem wiederholt dieselbe Option ausgewählt wird. Alle Items werden auf einer Fünf-Punkte-Skala beantwortet. Ein logistisches Regressionsmodell schätzt für jeden Befragungsmodus die Wahrscheinlichkeit, dass mindestens ein Fall von Straightlining festgestellt werden kann. Die logistische Regression geht von einer linearen Beziehung zwischen den unabhängigen Variablen und der logistischen Odds-Ratio der abhängigen Variable aus. Sie nimmt auch die Abwesenheit von Multikollinearität an, was bedeutet, dass die unabhängigen Variablen nicht stark korreliert sind. Multikollinearität liegt im Modell nicht vor, wie die zufriedenstellenden Werte des Variance Inflation Factor (VIF) innerhalb des Regressionsmodells anzeigen, die das Fehlen einer signifikanten Multikollinearität zwischen den unabhängigen Variablen bedeuten.

Unabhängige Variablen

Alle statistischen Modelle berücksichtigen dieselben Merkmale der Befragten und der Befragungssituation. Zu den Merkmalen der Befragten zählen Alter, Geschlecht, Migrationshintergrund, Deutschkenntnisse, Klassenstufe und Schultyp. Beim Schultyp werden drei Arten unterschieden: Gymnasium, Sekundarschule und Gesamtschule sowie Realschule und Hauptschule. Der Hauptunterschied besteht darin, dass der Abschluss der ersten Stufe die Absolventen für eine Hochschulausbildung qualifiziert, in der zweiten Stufe ist dies optional und in der dritten Stufe wird diese Option nicht angeboten. Tabelle 1 enthält deskriptive Statistiken zu den in den Analysen verwendeten Variablen.

Tabelle 1: Deskriptive Statistik

Variable	Mittelwert	SD	Min	Max	N
<i>Merkmale der Befragten</i>					
Alter	13,84	1,24	12	17	923

Geschlecht: weiblich	0,54	0,50	0	1	923
Sprachprobleme vorhanden	0,02	0,15	0	1	923
Migrationshintergrund vorhanden	0,41	0,49	0	1	923
Gesamt-/Sekundarschule	0,38	0,49	0	1	923
Gymnasium	0,49	0,50	0	1	923
Haupt-/Realschule	0,26	0,44	0	1	923
<i>Merkmale der Befragung</i>					
Online-Umfrage	0,22	0,41	0	1	923
Nur Lehrkraft anwesend	0,49	0,50	0	1	923
Lehrkraft & externe Aufsicht anwesend	0,38	0,49	0	1	923
<i>Datenqualitätsindikatoren</i>					
Abbruch der Befragung	0,04	0,19	0	1	923
Anzahl Item-Nonresponse	8,12	24,06	0	186	923
Befragungsdauer (Minuten)	21,03	10,76	2	56	923
Vorkommen von Straightlining	0,28	0,66	0	4	923

Die Befragungssituation ist gekennzeichnet durch die Aufsicht (Lehrkraft anwesend oder nicht, Aufsichtsperson anwesend oder nicht), den Ort der Befragung (Schule oder zu Hause) und die Form, in der der Fragebogen beantwortet wurde (Papier oder online). Es gab 84 Befragungsgruppen, und die Modelle kontrollieren potenzielle Heteroskedastizität, indem sie die Standardfehler in allen Modellen entsprechend clustern. Die Aufsicht kann drei verschiedene Formen annehmen. Die Zuweisung erfolgte entsprechend der von den Gesundheitsbehörden erlassenen Kontaktbeschränkungen. Während der Durchführung der Umfrage praktizierten einige Schulen den sogenannten „Wechselunterricht“ oder „alternierenden Distanzunterricht“. Die Klassen wurden in zwei Gruppen aufgeteilt, die eine Woche lang zu Hause oder in der Schule unterrichtet wurden und in der folgenden Woche die Plätze tauschten. Auf diese Weise konnten die Forschenden verschiedene Ansätze im selben Klassenraum testen. Die Zuordnung zu einer der beiden Gruppen erfolgte in der Regel alphabetisch nach dem Nachnamen. Im Feld wurden die folgenden drei Varianten verwendet:

