Low Research Data Availability in Articles Published in Educational Psychology Journals: No Indication of Effective Research Data Policies

Markus Huff[1,2] and Elke C. Bongartz[3]

[1]Department of Psychology, University of Tübingen, Germany

[2]Leibniz-Institut für Wissensmedien (IWM), Tübingen, Germany

[3]German Institute for Adult Education (DIE), Bonn, Germany

Author Note

We thank Chiara Grimm and Lara Kläffling for their help in coding the scientific publications. We have no conflicts of interest to disclose.

Correspondence concerning this article should be addressed to Markus Huff, Leibniz-Institut für Wissensmedien, Schleichstr. 6, D-72076 Tübingen, Germany. Email: m.huff@iwm-tuebingen.de

Abstract

Research data availability contributes to the transparency of the research process and the credibility of educational psychology research and science in general. Recently, there have been many initiatives to increase the availability and quality of research data. Many research institutions have adopted research data policies. This increased awareness might have raised the sharing of research data in empirical articles. To test this idea, we coded $N = 1242$ publications from five educational psychology journals and the psychological journal Cognition (as a baseline) published in 2018 and 2020. Research data availability was low (3.85% compared to 62.74% in Cognition) but has increased from 0.32% (2018) to 7.16% (2020). However, neither the data transparency level of the journal nor the existence of an official research data policy on the level of the corresponding author's institution was related to research data availability. We discuss the consequences of these findings for institutional research data management processes.

Keywords: educational psychology, research data sharing, research data policy, FAIR data principles, data transparency levels

Low Research Data Availability in Articles Published in Educational Psychology Journals: No Indication of Effective Research Data Policies

Research data availability is one of the keys to research transparency and credibility of scientific findings (Asendorpf et al., 2013; Bond-Lamberty, 2016; Hardwicke et al., 2018). From a societal perspective, reliable and transparent scientific findings are particularly critical when they are the basis for political and practical recommendations, such as in educational research (including educational psychology) (Fleming et al., 2021; Patall, 2021). Importantly, available research data do not only increase the comprehensibility of research results on which they are based, but also serve as a valuable source to further process them into secondary data and use them as a buildup for secondary data analysis (Weston et al., 2019). Thus, data sharing as part of open science acts as a scientific accelerator and contributes significantly to scientific progress in the tradition of scientific paradigms formulated by Popper (1959, 1963) and Merton (1973). Yet, the potential for growth in sharing research data is extensive (Bond-Lamberty, 2016). Earlier studies have shown that 73% of authors did not share research data from published studies (Wicherts et al., 2006). This is remarkable, since these authors, by publishing their scientific article, have agreed to share research data for reanalyzes as specified by the American Psychological Association's (APA) Certification of Compliance With APA Ethical Principles (American Psychological Association, 2001). Further, research data sharing relates to reported statistical quality (Wicherts et al., 2011). Because the availability of research data declines rapidly with article age (Tedersoo et al., 2021; Vines et al., 2014), in recent years, both scientific journals and research institutions have adopted policies on handling research data with the goal to archive research data in a sustainable way. Boccali et al. (2021) understand sustainability of research data to mean long-term preservation, accessibility, and interoperability. In the present study, we ask if research

data policies (on both the journal and the research institute's level) impact research data availability in educational psychology?

Not only since the 2014 series of papers published in the journal Lancet on the quality of biomedical research (Al-Shahi Salman et al., 2014; Chalmers et al., 2014; Chan et al., 2014; Glasziou et al., 2014; Ioannidis et al., 2014), and the replication crises in psychology (Open Science Collaboration, 2015), the educational sciences (Makel & Plucker, 2014; Plucker & Makel, 2021; Shaver & Norton, 1980), the social sciences (Camerer et al., 2018), cancer research (Errington et al., 2021), and economics (Camerer et al., 2016), the issue of sustainability has become increasingly important when it comes to research data management. Because the availability of research data is one of the building blocks for the credibility of science, we focus in this study on research data availability in the field of "education, teaching, and educational psychology," which has high practical relevance for society (Flake, 2021; Fleming et al., 2021; Gehlbach & Robinson, 2021; Patall, 2021; van der Zee & Reich, 2018). Evidence-based advice is an essential pillar for the advancement of educational settings. Clearinghouse studies, as an example, provide compiled empirical evidence to practitioners in a highly standardized way (Gersten & Hitchcock, 2009). Thus, the transparency and traceability of the research process are of high importance for these kinds of studies.

Several initiatives aim at overcoming the issue of low research data availability such as the GO FAIR initiative (Mons et al., 2020; Velterop & Schultes, 2020). Gradually, scientific journals have adapted their author's notes to recommendations for sharing research data and require authors - more or less concretely - whether and how research data should or must be shared (“Time to Recognize Authorship of Open Data,” 2022). One of many measures to overcome data transparency deficits is establishing high levels of data transparency, in line with the Guidelines for Promoting Transparency and Openness (TOP) (Center for Open Science, 2020; Haroz, 2018; Nosek et al., 2015). The TOP guidelines provide a template for

improving transparency in research published in scientific journals. Similarly, data transparency levels allow the classification of the journals' data policies into multiple categories with ascending levels of strictness; Level 0 corresponds to a non-implementation (i.e. if a journal just recommends data sharing but has not implemented a data policy yet); (1) an article must include a link to the research data; (2) data must be posted to a trusted repository and exceptions must be explicitly stated; and (3) data must be posted to a trusted repository and reported analyses will be reproduced independently prior to publication (see also Table 1). When the journal *Cognition* implemented its open data policy (DATA TRANSPARENCY LEVEL: 2) in March 2015, the proportion of articles that included a data availability statement skyrocketed from 25% to 78% (Hardwicke et al., 2018). We therefore included Cognition as a baseline. A measure that distinguishes the authors of an article directly and for all to see is awarding badges (e.g., "open data") to scientific articles that include a link to a data repository. When the Journal *Psychological Science* introduced the open science badges, research data availability significantly improved (Kidwell et al., 2016).

On a systemic level, research institutions (such as universities or non-university research institutions) have implemented research data policies, and funding agencies have formulated requirements for handling research data (European Research Council, 2019; German Research Foundation, 2022).

The costs and benefits of available research data

It is important to note that making research data available is related to costs (Perry & Netscher, 2022). Data curation costs and implementing transparency standards vary depending on - among others - disciplines, study design, data complexity, and the personal information included in the data (Hensel, 2021). Thus, these costs should be estimated in the best possible way in advance when planning a research project. Measuring costs in research data management is complex, and there are only approaches to it in research so far. Measuring

opportunity costs in the sense of the missed opportunity respective chance (e.g., investing in comprehensibility in argumentation instead) is much more difficult because the alternatives are not well known and hardly considered. Furthermore, the determination of the optimum in research data availability, characterized by the maximization of the difference between the benefits and the costs (the balance of marginal benefits and marginal costs from a microeconomic perspective) of research data-related measures in this respect, would be promising, but is still pending.

Despite all this, the benefits of making research data available are manifold and range from replicability of scientific findings (Camerer et al., 2018) to increased citations rates (Piwowar et al., 2007; Piwowar & Vision, 2013). Articles that include statements that link to data in a repository have an up to 25.36% (± 1.07%) higher citation impact on average (Colavizza et al., 2020).

