

3

Working Paper
2022

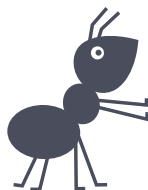
KonsortSWD



Konsortium für die
Sozial-, Verhaltens-, Bildungs- und
Wirtschaftswissenschaften

Nutzung von Logdaten in der empirischen Bildungsforschung - Eine Bedarfsanalyse

Benjamin Becker
Claudia Neuendorf
Malte Jansen



September 2022

www.konsortswd.de

Nutzung von Logdaten in der empirischen Bildungsforschung - Eine Bedarfsanalyse

Benjamin Becker,^{1*} Claudia Neuendorf,² Malte Jansen^{1,3}

September 2022

DOI: 10.5281/zenodo.7030996

¹ Institut zur Qualitätsentwicklung im Bildungswesen an der Humboldt Universität zu Berlin

² Universität Tübingen

³ Zentrum für internationale Bildungsvergleichsstudien

Autor:innen-Hinweis

Benjamin Becker <https://orcid.org/0000-0003-3074-0918>

Claudia Neuendorf <https://orcid.org/0000-0002-3024-0000>

Malte Jansen <https://orcid.org/0000-0001-7081-6505>

Wir danken unseren studentischen Mitarbeitenden Zulma Estrada und Katerina Blizkovska für die Rechercharbeiten auf Grundlage der Abstract- und Titel-Bände.

Abstract

In großen internationalen und nationalen Vergleichsstudien haben computer-basierte Leistungstests papier-basierte Leistungstests weitgehend abgelöst. Aus Forschungsperspektive ist ein zentraler Vorteil dieser Art von Leistungstests die Möglichkeit der Sammlung von sogenannten Log- oder Prozessdaten. Die Struktur dieser Daten ist jedoch häufig komplex und Möglichkeiten der Bereitstellung dieser Daten für Sekundärnutzende sind vielfältig, aber auch herausfordernd. So können Logdaten beispielsweise abstrahiert als Logdaten-Indikatoren (z. B. Bearbeitungszeiten) mit vergleichsweise geringem Aufwand in bestehende Datenstrukturen eingefügt werden oder möglichst verlustfrei als rohe Logdaten in separaten Datenstrukturen bereitgestellt werden. Die Bereitstellung roher Logdaten ist jedoch häufig mit substanziellem Aufwand verbunden. Darüber hinaus gibt es noch wenig gesicherte Erkenntnisse darüber, welche Art von Logdaten in der Forschungspraxis überhaupt genutzt werden.

Der vorliegende Beitrag möchte mit einer ersten Bedarfsanalyse helfen, diese Forschungslücke zu schließen. Er quantifiziert die aktuelle Nachfrage nach Logdaten im Feld der empirischen Bildungsforschung und berücksichtigt dabei insbesondere die Art der Logdaten. Hierzu wurden die Abstract-Bände verschiedener deutsch- und englischsprachiger Konferenzen im Zeitraum 2018 bis 2020 analysiert. Die Ergebnisse zeigen, dass es in der internationalen Forschung innerhalb dieses Zeitraums einen enormen Anstieg an Beiträgen auf Grundlage von Logdaten gab. Dabei wurden vor allem rohe Logdaten sowie Bearbeitungszeiten als spezifische Form von Logdaten-Indikatoren genutzt. Aus diesen Ergebnissen werden Empfehlungen für Forschungsdatenzentren und Datengeber:innen abgeleitet.

Keywords: Logdaten, Prozessdaten

1. Theoretischer Hintergrund

In den letzten Jahren zeichnet sich innerhalb der empirischen Bildungsforschung, insbesondere bezogen auf internationale Large-Scale Assessments, ein eindeutiger Trend ab: Während früher fast ausschließlich papier-basierte Leistungstests eingesetzt wurden (*Paper-Based Assessments*, PBA), ist dieses Format in den letzten Jahren größtenteils durch sogenannte Technologie-beziehungsweise Computer-basierte Assessments (*Technology/Computer-Based Assessments*, TBA/CBA) abgelöst worden. Zu den Vorreitern dieser Entwicklung gehören dabei unter anderem die Large-Scale Assessments PIAAC (OECD, 2013), PISA (OECD, 2016), TIMSS (Martin et al., 2020) und PIRLS (Martin et al., 2017). Durch den Einsatz dieser Assessment-Form ist es möglich, in Kompetenztests die sich wandelnden Anforderungen in einer immer stärker digitalisierten Umwelt abzubilden, indem beispielsweise innovative und/oder im digitalen Alltag eingebettete Item-Formate verwendet werden. Diese Studien machen sich aber auch weitere zentrale Vorteile von TBA gegenüber PBA zunutze: unter anderem das Sparen von Druckkosten, Möglichkeiten zur stärkeren Standardisierung des Assessmentverlaufs sowie Möglichkeiten zum adaptiven bzw. multi-stage Testen (Parshall et al., 2002). Einer der wenigen Hinderungsgründe, die diese Entwicklung momentan noch verlangsamen, ist lediglich die unzureichende technische Ausstattung zum Beispiel innerhalb der getesteten Schulen. Anfängliche Hinderungsgründe, wie fehlende Testsysteme beziehungsweise der hohe Aufwand, der mit der Entwicklung solcher Systeme verbunden ist, sowie Fragen zur Vergleichbarkeit von papier- und computer-basiertem Testen (z. B. Kröhne & Martens, 2011) scheinen hingegen in der aktuellen Large-Scale Assessment-Praxis nur noch eine untergeordnete Rolle zu spielen.

Da Datensätze aus Large-Scale Assessments vor allem an *Forschungsdatenzentren* (FDZs) bereitgestellt werden und beliebte Quellen für Sekundäranalysen sind (Hopfenbeck et al., 2018; Lenkeit et al., 2015), stellt sich daher auch die Frage an die Forschungsdateninfrastruktur, ob durch die Umstellung der Assessments auf TBA neue Anforderungen an die Daten-Archivierung entstehen. Im Folgenden sollen daher die neu entstandenen Bedarfe der Scientific Community analysiert und erste Lösungsansätze zum Umgang mit diesen Anforderungen diskutiert werden.

