


Open Science als Beitrag zur Qualität in der Bildungsforschung

Georg Krammer · Erich Svecnik 

Eingegangen: 13. November 2020 / Überarbeitet: 14. Dezember 2020 / Angenommen: 20. Dezember 2020 / Online publiziert: 8. Januar 2021
© Springer Fachmedien Wiesbaden GmbH, ein Teil von Springer Nature 2021

Zusammenfassung Ausgangspunkt dieses Beitrags sind Diskussionen um die Belastbarkeit empirischer Befunde in benachbarten Disziplinen, namentlich der Sozialpsychologie, die in der sog. „Replication Crisis“ gipfelten. Von derartigen Diskussionen um Replikationen und „Questionable Research Practices“ ist die Bildungsforschung bisher noch nicht in dem Maße betroffen, aber die Problemlagen sind in Teilbereichen ähnlich. Es mag daher nur eine Frage der Zeit sein, bevor diese Kontroversen auch in der Bildungsforschung aufkommen. Vor diesem Hintergrund argumentieren wir, wie Open Science einen Beitrag leisten kann, um die Belastbarkeit von Befunden der Bildungsforschung zu erhöhen. Im Besonderen greifen wir drei Open Science Praktiken auf: Präregistrierung, Open Materials und Open Data. Wir stellen diese vor und beleuchten, wie sie in der Bildungsforschung implementiert werden können. Wir thematisieren dabei die spezifischen Verhältnisse der Bildungsforschung im Vergleich zu Nachbardisziplinen, und gehen auf Limitationen und Besonderheiten der Bildungsforschung ein. Wir schließen mit einem Plädoyer für Transparenz.

Schlüsselwörter Empirische Bildungsforschung · Open Science · Replikationskrise · Transparenz · Questionable Research Practices

G. Krammer
Pädagogische Hochschule Steiermark, Hasnerplatz 12, 8010 Graz, Österreich

E. Svecnik (✉)
IQS – Institut des Bundes für Qualitätssicherung im österreichischen Schulwesen,
Hans-Sachs-Gasse 3, 8010 Graz, Österreich
E-Mail: erich.svecnik@iqs.gv.at

The contribution of Open Science to the quality of educational research

Abstract The starting point of this paper are the discussions of the robustness of empirical findings in related disciplines, namely social psychology, which culminated in the so-called “replication crisis”. These discussions about replication and “questionable research practices” have only started to reach the educational sciences. At the same time, parts of the educational sciences are prone to the same problems as related disciplines. Therefore, it may only be a matter of time before these controversies also arise in the educational sciences. Against this backdrop, we argue how Open Science can contribute to increasing the robustness of educational sciences’ findings. In particular, we suggest three Open Science practices: Pre-registration, Open Materials and Open Data. We present these practices and examine how researchers can implement these Open Science practices in the educational sciences. We discuss the specific conditions of the educational sciences in comparison to related disciplines and address the limitations and particularities of the educational sciences. We conclude with a plea for transparency.

Keywords Empirical educational science · Open science · Replication crisis · Transparency · Questionable research practices

1 Einleitung

In der akademischen Forschung wird es belohnt, viel und rasch zu publizieren. Zusammen mit teilweise intransparenten Publikationsprozessen kann dies zu systematischen Fehlern führen, die die Qualität der gewonnenen Erkenntnisse und der publizierten Studien in Frage stellen (Frith 2020; Smaldino und McElreath 2016; Smaldino et al. 2019; Wicherts et al. 2016). In anderen Disziplinen wird bereits thematisiert, ob und wie Forschungsergebnisse verzerrt und nicht belastbar sind (s. van der Zee und Reich (2018) für eine Übersicht: animal welfare, biomedicine, climate research, criminology, energie efficiency, hardware development, high-energy physics, information science, mass spectrometry, neuroscience, robotics, sex research). Eindringlich zeigt dies beispielsweise die Open Science Collaboration (2015) für die Psychologie. Das Forscherteam replizierte 100 Studien, die in hochrangigen wissenschaftlichen Fachzeitschriften publiziert waren. Von diesen 100 Studien waren ursprünglich 97 signifikant, in deren Replikationen aber nur mehr 36. Einen Schritt weiter ging Ioannidis (2005), welcher provokant schlussfolgert, dass die meisten publizierten Forschungsergebnisse schlichtweg falsch sein müssen. Die Bildungsforschung ist derzeit noch kaum von derartigen Diskussionen betroffen, aber es scheint ratsam, sich über Möglichkeiten zur Erhöhung der Qualität Gedanken zu machen. In einzelnen Teildisziplinen scheinen Replikationsstudien jedenfalls bereits zunehmendes Interesse zu genießen (z. B. Coyne et al. 2016; Travers et al. 2016; Schmidt 2009). Vor diesem Hintergrund setzen wir uns im vorliegenden Beitrag zwei Ziele: 1) das Bewusstsein für diese Thematik in der Bildungsforschung zu steigern und 2) Open Science als eine Möglichkeit aufzuzeigen, um die Belastbarkeit von Befunden aus der Bildungsforschung langfristig zu erhöhen.

1.1 Papier ist geduldig

Ein Blick in die Vergangenheit zeigt, dass die Bildungsforschung die einleitend genannte Kontroverse bezüglich der Qualität der gewonnenen Erkenntnisse eigentlich schon kennt, sie aber leicht vergisst. Es sei als Beispiel der Pygmalion-Effekt in Schulen genannt. Rosenthal und Jacobson (1966) zeigten, dass Intelligenztestergebnisse von Schülerinnen/Schülern durch die Erwartungen ihrer Lehrkräfte beeinflusst werden können. Aber zeigten sie das wirklich?

In seinem kritischen Überblicksartikel zu der Geschichte des Pygmalion-Effekts zeigt Spitz (1999) auf, wie nicht belastbare Befunde ihren Weg in Publikationen, in Lehrbücher und unsere Allgemeinbildung finden, und gleichzeitig widersprüchliche Ergebnisse und methodische Kritik ignoriert werden. U. a. bringt er dafür Zitate von Thorndike – „*In spite of anything I can say, I am sure it [Pygmalion] will become a classic widely referred to and rarely examined critically*“ (S. 208) – und Cronbach – „*This book [on the Pygmalion effect] taken as a whole is a masterpiece of confusion, aligned with a hypothesis that most of the data contradict*“ (S. 224). Will man den Pygmalion-Effekt aber nicht nur mit Anekdoten kritisieren, dann sei auf Studien von Rosenthal und Kolleginnen/Kollegen verwiesen, die ihren Effekt nicht nur nicht fanden (Conn et al. 1968), sondern sogar gegenläufig fanden (Rosenthal 1969). Es sei auch auf Re-Analysen der ursprünglichen Daten verwiesen, die den Effekt nicht bestätigen konnten und auf vielfache Probleme der Messung und des Designs verweisen (Elashoff und Snow 1971). Auch ist kritisch anzumerken, dass die Lehrerinnen/Lehrer der Studien im Nachhinein oft nicht mehr wussten, wer die vermeintlich hochbegabten Kinder überhaupt waren (z. B. Rosenthal und Jacobson 1968).