Study overview and research questions

We report a study in which we analyzed the availability of research data in articles published in six empirical journals covering topics in education research (i.e. educational psychology, learning, and education) in the years 2018 and 2020. As a baseline, we analyzed the research data availability in the journal *Cognition* (Hardwicke et al., 2018).

The general awareness of the importance of research data has substantially increased (Kidwell et al., 2016). The FAIR data concept was published in 2016 (Wilkinson et al., 2016). Because it takes time to adopt new concepts in research, we chose 2018 as start point. We thus ask as **Research question 1**, *is the availability of research data increasing between articles published in 2018 and 2020?*

The journal's policy regarding the handling of research data is the basis for preparing and handling the submission and, thus, should be instructive for both the authors and the editor. The journal Cognition adopted an open data policy in March 2015. Since then, the availability of research data has risen significantly, and almost all articles now include a data availability statement (Hardwicke et al., 2018). We thus expect, that data availability increases with increasing data transparency level of the journal also in the selected educational journals. Thus, **Research question 2** is, *do the data transparency levels of the scientific journals impact the availability of research data?*

Similarly, the fact that a university or research institution has officially adopted a research data policy should increase the researchers' awareness. Therefore, we assume that the availability of research data is higher for articles whose corresponding author is from an institution that has implemented a research data policy. Consequently, **Research question 3** is, *do the research data policies of the corresponding author's institution impact the availability of research data?*

Work published in the selected educational psychology journals might differ from the work published in the journal Cognition. This might explain observed differences concerning research data availability between the two research fields. For example, because educational psychology research also includes large panel studies such as the PISA (Programme for International Student Assessment) studies (e.g., Brunner et al., (2007), secondary data analyses might be more prevalent in the educational sciences than in work published in the journal Cognition. We thus analyze if research data availability is different for secondary data analyses published in the educational psychology journals and the journal Cognition. We therefore formulate **Research question 4** as, *is the work reported in the Educational*

psychology journals substantially different from the work reported in the journal Cognition in terms of secondary data analysis?

Method

We report how we determined our sample, all data exclusions, all measures, research questions, and analytic plans in the study.

Sample

The initial screening began with analyzing the publication lists (e.g., annual reports) of renowned German Leibniz institutes doing research in this field - the Leibniz Institute for Research and Information in Education (DIPF), the German Institute for Adult Education - Leibniz Centre for Lifelong Learning (DIE), the Leibniz Institute for Science and Mathematics Education (IPN), and the Leibniz-Institut für Wissensmedien (IWM). This resulted in a list of 15 journals. We analyzed the data transparency levels (see below and Table 2; in parentheses): British Journal of Educational Psychology (data transparency level in 2018: 2), Journal of Educational Psychology (data transparency level in 2018: 0), Journal of Computer Assisted Learning (data transparency level in 2018: 1), Instructional Science (data transparency level in 2018: 1), Zeitschrift für Weiterbildungsforschung = Journal for Research on Adult Education (data transparency level in 2018: 0), Zeitschrift für Pädagogische Psychologie = German journal of educational psychology (data transparency level in 2018: 0), Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie (data transparency level in 2018: 0), Zeitschrift für Erziehungswissenschaft (data transparency level in 2018: 0), Learning and Instruction (data transparency level in 2018: 0), Learning and Individual Differences (data transparency level in 2018: 0), Computers in Human Behavior (data transparency level in 2018: 0), Computers and Education (data transparency level in 2018: 0), Metacognition and Learning (data transparency level in 2018: 0), Cognition and

Instruction (data transparency level in 2018: 0), Applied Cognitive Psychology (data transparency level in 2018: 0).

We started with the idea to include one journal per data transparency level in the analysis sample. However, as there was no journal with data transparency level 3, we changed sampling considering both the 2018 Social Sciences Citation Index (SSCI) and data transparency level. We thus firstly included the highly influential journals Journal of Educational Psychology (data transparency level in 2018: 0; Journal Impact Factor rank 3/59 in the category Psychology, Educational) and Learning and Instruction (data transparency level in 2018: 0; Journal Impact Factor rank 5/59 in the category Psychology, Educational), and the German open access journal Zeitschrift für Weiterbildungsforschung = Journal for Research on Adult Education (which is currently still published predominantly in German; data transparency level in 2018: 0; not yet listed in SSCI). Next, we selected the only journals with data transparency level 1 (Journal of Computer Assisted Learning and Instructional Science; Journal Impact Factor rank 41/243 in the category Education & Educational Research). Finally, we included the only journal with data transparency level 2 (British Journal of Educational Psychology; Journal Impact Factor rank 14/59 in the category Psychology, Educational). As a baseline, we also analyzed articles published in the journal Cognition (Cognition, data transparency level in 2018: 2; Journal Impact Factor rank 11/88 in the category Psychology, Experimental) (Hardwicke et al., 2018). Note that no journal has had a change in data transparency level between 2018 and 2020.

Measures

Data transparency level. We analyzed the journals' editorial policies (i.e., author notes or instructions for authors). In particular, we applied the data transparency levels as suggested by Nosek et al. (2015) and Haroz (2018). Data transparency levels 0 was assigned if the policy just encourages data sharing or says nothing. As an example, *Learning and Instruction*'s

policy states “This journal encourages and enables you to share data that supports your research publication where appropriate, and enables you to interlink the data with your published articles.” (*Guide for Authors - Learning and Instruction*, 2021). Data transparency level 1 was assigned if the policy states whether data are available, if so, and where to access them. As an example, the *Journal of Computer Assisted Learning*’s policy states “The Journal of Computer Assisted Learning expects that anonymized data supporting the results reported in the paper will be archived in an appropriate public repository. Exceptions may be granted for sensitive information such as data which cannot be properly anonymized, at the discretion of the Editors. Authors will be able to complete a data accessibility statement which will be published with their paper.” (*Guide for Authors - Journal of Computer Assisted Learning*, 2021). Data transparency level 2 was assigned if the policy states that the data must be posted to a trusted repository and that exceptions must be identified. As an example, the *British Journal of Educational Psychology*’s policy states “The journal expects that where possible all data supporting the results in papers published are archived in an appropriate public archive offering open access and guaranteed preservation.” (*Guide for Authors - British Journal of Educational Psychology*, 2021). Finally, data transparency level 3 would have been assigned if a journal required the data to be posted to a trusted repository and reported analyses will be reproduced independently prior to publication. We did not identify a journal with this data transparency level. We discussed the contents of the research data policies and identified its data transparency level accordingly. We did not calculate the interrater reliability.

Research data policies. After identifying the corresponding author's institution website URL (e.g., www.die-bonn.de) we used the following search string on Google: "research data policy site:institution" (e.g., "research data policy site:www.die-bonn.de") and downloaded the research data policy, if available. Note, we only considered official policies or guidelines. In

particular, we did not consider statements concerning best practices or tips on research data management published on the institutions' websites.

To ensure that the policy was relevant for a specific article, we checked the release dates of each policy. If an article was published before the release date of the corresponding institution's research data policy, it was marked as published without policy. If there was more than one version of a research data policy (i.e. an updated version was published after the initial one), we always considered the first date. For five institutions, we could not retrieve the publication dates of their research data policies and inquired with the relevant offices. After five weeks, only two institutions responded to our inquiries, and we coded the availability of their research data policy accordingly. For the remaining institutions, availability was coded as "NA" (6 articles in total).