(A) Nur Lehrkraft: Ein Teil der Klassenräume wurde ohne externe Aufsicht untersucht. Die Lehrkräfte erhielten Papierfragebögen oder kurze Links zu einem Websurvey. Die Links waren auch als QR-Codes verfügbar. Sie hatten die Möglichkeit, ein vom Zentrum für Interdisziplinäre Regionalforschung (ZEFIR) bereitgestelltes Einführungsvideo zu zeigen oder das Verfahren selbst zu erklären. Für diese Option standen ihnen ein Einführungstext und ein Handbuch zur Verfügung, die vom Institut vorbereitet worden waren, um sicherzustellen, dass alle Befragten die gleichen Informationen erhielten. Die Befragten konnten entweder ihre persönlichen Geräte oder schulische Geräte verwenden, je nachdem, was die Lehrkraft unter Berücksichtigung von Ausstattung, Handyverboten und Netzabdeckung für die beste Lösung in ihrem Klassenzimmer hielt.

(B) Lehrkraft und externe Aufsicht: In einer weiteren Gruppe von Klassenräumen waren Lehrkräfte anwesend, die mehr oder weniger ständig über eine Videokonferenzsoftware mit einer externen Aufsicht im Auftrag des ZEFIR verbunden waren. Sie stellten das Befragungsprojekt vor und boten Hilfe bei Verständnisfragen an. Sie beobachteten auch die Situation, konnten jedoch aufgrund ihrer rein virtuellen Anwesenheit nicht immer wirksam eingreifen. Dieser Modus wurde mit Online-Befragungen kombiniert, die auf Schulcomputern oder den persönlichen Geräten der Befragten durchgeführt wurden.

(C) Nur externe Aufsicht: Einige Klassen standen für Interviews in der Schule nicht zur Verfügung. Wir nutzten den von den Schulen eingerichteten Zugang zum Distanzunterricht (Zoom oder Microsoft Teams), an den die Befragten zu diesem Zeitpunkt bereits gewöhnt waren. Sie waren alle online in einer

Videokonferenz und füllten zeitgleich einen Websurvey auf ihren eigenen Geräten zu Hause aus. In diesem Modus waren keine Lehrkräfte anwesend.

Ergebnisse

Die folgende Ergebnisdarstellung beschränkt sich auf die Effekte des Befragungsmodus, weiterführende Analysen und die vollständigen Regressionsmodelle sind in Stefes (2025) dargestellt und erläutert,

