Journal Of Open Research in Education 1 2026
© The Authors [CC-BY]
https://doi.org/10.58069/jore.2026.1.133

The Gap Between Open Access and Open Science in Educational Research. A Repository-Based Analysis of Data, Code, and Materials Sharing in Open-Access Papers

Markus Gebhardt
LMU München, ORCID 0000-0002-9122-0556
Nikola Ebenbeck
LMU München, ORCID 0000-0002-4167-981X
Jeremias Rieser
LMU München, ORCID 0009-0005-021

September 2026

Abstract

Open Science practices are increasingly promoted in educational research by funding agencies, scientific associations, and researchers. However, only a small proportion of published articles provide links to research data or other materials and fully adhere to Open Science principles. This article investigates the prevalence of repository links to OSF and GitHub in articles published in *Frontiers in Education* (Frontiers Media SA), *Education Sciences* (MDPI), *Cogent Education* (Taylor & Francis), *International Journal of Educational Technology in Higher Education* (IJETHE; Springer), *AERA Open* (AERA and Sage Publishing), *Sage Open* (Sage Publishing), and *Frontline Learning Research* (EARLI). Using the OpenAlex database, we identified and downloaded 15,953 open-access articles. Of these, only 223 contained a functional repository link that was related to the article, corresponding to 1.4% of the downloaded corpus. Among the articles with repository links, most employed a quantitative research design (75.3%), followed by mixed-methods designs (12.1%) and qualitative designs (4.9%). Review articles accounted for 3.1% of the sample, while conceptual articles, including didactic and instructional contributions, represented 2.7%. Methodological articles constituted the smallest category, comprising 1.8% of the included articles. We discuss these findings in relation to current Open Science practices in educational research.

Keywords: Open Science; educational research; data sharing; code sharing; research repositories; OSF; GitHub; open-access publications; reproducibility; transparency

1 Introduction

1.1 Open Science Practices and their Reasoning

Open science practices enable researchers to make the research process transparent and comprehensible to others. In educational research, findings are commonly disseminated through journal art-

icles or book chapters. While these formats remain central to scholarly communication, they are often insufficient for fully understanding empirical research, particularly when complex data structures, analytical procedures, instruments, software, or code are involved. This limitation is not only a matter of article length but also reflects the constraints of narrative publication formats. Additional research outputs, such as datasets, analysis scripts, software, instruments, and other study materials, are often necessary to understand how results were produced, to assess their robustness, and to build on previous work. In this sense, open science can be understood as an extension and deepening of established principles of good scientific practice (Nosek et al., 2022).

The central aim of open science is to make research and its results transparent, comprehensible, accessible, and reproducible (Cook et al., 2018). This broad understanding of openness applies to different stages and products of the research process. It includes the research question, methodology or analytical approach (*open methodology* or *open analysis*), data (*open data*), the published article (*open access*), the review process (*open peer review*), research products such as software or tools (*open source*), and educational materials (*open educational resources*). Only when these different components are made accessible can researchers and other stakeholders meaningfully understand, evaluate, and reuse research beyond the immediate context in which it was produced (Watson, 2015; van Dijk et al., 2021). In the 21st century, it should therefore be standard practice in educational research to examine, compare, and reuse the data, instruments, and products of colleagues in the field (Cook et al., 2018). Such practices are particularly relevant for inclusive and accessible research cultures, as they lower barriers to participation, verification, and cumulative knowledge production.

A further argument for open science concerns the public character of educational research. Educational research is frequently funded through public resources and is therefore, at least indirectly, supported by society. From this perspective, research funded by public money should exemplify the standards of open science and make as much information and material as possible accessible to the public. Funding agencies have increasingly incorporated open science requirements into their regulations, and open science practices are particularly visible in funded projects that are subject to such requirements (van Dijk et al., 2021). However, this also raises the question of whether and to what extent open science practices are implemented in educational research independently of funder mandates.

Limited transparency becomes a scientific problem when findings can only be verified, evaluated, or replicated to a restricted degree. In medicine and psychology, this problem has been discussed in terms of the *replication crisis*, as results obtained by one research group have often not been reproduced by others. In psychology, only 39% of replicated studies produced comparable results (Open Science Collaboration, 2015), while replication rates were 62% for laboratory experiments in economics (Camerer et al., 2016) and 12% in cancer research (Errington et al., 2021). Although comparable systematic evidence is still lacking for educational research, similar challenges can be assumed, particularly for empirical and intervention-based studies. Moreover, replicated findings are generally scarce in educational intervention research (Cock et al., 2018). In psychology, the lack of

open science practices has therefore been strongly criticized, and a broader cultural shift has been called for (Nosek et al., 2022). Universities and research institutes are expected to actively require and shape open science practices. However, according to an analysis of academic job postings, only 2.2% currently do so (Nosek et al., 2022).

Educational research differs from psychology in that experimental designs are less dominant and research more often includes qualitative, interpretive, design-based, or context-sensitive approaches. As a result, replication cannot always be understood in the same way across all areas of the field. Nevertheless, the broader problem remains: without access to data, materials, analytical decisions, code, instruments, or documentation, findings are difficult to scrutinize, reuse, or build upon. Greater transparency and a stronger culture of sharing are therefore desirable not only for replication in a narrow sense, but also for the cumulative development, critical evaluation, and practical relevance of educational research.