Please note that changes to Google's algorithm and/or changes at the institutional webpages might alter the discovery of the policies. This is beyond our control.

N all. The total number of articles published in a journal per year.

N emp. The total number of articles reporting original research such as experiments, field studies, reanalyses, meta-analyses, or a combination thereof. Non-empirical articles are – among others – editorials, corrigenda, review articles, opinion pieces, or theoretical articles not reporting any data.

N data. The total number of empirical articles (*N emp*), which mention a link to the research data.

N data available. The total number of empirical articles (*N emp*), which mention a link to the research data *and* the provided link is, eventually leading to the data (i.e. the link is valid).

Please note that we only checked the validity but not the persistence of the link. This means that we cannot exclude the possibility that a link loses its validity and thus the data are no longer accessible. Further, we also did not check the data's interoperability (i.e. the data's ability to integrate with other data and/or tools) in the sense of the FAIR data concept (Wilkinson et al., 2016) .

Prop data available. The number of articles with shared data relative to the number of empirical articles (*N data available/N emp*).

Research questions and analytic plans

Research questions 1 to 3 and its related analytic plans were determined before data collection and data analysis. The fourth research question and its analytic plan was developed during data analysis.

Data analysis

Overview. We coded $N = 1242$ scientific articles (including one retracted article, which we excluded before statistical analysis), of which 1167 (*N emp*) were empirical. In 535 of those 1167 empirical articles (45.84%), we found a link to a data repository, which was valid in 349 articles (29.99%). See also Figure 1 and Table 2.

Statistical tests. The analysis of the research questions is based on the articles published in the educational psychology journals (BJEP, JCAL, JEP, JLI, and ZfW). We fitted generalized linear models (GLM) with research data availability as the dependent measure (yes, no; binomial).

Results

Research question 1: *Is the availability of research data increasing between articles published in 2018 and 2020?*

The fitted GLM included the year (2018, 2020) as a fixed effect by controlling for the journal (results see Table 3). Confirming research question 1, data availability increased significantly from 0.32% (*N* = 1 of 314 articles) in 2018 to 7.16% (*N* = 24 of 335 articles) in 2020, OR = 28.33 (95% CI: 5.843, 510.72), *p* = .001. Thus, the odds of reporting research data is about 28 higher in 2020 than in 2018.

Research question 2: *Do the data transparency levels of the scientific journals impact the availability of research data?*

The fitted GLM included the data transparency levels (0, 1, 2) as a continuous fixed effect by controlling for the year (results see Table 4). Data sharing with transparency level 0 was 3.08% (*N* = 10 of 325 articles), 4.09% (*N* = 9 of 220 articles) with data transparency level 1, and 5.77% (*N* = 6 of 104 articles) with data transparency level 2. Not supporting research question 2, data transparency levels on the journal level was not associated with research data availability, OR = 1.25 (95% CI: 0.75, 2.05), *p* = .388.

Research question 3: *Do the research data policies of the corresponding author's institution impact the availability of research data?*

The fitted GLM included the information on whether the corresponding author's institution adopted a research data policy (yes, no) as a fixed effect by controlling for the year. Whether an institution official research data policy did not influence research data availability. Data sharing in institutions without implemented research data policy (*N* = 16 of 394 articles, 4.06%) was not different from data sharing in institutions with implemented research data policy (*N* = 9 of 255 articles, 3.53%), OR = 0.68 (95% CI: 0.28, 1.55), *p* = .372, thus, not supporting Research question 3 (results see Table 5 and Figure 2).

Research question 4: *Is the work reported in the Educational psychology journals substantially different from the work reported in the journal Cognition in terms of secondary data analysis?*

As a control, we explored the differences in data availability between the work published in the journal Cognition and the educational journals regarding secondary data analysis. One could argue that educational articles use secondary data more often than work published in the journal Cognition and that this might affect research data availability, consequently. Although there is a significantly higher proportion of secondary data analyses reported in the educational psychology journals (N = 120 of 649 articles, 18.49%) than in the journal Cognition (N = 23 of 518 articles, 4.44%), $X^2(1, N = 1167) = 51.59$, $p < .001$ (Yates' continuity correction), data availability for those articles is significantly higher in Cognition, (N = 14 of 23 articles, 60.87%), than in the educational psychology journals, (N = 6 of 120 articles, 5.00%), $X^2(1, N = 143) = 45.54$, $p < .001$ (see Figure 3). We can only speculate about the reasons. Access to large panel data (e.g., PISA) is highly restricted, which might be a reason for lower data sharing in educational psychology journals. However, this does not prevent adding a link to the metadata. In summary, we conclude that both studies reporting primary and secondary analyses can be reported such that they provide a (valid) link to the underlying research data.

Discussion

This study examined how the availability of research data in educational psychology has changed over time as a function of institutional context. Research data availability was generally low (3.85% compared to 62.74% in the journal Cognition) but has increased substantially from 0.32% in 2018 to 7.16% in 2020. Interestingly, the proportion with an invalid link is much lower for the journal Cognition (N = 129 of 454 articles reporting an

invalid link to research data, 28.41%) than for the Educational psychology journals (N = 56 of 81 articles reporting an invalid link to research data, 69.14%). In relation to all empirical articles in the examined Educational psychology journals, this means that more than 8% of the data (56 of 649 empirical articles) are lost at this point. A simple solution to increase data availability would be to check the validity of the included link to research data as part of the editorial process.

We did not observe an influence of the editorial policy (as operationalized via the journals' data transparency levels) nor the availability of a data management policy on the level of the corresponding author's institution. We controlled for type of analysis (primary vs. secondary data analysis) that might be different for the journal types (Cognition vs. Educational psychology) and thus explain the differences. Yet, the data showed that low data availability did not depend on the analysis type. In the following, we discuss these results in the context of current considerations in open science and make suggestions on how to implement changes at the individual as well as the institutional and journal level to improve research data availability.

The role of the author guidelines and the editorial actions

We analyzed the guide for authors of the evaluated journals and found that the data transparency level is not predictive for actual data availability. Whereas both the author guidelines of the journal Cognition and the British Journal of Educational Psychology (BJEP) qualify for the data transparency level 2, actual research data availability differs dramatically, Cognition: 62.74%, BJEP: 5.77%. Even when considering that implementing a research data policy takes time (note that BJEP's data transparency level in 2018 was already 2) and hence considering only the year 2020, this difference is still significant, Cognition: 58.05%, BJEP: 8.57%. However, the data transparency level 2 allows exemptions from mandatory data sharing, which must be made explicit at article submission. It might thus be, that there are just

more exemptions granted in BJEP than in the journal Cognition. Since we consider it unlikely that this large difference is due to these exemptions alone, we conclude that the contents of the author guidelines need to be implemented in various editorial processes. Beginning with the initial submission, the authors should be asked whether they have read and understood the recommendation regarding handling research data. At best, they should be guided on where to deposit the research data. Note, however, that such an approach could lead to conflicts between the publisher (with economic interests) and the idea that publicly financed research data referenced in scientific articles, should be publicly available and thus hosted in public data centers. We elaborate on the related FAIR data concept (Wilkinson et al., 2016) in the paragraph *Towards an integrated data management system* below. In the next step, the editors and reviewers should check research data availability by default and point the authors' attention to this aspect if necessary.