Item-Nonresponse

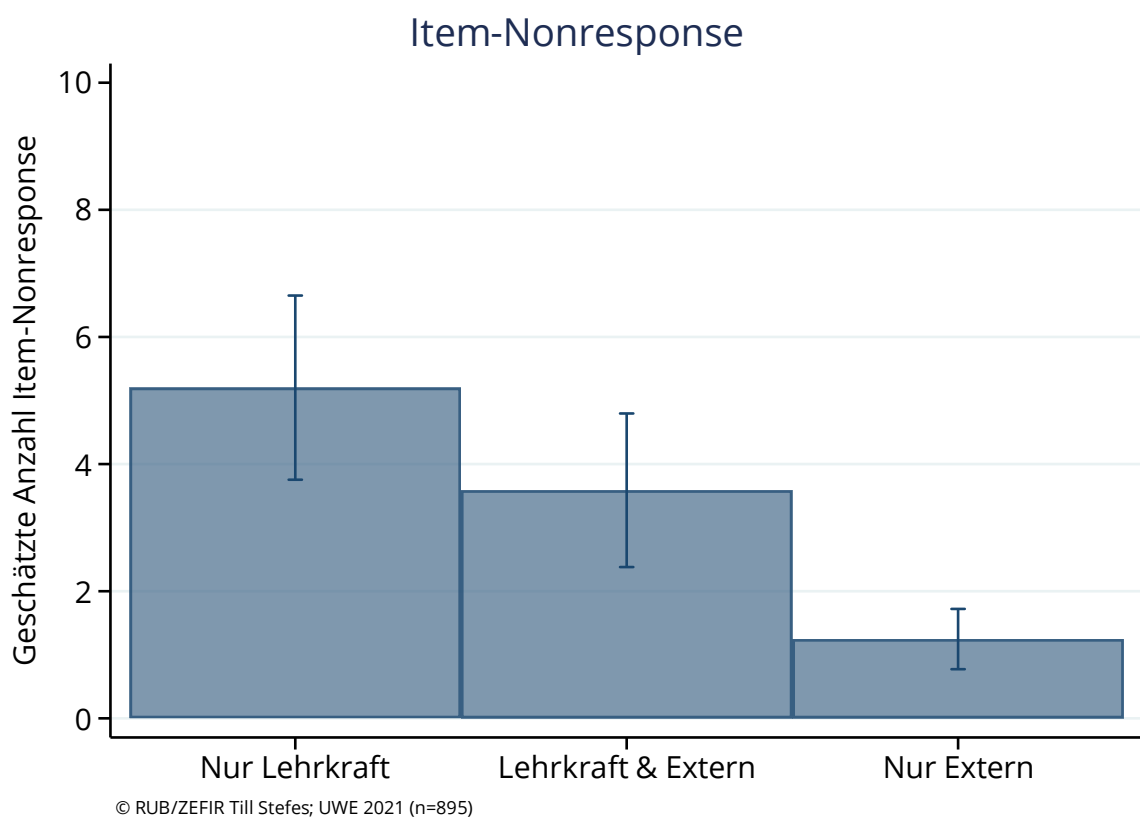


Abbildung 1: Geschätzte Item-Nonresponse nach Befragungsmodus (negative binomiale Regression)

Abbildung 1 zeigt die geschätzte Item-Nonresponse unter Berücksichtigung der drei Befragungsmodi auf der Basis eines negativen binomialen Regressionsmodells. Der realistische Maximalwert der Item-Nonresponse wäre 150. Dies ist die höchstmögliche Anzahl unbeantworteter Fragen für Beobachtungen, die nicht als Abbruch betrachtet werden. Während es keinen signifikanten Unterschied zwischen Modus (B) und einem der anderen gibt, unterscheiden sich die Modi, bei denen nur ein Erwachsener anwesend ist, hinsichtlich der Item-Nonresponse signifikant. Befragte, die nur von Lehrkräften beaufsichtigt wurden, hatten eine durchschnittliche Item-Nonresponse von fünf, während sie bei Befragten, die nur von externen Aufsichtspersonen beaufsichtigt wurden, weniger als zwei betrug. Dies unterstützt die Hypothesen 1b und 2a, wonach Lehrkräfte die Datenqualität beeinträchtigen, während externe Aufsicht einen klaren Vorteil darstellen.