1.2 Changing Publication Practices in Education

Digital infrastructures play a central role in making open science practices feasible. In particular, the Open Science Framework (OSF; Foster & Deardorff, 2017) supports this by providing integrated tools that cover different stages of the research life cycle, from the initial research idea and study design to the storage and analysis of collected data, as well as the creation and publication of reports or preregistrations. OSF projects can also be assigned a DOI, which allows them to be cited and archived as independent scholarly outputs. In this respect, OSF extends beyond the functions of many traditional university repositories, which have often focused primarily on documenting the final research output, usually in the form of a PDF. By contrast, OSF also enables researchers to document and share datasets, analysis code, study materials, teaching materials, presentations, and other project-related documents.

GitHub is another infrastructure for repository-based open science. Originally developed for software development and version control, it is increasingly referenced in scholarly publications and has become particularly relevant for sharing research code, computational workflows, and software (Escamilla et al., 2022). GitHub can support collaborative and reproducible research by making code, data, protocols, and project documentation publicly accessible and version-controlled (Braga et al., 2023). However, it is less directly tailored to the general research life cycle than OSF and requires greater technical familiarity. For many educational researchers without programming experience, GitHub may therefore appear complex, which may limit its adoption despite its potential value for transparent and reproducible research.

These infrastructural developments are connected to broader changes in the publication culture of educational research. The field has shifted from a tradition shaped largely by single-authored humanities-oriented work and monograph publications toward more empirical qualitative and quantitative research, often conducted by multiple authors and published in journals (Schmidt-Hertha et al., 2024). This shift increases the relevance of open science infrastructures, as empirical and collaborative research typically produces a wider range of research outputs than can be represented in a

conventional journal article alone.

At the same time, the economic conditions of scholarly publishing have changed substantially. The costs of journals and scientific publishing have risen sharply in recent years, even after adjusting for inflation, from an estimated $910.3 million in 2019 to $2.538 billion in 2023 (Haustein et al., 2024). The journal market has moved from a paper-based subscription model to an electronic subscription model, in which libraries pay for access to scientific journals, and increasingly toward an author-pays model, in which authors or their institutions pay article processing charges (APCs) for open-access publication. Depending on the journal, these fees commonly range from $1,000 to $4,000 per article (Haustein et al., 2024).

These developments are widely criticized because much of the scholarly labor on which academic publishing depends, including writing, reviewing, and editorial work, is provided free of charge by the scientific community. As a result, the costs charged by publishers are often considered disproportionate to the services provided, particularly when compared with the actual costs of publication and dissemination (Dougherty & Horne, 2022). Against this background, open science raises not only methodological and epistemic questions, but also questions about access, ownership, and the public value of research outputs.

1.3 Data Sharing Practices

Sharing data, code, and other research materials is essential for empirical educational research. Nevertheless, these practices are still rarely visible in scientific journals. While some researchers act as pioneers in this field, and while funded projects increasingly implement open science practices in response to specific funding requirements, these practices are far from being established as routine publication standards in educational research (van Dijk et al., 2021).

Related fields provide a mixed but informative point of comparison. Tedersoo et al. (2021) examined 875 articles published in Nature and Science and found that 54.2% included full or partial data. By contrast, Vanpaemel et al. (2015) found that, among 300 recent psychology articles, only four authors had made their data available in public repositories. Cross-disciplinary evidence from the MapOSR project, a mapping review dataset of empirical studies on open science, indicates increasing engagement with open science over time, although overall levels remain low (Lasser et al., 2022). These findings suggest that data sharing is growing across disciplines, but that its implementation remains uneven and field-specific.

One mechanism that has been used to promote open science is the introduction of open science badges by the Center for Open Science (osf.io). In Psychological Science, the introduction of badges increased the proportion of articles with openly available data from 3% to 23%, whereas comparison journals without badges did not show a comparable increase (Kidwell et al., 2016). Badges therefore appear to have an effect in their own right (Zong et al., 2023). This may be because they make open science practices visible and actively signal that data sharing, preregistration, or open materials are valued by the journal. In this sense, a badge is a more prominent and perform-

ative indicator than a general statement in submission guidelines, where information about open science may appear as only one requirement among many. However, the effect of badges is not automatic. In an experiment conducted with BMJ Open, badges alone did not persuade authors to share their data (Rowhani-Farid et al., 2020). Schneider et al. (2022) further showed that Open Science badges can increase trust among students and researchers, while they appear to be largely irrelevant for the general public. Open Science badges are currently used by some publishers, most prominently in psychology, but not yet in educational research (Center for Open Science, 2026). Correspondingly, open data practices have so far rarely been visible in education conferences or journal articles.

The limited adoption of open science is also reflected in the scarcity of replication studies in educational research. Makel and Plucker (2014) analyzed the complete publication history of the top 100 education journals and found that only 0.13% of articles were replications. This low proportion is problematic because replication is one important mechanism for testing the robustness and generalizability of research findings. At the same time, replication depends on access to information about the original study, including data, instruments, analytical procedures, and, where applicable, analysis code. The rarity of replication studies therefore points not only to a lack of replication culture, but also to broader limitations in the transparency and reusability of educational research.