The role of institutional research data management policies

A surprising finding of our analysis is that research data availability is not different for universities and research institutes that have implemented a research data policy and those that have not. We can only speculate about the reasons. First, the authors might not be aware of such a policy and, hence, do not know how research data should be handled. Second, although researchers know how to manage research data according to the institutional research data policy, they balk at the effort of storing them in a repository because the curation of research data is related to significant costs (Perry & Netscher, 2022). This is a reasonable strategy in case the non-adherence to such guidelines is not sanctioned. Such sanctions could include funding restrictions, for example. However, we do not believe that sanctions are constructive and practicable. Instead, we think that it is more purposeful to incentive researchers (Mellor, 2021). For that, it is essential to get an overview of research data practices. Thus, the research institutions should systematically monitor and document

research data output and outcome (such as usage, citations) similar to scientific publications. Researchers could add this information to their academic CV or webpage. Alternatively, awarding useful and FAIR data sets on the level of universities, research institutes, or research foundations could offer a way to boost data sharing (van der Zee & Reich, 2018).

Towards an integrated data management system

In the following, we identify four dimensions describing different but interrelated aspects to increase research data availability: (1) the researchers, (2) the research institutions, (3) the scientific journals (including editors and reviewers), and (4) how technical solutions could support increasing research data availability.

First, we assume that the most significant potential in the endeavor to increase the availability of research data lies with the researchers. Therefore, we consider it essential to improve data literacy (Ridsdale et al., 2015) and educate the researchers about the legal regulations and the benefits of shared research data. Such teaching units should include the requirements of the (national and international) funders, the institutional research data policies, and the guidelines for safeguarding good research practice (German Research Foundation, 2019; Science Europe, 2021). Further, such units should also teach the benefits of sharing research data. For example, scientific articles including statements linking to data in a repository have an up to 25.36% higher citation impact on average (Colavizza et al., 2020). More research is needed to study both the reasons and motives for (not) sharing research data on the levels of the individual researchers (Linek et al., 2017), possible interventions, and their potential effectiveness. The results of such studies are of high relevance to further developing the academic incentive system.

Second, we believe that institutional data management professionalization is key to optimal data curation and increasing data availability (Hardwicke et al., 2018; Vines et al., 2013). As requirements in data management have increased over the years (e.g., FAIR

criteria, Wilkinson, 2016), professional research data managers or data stewards should be an integral part of every research project. The tasks of such research data managers should include supporting the preparation of a research data management plan (including data descriptors). Recently, based on an idea from Science Europe (Science Europe, 2018, 2021), discipline-specific, standardized data management plans are being developed (*Domain Data Protocols for Educational Research*, 2022) and – at least in Germany – a transfer to other disciplines is planned via the connection to National Research Data Infrastructure Germany (NFDI), which is organized in a discipline-specific way (e.g., Consortium for the Social, Behavioural, Educational, and Economic Sciences – KonsortSWD). Note, the NFDI is the German counterpart to the European Open Science Cloud (EOSC). Further, research data managers should also assist in handling the data privacy policies (in case human subjects are involved) and the final deposition of the research data on a trusted repository. This last step should consider the FAIR (findable, accessible, interoperable, reusable) data criteria (Wilkinson et al., 2016). In particular, FAIR data must include metadata (i.e. data descriptors) that are complete, of high quality, and machine-actionable.

The FAIR data criteria advance the Open Data Concept, which traces back to 2006. According to the Open Knowledge Foundation's Open Data Handbook's definition (*Open Data*, 2022), "Data is open if it can be freely accessed, used, modified and shared by anyone for any purpose - subject only, at most, to requirements to provide attribution and/or share-alike." This comprises legal and technical openness. Over time, a similar but different data concept/typology became established, aiming at other/new purposes: FAIR data. The starting point for developing the FAIR principles in 2014 was joint academic and private stakeholders' interest in getting over data discovery and reuse obstacles. Subsequently, the Lorentz workshop in Leiden, Netherlands, elaborated four basic principles and 15 sub-principles, which later became known under the acronym FAIR (*The FAIR Data Principles - FORCE11*, 2014) representing findability, accessibility, interoperability, and reusability of

research (meta-)data for humans and machines. The principles were refined and improved by the FORCE 11 community members and finally published by Wilkinson et al. (2016). The FAIR principles are neither a new standard nor a requirement but provide guidelines for the sustainable reusability of research objects (Mons et al., 2020). Open data can but need not be FAIR data and vice versa. For example, a good portion of data in the educational sciences is sensitive (e. g., disclosive or confidential) and, therefore, despite public funding, often not available open access (Betancort Cabrera et al., 2020). Focusing the "A" in FAIR, the maxim "as open as possible, as closed as necessary" in accordance with the Open Research Data Pilot applies (European Commission Directorate-General for Research & Innovation, 2016). Accessibility counts under well-defined conditions. As Mons et al. (2017) clearly state, "The FAIR principles, although inspired by Open Science, explicitly and deliberately do not address moral and ethical issues pertaining to the openness of data. In the envisioned Internet of FAIR Data and Services, the degree to which any piece of data is available, or even advertised as being available (via its metadata) is entirely at the discretion of the data owner. FAIR only speaks to the need to describe a process – mechanised or manual – for accessing discovered data [...] None of these principles necessitate data being 'open' or 'free'." In a nutshell, compared to open data, the FAIR data concept is more adapted to special needs in the research cycle. We suggest that the FAIR data criteria guide institutional research data management and policies including data handling and final deposition on a trusted repository (Wilkinson et al., 2016).

We further suggest expanding institutional support for research data management. Professionalization in research data management could also address the problem of suboptimal data curation (Hardwicke et al., 2018). Over the past decade, a plethora of new job titles have emerged (Tammaro et al., 2019), whose job profile, roles, and required competencies differ on the national and the international level. Given the longstanding lack of common terminology, positions and skills still need to be more strongly elaborated and valued

at the institutional level. Among others, research data managers, data scientists, data librarians, data curators, and data stewards at the institutional level, for an example, can support researchers in data handling along the research data cycle (e.g., metadata creating, dealing with legal issues such as licensing, and data ingest).

Third, we believe that scientific journals have an essential role in increasing the availability of research data. This can be abstractly divided into editorial tasks and duties as well as infrastructural components (such as labeling scientific articles with published data). Most importantly, the chief editors should adopt a research data policy. We suggest at least data transparency level 2 (data must be posted to a trusted repository, exceptions must be identified at article submission). In the editorial process, all persons involved (i.e. editors, reviewers) must be aware of this policy and incorporate them into the overall decision-making process. During the review and revision process, attention should be paid to the availability of research data, and, if not available, appropriate advice should be given to the authors. This could also reduce the proportion of articles reporting invalid links to research data, which was a lot higher for the Educational psychology journals than for the journal Cognition. Regarding the infrastructural component, the journal editors should consider labeling the adherence to open science practices like sharing research data. One such method is to mark a published article with "open data" or to assign so-called "open science badges" (Blohowiak et al., 2013). The latter have been considerably successful in boosting the availability of research data from only 3% to 23% in a short time in the journal *Psychological Science* (Kidwell et al., 2016).