Befragungsdauer

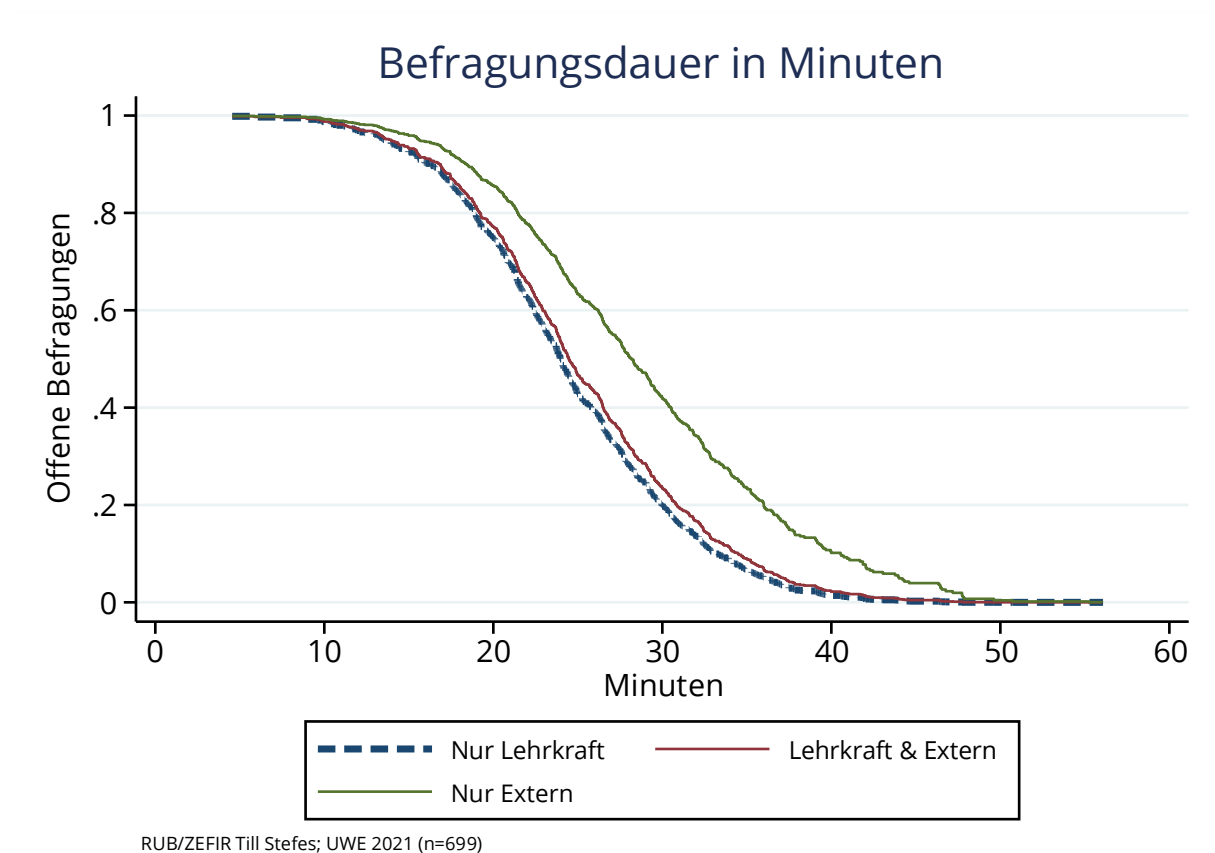


Abbildung 2: Vorhergesagte Interviewdauer in Minuten nach Art der Aufsicht (Cox-Regression)

Die in Abbildung 2 dargestellten „Survival-Kurven“ zeigen den geschätzten Anteil der Online-Befragungen an, die zu den auf der x-Achse angegebenen Zeitpunkten noch offen sind, basierend auf einer Cox-Regression. Die gestrichelte Linie steht für Befragungen, die nur von Lehrkräften beaufsichtigt wurden, die dünne rote Linie steht für die Befragungen mit Lehrkräften und zusätzlicher Unterstützung durch externe Aufsichtspersonen. Beide Verläufe sind kaum zu unterscheiden. Sie unterscheiden sich von der dritten, dicken grünen Linie. Befragte, die die Umfrage zu Hause unter virtueller Aufsicht durch externes Personal absolvierten, benötigten im Durchschnitt mehr Zeit für die Beantwortung der Fragen. Nach 30 Minuten hatten etwa 80 % der erstgenannten Gruppe die Umfrage abgeschlossen, während es in der zweiten Gruppe weniger als 60 % waren.