Researchers in education do not necessarily reject open science in principle. Makel et al. (2021) report that 20% to 50% of educational researchers express openness toward open science practices. However, implementation remains difficult. Commonly cited barriers include inadequate infrastructure, ethical concerns, the perception that open science constitutes an additional burden, and the view that open science is not always appropriate for particular research designs or contexts (Makel et al., 2026). Open science therefore remains in tension with parts of the existing research culture in education. This is particularly evident in qualitative research, where the release of data and broader forms of open research are often viewed critically because of concerns about confidentiality, contextual sensitivity, interpretation, and the protection of participants (Prosser et al., 2023).

At the same time, recent evidence complicates the assumption that qualitative research is necessarily less compatible with open science. Schneider (2024) surveyed 250 researchers from different disciplines about 13 open science practices: open methodology, open access publication, open data, open materials, science communication, open code, open peer review, public project plans, open software, open educational resources, preregistration, citizen science, and open source. The study identified four openness profiles: fully open (36%), open with low public participation (17%), potential for openness (36%), and maybe science communication (11%). These profiles were similar across research fields and disciplines. Notably, the only significant correlation was found between qualitative research and open science, a finding that contradicts common criticism of qualitative research as inherently resistant to openness.

Overall, more initiatives, professional associations, and funding agencies are calling for open science, but change in publication practices appears to be slow. A broader shift is needed because open science can advance research itself by increasing transparency, reusability, and cumulative know-

ledge production (van Dijk et al., 2021). Survey data from the social sciences indicate that open science has become increasingly familiar to researchers: the proportion of respondents who reported having used open science at least once in their careers rose from 47% in 2010 to 87% in 2020 (Ferguson et al., 2023). It can therefore be assumed that open science is now widely known among high-impact researchers. The key question is whether this growing awareness, and the methodological and cultural shifts associated with open science, are also reflected in actual publication practices in educational research.

2 Research Question

Taken together, the literature suggests a growing awareness of open science and an increasing availability of infrastructures that make open research practices technically feasible. At the same time, the adoption of these practices appears to remain uneven across disciplines, publication cultures, and research designs. For educational research, this tension is particularly relevant. The field has become increasingly empirical and journal-based, yet it remains unclear whether this shift has been accompanied by a corresponding increase in the public sharing of data, code, materials, preregistrations, or software. Based on these considerations, the study addresses the following research questions:

1. To what extent do articles in open-access education journals use open science practices and how has the use of open science practices changed over time?
2. What kinds of open science practices are used and do they differ by article type?
3. In which thematic areas of educational research are open science practices used?

3 Materials and Methods

3.1 Study Design

We use a bibliometric and content-analytical design to examine open science practices in selected open-access journals in the education sciences. The analysis focuses on the visibility, accessibility, and content of repository links in published journal articles. More specifically, we investigate whether articles provide links to OSF or GitHub and what types of research-related materials are made available through these repositories. Therefore, we combine automated full-text screening with manual verification and coding. The code and data are available on osf or github. https://osf.io/48yuw

https://github.com/nikola-ebenbeck/How-Open-are-the-Education-Sciences

Preregistration: https://osf.io/vq5nh/overview

3.2 Journal Selection

Since the analysis includes only open-access articles, hybrid journals with paywalled articles were excluded, and only open-access journals were included. We selected one open-access journal from each large publisher, as well as one from each of the two major academic associations in education (AERA, EARLI). Therefore, we selected the following journals: *Frontiers in Education* (Frontiers Media SA), *Education Sciences* (MDPI), *Cogent Education* (Taylor & Francis), *International Journal of Educational Technology in Higher Education* (*IJETHE*; Springer), *AERA Open* (AERA and Sage Publishing), *Sage Open* (Sage Publishing), and *Frontline Learning Research* (EARLI). These journals represent prominent open-access publication venues in education science and cover a broad range of empirical, methodological, theoretical, and applied educational research. The journal sample includes both general education journals and journals with a stronger focus on technology-enhanced learning, learning research, or interdisciplinary educational research.

The journals examined generally endorsed open-science practices and research-data sharing, although the scope and binding nature of their policies varied considerably. By the end of 2025, most of the journals required or explicitly encouraged authors to include a Data Availability Statement. However, none of the education journals examined imposed an unrestricted general requirement to make all underlying research data publicly available. Frontiers in Education and the International Journal of Educational Technology in Higher Education (IJETHE) had the most stringent formal data-availability requirements. Compared with the other journals examined, IJETHE placed particularly strong emphasis on data sharing by explicitly encouraging authors to deposit supporting datasets in publicly accessible repositories and provide them in machine-readable formats whenever possible. Education Sciences, Cogent Education, and Sage Open increasingly standardized the disclosure of data availability during the study period, although public data sharing generally remained conditional rather than mandatory. AERA Open maintained a comparatively cautious policy, encouraging data sharing whenever feasible without imposing a general open-data requirement. No substantial introduction of binding open-data requirements was identified for Frontline Learning Research.