Die Geschichte des Pygmalion-Effekts zeigt auf, dass Bildungsforscherinnen/ Bildungsforscher auch als Menschen nicht unfehlbar sind, und in ein akademisches System eingebettet sind, wo sie forschen. Die Studien von Rosenthal und Kolleginnen/Kollegen erlangten rasch mediale Präsenz und beeinflussten die Bildungspolitik. Sie suggerierten einfache Rezepte, um Herausforderungen im Bildungssystem schnell zu lösen. Ihre Studien brachten den Autoren viele Publikationen und große Bekanntheit. Fast ein halbes Jahrhundert später ist von der Kritik wenig geblieben. Es wurde belohnt, viel Aufsehenerregendes mit Neuigkeitswert zu publizieren. Doch wie publiziert man Aufsehenerregendes – am besten viel und schnell?

1.2 „Publish or perish“

Eine Möglichkeit viel zu publizieren ist es viel zu forschen und viel zu verschriftlichen. Dies reicht aber nicht aus, denn um viel zu publizieren, ist es hilfreich, viele (signifikante) Effekte zu finden (s. Publikationsbias: Dickersin 1990). Dieses Problem thematisierte Sterling bereits 1959: Herausgeberinnen/Herausgeber haben Widerstand Null-Effekte zu publizieren. Dreißig Jahre später zeigen Sterling et al. (1995) auf, dass sich in den verstrichenen Jahren diesbezüglich nichts verändert hat. Betrachtet man die Veröffentlichungen in wissenschaftlichen Fachzeitschriften (Fannelli 2010a) zeigt sich dies darin, dass überwiegend (z. B. Weltraumwissenschaft: 70,2 %) bis fast ausschließlich (z. B. Psychologie und Psychiatrie: 91,5 %) statistisch

signifikante Effekte publiziert werden. Vor allem für die sozialwissenschaftlichen Disziplinen sind dies erstaunliche Ergebnisse, da mit der gegebenen statistischen Power der publizierten Studien niemals so viele Effekte gefunden werden sollten (Szucs und Ioannidis 2017; vgl. dazu auch wahrscheinlichkeitstheoretische Überlegungen und Analysen für verschiedenen wissenschaftliche Disziplinen von Head et al. 2015). Der Druck zu publizieren scheint hier ausschlaggebend zu sein (Fanelli 2010b).

Dieser Druck zu publizieren spiegelt sich auch im Verhalten der Forscherinnen/Forscher wider. So zeigen Meta-Analysen, dass 1,97 % der befragten Forscher/innen angeben, bereits einmal Daten oder Ergebnisse fabriziert, gefälscht oder modifiziert zu haben (Fanelli 2009). Da solche Befunde aus Selbstberichten stammen, liegt nahe, dass sie konservative Unterschätzungen sind.

Gleichzeitig wirken auch wesentlich subtilere Mechanismen als das offensichtliche Fehlverhalten, Daten zu fälschen – bisweilen auch, ohne dass dies den Forschenden bewusst ist. Darunter fallen auch das „Abrunden“ von p -Werten (z. B. $p=0,051$ als $p<0,05$), das willkürliche Ausschließen von Datenpunkten (insbesondere nachdem man sich dessen Auswirkungen angesehen hat) oder das Fortsetzen/Abbrechen der Datenerhebungen, nachdem bereits ein Blick auf p -Werte geworfen wurde. Zu diesen subtileren Mechanismen – sogenannten „Questionable Research Practices“ (John et al. 2012) – zählt auch das Suchen nach statistischen Signifikanzen in einer Fülle von Hypothesentests und Analysen („ p -Hacking“). Damit geht auch einher, nicht signifikante abhängige oder unabhängige Variablen bzw. Kovariaten in Publikationen nicht mehr zu erwähnen oder das Aufstellen von Null- und Alternativhypothese, nachdem bereits Daten vorliegen und Analysen durchgeführt wurden („Hypothesizing After the Results are Known“ – HARKing: Kerr 1998), ohne eine Kreuzvalidierung geplant zu haben.

Studien zur Prävalenz solcher „Questionable Research Practices“ (z. B. John et al. 2012) zeigen, dass unter befragten Psychologinnen/Psychologen 63,4 % in einer Publikation nicht alle abhängigen Variablen berichtet haben; 45,8 % in einer Publikation nur jene Studien berichteten, die Effekte zeigten; 38,2 % Entscheidungen über Datenselektion erst trafen, nachdem die Auswirkungen auf die Ergebnisse geprüft wurden; 27 % in einer Publikation ein ihnen unerwartetes Ergebnis als von Anfang an intendiert darstellten.

Den möglichen Effekt dieser „Questionable Research Practices“ arbeiten Simmons et al. (2011) in ihrer Simulationsstudie heraus. Mit nur vier marginalen Modifikationen der Analysen (z. B. Geschlecht als Kovariate aufnehmen oder nicht), erreichten sie in 61 % der Simulationen statistische Signifikanz ($p<0,05$), obwohl es keinen Effekt gab. Untersuchungen zu publizierten Studien legen nahe, dass diese möglichen Effekte sich schon niedergeschlagen haben: Studien, bei welchen vor der Datenerhebung die Pläne für ihre Datenerhebung und -auswertung nicht veröffentlicht wurden, haben höhere Effektstärken (Median $r=0,36$ vs $0,16$: Schäfer und Schwarz 2019) und berichten mehr bestätigte Hypothesen (96 % vs. 44 %: Scheel et al. 2020) als jene, bei denen der Forschungsprozess von Beginn an offengelegt wurde.

Zusammengenommen gilt damit, was bereits seit langem vermutet wurde, nämlich dass die Gesamtheit aller publizierten Arbeiten nur ein verzerrtes Abbild der

Wirklichkeit darstellt (Sterling 1959). Diese Verzerrung wird in Meta-Analysen regelmäßig adressiert (s. Publikationsbias, z.B. Brohmer et al. 2019). Weiters zeigt sich diese Verzerrung aber auch darin, dass groß angelegte Replikationsstudien wiederholt zum Schluss kommen, dass sich publizierte Effekte nicht konsistent replizieren lassen (Psychologie: Open Science Collaboration 2015; Bildungsforschung: Makel und Plucker 2014; Pharmazie: Prinz et al. 2011; Krebsforschung: Begley und Ellis 2012). Auf der positiven Seite gibt es auch hier Gegenbeispiele wie die Persönlichkeitspsychologie, wo sich Effekte großteils replizieren lassen (Soto 2019).