Straightlining

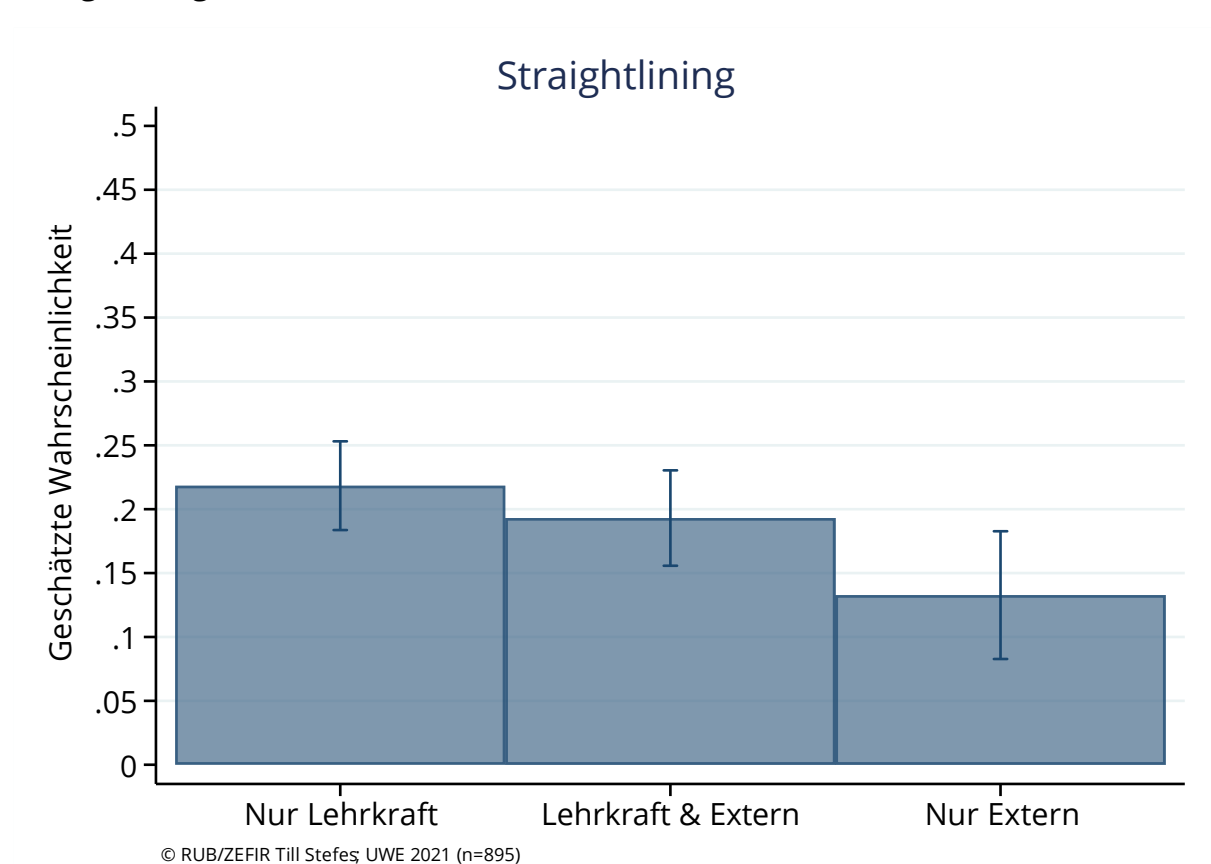


Abbildung 3: Geschätzte Wahrscheinlichkeit für mindestens ein Auftreten von Straightlining nach Befragungsmodus (logistische Regression)

Abbildung 3 zeigt die geschätzte Wahrscheinlichkeit für das Auftreten von Satisficing-Verhalten bei der Beantwortung von Fragebögen, das durch das Vorhandensein einer geraden Linie in mindestens einer von sechs Item-Batterien erkennbar ist. Im Durchschnitt wiesen die in Schulen ausgefüllten Fragebögen häufiger gerade Linien in den Rastern auf. Ein signifikanter Unterschied findet sich wie zuvor bei der Item-Nonresponse nur zwischen den von Lehrkräften und nur von Externen beaufsichtigten Befragten. Die Wahrscheinlichkeiten sind für die erstgenannte Gruppe mit über 20 % relativ hoch.

Diskussion

Ein erfolgreicher Rekrutierungsprozess kann als Voraussetzung für eine hohe Datenqualität angesehen werden, und die Literatur zeigt, dass Schulen dabei sehr effiziente Mittel sein können (Heath et al. 2009; Alibali und Nathan 2010; Bartlett et al. 2017; March et al. 2022; Ellonen et al. 2023). Diese Studie stellt dies nicht in Frage, rät jedoch zu einer sorgfältigen Prüfung der potenziellen Auswirkungen einer Beteiligung von Schulpersonal bei der Datenerhebung. Hatch et al. (2023) schlagen vor, so eng wie möglich mit Schulleitung und Lehrkräften zusammenzuarbeiten. Die Schlussfolgerung basiert in erster Linie auf Überlegungen zur Rücklaufquote. Die These ist berechtigt, da der Verlust einer einzigen Schule in diesem Prozess die Repräsentativität der gesamten Stichprobe gefährden kann (Newransky et al. 2020; Ellonen et al. 2023). Die Kernaussage dieses Beitrages lautet, dass die Beteiligung von Schulpersonal in

standardisierten Befragungen in Klassenräumen auf ein Minimum beschränkt werden sollte, von der Rekrutierung von Befragten abgesehen.

Bisherige Erkenntnisse aus der Literatur sprechen zwar dafür, dass jegliche Form von Aufsicht die Abschlussquoten und die Validität von Befragungsdaten bei der Befragung von Jugendlichen erhöhen kann (March et al. 2020; Bidonde et al. 2023), doch ist diese Aufsicht möglicherweise nicht in jedem Fall unproblematisch. Zwei wichtige Faktoren sprechen gegen die Beaufsichtigung von standardisierten Befragungen durch Lehrkräfte: Sie können unzuverlässig sein (Demkowicz et al. 2020; Rasberry et al. 2020; March et al. 2022) und sie können unbeabsichtigt die soziale Erwünschtheit oder damit verbundene Antwortverzerrungen erhöhen (Strange et al. 2003; Duncan und Magnuson 2013; Atkeson et al. 2014; Cops et al. 2016; Möhring und Schlütz 2019). Die Ergebnisse der Analysen in diesem Beitrag lassen den Schluss zu, dass die Datenqualität geringer ist, wenn allein Lehrkräfte für die Datenerhebung verantwortlich sind, was sich in den Unterschieden bei Item-Nonresponse und Satisficing-Mustern zeigt.