3.3 Data Collection

A preliminary list of potentially relevant articles was obtained from OpenAlex (Priem et al., 2022). OpenAlex is a large open bibliographic database that provides metadata on scholarly publications, including information on journals, publication years, publication types, open-access status, and digital object identifiers. Articles were included in the initial corpus if they met the following criteria: 1) The article was published in one of the selected journals, 2) The article was published between 2020 and 2025, 3) The article was available as an open-access publication, 4) The publication was an article or article-like scholarly contribution, 5) A full-text PDF could be identified and downloaded, and 6) the article was not an editorial, erratum, correction, book reviews, announcement, or other non-article publication format.

The corresponding full-text PDFs were downloaded from the publishers' websites using a Python-based retrieval workflow. Articles for which no PDF could be retrieved were excluded from full-text-based analyses. The downloaded PDFs were converted into machine-readable text using *GROBID* (GROBID, 2008). GROBID is a tool for extracting and structuring bibliographic and full-text information from scholarly PDF documents. The resulting text files were subsequently processed for link extraction. In addition, the R package *metacheck* (Lakens et al., 2026) was used to identify links to external repositories. The automated screening focused on links to OSF (Foster & Deardorff, 2017) and GitHub because both platforms are widely used for sharing research data, analysis code, study materials, preregistrations, and software.

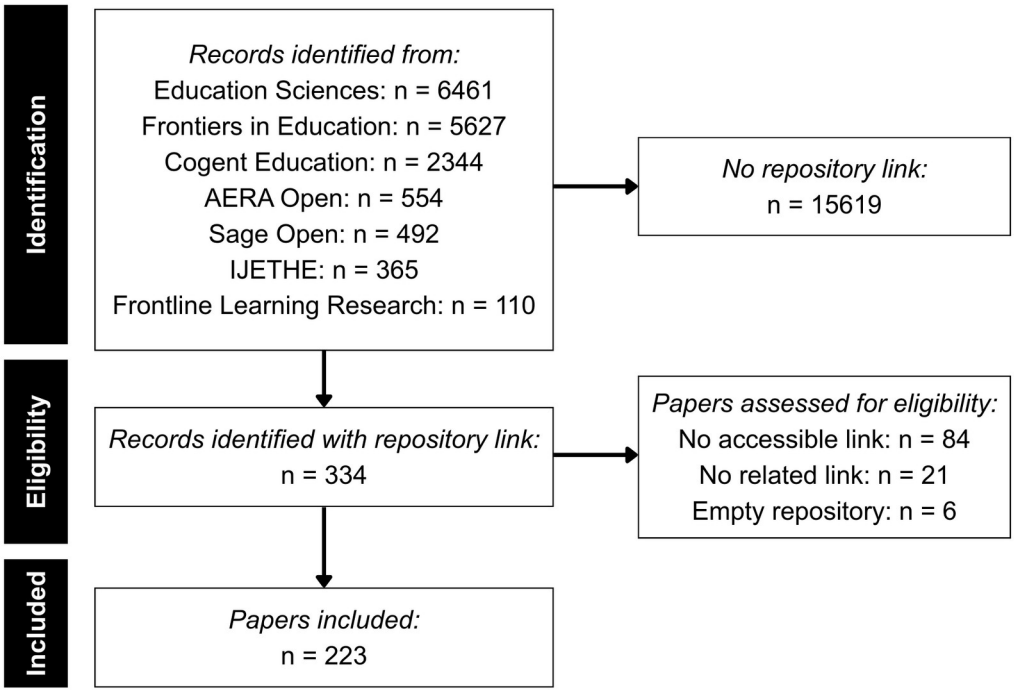


Figure 1: Workflow of article identification, inclusion and exclusion

The machine-readable full texts were searched for URLs containing references to OSF and GitHub. The automated link extraction included both explicit URLs and embedded links where they could be identified in the processed text. The resulting link list was cleaned by removing exact duplicates within the same article. Furthermore, all automatically identified OSF and GitHub links were manually inspected. We checked manually, whether the link was functional or invalid, whether the repository was publicly accessible and whether the repository contained materials directly related to the article. Papers were excluded, if they did not have a valid, open and related link. Figure 1 shows the article selection process. In total, 15,953 articles were identified and downloaded. Most articles were published in *Education Sciences* (n = 6,461; 40.50%) and *Frontiers in Education* (n = 5,627; 35.27%), followed by *Cogent Education* (n = 2,344; 14.70%), *AERA Open* (n = 554; 3.47%), *Sage Open* (n = 492; 3.08%), *IJETHE* (n = 365; 2.29%), and *Frontline Learning Research* (n = 110;

0.69%). Of these articles, 15619 did not contain an OSF or GitHub repository link and were excluded. A total of 334 articles were assessed for eligibility. During the manual review, 111 articles were excluded because their links were broken, led to restricted folders, or did not provide access to any article-related data, materials, or code. The final sample included 223 articles with a functional and article-related repository link, corresponding to 1.4% of all downloaded articles.

The included studies covered a broad range of geographical contexts, although the distribution was strongly concentrated in Europe and North America. The United States was the most frequently represented country (n = 50), followed by Germany (n = 39). Other frequently represented countries were the United Kingdom (n = 11), the Netherlands (n = 8), and Canada, China, France and Spain (each n = 7). At the continental level, most studies were conducted in European contexts (n = 95), followed by North America (n = 59). Asian countries were represented less frequently (n = 18), followed by South America (n = 10), Africa (n = 4), and Oceania (n = 2).