1.1. Logdaten

Neben den eingangs genannten inhaltlichen und organisatorischen Gründen, die im Idealfall dazu führen, dass TBA eine höhere Validität und eine höhere Reliabilität ermöglicht als PBA, bietet TBA darüber hinaus aus Forschungsperspektive noch einen weiteren, großen Vorteil: Die Möglichkeit des Sammelns von *Logdaten*. Logdaten, auch Prozessdaten genannt, sind Daten, die vom Testsystem während der Testbearbeitung gesammelt werden und das konkrete Bearbeitungsverhalten der Testteilnehmenden während der Testung beschreiben (Kröhne & Goldhammer, 2018). Mithilfe dieser Art von Daten können empirische Aussagen über das Testbearbeitungsverhalten getroffen werden, beispielsweise: Wie viel Zeit nehmen sich Testteilnehmende für eine bestimmte Aufgabe? In welcher Reihenfolge werden die Aufgaben beantwortet?

Diese Art von Informationen wird tatsächlich bereits teilweise für die Auswertung und Berichterlegung von Large-Scale Assessments herangezogen. In PIAAC werden beispielsweise fehlende Werte in Abhängigkeit von Antwortzeiten kodiert (OECD, 2013). Aktuelle Forschung beschäftigt sich darüber hinaus mit Fragen wie: Können Antwortzeiten fehlende Werte am Ende von Tests erklären (Pohl et al., 2019)? Können bestimmte, häufig auftretende Lösungsstrategien bei komplexen Aufgaben identifiziert werden (Ulitzsch et al., 2021)? Wie hängen mehrmaliges Aufrufen von Aufgaben mit Bearbeitungsgeschwindigkeit und Bearbeitungsgenauigkeit zusammen (Bezirhan et al., 2021)? Und wie häufig treten vorzeitige Testabbrüche auf und wie können diese interpretiert werden (Ulitzsch et al., 2020)?

Dabei ist festzuhalten, dass Logdaten nicht nur im Bereich von Large-Scale Assessments von Bedeutung sind. Auch in anderen Bereichen, wie zum Beispiel in der Software- oder Webentwicklung, werden Logdaten (dort häufig auch *Paradaten* genannt) schon seit längerem genutzt, um Rückschlüsse auf das Verhalten von Personen in Interaktion mit technischen Systemen aufzuzeichnen und zu analysieren (z. B., Cooley et al., 1997). Auch in der Survey-Forschung werden Logdaten verwendet, um das Bearbeitungs- und Antwortverhalten von Teilnehmenden zu erfassen (Stieger & Reips, 2010). Im Folgenden fokussieren wir uns jedoch spezifisch auf Logdaten im Kontext der empirischen Bildungsforschung.

1.2. Logdaten-Speicherung

Neben dem großen Potenzial bietet die Arbeit mit Logdaten jedoch auch einige Herausforderungen: Sowohl für Fragebogendaten als auch Kompetenzdaten haben sich in der Psychologie und empirischen Bildungsforschung Standards für die Datenspeicherung und Dokumentation entwickelt. Häufig werden diese Daten in rechteckigen Datenstrukturen gespeichert, wobei eine Zeile einer getesteten Person und eine Spalte einem eingesetzten Item oder einer gestellten Frage entspricht. Die Dokumentation dieser Daten erfolgt häufig über Metadaten, insbesondere Variablen- und Wertelabels, über die sich konkrete Fragen-Formulierungen und -Inhalte im Original abbilden lassen. Auch Abstraktionen der Roh-Daten, wie zum Beispiel Skalen-Mittelwerte, Summen-Scores oder Kompetenzschätzungen (z. B. Plausible Values, Braun & von Davier, 2017) können in solchen Datenstrukturen sinnvoll ergänzt werden.

Bei Logdaten ist die Speicherung und Dokumentation der Daten jedoch häufig deutlich komplexer. Dies ist darin begründet, dass unterschiedliche Testsysteme, mit denen computerbasierte Assessments durchgeführt werden, (a) unterschiedliche Logdaten sammeln und (b) Logdaten sehr unterschiedlich strukturieren und bezeichnen. Dabei gibt es bisher kaum verbreitete Standards, welche Logdaten gesammelt werden sollten und auch ein eindeutiges Vokabular für die Bezeichnungen von Logdaten fehlt (für einen der wenigen allgemeineren Ansätze siehe z. B., Kröhne & Goldhammer, 2018).

Eine weitere zentrale Herausforderung ist, dass die Bedeutung insbesondere von rohen Logdaten nur schwer über herkömmliche Metadaten beschrieben werden kann. Welche Bedeutung ein Mausklick verortet in einem Koordinatensystem inhaltlich hat, erschließt sich

nur bei Betrachtung des dargestellten Screens. Eine vollständige Beschreibung des Screens (Welche bedienbaren Buttons werden angezeigt? Welche grafischen Elemente sind ansonsten sichtbar?) in Form eines Variablenlabels ist quasi unmöglich. Die Veröffentlichung der zugrundeliegenden Items, die eine Interpretation der Logdaten ermöglichen würden, ist hingegen aus Gründen der Testsicherheit häufig keine Option. Da Items in Large-Scale Assessments oftmals in mehreren Testdurchgängen verwendet werden, würden solche Veröffentlichungen die Validität späterer Testdurchgänge gefährden. Doch selbst wenn solche Veröffentlichungen möglich sind, sind entsprechende Dokumentationsformen komplex und bisher selten erprobt.

Schlussendlich besteht eine weitere, zentrale Herausforderung darin, dass ein Log-Event in der Regel nicht in Form einer einzelnen Variable dargestellt werden kann (Kröhne & Goldhammer, 2018). Ein *Log-Event* (eine einzelne, aufgezeichnete Interaktion eines Testteilnehmenden mit dem Testsystem) kann beispielsweise ein Mausklick oder ein Tastaturanschlag sein. Ein Mausklick zeichnet sich dadurch aus, in welchem Item oder in welcher Fragebogeneinheit er stattfindet, wann er zeitlich stattfindet und mit welchen Screen-Koordinaten. Ein Tastaturanschlag verfügt über einige überlappende Attribute (Zeitpunkt, Item/Fragebogeneinheit), andere sind jedoch nicht anwendbar (Koordinaten) und einige besondere Eigenschaften kommen hinzu (z. B.: welche Taste wurde gedrückt?).

Unter anderem auch diese Herausforderungen führen dazu, dass es im Grunde zwei verschiedene Arten gibt, Logdaten zu speichern: (1) das Speichern der rohen Logdaten-Events und (2) das Speichern von Logdaten-Indikatoren (Goldhammer et al., 2021; Kröhne & Goldhammer, 2018). Im Folgenden sollen die Grundideen beider Vorgehensweisen, sowie ihre Vor- und Nachteile kurz erläutert werden.