Fourth, we believe that low-threshold technical solutions for research data management should support the whole research process. Currently, there are some promising approaches, such as the Research Data Management Organiser – RDMO (*RDMO Research Data Management Organiser*, 2022) or ZPID's DataWiz (*DataWiz*, 2022). While RDMO is customizable to discipline-specific and institutional needs (*Domain Data Protocols for Educational Research*, 2022) DataWiz is already tailored to the needs of psychology. Both

support the researchers during a research project. Thus, all data-related concerns are taken into account. Such technical solutions should be part of the researchers' training and incorporated into the research projects.

Limitations

The present study is based on publicly available data published in scientific articles. Yet, research data availability might also be influenced by background factors such as the journals' submission process structure, reviewer comments such as the Peer Reviewers' Openness Initiative (Morey et al., 2016), or individual editorial actions. All these factors are beyond our analysis. In particular, we did not analyze the structure and the affordances of the journals' submission processes. The underlying technical systems and the implemented process have the power to influence research data availability. For example, the inclusion of a question if research data have been made available according to the guide for authors might be a first step in this process.

A further limitation is the selection of the analyzed journals. Prescreening criteria were the relevance to internationally renowned research institutes of the Leibniz Association in educational sciences and educational psychology. The final selection criterium was data transparency level (at least one journal from each level; Note that we did not find a journal with data transparency level 3). Although – in our opinion – the prescreening and the selection reflect the field, and the results are thus of high relevance, we do not exclude the possibility that the results might be different for other journals in the field.

This study analyzed data availability as a function of research data policies on the journals' and the corresponding author's affiliation level. We did not analyze recommendations regarding data sharing of further stakeholders, such as funding agencies and professional societies. As data sharing is at a very low level, we consider the influence of

recommendations and regulations from those stakeholders at best to be minimal. Further research is needed to study those influences on data sharing activities.

We identified the data transparency level while reading the journal's guide for authors as a group. Although we believe that the categories are clear, it is important to mention that we did not have an independent rating for the data transparency level.

Finally, this analysis only covered two years, namely 2018 and 2020. While we observed increased data sharing in 2020 than in 2018, it would be interesting to monitor the impact of research data policies on data sharing behavior over a more extended period. We consider the present analysis as a starting point. The research data (including analysis scripts) are freely accessible and can be easily updated in the future.

Conclusion and outlook

As part of transparent and open science, data sharing serves as a scientific accelerator and contributes significantly to scientific progress. It is thus in the tradition of scientific paradigms as formulated by Popper (1959, 1963) and Merton (1973) as leading representatives of the philosophy and sociology of science. Yet, data sharing in educational research is low and – as the present study shows – neither influenced by the institutional research data policies nor the journals' guidelines. We outlined an idea for developing an integrated and comprehensive data management system, which not only focuses on the individual researchers but also involves the various stakeholders (such as research infrastructure institutions). In summary, our approach complements the idea of *open education science* (van der Zee & Reich, 2018; van Dijk et al., 2021) by adding an infrastructural emphasis. Since *education research* is particularly characterized by its diversity of research methods, high sensitivity of data, and high variety of data types, the sharing principles are highly important and, thus, must be based on a solid foundation. In addition to providing mere data storage, research infrastructure institutions such as research

data centers and research libraries play an essential role in assisting, guiding, and teaching the researchers in the complex processes that finally lead to the successful sharing of FAIR data.

In compliance with van der Zee and Reich (2018) and Mellor (2021), we consider incentives for data sharing as key to boosting sharing rates. Behavioral and meta-scientific research is needed to find the best and most promising ways. The possibilities include mandatory research data sharing to receive a full peer review from authors following the Peer Reviewers' Openness initiative (Morey et al., 2016) and compulsory uploading the research data in the submission system. In our opinion, however, researchers themselves should realize the benefits of data sharing and act on their own accordingly. Thus, we propose teaching the researchers about legal aspects (e.g., licensing, founder requirements), technical solutions (e.g., RDMO), improvement of efficiency of the individual researcher and scientific discovery, protection against data loss, funding opportunities (Klein et al., 2018), and potential citation benefits (Colavizza et al., 2020). Finally, making shared data sets an official part of the academic CV would enable hiring committees to consider data sharing activities as part of transparent and open science.

We started this research with the expectation to observe an effect of institutional policies on research data sharing. Yet, although requirements are met (at least from the perspective of the research infrastructure), data sharing in educational psychology is low and not affected by policies and guidelines. In our view, only a change within the scientific system that recognizes research data as a crucial part of the scientific process can lead to a substantial increase in the share of shared research data. Societal relevant fields such as educational psychology should have a particular interest in it.

Data availability statement

Data and analyses scripts (for the statistical programming language R) have been made publicly available via the Open Science Framework (OSF) and can be accessed at

https://osf.io/6mw7a/?view_only=37365f5d55084eb0ba9ac85ec193b061 *[NOTE: This is an anonymous view-only link for the review process. We will make the project public once the manuscript is accepted for publication.]* After the final publication of this article, we will also publish the research data via the German Network of Educational Research Data (VerbundFDB: https://www.forschungsdaten-bildung.de/en/studies/search).

Author Contributions

Research idea: MH, ECB; Data collection: CG and LK under the supervision of MH; Data analysis and curation: MH; Writing: MH, ECB

References

Al-Shahi Salman, R., Beller, E., Kagan, J., Hemminki, E., Phillips, R. S., Savulescu, J., Macleod, M., Wisely, J., & Chalmers, I. (2014). Increasing value and reducing waste in biomedical research regulation and management. *Lancet*, *383*(9912), 176–185. https://doi.org/10.1016/S0140-6736(13)62297-7

American Psychological Association. (2001). Ethical standards for the reporting and publishing of scientific information. In *Publication manual of the American Psychological Association* (5th ed., pp. 387–396). APA.

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., Fiedler, S., Funder, D. C., Kliegl, R., Nosek, B. A., Perugini, M., Roberts, B. W., Schmitt, M., van Aken, M. A. G., Weber, H., & Wicherts, J. M. (2013). Recommendations for Increasing Replicability in Psychology. *European Journal of Personality*, *27*(2), 108–119. https://doi.org/10.1002/per.1919

Betancort Cabrera, N., Bongartz, E. C., Dörrenbächer, N., Goebel, J., Kaluza, H., & Siegers, P. (2020). White Paper on implementing the FAIR principles for data in the Social, Behavioural, and Economic Sciences. *RatSWD Working Paper Series*. https://doi.org/10.17620/02671.60

Blohowiak, B. B., Cohoon, J., de-Wit, L., Eich, E., Farach, F. J., Hasselman, F., Holcombe, A. O., Humphreys, M., Lewis, M., & Nosek, B. A. (2013, February 19). *Badges to Acknowledge Open Practices*. OSF. https://osf.io/tvyxz/

Boccali, T., Sølsnes, A. E., Thorley, M., Winkler-Nees, S., & Timmermann, M. (2021). *Practical Guide to Sustainable Research Data*. https://doi.org/10.5281/zenodo.4769703

Bond-Lamberty, B. (2016). Data sharing and scientific impact in eddy covariance research. *Reviews of Geophysics*, 1440–1443. https://doi.org/10.1002/2018JG004502

Brunner, M., Artelt, C., Krauss, S., & Baumert, J. (2007). Coaching for the PISA test. *Learning and Instruction*, *17*(2), 111–122. https://doi.org/10.1016/j.learninstruc.2007.01.002

Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeiffer, T., Razen, M., & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*, *351*(6280), 1433–1436. https://doi.org/10.1126/science.aaf0918

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B. A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behaviour*, 1. https://doi.org/10.1038/s41562-018-0399-z

Center for Open Science. (2020). *Transparency and Openness Promotion (TOP) Guidelines*. OSF. https://osf.io/9f6gx/

Chalmers, I., Bracken, M. B., Djulbegovic, B., Garattini, S., Grant, J., Gülmezoglu, A. M., Howells, D. W., Ioannidis, J. P. A., & Oliver, S. (2014). How to increase value and reduce waste when research priorities are set. *The Lancet*, *383*(9912), 156–165. https://doi.org/10.1016/S0140-6736(13)62229-1

Chan, A.-W., Song, F., Vickers, A., Jefferson, T., Dickersin, K., Gøtzsche, P. C., Krumholz, H. M., Ghersi, D., & Worp, H. B. van der. (2014). Increasing value and reducing waste: Addressing inaccessible research. *The Lancet*, *383*(9913), 257–266. https://doi.org/10.1016/S0140-6736(13)62296-5

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, *15*(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

DataWiz. (2022). DataWiz. https://datawiz.leibniz-psychology.org/DataWiz/?datawiz_locale=en

Domain data protocols for educational research. (2022). https://ddp-bildung.org/

Errington, T. M., Mathur, M., Soderberg, C. K., Denis, A., Perfito, N., Iorns, E., & Nosek, B. A. (2021). Investigating the replicability of preclinical cancer biology. *ELife*, *10*, e71601. https://doi.org/10.7554/eLife.71601

European Commission Directorate-General for Research & Innovation. (2016). *H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020*. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

European Research Council. (2019). *Open research data and data management plans—Information for ERC grantees*. European Commission. https://erc.europa.eu/sites/default/files/document/file/ERC_info_document-Open_Research_Data_and_Data_Management_Plans.pdf

Flake, J. K. (2021). Strengthening the foundation of educational psychology by integrating construct validation into open science reform. *Educational Psychologist*, *56*(2), 132–141. https://doi.org/10.1080/00461520.2021.1898962

Fleming, J. I., Wilson, S. E., Hart, S. A., Therrien, W. J., & Cook, B. G. (2021). Open accessibility in education research: Enhancing the credibility, equity, impact, and efficiency of research. *Educational Psychologist*, *56*(2), 110–121. https://doi.org/10.1080/00461520.2021.1897593

Gehlbach, H., & Robinson, C. D. (2021). From old school to open science: The implications of new research norms for educational psychology and beyond. *Educational Psychologist*, *56*(2), 79–89. https://doi.org/10.1080/00461520.2021.1898961

German Research Foundation. (2019). *Guidelines for safeguarding good research practice. Code of conduct*. https://doi.org/10.5281/zenodo.3923602

German Research Foundation. (2022). *Handling of Research Data*. Handling of Research Data - Information on the Resources Available. https://www.dfg.de/en/research_funding/principles_dfg_funding/research_data/resources_available/index.html

Gersten, R., & Hitchcock, J. (2009). What Is Credible Evidence in Education?: The Role of the What Works Clearinghouse in Informing the Process. In S. Donaldson, C. Christie, & M. Mark, *What counts as credible evidence in applied research and evaluation practice?* (pp. 78–95). SAGE Publications, Inc. https://doi.org/10.4135/9781412995634.d11

Glasziou, P., Altman, D. G., Bossuyt, P., Boutron, I., Clarke, M., Julious, S., Michie, S., Moher, D., & Wager, E. (2014). Reducing waste from incomplete or unusable reports of biomedical research. *The Lancet*, *383*(9913), 267–276. https://doi.org/10.1016/S0140-6736(13)62228-X

Guide for Authors—British Journal of Educational Psychology. (2021). https://bpspsychub.onlinelibrary.wiley.com/hub/journal/20448279/homepage/forauthors.html

Guide for Authors—Journal of Computer Assisted Learning. (2021). https://onlinelibrary.wiley.com/page/journal/13652729/homepage/forauthors.html

Guide for Authors—Learning and Instruction. (2021). https://www.elsevier.com/journals/learning-and-instruction/0959-4752/guide-for-authors

Hardwicke, T. E., Mathur, M. B., MacDonald, K., Nilsonne, G., Banks, G. C., Kidwell, M. C., Hofelich Mohr, A., Clayton, E., Yoon, E. J., Henry Tessler, M., Lenne, R. L., Altman, S., Long, B., & Frank, M. C. (2018). Data availability, reusability, and analytic reproducibility: Evaluating the impact of a mandatory open data policy at the journal Cognition. *Royal Society Open Science*, *5*(8), 180448. https://doi.org/10.1098/rsos.180448

Haroz, S. (2018). Open Practices in Visualization Research: Opinion Paper. *2018 IEEE Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, 46–52. https://doi.org/10.1109/BELIV.2018.8634427

Hensel, P. G. (2021). Dissecting the tension of open science standards implementation in management and organization journals. *Accountability in Research*, *0*(0), 1–26. https://doi.org/10.1080/08989621.2021.1981870

Ioannidis, J. P. A., Greenland, S., Hlatky, M. A., Khoury, M. J., Macleod, M. R., Moher, D., Schulz, K. F., & Tibshirani, R. (2014). Increasing value and reducing waste in research design, conduct, and analysis. *The Lancet*, *383*(9912), 166–175. https://doi.org/10.1016/S0140-6736(13)62227-8

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L.-S., Kennett, C., Slowik, A., Sonnleitner, C., Hess-Holden, C., Errington, T. M., Fiedler, S., & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. *PLOS Biology*, *14*(5), e1002456. https://doi.org/10.1371/journal.pbio.1002456

Klein, O., Hardwicke, T. E., Aust, F., Breuer, J., Danielsson, H., Mohr, A. H., IJzerman, H., Nilsonne, G., Vanpaemel, W., & Frank, M. C. (2018). A practical guide for transparency in psychological science. *Collabra: Psychology*, *4*(1), 20. https://doi.org/10.1525/collabra.158

Linek, S. B., Fecher, B., Friesike, S., & Hebing, M. (2017). Data sharing as social dilemma: Influence of the researcher's personality. *PLOS ONE*, *12*(8), e0183216. https://doi.org/10.1371/journal.pone.0183216

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

Mellor, D. (2021). Improving norms in research culture to incentivize transparency and rigor. *Educational Psychologist*, *56*(2), 122–131. https://doi.org/10.1080/00461520.2021.1902329

Merton, R. K. (1973). *The sociology of science: Theoretical and empirical investigations*. University of Chicago Press.