Wie prävalent diese Verzerrungen in der Bildungsforschung sind, ist derzeit noch offen. Auch wissen wir nicht, inwiefern die Belastbarkeit der Befunde in der Bildungsforschung darunter leidet. Die Kontroversen in anderen Disziplinen legen aber nahe, dass dieselben Herausforderungen auch auf die Belastbarkeit der Befunde in der Bildungsforschung zutreffen. Wir finden bereits einige Hinweise auf ähnliche Problemlagen in der Bildungsforschung (Makel und Plucker 2014; Pigott et al. 2013; Rost und Bienefeld 2019). Es finden sich auch ähnliche Forderungen wie in anderen Disziplinen, nämlich, dass Bildungsforscherinnen/Bildungsforscher transparenter und nachvollziehbarer forschen sollten (van der Zee und Reich 2018). Vor diesem Hintergrund möchten wir im vorliegenden Beitrag eine Möglichkeit aufzeigen, um die Belastbarkeit von Befunden der Bildungsforschung langfristig zu sichern: Open Science.

2 Open Science: Begrifflichkeiten und Praktiken

Unter dem Begriff *Open Science* wird Verschiedenes verstanden. Folgt man der OECD (2015), liegt das Hauptaugenmerk von Open Science darauf, mit öffentlich Geldern finanzierte Forschungen auch öffentlich zugänglich zu machen („*to make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction*“ p. 7). Fecher und Friesike (2014) fassen Open Science breiter, indem sie fünf Grundgedanken unterscheiden. Diese fünf Grundgedanken sind: (1) Demokratie – alle haben dasselbe Anrecht auf das generierte Wissen; (2) Pragmatik – Wissen wird effizienter gemeinsam generiert; (3) Infrastruktur – Unterstützung des Forschungsprozesses durch technische Lösungen; (4) Öffentlichkeit – der Öffentlichkeit das generierte Wissen verständlich näherzubringen; und (5) Messung – herkömmliche Metriken für wissenschaftliche Performanz müssen neu konzipiert werden. An anderen Stellen wieder wird versucht, Open Science als ein Phänomen zu beschreiben dessen Fokus auf Transparenz, Zugänglichkeit und Zusammenarbeit liegt (Vicente-Saez und Martinez-Fuentes 2018, p. 435: „*Open Science is the transparent and accessible knowledge that is shared and developed through collaborative networks*“). Als konkrete Ausformungen des relativ abstrakten und globalen Begriffs Open Science findet man eine große Vielzahl von Einzelaspekten, wie z.B. Open Access, Open Data, Open Reproducible Research, Open Science Evaluation, Open Science Tools etc. (vgl. z.B. Knoth und Pontika 2015).

Im vorliegenden Beitrag fokussieren wir auf drei Open Science Praktiken: *Präregistrierung*, *Open Data* und *Open Materials*. Dies sind jene Teile von Open Science,

Abb. 1 Open Science Badges
(bezogen von <https://osf.io/tvyxz/files/>)



1) die im eigenen Forschungsprozess gut umgesetzt werden können, 2) die Forscherinnen/Forscher in die Eigenverantwortung nehmen, transparenter zu arbeiten, und 3) damit die einleitend diskutierten Herausforderungen für die Belastbarkeit von Forschung adressieren. Schließlich sind es auch diese drei Teile von Open Science, die in wissenschaftlichen Fachzeitschriften mehr und mehr umgesetzt werden. Beispielsweise vergeben manche wissenschaftliche Fachzeitschriften bereits sogenannte „Open Science Badges“. Eine Publikation bekommt einen Open Science Badge, wenn *Präregistrierung*, *Open Data* und/oder *Open Materials* umgesetzt wurde. Das Einführen von Open Science Badges kann genutzt werden, um gezielt die Umsetzung von Open Science in Publikationen zu unterstützen (Kidwell et al. 2016). Diese Open Science Badges sind in Abb. 1 dargestellt.

2.1 Präregistrierung

Präregistrierung bedeutet, dass bei einer Studie Fragestellungen, bei konfirmatorischen Studien die Hypothesen, die Stichprobe, die Erhebungsinstrumente, das Studiendesign und die geplanten Auswertungen vorab festgelegt werden. Dies entspricht dem allgemein akzeptierten Ablauf, wie Studien umzusetzen sind: zuerst wird geplant, dann durchgeführt (z.B. Döring und Bortz 2016, oder andere Methodenlehrbücher). Wie weiter oben erwähnt, wird diese Reihenfolge im Forschungsalltag nicht immer restlos eingehalten. Eine Präregistrierung bedeutet für Forscherinnen/Forscher auch im teilweise hektischen Forschungsalltag die Planung von Studien sorgfältig abzuschließen, ehe Daten erhoben werden.

Studien zeigen bereits auf, dass es systematische Unterschiede zwischen publizierten Studien mit und ohne Präregistrierung gibt. Dies zeigt sich darin, dass Effektstärken geringer ausfallen, wenn eine Präregistrierung erfolgte (Schäfer und Schwarz 2019). Diese systematische Verzerrung der publizierten Effekte wird auf mehrere Dinge zurückgeführt. Zuerst verfolgen Präregistrierungen das Ziel, dass explorative Studien zur Generierung von Hypothesen klar getrennt werden können von Studien zur Überprüfung von Hypothesen (für eine vertiefende Diskussion s. Nosek et al. 2018). Eine Präregistrierung einer Studie erschwert damit, dass unerwartete Ergebnisse ohne vorherige Annahmen im Nachhinein als geplant dargestellt werden. Sie erschwert auch, dass eine Vielzahl von Daten erhoben werden und danach nur ausgewählte publiziert werden. Darüber hinaus erschwert Präregistrierung, dass aus verschiedenen statistischen Auswertungen jene mit bevorzugtem Ergebnis gewählt wird, oder Stichproben gar systematisch selektiert werden (s. „researcher degrees of freedom“: Simmons et al. 2011).

Eine Präregistrierung kann von jeder Forscherin/jedem Forscher selbst durchgeführt werden. Eine Präregistrierung kann für quantitative und für qualitative Studien durchgeführt werden, wobei diese für qualitativen Studien jedoch anders konzipiert werden als für quantitative Studien (s. Haven und Van Grootel 2019). Schließlich ist eine Präregistrierung von Analysen auch bei Sekundäranalysen möglich; hier gilt es zu unterscheiden, ob die Daten bereits zugänglich waren oder nicht. Für Präregistrierung bieten sich Plattformen wie <https://osf.io> des *Center of Open Science* oder <https://aspredicted.org/> des *Wharton Credibility Lab* an der *University of Pennsylvania* an. *AsPredicted* versucht dem Forscher/der Forscherin entgegenzukommen, indem er/sie anhand von neun Fragen zu einer Präregistrierung geleitet wird.