1.2.1. Rohe Log-Events

Die Speicherung roher Logevents bedeutet, dass rohe Logdaten ohne zusätzlichen Abstraktionsprozess gespeichert werden. Falls das Testsystem sehr umfangreiche Logdaten speichert, kann mithilfe dieser Daten theoretisch sogar der vollständige Testverlauf einer Person reproduziert werden. Eine ausführliche Taxonomie wird beispielsweise in Kröhne und Goldhammer (2018) vorgestellt. Ein Beispiel für eine solche Datenspeicherung und Nutzbarmachung für Sekundär-Analysen sind die Logdaten des PIAAC 2012 Assessments. Diese wurden in einem sehr aufwendigen Prozess aufbereitet und im *PIAAC LogDataAnalyzer* Sekundärdaten-Nutzenden zur Verfügung gestellt (OECD, 2013). Eine Übersicht zu Forschungsarbeiten, die bisher auf der Basis des öffentlich verfügbaren Datensatzes erfolgt, findet sich in Goldhammer et al. (2020).

Der Vorteil einer solchen Datenbereitstellung ist, dass den Forschenden große Freiheitsgrade bei der Analyse der Prozessdaten bleiben. Forschende können selbst entscheiden, welche Logdaten-Indikatoren geeignet sind, um ihre Fragestellungen zu beantworten. Zusätzlich ermöglichen diese Arten von Daten tiefe Einblicke in die Ausgestaltung der betroffenen Assessments, was wiederum tieferegehende Qualitätskontrollen ermöglicht. Nachteile einer

solchen Datenbereitstellung sind, dass eine professionelle Aufbereitung und Dokumentation der Daten häufig sehr zeitaufwendig und umfangreich ist. Das resultierende Datenvolumen kann oftmals groß sein und die spezifische Datenstruktur verlangt von Datennutzenden Expertise im Verständnis und Umgang mit Logdaten. Sogar für technisch und inhaltlich versierte Nutzende sind häufig große zeitliche Ressourcen für die Einarbeitung in spezifische Datenstrukturen notwendig. Auch ist beispielsweise eine gründliche Anonymisierung von Prozessdaten von offenen Antwortfeldern in der Praxis quasi unmöglich. Die Herausforderungen im Umgang mit rohen Log-Events können auch daran erkannt werden, dass obwohl die PIAAC 2012 Logdaten frei verfügbar und ausführlich dokumentiert sind, fast alle bisher auf Basis der Daten erfolgten Publikationen (Co-)Autor:innen aufweisen, die direkt an der Durchführung der PIAAC-Studie beteiligt waren (Goldhammer et al., 2020).

1.2.2. Logdaten-Indikatoren

Logdaten-Indikatoren bezeichnen im Gegenzug die Abstraktion von rohen Log-Events, häufig auf Item-Ebene (Goldhammer & Zehner, 2017). Beispiele hierfür wären Antwortzeiten, Anzahl von Seitenbesuchen, oder Anzahl von Antwortwechseln. Konkrete Umsetzungen finden sich beispielsweise in den Logdaten-Indikatoren im Rahmen der PISA-Assessments, die in den öffentlich verfügbaren *Scientific Use Files* (SUF) zu finden sind (OECD, 2016).

Vorteile dieses Datentyps sind, dass sie sich unkompliziert als Variablen in vorhandene Datenstrukturen einfügen lassen. Ihre Dokumentation kann häufig erfolgreich über vorhandene Metadaten-Strukturen (z.B. Variablenlabels) erfolgen und ihre Nutzung ist für die meisten Forschenden intuitiver. Ein bedeutsamer Nachteil dieser Art von Daten sind jedoch die deutlich eingeschränkten Freiheitsgrade von Forschenden. Bei der Bereitstellung von Indikatoren müssen die Daten-Bereitstellenden im Grunde vorwegnehmen, welche Arten von Fragestellungen mithilfe der von ihnen bereitgestellten Daten bearbeitet werden können. Für solche Entscheidungen mangelt es jedoch aktuell noch einer fundierten Datengrundlage, welche Art von Logdaten von der Forschungsgemeinde am stärksten beforscht werden.

2. Fragestellung

Die Erhebung und Bereitstellung von Logdaten ist für viele Forschende in der empirischen Bildungsforschung, aber auch für Forschungsdatenzentren, noch Neuland. Es gibt bisher wenige Erkenntnisse, welche Arten von Fragestellungen von Forschenden mithilfe von Logdaten bearbeitet werden und welche Art der Bereitstellung von Logdaten daher geboten ist. Der vorliegende Artikel möchte genau diese Forschungsfragen beantworten: 1) Wie viel Forschungsinteresse erfahren Logdaten momentan allgemein in der Psychometrie und der empirischen Bildungsforschung und welche Priorität sollte ihre Bereitstellung beispielsweise in Forschungsdatenzentren erfahren? 2) Welche Arten von Logdaten werden momentan vor allem genutzt? 3) Welche Empfehlungen lassen sich daraus für Datengeber:innen und Forschungsdatenzentren ableiten?

Um diese Fragestellungen zu beantworten, wurden die Inhalte vorgestellter Poster sowie gehaltener Vorträge und Symposien auf aktuellen Forschungskonferenzen der Bildungsforschung systematisch analysiert. Da insbesondere auf Konferenzen aktuelle Forschungsthemen vorgestellt werden, erwarten wir, dass diese im Vergleich zu Publikationen in wissenschaftlichen Fachzeitschriften eine größere Sensitivität für Wechsel in Forschungsschwerpunkten der Forschungsgemeinde bieten.

3. Methode

Um einen exemplarischen Einblick in die aktuellen Forschungsfelder der Bereiche Psychometrie und empirischer Bildungsforschung, sowohl auf nationaler als auch internationaler Ebene zu erhalten, wurden folgende Konferenzen im Zeitraum 2018-2020 analysiert: (1) die *Tagung der Gesellschaft für Empirische Bildungsforschung* (GEBF) in den Jahren 2018 und 2019, (2) das *Annual Meeting des National Council on Measurement in Education* (NCME) in den Jahren 2018, 2019, und das vorläufige Programm 2020¹, (3) die *Konferenz der European Association for Research on Learning and Instruction* (EARLI) im Jahr 2019.

Analysiert wurden sowohl die Titel der Vorträge und Symposien als auch deren Abstracts². Die Stichwortsuche wurde von zwei verschiedenen studentischen Hilfskräften durchgeführt und alle Eintragungen durch den Erstautor des Manuskripts überprüft und eingeordnet. Eine Liste der verwendeten Stichwörter findet sich im Anhang. Wurde ein Beitrag in einer Session als Beitrag basierend auf Logdaten identifiziert, wurden sämtliche weiteren Beiträge innerhalb dieser Session händisch inspiziert.