Mons, B., Neylon, C., Velterop, J., Dumontier, M., da Silva Santos, L. O. B., & Wilkinson, M. D. (2017). Cloudy, increasingly FAIR; revisiting the FAIR Data guiding principles for the European Open Science Cloud. *Information Services & Use*, *37*(1), 49–56. https://doi.org/10.3233/ISU-170824

Mons, B., Schultes, E., Liu, F., & Jacobsen, A. (2020). The FAIR Principles: First Generation Implementation Choices and Challenges. *Data Intelligence*, *2*(1–2), 1–9. https://doi.org/10.1162/dint_e_00023

Morey, R. D., Chambers, C. D., Etchells, P. J., Harris, C. R., Hoekstra, R., Lakens, D., Lewandowsky, S., Morey, C. C., Newman, D. P., Schönbrodt, F. D., Vanpaemel, W., Wagenmakers, E.-J., & Zwaan, R. A. (2016). The Peer Reviewers' Openness Initiative: Incentivizing open research practices through peer review. *Royal Society Open Science*, *3*(1), 150547. https://doi.org/10.1098/rsos.150547

Nosek, B. A., Alter, G., Banks, G. C., Borsboom, D., Bowman, S. D., Breckler, S. J., Buck, S., Chambers, C. D., Chin, G., Christensen, G., Contestabile, M., Dafoe, A., Eich, E., Freese, J., Glennerster, R., Goroff, D., Green, D. P., Hesse, B., Humphreys, M., …

Yarkoni, T. (2015). Promoting an open research culture. *Science*, *348*(6242), 1422–1425. https://doi.org/10.1126/science.aab2374

Open Data. (2022). https://opendatahandbook.org/glossary/en/terms/open-data/

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. https://doi.org/10.1126/science.aac4716

Patall, E. A. (2021). Implications of the open science era for educational psychology research syntheses. *Educational Psychologist*, *56*(2), 142–160. https://doi.org/10.1080/00461520.2021.1897009

Perry, A., & Netscher, S. (2022). Measuring the time spent on data curation. *Journal of Documentation*. https://doi.org/10.1108/JD-08-2021-0167

Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing Detailed Research Data Is Associated with Increased Citation Rate. *PLOS ONE*, *2*(3), e308. https://doi.org/10.1371/journal.pone.0000308

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. *PeerJ*, *1*, e175. https://doi.org/10.7717/peerj.175

Plucker, J. A., & Makel, M. C. (2021). Replication is important for educational psychology: Recent developments and key issues. *Educational Psychologist*, *56*(2), 90–100. https://doi.org/10.1080/00461520.2021.1895796

Popper, K. (1959). *The logic of scientific discovery.* Hutchinson.

Popper, K. (1963). *Conjectures and refutations: The growth of scientific knowledge*. Routledge & Kegan Paul.

RDMO Research Data Management Organiser. (2022). RDMO Research Data Management Organiser. https://rdmorganiser.github.io/en/

Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). *Strategies and Best Practices for Data*

Literacy Education Knowledge Synthesis Report. SSHRC. https://doi.org/10.13140/RG.2.1.1922.5044

Science Europe. (2018). *Guidance document presenting a framework for discipline-specific research data management*. https://doi.org/10.5281/zenodo.4925907

Science Europe. (2021). *Practical guide to the international alignment of research data management—Extended edition*. https://doi.org/10.5281/zenodo.4915862

Shaver, J. P., & Norton, R. S. (1980). Randomness and Replication in Ten Years of the American Educational Research Journal. *Educational Researcher*, *9*(1), 9–15. https://doi.org/10.3102/0013189X009001009

Tammaro, A. M., Matusiak, K. K., Sposito, F. A., & Casarosa, V. (2019). Data Curator's Roles and Responsibilities: An International Perspective. *Libri*, *69*(2), 89–104. https://doi.org/10.1515/libri-2018-0090

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., Pedaste, M., Raju, M., Astapova, A., Lukner, H., Kogermann, K., & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data*, *8*(1), 192. https://doi.org/10.1038/s41597-021-00981-0

The FAIR Data Principles—FORCE11. (2014). FORCE11 The Future of Research Communications and E-Scholarship. https://force11.org/info/the-fair-data-principles/

Time to recognize authorship of open data. (2022). *Nature*, *604*(7904), 8–8. https://doi.org/10.1038/d41586-022-00921-x

van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, *4*(3), 2332858418787466. https://doi.org/10.1177/2332858418787466

van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open Science in Education Sciences. *Journal of Learning Disabilities*, *54*(2), 139–152. https://doi.org/10.1177/0022219420945267

Velterop, J., & Schultes, E. (2020). An Academic Publishers' GO FAIR Implementation Network (APIN). *Information Services & Use*, *40*(4), 333–341. https://doi.org/10.3233/ISU-200102

Vines, T. H., Albert, A. Y., Andrew, R. L., Débarre, F., Bock, D. G., Franklin, M. T., Gilbert, K. J., Moore, J.-S., Renaut, S., & Rennison, D. J. (2014). The availability of research data declines rapidly with article age. *Current Biology*, *24*(1), 94–97.

Vines, T. H., Andrew, R. L., Bock, D. G., Franklin, M. T., Gilbert, K. J., Kane, N. C., Moore, J.-S., Moyers, B. T., Renaut, S., Rennison, D. J., Veen, T., & Yeaman, S. (2013). Mandated data archiving greatly improves access to research data. *The FASEB Journal*, *27*(4), 1304–1308. https://doi.org/10.1096/fj.12-218164

Weston, S. J., Ritchie, S. J., Rohrer, J. M., & Przybylski, A. K. (2019). Recommendations for increasing the transparency of analysis of preexisting data sets. *Advances in Methods and Practices in Psychological Science*, *2*(3), 214–227. https://doi.org/10.1177/2515245919848684

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*(11), e26828. https://doi.org/10.1371/journal.pone.0026828

Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*(7), 726–728. https://doi.org/10.1037/0003-066X.61.7.726

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Table 1
Data transparency levels.

Data Transparency Level	Description
0 (not implemented)	Journal encourages data sharing or says nothing.
1	The article states whether data are available, and if so, where to access them.
2	Data must be posted to a trusted repository. Exceptions must be identified at article submission.
3	Data must be posted to a trusted repository and reported analyses will be reproduced independently prior to publication.

Table 2

Detailed overview of the N = 1242 analyzed articles. Note that we removed one retracted article (doi: https://doi.org/10.1111/jcal.12443) before data analysis.

Journal	Year	Data Transparency Level	N all	N emp	N data	N data available	Prop data available
Cognition	2018	2	236	220	193	152	0.69
	2020	2	307	298	261	173	0.58
British Journal of Educational Psychology	2018	2	37	34	0	0	0.00
	2020	2	71	70	7	6	0.09
Instructional Science	2018	1	42	35	0	0	0.00
	2020	1	28	27	1	1	0.04
Journal of Computer Assisted Learning	2018	1	90	86	0	0	0.00
	2020	1	75	72	8	8	0.11
Journal of Educational Psychology	2018	0	71	67	0	0	0.00
	2020	0	100	98	4	2	0.02
Learning and Instruction	2018	0	88	84	30	1	0.01
	2020	0	56	53	30	6	0.11
Zeitschrift für Weiterbildungsforschung = Journal for Research on Adult Education	2018	0	13	8	0	0	0.00
	2020	0	27	15	1	1	0.07

N all. The total number of articles published in a journal per year; *N emp.* The total number of articles published reporting original, empirical work; *N data.* The total number of empirical articles (*N emp*), which explicitly share (meta-)data; *N data available.* The total number of empirical articles (*N emp*), which explicitly share (meta-)data *and* the provided link is valid; *Prop data available.* The number of articles with shared data relative to the number of empirical articles (*N data available/N emp*).

Table 3

Results of the GLM fitted to study Research question 1 ("Is the availability of research data increasing between articles published in 2018 and 2020?").