3.4 Coding of Open Science Practices

For each article, the contents of the linked repositories were coded manually. The aim was to describe which types of research outputs were shared. Because a single article could provide several types of materials, multiple coding was allowed. We coded whether a link contained open data, open code, open analysis (meaning any analysis related protocols which are not code-based), open material (meaning study-related material such as questionnaires), preregistration material or open software.

3.5 Coding of Article Type

Each article was assigned to one primary article type. The article type was determined manually based on the research design, the data used, and the main contribution of the article. We distinguished between quantitative empirical article, qualitative empirical article, mixed-methods empirical article, review article, conceptual or theoretical article, and methodological article.

Articles were coded as quantitative if they primarily relied on numerical data and statistical analyses. Articles were coded as qualitative if they primarily relied on qualitative data and interpretive analysis. Articles were coded as mixed methods if both quantitative and qualitative empirical components were central to the study. Review articles included systematic reviews, scoping reviews, meta-analyses, and narrative reviews. Conceptual articles included theoretical, didactic, or argumentative contributions without original empirical data. Methodological articles focused primarily on the development, evaluation, or discussion of research methods, instruments, tools, or analytical procedures. If an article combined several elements, the dominant methodological orientation was coded as the primary article type.

3.6 Coding of Thematic Categories

Thematic categories of the articles were coded using LLM-Assisted Inductive Categorization

(LAIC, Arlinghaus et al., 2024), using ChatGPT model 5.5 thinking (Open AI, 2026) with 10 iterations. We based the LAIC process only on the papers' title and keywords, as the goal was a broad categorization of the topic and scientific community rather than a very specific and detailed topic analysis.

Because education science articles often address more than one topic, multiple thematic assignments were allowed for the LLM model. Each article was assigned to all fitting categories. If the 10 iterations had an alignment of less than 70% for categorization, we manually categorized the respective article instead.

3.7 Statistical Analysis

Descriptive statistics were used to report the number and percentage of articles in the full corpus by journal and publication year, the distribution of article types, the types of shared materials, and the thematic categories of empirical studies. Article types and open science practices were cross-tabulated.

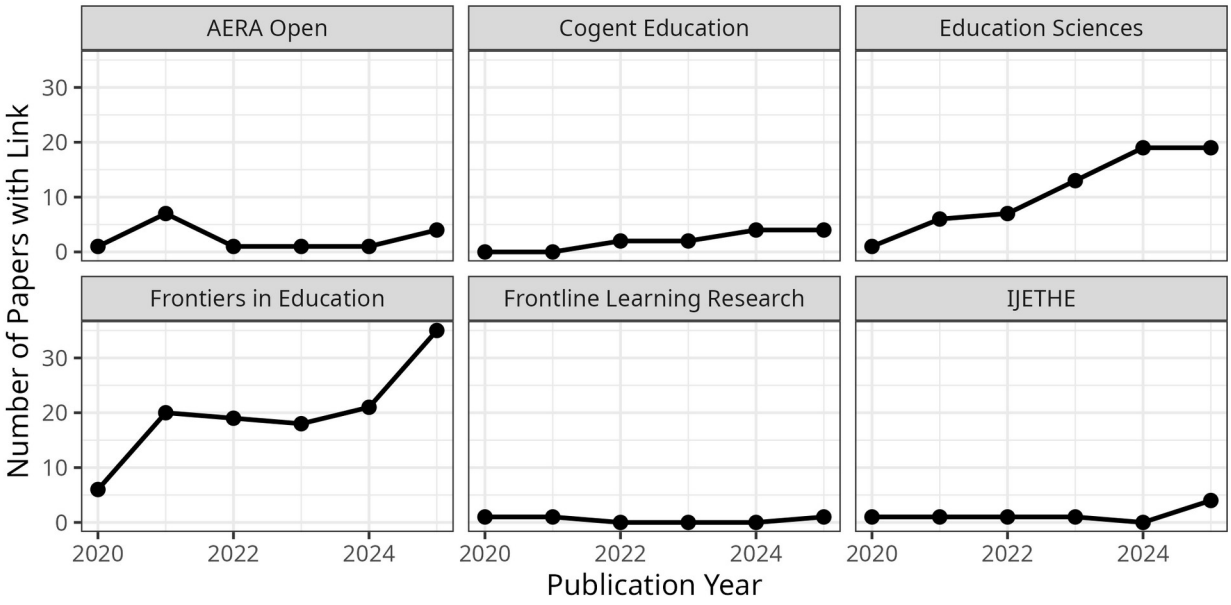


Figure 2: Development of repository-linked article numbers (Note: Sage Open is not included in the figure, as there is only one paper included.)

Further, to estimate whether the probability of a paper containing a repository link changed over time, we fitted a Bayesian multilevel logistic regression with random intercepts and slopes per journal, assuming that each journal has a different innate probability of linking to data, and that each journal also has a different rate of change in this behavior. A Bayesian approach was chosen because repository links were expected to be rare and because the number of repository-linked articles differed substantially across journals.