3.1. Einschlusskriterien

In einem ersten Schritt wurden Beiträge identifiziert, deren Forschungsfragestellungen anhand von Logdaten bearbeitet wurden. Dabei wurde nicht nach Kontext (zum Beispiel technologiebasiertes Assessment oder experimentelle Studien) unterschieden. Zudem war es unerheblich, ob Beiträge ausschließlich auf Logdaten basierten oder noch weitere Datentypen verwendet wurden. Explizit eingeschlossen wurden Beiträge basierend auf Eye-Tracking-Daten. Beiträge, die exklusiv auf physiologischen oder neurophysiologischen Maßen (EEG, MRT, CT, oder ähnliches) beruhten, wurden hingegen nicht berücksichtigt.

3.2. Kategorisierung

Die identifizierten Vorträge und Symposien wurden hinsichtlich der Art der genutzten Daten (Indikatoren vs. rohe Logdaten) unterteilt. Dabei wurde nicht unterschieden, ob Forschungsarbeiten berichteten, dass Indikatoren für die Forschungsarbeit gebildet worden waren oder ob diese vorab schon vorlagen. Vielmehr wurde dahingehend kategorisiert, ob sich die beschriebenen Logdaten mit Logdaten-Indikatoren in bereits veröffentlichten Datensätzen

¹ Die NCME 2020 wurde, inklusive vollständig feststehendem Programm, in Präsenz geplant, jedoch in dieser Form nie durchgeführt.

² Für die GEBF Konferenz 2019 waren online lediglich Titel und keine Abstracts verfügbar, für die NCME 2018 waren zwar für Einzelbeiträge Abstracts verfügbar, jedoch nicht für Vorträge im Rahmen von Symposien.

deckten. Als Basis zur Kategorisierung von Logdaten-Indikatoren wurden die Logdaten-Indikatoren, die in den Scientific-Use-Files von PISA enthalten sind, herangezogen (OECD, 2016). Diese umfassen: Antwortzeiten (wieviel Zeit wurde für die Beantwortung eines Items verwendet oder wie viel Zeit wurde auf einer Seite verbracht), Anzahl Aktionen auf einer Seite und Anzahl der Besuche einer spezifischen Seite. Wurden andere Arten von Logdaten für die Beantwortung von Fragestellungen verwendet, muss davon ausgegangen werden, dass Forschende die entsprechenden Logdaten-Indikatoren selbst bilden mussten.

Wenn sich aus Titel und Abstract nicht ableiten ließ, welche Art von Logdaten verwendet wurden, aber explizit die Nutzung von Logdaten erwähnt wurde, wurde keine Kategorisierung des Beitrags vorgenommen, der Beitrag aber trotzdem als Beitrag basierend auf Logdaten gewertet. Zudem waren Mehrfachkategorisierungen von Beiträgen möglich, zum Beispiel wenn ein Beitrag Antwortzeiten und rohe Logdaten verwendete.

4. Ergebnisse

Die vollständigen Ergebnisse der Recherchen und Stichwortsuchen finden sich im Online-Supplement als Excel-Tabelle (<https://osf.io/tu679/>). Zudem findet sich dort die R-Syntax zur Analyse und Ergebnisdarstellung der Rechercheergebnisse.

KONFERENZ	LOGDATEN GESAMT	ROHE LOGDATEN	ANTWORTZEITEN	EYE-TRACKING
GEBF 2018	15	6	6	4
GEBF 2019	8	5	1	2
NCME 2018	35	18	16	0
NCME 2019	70	15	48	0
NCME 2020	115	45	63	0
EARLI 2019	96	40	16	48

Tabelle 1: Anzahl Beiträge zu Logdaten auf Konferenzen gesamt und aufgeschlüsselt nach Art der Logdaten.

Anmerkung: Aufgrund von mehrfachen Kategorisierungen und fehlenden Kategorisierungen ergibt die Summe der einzelnen Logdaten-Kategorien nicht zwangsläufig die Gesamtzahl der Beiträge basierend auf Logdaten.

Eine Übersicht der Ergebnisse bezüglich der Nutzung von Logdaten insgesamt sowie der Nutzung von Logdaten-Indikatoren und rohen Logdaten ist in Tabelle 1 abgebildet. Dargestellt sind jeweils die absolute Anzahl an Beiträgen basierend auf Logdaten (*Logdaten gesamt*), die Anzahl der Beiträge basierend auf rohen Logdaten (*rohe Logdaten*) sowie die Anzahl der Beiträge basierend auf Antwortzeiten (*Antwortzeiten*) und auf Eye-Tracking-Daten (*Eye-Tracking*). Die Ergebnisse verdeutlichen, dass die Anzahl der Vorträge, die sich mit dem Thema

Logdaten beschäftigen, im internationalen Kontext stark zugenommen hat. Im Rahmen der NCME kam es von 2018 zu 2019, sowie von 2019 zu 2020 jeweils fast zu Verdopplungen der absoluten Anzahl der Beiträge, die auf Logdaten basieren. Im Rahmen der GEBF ergibt sich ein gegensätzliches Bild, wobei es sowohl 2018 als auch 2019 absolut sehr wenige Beiträge basierend auf Logdaten gab.

Bezüglich der kategorisierten Logdaten-Indikatoren sind lediglich Antwortzeiten ausgewiesen, da die anderen beschriebenen Indikator-Typen (Anzahl Aktionen, Anzahl Seitenbesuche) kein einziges Mal genannt wurden. Weiterhin scheinen sich Antwortzeiten insbesondere im Rahmen der NCME, also im Rahmen methodischer beziehungsweise psychometrischer Forschung, großer Beliebtheit zu erfreuen und dort auch häufiger genutzt zu werden als rohe Logdaten. Auf Konferenzen mit stärkerem inhaltlichem Fokus ergibt sich hingegen ein leicht gegenteiliges Bild. Sowohl im Rahmen der GEBF als auch der EARLI werden rohe Logdaten stärker genutzt als Antwortzeiten.

Da die Nutzung von Eye-Tracking-Daten insbesondere im Rahmen der EARLI vermehrt auftrat, wurde diese Art von Daten ebenfalls explizit in Tabelle 1 ausgewiesen. Dabei ist darauf hinzuweisen, dass Eye-Tracking-Daten keine klassische Form von Logdaten sind, da solche Daten nicht im Rahmen eines Testsystems beiläufig anfallen, sondern explizit mit einem spezifischen Eye-Tracking-Gerät erhoben werden. Nichtsdestotrotz gibt es einige Ähnlichkeiten in der Art und Struktur der Daten, insbesondere im Vergleich mit rohen Logdaten. Auch hier fallen viele Events (Blickbewegungen beziehungsweise Mausbewegungen, Fixationen beziehungsweise Klicks) mit entsprechenden Zeitstempeln an (siehe z. B., Yaneva et al., 2021). Dabei ist jedoch hervorzuheben, dass Eye-Tracking-Daten im Rahmen der NCME gar nicht genutzt wurden.