		Data_Available				
Predictors	*Odds Ratios*	*SE*	*CI*	*standardized CI*	*z*	*p*
(Intercept)	0.00	0.00	0.00 – 0.02	0.00 – 0.02	-5.23	**<0.001**
Year [2020]	28.33	28.33	5.84 – 510.72	5.84 – 510.72	3.25	**0.001**
Journal [Instructional Science]	0.40	0.40	0.02 – 2.49	0.02 – 2.49	-0.83	0.405
Journal [Journal of Computer Assisted Learning]	1.29	1.29	0.43 – 4.12	0.43 – 4.12	0.45	0.651
Journal [Journal of Educational Psychology]	0.22	0.22	0.03 – 0.99	0.03 – 0.99	-1.81	0.070
Journal [Learning and Instruction]	1.54	1.54	0.48 – 5.08	0.48 – 5.08	0.74	0.461
Journal [Zeitschrift für Weiterbildungsforschung = Journal for Research on Adult Education]	0.76	0.76	0.04 – 4.94	0.04 – 4.94	-0.24	0.807
Observations	649					

Table 4

Results of the GLM fitted to study Research question 2 ("Do the data transparency levels of the scientific journals impact the availability of research data?").

	Data_Available					
Predictors	*Odds Ratios*	*std. Beta*	*CI*	*standardized CI*	*z*	*p*
(Intercept)	0.00	0.00	0.00 – 0.01	0.00 – 0.01	-5.79	**<0.001**
Data Transparency Level	1.25	1.18	0.75 – 2.05	0.80 – 1.70	0.86	0.388
Year [2020]	23.52	23.52	4.92 – 422.08	4.92 – 422.08	3.08	**0.002**
Observations	649					

Table 5

Results of the GLM fitted to study Research question 3 ("Do the research data policies of the corresponding author's institution impact the availability of research data?").

	Data_Available					
Predictors	*Odds Ratios*	*std. Beta*	*CI*	*standardized CI*	*z*	*p*
(Intercept)	0.00	0.00	0.00 – 0.02	0.00 – 0.02	-5.59	**<0.001**
RD Policy Institution [1]	0.68	0.68	0.28 – 1.55	0.28 – 1.55	-0.89	0.372
Year [2020]	25.29	25.29	5.28 – 454.18	5.28 – 454.18	3.15	**0.002**
Observations	649					

Figure 1

Flow chart depicting the different stages of article selection. Of the initial 1242 articles, only 350 included a valid link to the corresponding research data. Note: "All articles" includes one retracted article. See also Table 2.

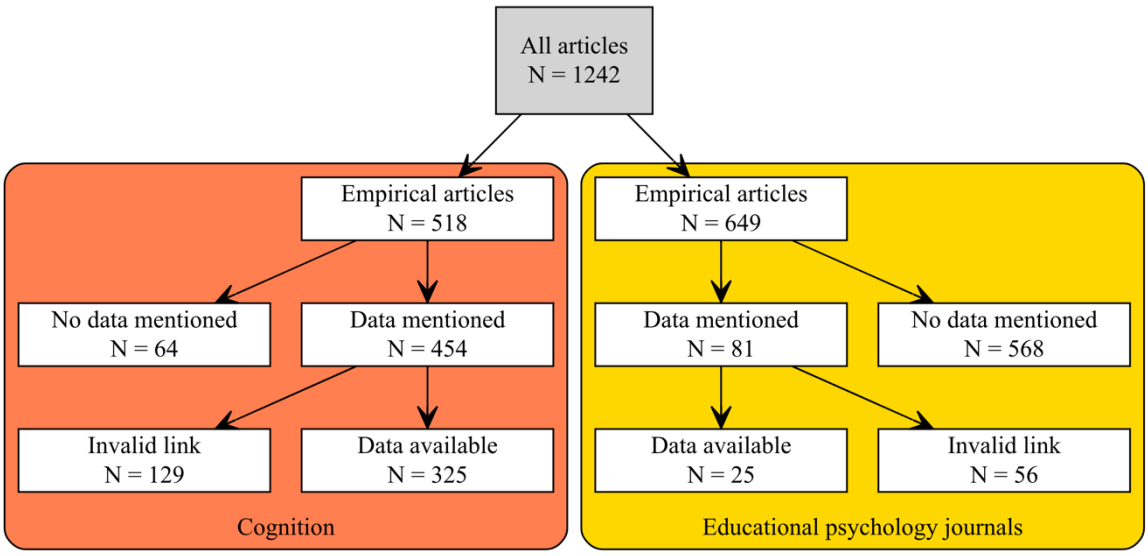

Figure 2

Flow chart depicting the availability of research data as a function of whether the corresponding author's institution has adopted an official research data policy or not for the years 2018 (left) and 2020 (right).

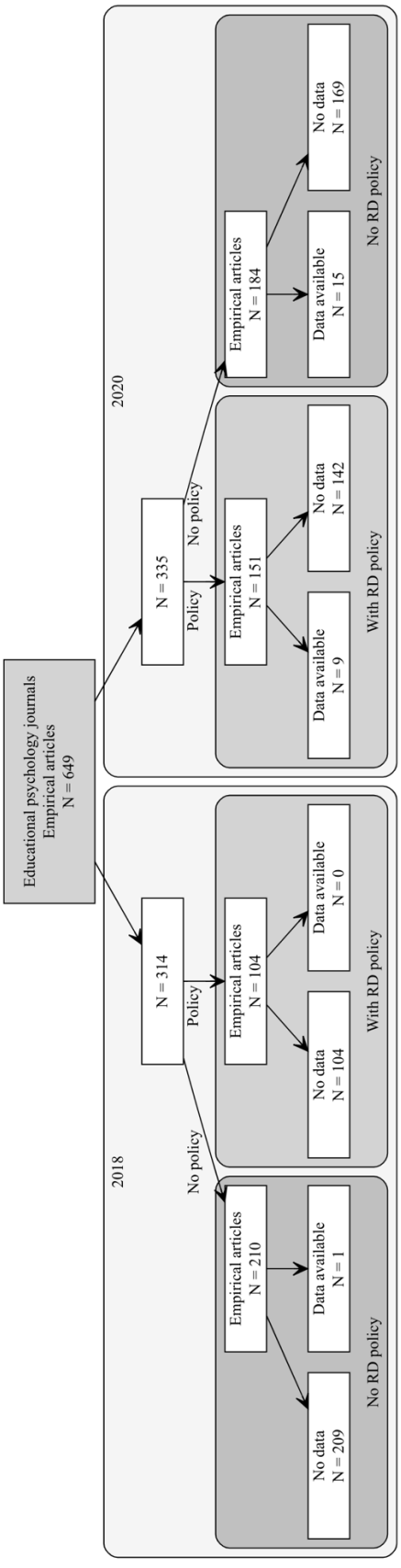
Educational psychology journals
Empirical articles
N = 649
2018
N = 314
Policy
No policy
With RD policy
Empirical articles
N = 104
Data available
N = 0
No data
N = 104
No RD policy
Empirical articles
N = 210
Data available
N = 1
No data
N = 209
2020
N = 335
Policy
No policy
With RD policy
Empirical articles
N = 151
Data available
N = 9
No data
N = 142
No RD policy
Empirical articles
N = 184
Data available
N = 15
No data
N = 169

Figure 3

Flow chart depicting the distribution of primary and secondary data analyses in the journal Cognition (orange) and the educational psychology journals (yellow).