Externe Aufsichtspersonen können einer möglichen Beeinträchtigung der Datenqualität entgegenwirken. Es konnte gezeigt werden, dass externe Aufsichtspersonen die Motivation erhöhen (Demkowicz et al. 2020; March et al. 2022) und zu einer neutralen Atmosphäre beitragen (Demkowicz et al. 2020), wodurch der Erhebungsprozess potenziell verbessert wird. Während Lehrkräfte in Schulen in der Regel ohnehin anwesend sind, ist (zusätzliche) externe Aufsicht kostenintensiv (Walser und Killias 2012; Bartlett et al. 2017; March et al. 2022). Daher muss die Entscheidung, sie einzusetzen, gut begründet werden. Der Befund, dass eine ergänzende, virtuelle Aufsicht im Vergleich zu Modi mit ausschließlich Lehrkräften keinen Unterschied in der Datenqualität bewirkt, deutet darauf hin, dass diese Form der Intervention nicht ausreichend ist. In dieser Studie wurden zwei mögliche Wege getestet, externe Aufsicht sicherzustellen. Zusätzliche Aufsichtspersonen waren virtuell in einer Videokonferenz anwesend, während die Lehrkräfte im Klassenraum präsent waren. Dieser Modus, als „Modus B“ bezeichnet, zeigte nur minimale Effekte auf die Datenqualität. Der Aufwand zur Einrichtung dieses Modus erwies sich im Vergleich zu seinem Einfluss auf die Datenqualität als unverhältnismäßig hoch. „Modus C“ erzielte deutlich bessere Ergebnisse und involvierte Lehrkräfte und Klassenräume zunächst gar nicht. Die Unterschiede in Item-Nonresponse und Antwortmustern sind stichhaltige Argumente dafür, dass sich zukünftige Forschungen bei der Durchführung standardisierter Befragungen in Klassenzimmern nicht allein auf Lehrkräfte stützen sollten.

Literatur

- Alibali, M. W., und M. J. Nathan. 2010. Conducting Research in Schools: A Practical Guide. *Journal of Cognition and Development* 11(4):397–407.
- Atkeson, L. R., A. N. Adams, und M. Alvarez. 2014. Nonresponse and Mode Effects in Self- and Interviewer-Administered Surveys. *Political Analysis* 22(3):304–320.
- Bartlett, R., T. Wright, T. Olarinde, T. Holmes, E. R. Beamon und D. Wallace. 2017. Schools as Sites for Recruiting Participants and Implementing Research. *Journal of Community Health Nursing* 34(2):80–88.
- Bidonde, J., J. F. Meneses-Echavez, E. Hafstad, G. S. Brunborg und L. Bang. 2023. Methods, Strategies, and Incentives to Increase Response to Mental Health Surveys Among Adolescents: A Systematic Review. *BMC Medical Research Methodology* 23(1):270.
- Brown, B., und J. Larson. 2009. Peer Relationships in Adolescence. In *Handbook of Adolescent Psychology*, Hrsg. R. Lerner und L. Steinberg, 74–103. Hoboken, NJ: Wiley.
- Cops, D., A. De Boeck und S. Pleysier. 2016. School vs. Mail Surveys: Disentangling Selection and Measurement Effects in Self-reported Juvenile Delinquency. *European Journal of Criminology* 13:92–110.