KONFERENZ	NAVIGATIONSSEQUENZEN	TASTATUR-LOGS
GEBF 2018	2	0
GEBF 2019	2	0
NCME 2018	3	5
NCME 2019	0	6
NCME 2020	6	10
EARLI 2019	8	26

Tabelle 2: Exploratorische Analysen zu Logdaten-Themen auf Konferenzen: Betrachtung von Navigationssequenzen und Tastatur-Anschlägen.

Tabelle 2 erweitert die beschriebenen Befunde um exploratorische Analysen. Hier wird explizit die Gruppe der rohen Logdaten näher in den Blick genommen, wobei *Navigationssequenzen*

und Tastatur-Logs (*Key-Logs*) vermehrt genannt wurden. Key-Logs sind Logdaten, die bei der Tastatur-Benutzung aufgezeichnet werden. Dabei ist eine mögliche Darstellungsform, jeden einzelnen Tastaturanschlag inklusive Zeitstempel, ausgewähltem Textfeld und gedrückter Taste aufzuzeichnen. Optional kann sogar die Stärke des Tastaturanschlags aufgezeichnet werden (Rodrigues et al., 2013). Auf Basis dieser Logdaten können Fragen beantwortet werden wie: Wie erschöpft ist eine tippende Person (Ulinskas et al., 2017)? Wie gestresst ist eine tippende Person (Freihaut & Göritz, 2021; Liu et al., 2018)? Auch auf Basis roher Key-Logs können Logdaten-Indikatoren extrahiert werden, dabei scheinen diese jedoch noch wenig standardisiert beziehungsweise sehr spezifisch für einzelne Fragestellungen (z. B., Liu et al., 2018).

Mit Navigationssequenzen ist hingegen das Navigationsverhalten einer Person innerhalb eines Testsystems gemeint. Navigationssequenzen können dabei sowohl Aufgaben-übergreifendes Navigationsverhalten (Wird der Test linear bearbeitet? Werden Aufgaben mehrmals besucht?) als auch Navigationsverhalten innerhalb einzelner, komplexer Aufgaben (Welche Hyperlinks werden in welcher Reihenfolge verwendet? Welche Seiten werden besucht? Siehe z. B., Hahnel et al., 2016) bezeichnen. Bei der Abstraktion von rohen Logdaten zu Indikatoren für Navigationsverhalten ist zu beachten, dass solche Indikatoren deutlich spezifischer für bestimmte Items und Item-Formate wären als beispielsweise Antwortzeiten. Uns sind keine öffentlich verfügbaren Datensätze bekannt, die Navigationssequenzen oder Key-Logs als Logdaten-Indikatoren zur Verfügung stellen, beziehungsweise entsprechende Konventionen zur Bereitstellung vorschlagen.

5. Diskussion

In den letzten Jahren sind in der Psychologie und in der empirischen Bildungsforschung Computer-basierte beziehungsweise Technologie-basierte Assessments stark auf dem Vormarsch. Da diese Erhebungsform das Sammeln von Logdaten ermöglicht, bieten sich hierdurch große Potenziale, sowohl methodisch als auch inhaltlich neuartige Fragestellungen zu beantworten. Allerdings sind mit der Sammlung und Analyse von Logdaten nicht nur große Chancen, sondern auch große Herausforderungen verbunden. So gibt es bisher kaum gemeinsame Standards bezüglich der Sammlung, Speicherung und Bereitstellung von Logdaten.

Das vorliegende Manuskript beschäftigt sich daher mit der Frage, welche Arten von Logdaten in welcher Häufigkeit in der aktuellen Forschung genutzt werden. So sollen Forschungsdatenzentren und Forschende in die Lage versetzt werden, informierte Entscheidungen zu treffen, welche Art von Daten für Sekundäranalysen zur Verfügung stehen sollten. Unsere Analyse der Konferenzen GEBF, NCME und EARLI im Zeitraum 2018-2020 zeigt, dass sich insgesamt die Beforschung von Logdaten einer stark steigenden Beliebtheit erfreut. Insbesondere werden dabei rohe Logdaten sowie Antwortzeiten als Datengrundlage genutzt. Die methodische und internationale Forschung scheint dabei in der Nutzung von Logdaten der inhaltlichen und nationalen Forschung voraus zu sein. Insgesamt erscheint es jedoch plausibel

zu erwarten, dass auch auf nationaler und inhaltlicher Ebene das Interesse an Forschung basierend auf Logdaten in nächster Zeit zunehmen wird, wie beispielsweise auch ein Call for Papers der Zeitschrift für Psychologie mit dem Titel „Innovations in Exploring Sequential Process Data“ (<https://doi.org/10.1027/2151-2604/a000487>) verdeutlicht.

Besonders erwähnenswert ist neben diesen Befunden die Vielzahl der Kontexte, aus denen Logdaten stammen. Diese umfassen Bereiche wie technologie-basierte Large-Scale Assessments, experimentelle Studien der Lehr-Lern-Forschung oder Forschung zu *kollaborativem Problemlösen*, in welcher Interaktionen von Personen mit anderen Personen via technologischer Systeme erfasst werden. Darüber hinaus treten jedoch auch sehr innovative Bereiche in den Vordergrund, wie *Learning Analytics*, in dem Verhalten auf Lernplattformen in Form von Logdaten erfasst wird oder *Game-Based Assessments*, in denen Verhalten beispielsweise innerhalb einer komplexen Computer-Simulation erfasst wird.

5.1. Limitationen

Einige zentrale Limitationen des beschriebenen Vorgehens sollen im Folgenden thematisiert werden. Es ist offensichtlich, dass die von uns durchgeführte Recherche nur einen kleinen Ausschnitt an Konferenzen untersucht. Weitere Konferenzen mit ähnlichem Fokus wären beispielsweise das *International Meeting of the Psychometric Society* (IMPS), die *Konferenz der European Association of Methodology* (EAM), die *Tagung der Fachgruppe Methoden und Evaluation der Deutschen Gesellschaft für Psychologie* (FGME) sowie die *Tagung der Fachgruppe Pädagogischen Psychologie der Deutschen Gesellschaft für Psychologie* (PaePsy). Nichtsdestotrotz sind wir der Meinung, dass bereits die ausgewählten Konferenzen einen klaren Trend in der Nutzung von Logdaten erkennen lassen.