Eine Präregistrierung muss aber nicht nur von den Forscherinnen/Forscher selbst ausgehen, sondern kann auch von wissenschaftlichen Fachzeitschriften in den Peer-Review-Prozess integriert werden. Diese Integration liegt der Beitragsart „Registered Reports“ (Chambers 2019) zugrunde. Registered Reports sehen Peer-Reviews zu zwei Stadien des Forschungsprozesses vor. Ein Manuskript wird eingereicht, wenn die Studie und deren Auswertung geplant, und somit quasi die Einleitung und die Methode verschriftlich sind („Stage 1 manuscript“). Nach Begutachtung im Peer-Review-Prozess kann dieses Manuskript vorläufig zur Publikation angenommen werden und die Datenerhebung beginnen. Eine vorläufige Annahme bedeutet hier, dass die Studie die Aussicht hat, publiziert zu werden *unabhängig* von den tatsächlichen Ergebnissen. Diese vorläufige Annahme gilt selbstverständlich nur solange die Studie wie beschrieben durchgeführt wird, die Ergebnisse wie beschrieben ausgewertet und interpretiert werden, und die Studie entsprechend verschriftlich wird. Wenn die Studie abgeschlossen und verschriftlich ist, wird sie abermals eingereicht („Stage 2 manuscript“) und geht erneut in den Peer-Review-Prozess. Laut dem Center für Open Science erlauben derzeit (28.10.2020) 265 wissenschaftliche Fachzeitschriften Registered Reports, so zum Beispiel der *British Journal of Educational Psychology*, *Frontiers in Education*, *Learning and Instruction* oder *Mind, Brain and Education*. Derzeit wird auch bei der *Zeitschrift für Bildungsforschung* an der Implementation von Registered Reports gearbeitet. Registered Reports sollen ab März 2021 eingereicht werden können.

Abschließend wollen wir noch eine Implikation von Registered Reports für Jungwissenschaftlerinnen/Jungwissenschaftler hervorheben. Registered Reports ermöglichen es, auch Null-Effekte leichter zu publizieren. Der Widerstand von Herausgeberinnen/Herausgebern von Zeitschriften, Null-Effekte zu publizieren, ist ein schon lang bekanntes Problem (Sterling 1959). Dies übt bereits einen Druck auf Jungwissenschaftlerinnen/Jungwissenschaftler aus, welche z. B. in kumulativen Abschlussarbeiten darauf angewiesen sind, erhobene Daten zu verwerten. Dieser Druck schlägt sich in Qualifizierungsarbeiten in der Bildungsforschung nieder: Pigott et al. (2013) analysierten 9530 Dissertationen zwischen 2001 und 2005 (publiziert bis spätestens 2011). Sie fanden, dass knapp die Hälfte der publizierten Qualifizierungsarbeiten signifikante Effekte berichten, wobei im Mittel mehr als zwanzig statistische Tests pro Arbeit zu finden waren. Pigott et al. fanden weiters, dass Publikationen, die Daten dieser Qualifizierungsarbeiten verwendeten, nur ca. die Hälfte der ursprünglichen Variablen berichteten. Inkludieren kumulative Abschlussarbeiten hingegen Registered Reports, kann bereits vor der Datenerhebung eine Publikation vorläufig

angenommen werden. Es gibt damit mehr Planungssicherheit für Abschlussarbeiten. Gleichzeitig kann sich das positiv auf die Disziplin auswirken, da die dadurch publizierten Null-Effekte gegen die oben beschriebene vermutete Masse an falsch-positiven Studienergebnissen in der Literatur wirken würden (s. Sterling et al. 1995).

2.2 Open Data

Open Data bedeutet, dass Daten unter Wahrung des Datenschutzes öffentlich zugänglich gemacht werden. Dies kann sich auf Datensätze an sich beziehen, oder auf Daten, die einer spezifischen Publikation zugrunde liegen.

Die Offenlegung der Daten bringt mehrere Vorteile. Erstens nimmt es Forscherinnen/Forscher in die Eigenverantwortung, die eigenen Daten sachgemäß aufzubereiten und zu dokumentieren. Diese Aufbereitung kann leicht im teilweise hektischen Forschungsalltag zu Gunsten von dringenderen Verpflichtungen zu kurz kommen. Deshalb liegen viele Daten in sprichwörtlichen Schreibtischschubladen der Forscherteams und man ist eventuell um eine Antwort verlegen, wenn nach der (Daten-)Grundlage eines spezifischen Ergebnisses einer Publikation von vor fünf Jahren gefragt wird.

Zweitens ermöglicht die Offenlegung der Daten, dass publizierte Ergebnisse nachvollzogen und überprüft werden können. Es stärkt damit das Vertrauen in Publikationen, da Forscherinnen/Forscher den Weg zu ihren Ergebnissen offenlegen. Eine Stärkung dieses Vertrauens sollte grundsätzlich angestrebt werden: das Nicht-Offenlegen von Daten geht oft einher mit evidenten Fehlern in Publikationen (Wichert et al. 2011). Nicht außer Acht zu lassen ist hierbei auch die Lernmöglichkeit für Jungwissenschaftlerinnen/Jungwissenschaftler oder Studierende. Diese haben durch die Offenlegung der Daten die Möglichkeit Analysen erfahrener Forscherinnen/Forschern zu replizieren.

Drittens macht es die Daten für weitere Forschungsvorhaben nutzbar. Die Daten können für größer angelegte Projekte, wie beispielsweise Meta-Analysen, genutzt werden. Darüber hinaus ermöglicht Open Data es anderen Forscherinnen/Forschern, bestehende Daten für Erkenntnisgewinn zu nutzen, an den ursprünglich noch nicht gedacht wurde oder der ursprünglich noch gar nicht möglich war. Open Data kann somit einen Beitrag dazu leisten, eine Disziplin als Ganzes vorwärts zu bringen.

Für die Umsetzung von Open Data gibt es viele Möglichkeiten. Forscherinnen/Forscher können ihre Daten selbst offenlegen, indem sie bestehende Plattformen nutzen. Als Beispiele seien hier Zenodo (<https://zenodo.org/>) von CERN oder OSF (<https://osf.io>) des *Center of Open Science* genannt. Letzteres erlaubt es auch, aus Servern in verschiedenen Regionen zu wählen, auch innerhalb der EU. Es gibt auch wissenschaftliche Zeitschriften, die sich der Publikation von Daten widmen, wie *Data in Brief* oder *The Journal of Open Psychology Data*. Weiters gibt es in Österreich Anlaufstellen für Open Data, die Forscherinnen/Forscher in diesem Prozess unterstützen, wie zum Beispiel AUSSDA (Austrian Social Science Data Archive: <https://data.aussda.at/>). Eine tiefere Diskussion zur Weiter- und Wiederverwendung von Daten ist in den *FAIR Richtlinien* (David et al. 2020; Wilkinson et al. 2016) zu finden.

2.3 Open Materials

Open Materials stellt den Anspruch, die Teile des Forschungsprozesses zugänglich zu machen, die für die Reproduktion 1) des berichteten Verfahrens und 2) der Analyse erforderlich sind. Ersteres muss es anderen Forscherinnen/Forschern ermöglichen, die Studie mit demselben Material durchführen zu können (z. B. Stimuli). Einschränkungen gelten hier für kommerziell geschützte Materialien (z. B. lizenzierte Fragebögen), die entsprechend deren Rechte zur Verbreitung angegeben werden sollten. Zweiteres muss es anderen Forscherinnen/Forschern ermöglichen, die Daten auf dieselbe Art auswerten zu können. Am einfachsten können dafür zusätzlich zu Open Data annotierte SPSS-Syntax-Files oder R-Skripte zugänglich gemacht werden. Für weiterführende Möglichkeiten sei die interessierte Leserin/der interessierte Leser auf Software-Lösungen wie „R Markdown“ verwiesen (z. B. Gandrud 2015).