4 Results

4.1 Development of Repository-Linked Articles Over Time

In absolute terms, clear increases in the number of papers using repository links were observed only for *Education Sciences* and *Frontiers in Education* (Figure 2). In relative terms, however, there were no significant changes in the use of OSF and GitHub links in the articles. The prevalence of repository-linked articles was low across the selected journals. The estimated intercept indicated a baseline probability of approximately 1.17 % for an article containing a functional and article-related OSF or GitHub link (β = -4.44) across all journals. Although some journals appeared to show increases in linked articles over time, these patterns are inconsistent and vary between journals. The estimated effect of the publication year was small and positive but uncertain (β = 0.04, 95% CrI [-0.13, 0.23]). The model therefore does not support a temporal increase in the linking behaviour across the articles.

5 Open Science Practices

Most articles with repository links used a quantitative design (75.3%), 12.1% used a mixed-methods design, and 4.9% were qualitative. Review articles accounted for 3.1% of the sample. Conceptual articles, including didactic and instructional contributions, made up 2.7% of the sample. Methodological articles represented the smallest group, with 1.8% of all articles included.

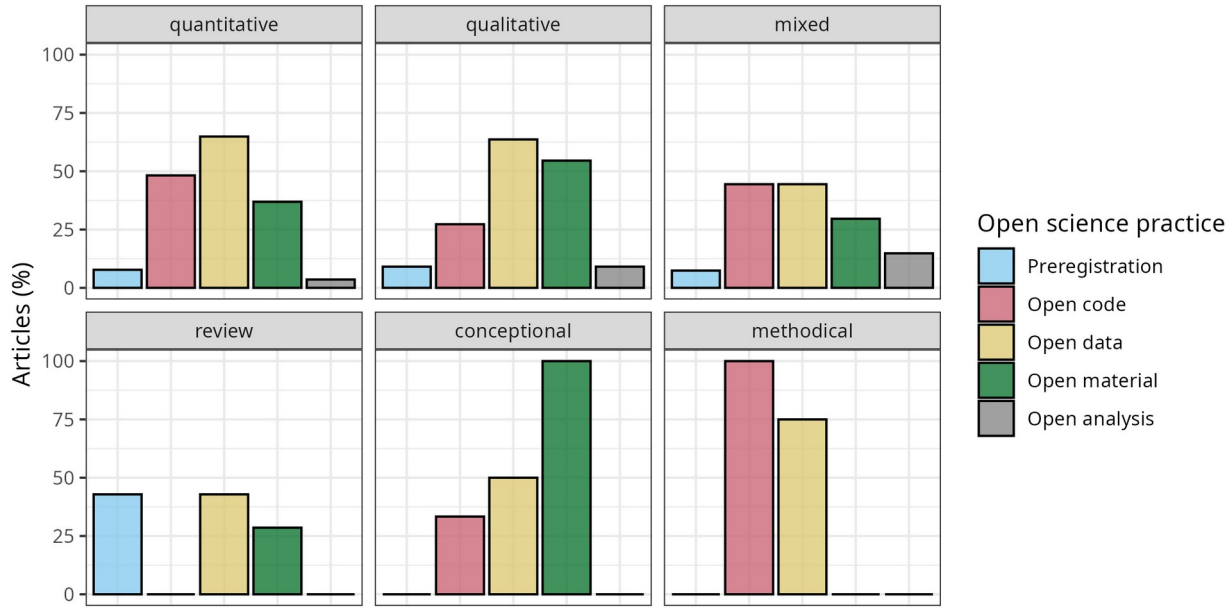


Figure 3: Open science practices of different article types

Thereby, open science practices also differed by article type (Figure 3). In quantitative articles, open data was the most common practice, followed by open code and open materials. Mixed-methods articles showed a similar pattern. In qualitative articles, open data and open materials were also rel-

atively frequent, whereas open code was less common. Review articles differed from empirical articles and most frequently included preregistrations, while open data, materials and code were less common. Conceptual articles mainly shared open materials. Methodological articles most often provided open code and open data.

6 Thematic Categories

A total of 14 thematic categories were identified within the articles. Articles could be assigned to more than one category, as many studies addressed multiple thematic areas (Table 1). Therefore, the frequencies represent thematic assignments rather than mutually exclusive article groups.

Repository-based open science practices were particularly visible in research areas that often involve instruments, measurement procedures, intervention materials, or quantitative data (i.e., Assessment, Psychometrics and Evaluation, Learning Psychology, Higher and Graduate Education, Curriculum and Instruction). Interestingly, articles with a more technical focus (i.e., Educational Technology, Digital Learning) were more in the middle field and not on top of the ranking.