Bezüglich der Klassifizierungen der Studien in die vorgestellten Kategorien ist festzuhalten, dass solche Kategorisierungen - lediglich auf Basis von Abstracts und Titeln - ungenau sein können. Da viele der Konferenzbeiträge noch nicht in veröffentlichten Publikationen gemündet haben, lässt sich teilweise die Art der genutzten Daten nicht mit vollständiger Sicherheit bestimmen. Wir haben jedoch versucht, ein möglichst konservatives Vorgehen zu wählen, um Beiträge nur bei großer Sicherheit als Beiträge basierend auf Logdaten zu klassifizieren und bei fehlenden weiteren Informationen keine Kategorisierung vorzunehmen.

5.2. Empfehlungen

Auf Basis unserer Recherchen und Analysen empfehlen wir Datengeber:innen und Forschungsdatenzentren in der empirischen Bildungsforschung folgendes Vorgehen bezüglich der Bereitstellung von Logdaten:

(1) Nachfrage: Generell erscheint die Bereitstellung von Logdaten lohnenswert. Logdaten werden in immer stärkerem Maße beforscht und es ist anzunehmen, dass die Nachfrage nach solchen Daten in nächster Zeit noch weiter steigen wird, da beispielsweise neue Modelle aus der methodischen Forschung in der inhaltlichen Forschung angewendet werden.

(2) Datengenerierung: Von Primärforschenden erfordert insbesondere die Bereitstellung roher Logdaten, aber auch die Bereitstellung von Logdaten-Indikatoren ein frühes Mitdenken dieser Anforderungen bei der Entwicklung von technologie-basierten Testsystemen oder Lernplattformen. Die Logdaten-Erhebung, Strukturierung und Benennung muss bei der Programmierung des Testsystems mitgedacht werden, idealerweise so, dass spätere Aufbereitungsschritte effizient automatisiert werden können. Des Weiteren erfordert die Bereitstellung bestimmter Logdaten teils spezifische Layout-Entscheidungen. Ein Beispiel hierfür wären Antwortzeiten und Item-weise Darstellung von Aufgaben und Fragenbogen-Batterien. Wenn mehrere Items auf einer Seite dargestellt werden, können Seiten-Verweildauern beispielsweise nicht mehr als Antwortzeiten interpretiert werden (Kröhne & Goldhammer, 2018). Hier können unter Umständen auch Entscheidungen über die Responsivität einer Testumgebung gegenüber unterschiedlichen Endgeräten betroffen sein.

(3) Bereitstellung von Indikatoren: Die Bereitstellung von ausgewählten Logdaten-Indikatoren erscheint äußerst wünschenswert. Aktuell nutzt ein Großteil der Logdatenforschung Antwortzeiten. Eine konsistente Bereitstellung dieser Daten scheint aus technischer Perspektive unaufwendig und bezüglich der Nachfrage lohnenswert.

(4) Bereitstellung roher Logdaten: Darüber hinaus scheint aber auch, gerade bei größeren Datensätzen, die Bereitstellung der rohen Logdaten wünschenswert. Durch eine solche Datenbereitstellung werden Sekundärdaten-Nutzende in die Lage versetzt, eigene Logdaten-Indikatoren, passend zur eigenen Fragestellung zu bilden. Momentan gibt es noch sehr wenige Datensätze, die diese Möglichkeit bieten und die Mehrzahl der in den Vorträgen verwendeten Daten ist nicht öffentlich zugänglich. Darin spiegelt sich der Fakt, dass in vielen Forschungsdatenzentren noch wenig Expertise in der Verfügbarmachung von Logdaten besteht. Gerade Forschungsdatenzentren, die den Anspruch haben, Daten FAIR, (d. h., auffindbar, zugänglich, interoperabel und wiederverwendbar) bereitzustellen und ihre Langzeitarchivierbarkeit zu sichern, haben besonderes Interesse an Leitlinien für die Archivierung und Verfügbarmachung von Logdaten.

(5) Auffindbarkeit: Die Auffindbarkeit von Logdaten erfordert die Vergabe standardisierter Metadaten. In bisherigen kontrollierten Vokabularen für Daten kommen Logdaten meist nicht als eigene Kategorie vor. Erste Anstöße für ein differenzierteres, kontrolliertes Vokabular, welches sich spezifisch auf Logdaten bezieht, liefert beispielsweise die Arbeit von Kröhne und Goldhammer (2018). Hier müssten von Seiten der Forschungsdatenzentren jedoch noch konzeptuelle Arbeiten an entsprechenden Metadatenschemata erfolgen.

(6) Zugänglichkeit für Sekundärnutzende: Bei der Zugänglichkeit ist insbesondere zu beachten, dass, wie zuvor bereits angesprochen, manche Typen von Logdaten nur mit erheblichem Aufwand anonymisierbar sind (dazu zählen insbesondere Tastaturanschläge). Für diese Daten müsste aus Datenschutzgründen eine stärker geschützte und kontrollierte Zugangsstufe gewählt werden. Andere Logdatentypen können hingegen als absolut anonym betrachtet werden und möglicherweise zum freien Download zur Verfügung gestellt werden.

(7) Interoperabilität: Die Interoperabilität kann sowohl Daten als auch Metadaten betreffen. Metadaten müssen, um interoperabel zu sein, in weithin verwendeten Metadatenschemata (z. B. VerbundFDB-Metadatenschema, da|ra-Metadatenschema) als kontrolliertes Vokabular angelegt sein. Damit die Daten selbst interoperabel werden, müssen sie in andere Datenstrukturen integrierbar sein. Das bedeutet, dass zum einen eine Verknüpfung von Logdaten und anderen Daten, zum Beispiel über ID-Variablen, möglich bleiben und zum anderen Logdaten mit vorhandener Software und Auswertungsroutinen verarbeitbar sein sollten. Ansätze für die Verarbeitung von Logdaten finden sich beispielsweise in den R-Paketen logFSM (Kroehne, 2021), LOGAN (Reis Costa & Leoncio, 2019) und ProcData (Tang et al., 2021).