Open Materials bergen direkte und indirekte Vorteile für die Forscherinnen/Forscher und auch für die Disziplin. Am augenscheinlichsten erleichtern Open Materials anderen Forscherinnen/Forscher eine Replikation bestehender Ergebnisse. Open Materials ermöglichen darüber hinaus im Detail nachzuvollziehen, wie Analysen durchgeführt wurden. Sowohl hoch-aufwändige qualitative Auswertungen als auch hoch-komplexe quantitative Verfahren können in Publikationen oft nicht im gewünschten Detail dargestellt werden. Mit Open Materials ist dies allerdings möglich. Damit sind andere Forscherinnen/Forscher überhaupt erst in die Lage, publizierte Analysen im Detail nachvollziehen zu können. Dasselbe gilt auch für Gutachterinnen/Gutachter im Peer-Review-Prozess.

Schließlich nehmen Open Materials Forscherinnen/Forscher in die Eigenverantwortung, die eigenen Analysen verständlich und nachvollziehbar zu dokumentieren, am besten zusammen mit Open Data. Forscherinnen/Forscher sollten dadurch in der Lage sein, auch Jahre später Auskunft darüber geben zu können, wie spezifische Ergebnisse zustanden kamen bzw. sollte dies anhand des veröffentlichten Materials für andere selbsterklärend sein.

2.4 Ein Beispiel

Im Folgenden soll ein Beispiel aus der Bildungsforschung veranschaulichen, wie der Forschungsprozess unter Einhaltung von Präregistrierung, Open Data und Open Materials ausschaun kann. Als Beispiel dient eine aktuelle Studie zur Online-Lehre zu Beginn von COVID-19, die ob ihres Mixed-Methods Zugangs sowohl qualitative als auch quantitative Teilstudien enthielt. Ein Pre-Print der Studie (Krammer et al. 2020a) wurde vor der eigentlichen Studie (Krammer et al. 2020b) veröffentlicht.

Präregistrierung, Open Data und Open Materials wurden mithilfe der Plattform OSF durchgeführt. Auf OSF finden sich auch detaillierte Anleitungen zur Umsetzung von Präregistrierungen (z. B. <https://help.osf.io/hc/en-us/articles/360019930893-Register-Your-Project>: OSF Support → Register Your Project). Das Forschungsprojekt wurde auf OSF zuerst nur privat zugänglich gemacht und wurde erst nach der Publikation öffentlich verfügbar. Weiters wurden anonymisierte Links ohne Bearbeitungsrechte erstellt, um im Peer-Review-Prozess die Anonymität zu wahren.

Im ersten Schritt wurde das Forschungsprojekt auf OSF erstellt (<https://osf.io/sujry/> bzw. https://osf.io/sujry/?view_only=556864b88bcd4237a32b9fb1660ca850) und wurde in seine zwei Teilstudien, die qualitative Studie 1 (<https://osf.io/7knhj/> bzw. https://osf.io/7knhj/?view_only=6fdb64dadff4d7484c6306b45c301df) und die quantitative Studie 2 (<https://osf.io/87v5y/> bzw. https://osf.io/87v5y/?view_only=bca9dc18dd8847c5bcc96a17d6383ac7) verzweigt.

Präregistrierung Beide Teilstudien wurden vor ihrer jeweiligen Datenerhebung im offenen Format präregistriert („Open-Ended Registration“). Die Präregistrierungen wurden mit einem Embargo für vier Jahre belegt. Innerhalb des Embargos sind die Inhalte nicht öffentlich zugänglich, sie können aber auch vorzeitig vom Team öffentlich gemacht werden. Für die qualitative Studie 1 beschrieb die Präregistrierung (<https://osf.io/438p6> bzw. https://osf.io/438p6/?view_only=bc6291b1dd548d89fe0570070b397fe) u. a. die Stichprobe, die offenen Fragen, die Datenerhebung, und schließlich die qualitative Inhaltsanalyse mit induktiver Kategorienbildung. Für die quantitative Studie 2 beschrieb die Präregistrierung (<https://osf.io/rj5f9> bzw. https://osf.io/rj5f9/?view_only=7fbbdc1c3955472b83d1431357dedbe0) die Rekrutierung (Beschreibung und Ausschlusskriterien), die experimentelle Manipulation, den Fragebogen und die statistischen Analysen. Für letzteres wurde entweder das konkrete Verfahren beschrieben (z. B. BEST mit Markov Chain Monte Carlo Länge 100.000, keine Ausdünnung, uninformierte Priors, im R Paket BEST) oder, wenn Details der Auswertung von den Daten abhingen, wie vorgegangen werden wird (z. B. der Schätzer in Abhängigkeit der Verteilung der Daten).

Open Data Nach den Datenerhebungen wurden für beide Teilstichproben alle Rohdaten auf OSF hochgeladen. Für die qualitative Studie 1 waren dies 75 txt-Dokumente, wo für jede befragte Person deren offene Antworten enthalten sind. Für die quantitative Studie 2 war dies ein Excel-Dokument zur Beschreibung aller Spalten des Datensatzes und der Datensatz als ein txt-Dokument.

Open Materials Für die qualitative Studie 1 wurde ein Excel-Dokument hochgeladen, wo für jede der textnahen induktiv-gebildeten Kategorie deren Häufigkeiten und deren Zusammenführung zu Ober-Kategorien aufgelistet sind. Für die quantitative Studie 2 wurden die vollständigen R-Skripte mit Annotationen hochgeladen.

3 Spezifika und Limitationen von Open Science in der Bildungsforschung

Bildungsforschung unterscheidet sich grundsätzlich von den „harten“ Naturwissenschaften, die in ihren experimentellen Designs zumeist eindeutige Treatments und kontrollierbare (und daher kontrollierte) Kontextbedingungen haben (vgl. z. B. Berliner 2002). Folglich stellt sich für uns die Frage, ob es in der Bildungsforschung auch Spezifika gibt, die berücksichtigt werden sollten, wenn Open Science Praktiken von anderen Disziplinen übertragen werden.

Ein wesentliches Spezifikum der Bildungsforschung besteht im methodischen Zugang. Häufig kommen qualitative Methoden zum Einsatz, oft auch kombiniert und integriert mit quantitativen in Form von Mixed Methods (z.B. Creamer 2018; Creswell und Creswell 2018; Kuckartz 2014). Aspekte von Open Science, insbesondere Präregistrierung und Open Materials, sind aber durchaus im Bereich des Möglichen und Sinnvollen (Haven und Van Grootel 2019). Für Open Data gilt es zu berücksichtigen, dass die generierten Daten zumeist textliche Äußerungen oder Beobachtungen sind, deren Anonymisierung bzw. Pseudonymisierung nur mit hohem Aufwand möglich ist.