Category	**Definition**	**N Articles**
Assessment, Psychometrics and Evaluation	Research on educational assessment, testing, measurement, psychometric properties of instruments, validation studies, feedback, grading, and the evaluation of learning outcomes or educational interventions.	65
Learning Psychology	Studies examining learners' motivation, engagement, emotions, self-regulated learning, metacognition, well-being, social-emotional learning, and related psychological factors influencing educational processes and outcomes.	49
Curriculum and Instruction	Research on curriculum development, pedagogical approaches, instructional design, teaching methods, learning materials, content knowledge, and the organization or sequencing of educational content.	47
Higher and Graduate Education	Studies situated in higher education, university teaching and learning, graduate education, academic development, student experiences in tertiary education, and institutional practices in colleges and universities.	44
Inclusion, Diversity and Special Education	Studies addressing educational equity, inclusion, diversity, accessibility, social justice, underrepresented groups, disability, special educational needs, and inclusive educational practices.	40
Teacher Education and Professional Learning	Research on pre-service and in-service teacher education, teacher professional development, teacher beliefs, competencies, instructional practices, and professional learning communities.	38
Educational Technology, AI and Learning Analytics	Studies focusing on educational technologies, artificial intelligence, adaptive systems, learning analytics, data-driven educational tools, and the use of digital systems to support, monitor, or personalize learning and teaching.	38
STEM, Science and	Research on teaching and learning in science, technology,	36

Mathematics Education	engineering, and mathematics, including discipline-specific instruction, conceptual understanding, scientific reasoning, mathematical thinking, and STEM-related educational interventions.	
Literacy, Language, Reading and Multilingual Education	Research on literacy development, reading, writing, language learning, communication, bilingual or multilingual education, second-language acquisition, and language-related instructional practices.	36
Education Policy, Reform, Systems and Program Evaluation	Studies examining education policy, governance, institutional or system-level reform, educational programs, implementation processes, accountability, and large-scale or organizational evaluation.	28
Digital, Online and Immersive Learning Environments	Studies on online learning, blended learning, distance education, virtual learning environments, MOOCs, simulations, virtual or augmented reality, and other immersive or digitally mediated learning settings.	27
Early Childhood and Child Development	Research on early childhood education, child development, family involvement, parenting, home learning environments, developmental processes, and educational experiences in early developmental stages.	19
Informal Learning, Science Communication and Learning Networks	Studies on informal, non-formal, and out-of-school learning, science communication, museums, community-based learning, peer learning, social networks, and learning across distributed or networked contexts.	13
Open Science	Studies related to open science practices, including open data, open materials, preregistration, reproducibility, transparency, open educational resources, and the sharing of research outputs or educational resources.	4

Table 1: Thematic Categories of Articles

7 Discussion

Repository-based open science practices have so far been adopted only to a limited extent in educational research. Across the selected open-access journals, only a small proportion of articles included links to public repositories. Of 15,953 articles, 334 contained an OSF or GitHub link, and only 223 articles provided a functional and article-related repository link, corresponding to 1.4% of the full corpus. This low prevalence was consistent across the journals examined between 2020 and 2025. Thus, the strong recent increase in open-access publishing should not be interpreted as evidence of a similarly strong increase in broader open science practices. Articles may be openly accessible while the underlying data, code, materials, and analytical documentation remain unavailable. In this sense, our findings support the assumption that open science is still often perceived as an additional requirement or burden rather than as an established part of publication practice in educational research (Makel et al., 2026).

The limited availability of repositories has direct implications for transparency and reproducibility. Most articles in our corpus did not provide supporting materials beyond the article itself and can

therefore be verified, reused, or replicated only to a limited extent. Even among the articles with repository links, openness was often partial rather than comprehensive. Many repositories contained only one type of open science practice rather than a complete set of data, code, materials, preregistrations, and analytical documentation. Open data was the most common practice, but data alone are often insufficient for understanding and reproducing empirical findings. In complex studies, information about instruments, tasks, questionnaires, intervention manuals, codebooks, analytical decisions, and analysis code is essential. Without such documentation, open data may have limited value for secondary analysis, replication, or cumulative research.

The distribution of repository-linked articles also reflects differences between research designs and research cultures. Quantitative articles were most strongly represented among articles with repository links, whereas qualitative research appeared less frequently. This pattern is consistent with previous findings indicating that open science practices are more established in quantitatively oriented research communities (Makel et al., 2021). However, the lower visibility of qualitative research in repository-based open science should not be interpreted as evidence that openness is irrelevant or impossible for qualitative studies. Even when primary qualitative data cannot be shared for ethical or confidentiality reasons, other materials can often be made available, including interview guides, coding schemes, codebooks, reflection logs, analytic memos, consent procedures, or documentation of analytical decisions. Open science in qualitative research therefore requires field-sensitive standards rather than a simple transfer of quantitative data-sharing norms.

The thematic distribution of repository-linked articles further suggests that open science practices are unevenly embedded within educational research communities. Articles from *Assessment, Psychometrics, and Evaluation* were particularly frequent. This indicates that areas of educational research that overlap with psychology and measurement research may be more likely to adopt repository-based openness, possibly because psychology has developed a more intensive discourse on transparency, replication, and open science practices. By contrast, technically oriented topics, such as Educational Technology, AI, Learning Analytics, and Digital Learning Environments, were not among the leading categories. This is noteworthy because these areas might be expected to share code, software, or digital materials more frequently. The findings therefore suggest that technical proximity alone does not necessarily translate into stronger open science practices within educational research.