(8) Dokumentation: Die Wiederverwendbarkeit von Logdaten ist schließlich eng mit der Güte der Dokumentation verbunden. Diese stellt, wie weiter oben beschrieben, eine der größten Herausforderungen bei der Bereitstellung von Logdaten dar. Bisher kommen selbst Logdaten-Indikatoren teilweise gar nicht in den Skalenhandbüchern von Studien vor, sondern sind lediglich im Datensatz selbst dokumentiert. Selbst diese Indikatoren sind jedoch nicht immer vollständig selbsterklärend. So gibt es beispielsweise verschiedene Möglichkeiten, eine ItemBearbeitungszeit zu berechnen. In dem Fall, dass die Testteilnehmenden eine Seite nur einmal sehen können, wird meist der Zeitpunkt, an dem die nächste Seite aufgerufen wird, als Zeitstempel zur Berechnung der Bearbeitungszeit genutzt. Bei Tests, die eine Navigation zwischen Items erlauben, könnte beispielsweise die kumulative Verweildauer auf einer Seite genutzt werden. Oder aber die kumulative Verweildauer bis zur letzten auf der Seite vorgenommenen Aktion. Diese detaillierten Informationen zur Berechnung eines Indikators sollten Nutzenden an die Hand gegeben werden. Bei der Bereitstellung roher Logdaten ist die Herausforderung noch größer und eine schriftliche Dokumentation reicht mitunter nicht aus. Hier wird eher auf komplexe Systeme zurückgegriffen, in denen Log-Sequenzen beispielsweise wiederabgespielt werden können, um das Zustandekommen der Log-Einträge zu verdeutlichen. Dies erfordert jedoch entweder die Veröffentlichung der Testinhalte oder die Schaffung von *Mock-Items*. Mock-Items stellen schematisch die Funktionsweise von Items dar, ohne spezifische Inhalte zu beinhalten.

(9) Langzeitarchivierung: Wurden Wege gefunden, Logdaten auf diese Weise FAIR zur Verfügung zu stellen, bleibt schließlich noch die Frage nach zusätzlichen Schritten, um eine Langzeitverfügbarkeit zu gewährleisten. Dazu gehört in der Regel, Daten und Dokumentation in haltbare Formate umzuwandeln. Daten können in der Regel in irgendeiner Form in Text- oder csv-Dateien umgewandelt werden. Die Dokumentation kann jedoch insbesondere bei sehr komplexen Datenstrukturen, die nur durch einen Re-play in der originalen Testumgebung dokumentierbar sind, herausfordernd sein. Möglicherweise könnte dies die Kooperation unterschiedlicher Archive notwendig machen (z.B. mit dem Testarchiv des FDZ Bildung am DIPF, Archive für Quellcode, wie z. B. GitHub).

5.3. Schlussfolgerung

Insgesamt erfreuen sich Logdaten in der Psychologie und empirischen Bildungsforschung einer wachsenden Beliebtheit. In Zukunft sollte die Bereitstellung solcher Daten für Sekundäranalysen dieser Nachfrage Rechnung tragen. Aktuell erscheint insbesondere die Bereitstellung von Antwortzeiten technisch umsetzbar und lohnenswert, während bezüglich der Bereitstellung von rohen Logdaten wohl noch erheblicher Forschungsbedarf bezüglich optimaler Datenstrukturen und Dokumentation der Daten besteht. Bevor diese Bereitstellung durch Forschungsdatenzentren in der Breite sinnvoll ist, sollte zunächst innerhalb der Forschungscommunity Konsens hergestellt werden, wie solche Daten erhoben, aufbereitet und gespeichert werden können und sollten. Es wäre denkbar, hier Workshops mit Forschenden zu organisieren, um diesen Prozess voranzutreiben.

Literaturverzeichnis

Bezirhan, U., von Davier, M. & Grabovsky, I. (2021). Modeling Item Revisit Behavior: The Hierarchical Speed–Accuracy–Revisits Model. *Educational and Psychological Measurement*, 81(2), 363–387. <https://doi.org/10.1177/0013164420950556>

Braun, H. & von Davier, M. (2017). The use of test scores from large-scale assessment surveys: Psychometric and statistical considerations. *Large-scale Assessments in Education*, 5(1), 1–16. <https://doi.org/10.1186/s40536-017-0050-x>

Cooley, R., Mobasher, B., & Srivastava, J. (1997). Web mining: Information and pattern discovery on the World Wide Web. *Proceedings Ninth IEEE International Conference on Tools with Artificial Intelligence*, 558–567. <https://doi.org/10.1109/TAI.1997.632303>

Freihaut, P., & Göritz, A. S. (2021). Does Peoples' Keyboard Typing Reflect Their Stress Level? *Zeitschrift für Psychologie*, 229 (4), 245–250. <https://doi.org/10.1027/2151-2604/a000468>

Goldhammer, F., Hahnel, C., & Kröhne, U. (2020). Analysing log file data from PIAAC. In D. B. Maehler & B. Rammstedt (Hrsg.), *Large-Scale Cognitive Assessment* (S. 239–269). Springer, Cham. https://doi.org/10.1007/978-3-030-47515-4_10

Goldhammer, F., Hahnel, C., Kröhne, U. & Zehner, F. (2021). From byproduct to design factor: on validating the interpretation of process indicators based on log data. *Large-scale Assessments in Education*, 9 (1), 1–25. <https://doi.org/10.1186/s40536-021-00113-5>

Goldhammer, F. & Zehner, F. (2017). What to make of and how to interpret process data. *Measurement: Interdisciplinary Research and Perspectives*, 15(3-4), 128–132. <https://doi.org/10.1080/15366367.2017.1411651>

Hahnel, C., Goldhammer, F., Naumann, J. & Kröhne, U. (2016). Effects of linear reading, basic computer skills, evaluating online information, and navigation on reading digital text. *Computers in Human Behavior*, 55, 486–500. <https://doi.org/10.1016/j.chb.2015.09.042>

Hopfenbeck, T. N., Lenkeit, J., El Masri, Y., Cantrell, K., Ryan, J. & Baird, J.-A. (2018). Lessons learned from PISA: A systematic review of peer-reviewed articles on the programme for international student assessment. *Scandinavian Journal of Educational Research*, 62 (3), 333–353. <https://doi.org/10.1080/00313831.2016.1258726>

Kroehne, U. (2021). *LogFSM: Analyzing Log Data from Educational Assessments using Finite-State Machines (LogFSM)* [R package version 0.4.5.15].