Weitere Besonderheiten der Bildungsforschung liegen darin begründet, dass experimentelle Designs in der Praxis äußerst schwierig bzw. unmöglich umzusetzen sind. „*The RCT design is a theoretical construct of considerable interest, but it has essentially zero practical application to the field of human affairs.*“ (Scriven 2008, S. 12). Qualitätsmerkmale naturwissenschaftlicher Forschung wie Doppel- oder auch nur Einfachverblindung, randomisierte Allokation zu Versuchs- und Vergleichsgruppen (wobei auch die Definition einer Kontrollgruppe in der Bildungsforschung hinterfragbar wäre: was wäre ein geeignetes Placebo?), prospektives Studiendesign usw. sind schlichtweg nur in wenigen Einzelfällen möglich. Wenn sie annähernd möglich sind, so zeigen sich deutliche Einflüsse methodischer Details auf Effektstärken (z. B. publizierte vs. unpublizierte Studien, oder kleine vs. große Stichproben: Cheung und Slavin 2016).

Im Bemühen, die Kontextbedingungen, wenn schon nicht kontrollieren zu können, so doch wenigstens deren Einfluss auf die Zielvariablen statistisch einzuschätzen, wird oft eine Vielzahl an Variablen erhoben – dies auch im Hinblick auf potenzielle Sekundäranalysen der Daten. Diese Kontextvariablen sind meist zu einer sachgemäßen Interpretation der Treatments notwendig. Beispielsweise haben Eder et al. (2015) im Rahmen der Evaluation des Modellversuchs *Neue Mittelschule* (NMS) parallel zu den Erhebungen der Zielvariablen auch Schulleitungen und Lehrpersonen zu konkreten Umsetzungen des Modells NMS an ihren Standorten befragt. Obwohl bei oberflächlicher Betrachtung das Treatment vordergründig als „Zugehörigkeit zur Versuchsgruppe der NMS-Pilotschulen“ klar definiert war, stellte sich heraus, dass das Modell NMS an den Versuchsschulen mit unterschiedlichen Schwerpunktsetzungen und unterschiedlicher Intensität umgesetzt wurde. Aus den Daten der Begleiterhebungen zu den konkreten Umsetzungen wurden vier Cluster von Schulen identifiziert und bei differenzierter Betrachtung dieser Cluster zeigten sich teilweise deutlich unterschiedliche Effekte (Eder et al. 2015, S. 179–202). Dies illustriert eindrücklich die Sinnhaftigkeit und Notwendigkeit der Erhebung von Kontextvariablen, würde es durch Nutzung von Open Data aber auch anderen Forscherinnen/Forschern ermöglichen, weitere Analysen anzustellen und Hypothesen zu prüfen.

Allerdings verlockt die bloße Verfügbarkeit dieser Vielzahl von Variablen auch zum HARKing, im angloamerikanischen Sprachraum auch als „Texas Sharpshooter Fallacy“ bekannt: Die Forscherin/der Forscher sieht sich explorativ Daten an und beleuchtet sie auf eventuelle Zusammenhänge. Die meisten solcher Studien der Bildungsforschung fallen damit nicht mehr unter die Kategorie eines Zufallsexperiments, das a priori formulierte Hypothesen auf statistische Signifikanz prüft. Solange

die so gewonnenen Befunde nicht durch direkte Replikationsstudien validiert wurden, müssen die hierbei aufgestellten ad-hoc Hypothesen allerdings als ungeprüft gelten (Ulrich et al. 2016). Klare Hypothesenformulierungen und detaillierte Planungen der Datenanalysestrategien sind beim Umgang mit großen Variablenmengen eine Notwendigkeit. Präregistrierung verlangt von Forscherinnen/Forschern bereits vor Datenerhebung alle Details der Studie festzulegen und kann damit HARKing entgegenwirken.

Dass diese vorherige Offenlegung der Planung in der Bildungsforschung nicht immer umsetzbar ist, zeigen Sekundäranalysen von großangelegten Erhebungen wie PISA oder den Bildungsstandardsüberprüfungen. Hier gibt es für die Erstanalysen klare Zielsetzungen und Fragestellungen, die vor der Datenerhebung formuliert wurden; für weiterführende (Sekundär-)Analysen gilt dies nur mehr eingeschränkt. In diesen Fällen sind die Daten schon vorhanden und erste Ergebnisse publiziert, wenn Bildungsforscherinnen/Bildungsforscher sich an die Sekundäranalysen machen, womit die geforderte Abfolge von Hypothesenformulierung und Datenerhebung nicht mehr einzuhalten ist. In diesen Fällen gilt es, noch kritischer die Stringenz der selbst vorgebrachten theoretischen Fundierung zu prüfen und die explorative Natur der Analysen anzuerkennen. Eine Nachvollziehbarkeit der für die entsprechenden Publikationen angestellten Analysen ist in diesen Fällen aufgrund der grundsätzlichen Offenheit der Daten meist unmittelbar gegeben, aber wegen der Weiterverwendung der Erhebungsinstrumente für Zeitvergleiche sind Open Materials nur eingeschränkt möglich.

Neben den hier skizzierten Einschränkungen gibt es bereits erste Versuche, Präregistrierung für die Bildungsforschung durch spezifische Leitlinien zu adaptieren und nutzbringend zu gestalten. Dazu gehören Empfehlungen für die individuellen Forscherinnen/Forscher („*Pre-register hypothesis-testing studies, Be transparent in describing the study design, Leave only one path (or process) per hypothesis, Split up the results section*“), aber auch für das wissenschaftliche Fachgebiet („*Encourage pre-registration, Borrow liberally from other disciplines and fields, Allow for iterations in the pre-registration process*“) (Gehlbach und Robinson 2018, S. 304).

4 Plädoyer für Transparenz

In diesem Beitrag setzten wir uns einerseits Bewusstseinsbildung und Sensibilisierung für Forschungspraktiken zum Ziel, und andererseits die Thematisierung von Open Science Praktiken in der Bildungsforschung. Wir sprechen uns für drei Open Science Praktiken aus: *Präregistrierung*, *Open Data* und *Open Materials*. Präregistrierung sichert, dass Forscherinnen/Forscher die Planung von Studien sorgfältig abschließen, ehe sie Daten erheben und/oder analysieren und kann damit Questionable Research Practices wie HARKing und *p*-Hacking entgegenwirken. Open Data sichert, dass Forscherinnen/Forscher Daten sachgemäß aufbereiten und dokumentieren, und ermöglicht, publizierte Ergebnisse nachzuvollziehen und zu überprüfen, und macht Daten für weitere Forschungsvorhaben nutzbar. Open Materials erleichtert Replikationen und steigert die Nachvollziehbarkeit von Analysen. Wir sprechen uns dafür aus, dass diese Open Science Praktiken dadurch einen Beitrag dazu leisten

können, dass langfristig die Belastbarkeit der Befunde der Bildungsforschung erhöht wird.