- Demkowicz, O., E. Ashworth, R. Mansfield, E. Stapley, H. Miles, D. Hayes, D. Burrell, A. Moore und J. Deighton. 2020. Children and Young People's Experiences of Completing Mental Health and Wellbeing Measures for Research: Learning from Two School-Based Pilot Projects. *Child and Adolescent Psychiatry and Mental Health* 14:1–18.
- Duncan, G. J., und K. Magnuson. 2012. Socioeconomic Status and Cognitive Functioning: Moving from Correlation to Causation. *Wiley Interdisciplinary Reviews: Cognitive Science* 3(3):377–386.
- Ellonen, N., T. Pösö, L. Mielityinen und E. Paavilainen. 2023. Using Self-Report Surveys in Schools to Study Violence in Alternative Care: A Methodological Approach. *Child Abuse Review* e2814.
- Epstein, D. 1998. "Are you a girl or are you a teacher?" The 'Least Adult' Role in Research about Gender and Sexuality in a Primary School. In *Doing Research About Education*, Hrsg. G. Walford. London: Falmer Press.
- Felderer, B., A. Kirchner und F. Kreuter. 2019. The Effect of Survey Mode on Data Quality: Disentangling Nonresponse and Measurement Error Bias. *Journal of Official Statistics* 35(1):93–115. doi:10.2478/JOS-2019-0005
- Gummer, T., und J. Roßmann. 2015. Explaining Interview Duration in Web Surveys: A Multilevel Approach. *Social Science Computer Review* 33(2):217–234.
- Kivivuori, J., V. Salmi und S. Walser. 2013. Supervision Mode Effects in Computerized Delinquency Surveys at School: Finnish Replication of a Swiss Experiment. *Journal of Experimental Criminology* 9:91–107.
- Leiner, D. J. 2019. Too Fast, too Straight, too Weird: Non-Reactive Indicators for Meaningless Data in Internet Surveys. *Survey Research Methods* 13(3):229–248.
- Little, R. J. 1988. A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 83(404):1198–1202.
- Little, R. J. 1992. Regression with Missing X's: A Review. *Journal of the American Statistical Association* 87(420):1227–1237.
- Lucia, S., L. Herrmann, und M. Killias. 2007. How important are interview methods and questionnaire designs in research on self-reported juvenile delinquency? An experimental comparison of Internet vs paper-and-pencil questionnaires and different definitions of the reference period. *Journal of Experimental Criminology* 3:39–64.
- March, A., E. Ashworth, C. Mason, J. Santos, R. Mansfield, E. Stapley, J. Deighton, N. Humphrey, N. Tait und D. Hayes. 2022. 'Shall We Send a Panda?' A Practical Guide to Engaging Schools in Research: Learning from Large-Scale Mental Health Intervention Trials. *International Journal of Environmental Research and Public Health* 19(6):3367.
- Möhring, W., und D. Schlütz. 2019. Das Interview als soziale Situation. In *Die Befragung in der Medien- und Kommunikationswissenschaft*, Hrsg. W. Möhring und D. Schlütz. 41–67. Wiesbaden: Springer VS.
- Mühlböck, M., N. Steiber und B. Kittel. 2017. Less Supervision, More Satisficing? Comparing Completely Self-Administered Web-Surveys and Interviews Under Controlled Conditions. *Statistics, Politics, and Policy* 8:13–28.
- Rasberry, C. N., I. Rose, E. Kroupa, A. Hebert, A. Geller, E. Morris und C. A. Lesesne. 2018. Overcoming Challenges in School-wide Survey Administration. *Health Promotion Practice* 19(1):110–118.
- Revilla, M., und C. Ochoa. 2015. What are the Links in a Web Survey among Response Time, Quality, and Auto-evaluation of the Efforts done? *Social Science Computer Review* 33(1):97–114.
- Stefes, T. 2023. *Umwelt, Wohlbefinden und Entwicklung von Kindern und Jugendlichen (UWE) Befragung 2021*. GESIS, Köln. Datenfile Version 1.0.0. <https://doi.org/10.7802/2613>
- Stefes, T. 2024. Data Quality of Different Modes of Supervision in Classroom Surveys. *Education Sciences*, 14(3), 299.
- Stefes, T., A. Lemke, V. Gaffron, K. Knüttel, J. Schuchardt, und S. Petermann. 2023. Konzeption und Durchführung der UWE-Befragung 2021. *ZEFIR Materialien* Band 22. Bochum: Zentrum für interdisziplinäre Regionalforschung (ZEFIR).

- Strange, V., S. Forest, A. Oakley und Ripple Study Team. 2003. Using Research Questionnaires with Young People in Schools: The Influence of the Social Context. *International Journal of Social Research Methodology* 6(4):337–346.
- Tourangeau, R., M. P. Couper und F. G. Conrad. 2013. *The Science of Web Surveys*. Oxford: Oxford University Press.
- Walser, S., und M. Killias. 2012. Who Should Supervise Students During Self-report Interviews? A Controlled Experiment on Response Behavior in Online Questionnaires. *Journal of Experimental Criminology* 8(1):17–28.