Kröhne, U. & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, 45(2), 527–563. <https://doi.org/10.1007/s41237-018-0063-y>

- Kröhne, U. & Martens, T. (2011). Computer-based Competence Tests in the National Educational Panel Study: The Challenge of Mode Effects. *Zeitschrift für Erziehungswissenschaften*, 14, 169–186. <https://doi.org/10.1007/s11618-011-0185-4>
- Lenkeit, J., Chan, J., Hopfenbeck, T. N. & Baird, J.-A. (2015). A review of the representation of PIRLS related research in scientific journals. *Educational Research Review*, 16, 102–115. <https://doi.org/10.1016/j.edurev.2015.10.002>
- Liu, H., Fernando, O. N. N. & Rajapakse, J. C. (2018). Predicting affective states of programming using keyboard data and mouse behaviors. *15th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 1408–1413. <https://doi.org/10.1109/icarcv.2018.8581248>
- Martin, M. O., Mullis, I. V. & Hooper, M. (2017). *Methods and Procedures in PIRLS 2016*. TIMSS & PIRLS International Study Center.
- Martin, M. O., von Davier, M. & Mullis, I. V. (2020). *Methods and Procedures: TIMSS 2019 Technical Report*. TIMSS & PIRLS International Study Center.
- OECD. (2013). *Technical Report of the Survey of Adult Skills PIAAC (Second Edition)*. OECD Publishing.
- OECD. (2016). *PISA 2015 assessment and analytical framework*. OECD Publishing. <https://doi.org/10.1787/19963777>
- Parshall, C. G., Spray, J. A., Kalohn, J. & Davey, T. (2002). *Practical considerations in computer-based testing*. Springer Science & Business Media. <https://doi.org/10.1007/978-1-4613-0083-0>
- Pohl, S., Ulitzsch, E. & von Davier, M. (2019). Using Response Times to Model Not-Reached Items Due to Time Limits. *Psychometrika*, 84(3), 892–920. <https://doi.org/10.1007/s11336-019-09669-2>
- Reis Costa, D. & Leoncio, W. (2019). *LOGAN: Log File Analysis in International Large-Scale Assessments* [R package version 1.0.0]. <https://CRAN.R-project.org/package=LOGAN>
- Rodrigues, M., Goncalves, S., Carneiro, D., Novais, P. & Fdez-Riverola, F. (2013). Keystrokes and clicks: Measuring stress on e-learning students. In J. Casillas, F. J. Martinez-Lopez & F. Vicari Rosaand De la Prieta (Hrsg.), *Management Intelligent Systems* (S. 119–126). Springer. https://doi.org/10.1007/978-3-319-00569-0_15
- Stieger, S. & Reips, U.-D. (2010). What are participants doing while filling in an online questionnaire: A paradata collection tool and an empirical study. *Computers in Human Behavior*, 26(6), 1488–1495. <https://doi.org/10.1016/j.chb.2010.05.013>
- Tang, X., Zhang, S., Wang, Z., Liu, J. & Ying, Z. (2021). *ProcData: Process Data Analysis* [R package version 0.3.2]. <https://CRAN.R-project.org/package=ProcData>

Ulinskas, M., Woźniak, M. & Damaševičius, R. (2017). Analysis of keystroke dynamics for fatigue recognition. *International Conference on Computational Science and Its Applications*, 235–247. https://doi.org/10.1007/978-3-319-62404-4_18

Ulitzsch, E., He, Q., Ulitzsch, V., Molter, H., Nichterlein, A., Niedermeier, R. & Pohl, S. (2021). Combining clickstream analyses and graph-modeled data clustering for identifying common response processes. *Psychometrika*, 86(1), 190–214. <https://doi.org/10.1007/s11336-020-09743-0>

Ulitzsch, E., von Davier, M. & Pohl, S. (2020). A multiprocess item response model for not-reached items due to time limits and quitting. *Educational and Psychological Measurement*, 80(3), 522–547. <https://doi.org/10.1177/0013164419878241>

Yaneva, V., Clauser, B. E., Morales, A., & Paniagua, M. (2021). Using Eye-Tracking Data as Part of the Validity Argument for Multiple-Choice Questions: A Demonstration. *Journal of Educational Measurement*, 58(4), 515–537. <https://doi.org/10.1111/jedm.12304>

Anhang

Stichwörter

Log-Daten, Log-D, Logdaten, Prozessdaten, Prozessd, process da, action da, Reponse Times, Antwortzeit, Bearbeitungszeit, ungszeit, time spent, Keystrokes, strokes, Tastenan, writing pr, Navigation, Sequenzen, Verhaltensmuster, explorativ, Eye, Tracking, Augenbew, Fixationszeiten, Calculator, Taschenrechn, Action, visit, Aktion, Besuch

Anmerkungen. Die Stichwörter beinhalten deutsche sowie englische Begriff. Des Weiteren wurden bewusst teilweise Wortstämme zur Suche verwendet, um Nicht-Treffer aufgrund von Deklinationen zu vermeiden.

Impressum

Kontakt:

Institut zur Qualitätsentwicklung im Bildungswesen (IQB) – Wissenschaftliche Einrichtung der Länder an der Humboldt-Universität zu Berlin e.V.

Luisenstraße 56

10117 Berlin

Berlin, September 2022

KonsortSWD Working Paper:

KonsortSWD baut als Teil der Nationalen Forschungsdateninfrastruktur Angebote zur Unterstützung von Forschung mit Daten in den Sozial-, Verhaltens-, Bildungs- und Wirtschaftswissenschaften aus. Unsere Mission ist es, die Forschungsdateninfrastruktur zur Beforschung der Gesellschaft zu stärken, zu erweitern und zu vertiefen. Sie soll nutzungsorientiert ausgestaltet sein und die Bedürfnisse der Forschungscommunities berücksichtigen. Wichtiger Grundstein ist dabei das seit über zwei Jahrzehnten durch den Rat für Sozial- und Wirtschaftsdaten (RatSWD) aufgebaute Netzwerk von Forschungsdatenzentren.

In dieser Reihe erscheinen Beiträge rund um das Forschungsdatenmanagement, die im Kontext von KonsortSWD entstehen. Beiträge, die extern und doppelblind begutachtet wurden sind entsprechend gekennzeichnet.

KonsortSWD wird im Rahmen der NFDI durch die Deutsche Forschungsgemeinschaft (DFG) gefördert – Projektnummer: 442494171.



Diese Veröffentlichung ist unter der Creative-Commons-Lizenz (CC BY 4.0) lizenziert:

<https://creativecommons.org/licenses/by/4.0/>

DOI: 10.5281/zenodo.7030996

Zitationsvorschlag:

Becker, B., Neuendorf, C., & Jansen, M. (2022). *Nutzung von Logdaten in der empirischen Bildungsforschung - Eine Bedarfsanalyse*. KonsortSWD Working Paper 3/2022. Berlin. Institut zur Qualitätsentwicklung im Bildungswesen (IQB). <https://doi.org/10.5281/zenodo.7030996>