Gleichzeitig haben wir gezeigt, dass auch Open Science kein Allheilmittel sein kann, kritisch hinterfragt werden muss und in Teilbereichen nicht sinnvoll anwendbar ist. Die Umsetzung von Open Science muss für jede Disziplin neu gedacht werden. Praktiken, die in einer Disziplin umsetzbar sind, mögen in einer anderen Disziplin (derzeit) nicht möglich sein oder auch nichts zum Besseren verändern. Vor diesem Hintergrund gilt es zu berücksichtigen, welche Spezifika wir in der Bildungsforschung antreffen und wo Open Science hier ansetzen kann.

Trotz diesen Einschränkungen möchten wir ein Plädoyer für Transparenz aussprechen. Wo möglich sollten Bildungsforscherinnen/Bildungsforscher sich selbst in die Verantwortung nehmen, transparenter zu forschen. Wir möchten dazu anregen, dass jede/jeder sich selbst fragt: Wenn ich die nächste Studie durchführe, weshalb soll ich die Studie nicht vorab fertig planen (*Präregistrierung*) und weshalb soll ich nicht danach die Daten und Materialien für andere offenlegen (*Open Data* und *Open Materials*) – mit dem Nebeneffekt, sie auch für mich selbst gut dokumentiert und auffindbar (FAIR, Wilkinson et al. 2016) zu halten?

Die Botschaft von Open Science darf es auch nicht sein, dass die Verantwortung für Veränderungen in den Forschungspraktiken ausschließlich bei den Forscherinnen/Forschern liegt. Um langfristig die Belastbarkeit von Forschung zu erhöhen, muss sich Entscheidendes auf der System-Ebene ändern; bei Institutionen, Geldgebern und wissenschaftlichen Fachzeitschriften. Mögliche Wege wurden hierfür schon an mehreren Stellen aufgezeigt. Vorschläge reichen von einer Würdigung von Open Science Praktiken bei Berufungen (s. Munafò et al. 2017) bis hin zu einem grundsätzlichen Hinterfragen, wie das heutige System der Wissenschaft sich zu einem nicht mehr nachhaltigen System entwickelt hat (s. „Slow Science“: Stengers 2018). Auf Slow Science aufbauend gibt es spannende Ansätze, wie eine Beschränkung der Anzahl der von einer Person haltbaren Forschungsfördermittel und von Publikationen pro Person und Jahr (Frith 2020). Veränderungen wurden auch schon umgesetzt, zum Beispiel darin, dass an manchen Institutionen Berufungskommission nicht nur zählen (d. h. Bibliometrie evaluieren), sondern den eigentlichen und zukünftigen Beitrag von Bewerberinnen/Bewerbern beurteilen sollen (Benedictus et al. 2016).

Zusammengefasst sollte es auf der individuellen Ebene Bemühungen geben transparenter zu forschen, aber auch auf der System-Ebene Veränderungen geben, die belastbarere Forschung ermöglichen und belohnen. Es muss uns um eine Wissenschaft gehen, die Wissen schafft.

Danksagung Wir möchten uns bei unseren geschätzten Kolleginnen/Kollegen und Freundinnen/Freunden bedanken, welche uns für diesen Beitrag mit ihrer Zeit und Expertise geholfen haben. Danke, Hilmar Brohmer, Barbara Pflanzl und Hubert Schapp (in alphabetischer Reihenfolge)!

Literatur

Alliance—SHARing Reward and Credit (SHARC) Interest Group, T. R. D., David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., et al. (2020). FAIRness literacy: the Achilles' heel of applying FAIR principles. *Data Science Journal*, 19, 32. <https://doi.org/10.5334/dsj-2020-032>.

- Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*, 531–533.
- Benedictus, R., Miedema, F., & Ferguson, M. W. (2016). Fewer numbers, better science. *Nature*, *538*(7626), 453–455.
- Berliner, D. C. (2002). Comment: educational research: the hardest science of all. *Educational Researcher*, *31*(8), 18–20. <https://doi.org/10.3102/0013189X031008018>.
- Brohier, H., Corcoran, K., Aert, R. C. M. van, & Eckerstorfer, L. V. (2019, August 2). *Is there Inspiration Through Observation? A Meta-analysis on Goal Contagion* (PREPRINT). <https://doi.org/10.31219/osf.io/xydggf>.
- Chambers, C. (2019). What's next for registered reports? *Nature*, *573*, 187–189. <https://doi.org/10.1038/d41586-019-02674-6>.
- Cheung, A. C. K., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, *45*(5), 283–292. <https://doi.org/10.3102/0013189X16656615>.
- Conn, L. K., Edwards, C. N., Rosenthal, R., & Crowne, D. (1968). Perception of emotion and response to teachers' expectancy by elementary school children. *Psychological Reports*, *22*, 27–34.
- Coyne, M. D., Cook, B. G., & Therrien, W. J. (2016). Recommendations for replication research in special education: a framework of systematic, conceptual replications. *Remedial and Special Education*, *37*(4), 244–253. <https://doi.org/10.1177/0741932516648463>.
- Creamer, E. G. (2018). *An introduction to fully integrated mixed methods research*. Los Angeles: SAGE.
- Creswell, J. W., & Creswell, J. D. (2018). *Research design: qualitative, quantitative, and mixed methods approaches* (5. Aufl.). Los Angeles: SAGE.
- Dickersin, K. (1990). The existence of publication bias and risk factors for its occurrence. *JAMA*, *263*(10), 1385–1389.
- Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (Springer-Lehrbuch). Berlin, Heidelberg: Springer. <https://doi.org/10.1007/978-3-642-41089-5>.
- Eder, F., Altrichter, H., Hofmann, F., & Weber, C. (2015). *Evaluation der Neuen Mittelschule (NMS) Befunde aus den Anfangskohorten; Forschungsbericht*. Graz: Leykam.
- Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0005738>.
- Fanelli, D. (2010a). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0010068>.
- Fanelli, D. (2010b). Do pressures to publish increase scientists' bias? An empirical support from US states data. *PLoS ONE*. <https://doi.org/10.1371/journal.pone.0010271>.
- Fecher, B., & Friesike, S. (2014). Open science: one term, five schools of thought. In *Opening science* (S. 17–47). Cham: Springer.
- Frith, U. (2020). Fast lane to slow science. *Trends in Cognitive Sciences*, *24*, 1–2. <https://doi.org/10.1016/j.tics.2019.10.007>.
- Gandrud, C. (2015). *Reproducible research with R and RStudio* (2. Aufl.). Chapman & Hall/CRC the R series. Boca Raton: CRC Press, Taylor & Francis Group.
- Gehlbach, H., & Robinson, C. D. (2018). Mitigating illusory results through preregistration in education. *Journal of Research on Educational Effectiveness*, *11*(2), 296–315. <https://doi.org/10.1080/19345747.2017.1387950>.
- Haven, T. L., & Van Grootel, L. (2019). Preregistering qualitative research. *Accountability in Research*, *26*(3), 229–244. <https://doi.org/10.1080/08989621.2019.1580147>.
- Head, M. L., Holman, L., Lanfear, R., Kahn, A. T., & Jennions, M. D. (2015). The extent and consequences of P-hacking in science. *PLoS Biology*, *13*(3), e1002106. <https://doi.org/10.1371/journal.pbio.1002106>.
- Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*, *2*(8), e124.
- John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532. <https://doi.org/10.1177/0956797611430953>.
- Kerr, N. L. (1998). HARKing: hypothesizing after the results are known. *Personality and Social Psychology Review*, *2*(3), 196–217.
- Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., Nosek, B. A., et al. (2016). Badges to acknowledge open practices: a simple, low-cost, effective method for increasing transparency. *PLoS Biology*, *14*(5), 1–15. <https://doi.org/10.1371/journal.pbio.1002456>.
- Knuth, P., & Pontika, N. (2015). Open Science Taxonomy. *figshare. Figure*. <https://doi.org/10.6084/m9.figshare.1508606.v3>.

- Krammer, G., Pflanzl, B., & Matischek-Jauk, M. (2020a). 39 Aspekte der Online-Lehre zur Förderung positiven Erlebens und/oder Motivation bei Lehramtsstudierenden: Mixed-Method Befunde zu Beginn von COVID-19. *September*. <https://doi.org/10.35542/osf.io/6v2yr>.
- Krammer, G., Pflanzl, B., & Matischek-Jauk, M. (2020b). Aspekte der Online-Lehre und deren Zusammenhang mit positivem Erleben und Motivation bei Lehramtsstudierenden: Mixed-Method Befunde zu Beginn von COVID-19. *Z f Bildungsforsch*. <https://doi.org/10.1007/s35834-020-00283-2>.
- Kuckartz, U. (2014). *Mixed methods*. Wiesbaden: Springer. <https://doi.org/10.1007/978-3-531-93267-5>.
- Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: replication in the education sciences. *Educational Researcher*, 43(6), 304–316. <https://doi.org/10.3102/0013189X14545513>.
- Munafò, M. R., Nosek, B. A., Bishop, D. V., Button, K. S., Chambers, C. D., Du Sert, N. P., & Ioannidis, J. P. (2017). A manifesto for reproducible science. *Nature human behaviour*, 1(1), 1–9.
- Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115, 2600–2606. <https://doi.org/10.1073/pnas.1708274114>.
- OECD (2015). *Making open science a reality*. Paris: OECD Publishing. <https://doi.org/10.1787/5jrs2f963zs1-en>.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*. <https://doi.org/10.1126/science.aac4716>.
- Pigott, T. D., Valentine, J. C., Polanin, J. R., Williams, R. T., & Canada, D. D. (2013). Outcome-reporting bias in education research. *Educational Researcher*, 42(8), 424–432. <https://doi.org/10.3102/0013189X13507104>.
- Prinz, F., Schlange, T., & Asadullah, K. (2011). Believe it or not: How much can we rely on published data on potential drug targets? *Nature Reviews Drug Discovery*, 10, 712–713.
- Rosenthal, R. (1969). Interpersonal expectations: Effects of the experimenter's hypothesis. In R. Rosenthal & R. L. Rosnow (Hrsg.), *Artifact in behavioral research* (S. 181–277). New York: Academic Press.
- Rosenthal, R., & Jacobson, L. (1966). Teachers' expectancies: determinants of pupils' IQ gains. *Psychological reports*, 19(1), 115–118.
- Rosenthal, R., & Jacobson, L. (1968). *Pygmalion in the classroom: teacher expectation and pupils' intellectual development*. New York: Holt, Rhinehart & Winston.
- Rost, D. H., & Bienefeld, M. (2019). Nicht replizieren: publizieren!? *Zeitschrift für Pädagogische Psychologie*, 33(3–4), 163–176. <https://doi.org/10.1024/1010-0652/a000253>.
- Schäfer, T., & Schwarz, M. A. (2019). The meaningfulness of effect sizes in psychological research: differences between sub-disciplines and the impact of potential biases. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2019.00813>.
- Scheel, A. M., Schijen, M., & Lakens, D. (2020, February 5). *An excess of positive results: Comparing the standard Psychology literature with Registered Reports*. <https://doi.org/10.31234/osf.io/p6e9c>.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100. <https://doi.org/10.1037/a0015108>.
- Scriven, M. (2008). A Summative evaluation of RCT methodology: an alternative approach to causal research. *Journal of Multidisciplinary Evaluation*, 5(9), 11–24.
- Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological science*, 22(11), 1359–1366.
- Smaldino, P. E., & McElreath, R. (2016). The natural selection of bad science. *Royal Society Open Science*, 3, 160384. <https://doi.org/10.1098/rsos.160384>.
- Smaldino, P. E., Turner, M. A., & Contreras Kallens, P. A. (2019). Open science and modified funding lotteries can impede the natural selection of bad science. *Royal Society open science*, 6, 190194. <https://doi.org/10.1098/rsos.190194>.
- Soto, C. J. (2019). How replicable are links between personality traits and consequential life outcomes? The life outcomes of personality replication project. *Psychological Science*, 30, 711–727. <https://doi.org/10.1177/0956797619831612>.
- Stengers, I. (2018). *Another science is possible: a manifesto for slow science*. Cambridge, Medford: Polity Press.
- Sterling T. D. (1959). Publication Decisions and their Possible Effects on Inferences Drawn from Tests of Significance—or Vice Versa. *Journal of the American Statistical Association*, 54(285), 30–34.
- Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: the effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, 49(1), 108–112.

- Szucs, D., & Ioannidis, J.P.A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. *PLoS Biology*, *15*(3), 1–18. <https://doi.org/10.1371/journal.pbio.2000797>.
- Travers, J. C., Cook, B. G., Therrien, W. J., & Coyne, M. D. (2016). Replication research and special education. *Remedial and Special Education*, *37*(4), 195–204. <https://doi.org/10.1177/0741932516648462>.
- Ulrich, R., Erdfelder, E., Deutsch, R., Strauß, B., Brüggemann, A., Hannover, B., & Rief, W. (2016). Inflation von falsch-positiven Befunden in der psychologischen Forschung. *Psychologische Rundschau*, *67*, 163–174.
- Vicente-Saez, R., & Martinez-Fuentes, C. (2018). Open Science now: a systematic literature review for an integrated definition. *Journal of Business Research*, *88*, 428–436. <https://doi.org/10.1016/j.jbusres.2017.12.043>.
- Wicherts, J.M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS ONE*, *6*, 1–7. <https://doi.org/10.1371/journal.pone.0026828>.
- Wicherts, J.M., Veldkamp, C.L.S., Augusteijn, H.E.M., Bakker, M., van Aert, R.C.M., & van Assen, M.A.L.M. (2016). Degrees of freedom in planning, running, analyzing, and reporting psychological studies: a checklist to avoid P-hacking. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2016.01832>.
- Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J., Appleton, G., Axton, M., Baak, A., Bouwman, J., et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific data*, *3*, 1–9.
- Zee, T. van der, & Reich, J. (2018). Open education science. *AERA Open*. <https://doi.org/10.1177/2332858418787466>.