Overall, the study provides little evidence of a substantial temporal shift toward greater repository-based openness between 2020 and 2025. Although absolute numbers increased in some journals, particularly in Education Sciences and Frontiers in Education, the proportional development did not indicate a significant overall increase in the use of OSF or GitHub links. This suggests that change is still slow and uneven. Academic communities play an important role in this process, but journals are equally important gatekeepers. Clear journal policies can normalize expectations regarding data, code, materials, and preregistrations. In this context, statements such as "data available upon request" should no longer be treated as the default form of openness, but only as an exception that requires justification. Some journals and academic communities are beginning to move in this direc-

tion. In addition to clearer guidelines, journals could support open science through dedicated budgets, reduced article processing charges for preregistered studies, or visible incentives such as badges, which have been shown to increase open science practices in some contexts (Kidwell et al., 2016; Zong et al., 2023). Since many education journals are only beginning to implement explicit open science requirements, future publication years will be particularly important for assessing whether these policies lead to measurable changes. Future research should examine how the prevalence of preregistrations, links to research data and preprints, and replication studies develops over time, as well as the extent to which emerging journals that actively promote or mandate open-science practices contribute to these developments.

8 Limitations

Several limitations must be considered. First, the analysis primarily identified studies from the United States and Europe. This pattern may reflect differences in research output, but it may also indicate that open science is more visible or more strongly institutionalized in these regions. This raises broader questions about the global distribution of open science practices, especially because open science could be particularly valuable for researchers and institutions in contexts with less developed research infrastructures. Second, the study focused on selected open-access journals and on OSF and GitHub links. Authors may share data, materials, or code through other routes, including supplementary files, Zenodo, Figshare, Dataverse, institutional repositories, or controlled-access procedures. The present study may therefore underestimate the overall level of openness in educational research. Third, the focus on visible repository links does not capture the quality, completeness, documentation, or actual reusability of the shared materials. A repository link indicates the presence of publicly accessible materials, but it does not necessarily mean that these materials are sufficient for replication or secondary use. Finally, repository accessibility may change over time, as links can become inactive, private, or disconnected from the article.

Despite these limitations, repository-based analyses provide a useful approach for monitoring open science practices in educational research. As open science standards become more common, such analyses may become easier and more informative. The central question will not only be whether articles include repository links, but also what standards reviewers, editors, journals, and professional associations use to evaluate the completeness and quality of shared materials before publication. Educational research needs field-specific open science standards that take into account different research designs, data types, ethical requirements, and epistemological traditions. Such standards should be developed by academic communities and professional associations and supported by clear journal policies (van Dijk et al., 2021). In addition, open science practices should be recognized in academic career structures, including hiring, promotion, and funding decisions. If openness remains an optional extra, it is unlikely to become a stable part of publication culture. If it becomes part of how research quality is evaluated, however, open science may contribute to a more transparent, reusable, and cumulative education science.

9 Literature

Braga, P. H. P., Hébert, K., Hudgins, E. J., Scott, E. R., Edwards, B. P., Sánchez Reyes, L. L., ... & Crystal-Ornelas, R. (2023). Not just for programmers: How GitHub can accelerate collaborative and reproducible research in ecology and evolution. Methods in Ecology and Evolution, 14(6), 1364–1380.

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., ... & Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. Science, 351(6280), 1433–1436.

Center for Open Science. (2026). Open Science Badges enhance openness, a core value of scientific practice. https://www.cos.io/initiatives/badges

Cook, B. G., Lloyd, J. W., Mellor, D., Nosek, B. A., & Therrien, W. J. (2018). Promoting open science to increase the trustworthiness of evidence in special education. Exceptional Children, 85(1), 104-118.

Dougherty, M. R., & Horne, Z. (2022). Citation counts and journal impact factors do not capture some indicators of research quality in the behavioural and brain sciences. Royal Society Open Science, 9(8), 220334.

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.

Escamilla, E., Klein, M., Cooper, T., Rampin, V., Weigle, M. C., & Nelson, M. L. (2022). The rise of GitHub in scholarly publications. In International Conference on Theory and Practice of Digital Libraries (pp. 187–200). Springer International Publishing.

Ferguson, J., Littman, R., Christensen, G., Paluck, E. L., Swanson, N., Wang, Z., ... & Pezzuto, J. H. (2023). Survey of open science practices and attitudes in the social sciences. Nature Communications, 14(1), 5401.

Foster, E. D., & Deardorff, A. (2017). Open Science Framework (OSF). Journal of the Medical Library Association: JMLA, 105(2), 203–206. https://doi.org/10.5195/jmla.2017.88

GitHub, Inc. (2026). GitHub [Computer software]. https://github.com

GROBID. (2008). GROBID [Computer software]. https://github.com/grobidOrg/grobid

Haustein, S., Schares, E., Alperin, J. P., Hare, M., Butler, L. A., & Schönfelder, N. (2024). Estimating global article processing charges paid to six publishers for open access between 2019 and 2023 (arXiv:2407.16551). arXiv.

Kidwell, M. C., Lazarević, L. B., Baranski, E., Hardwicke, T. E., Piechowski, S., Falkenberg, L. S., ... & Nosek, B. A. (2016). Badges to acknowledge open practices: A simple, low-cost, effective method for increasing transparency. PLoS Biology, 14(5), e1002456.

Lakens, D., Werner, J., & DeBruine, L. (2026). Scienceverse/metacheck [R]. ScienceVerse. https://github.com/scienceverse/metacheck

Lasser, J., Schneider, J., Lösch, T., Röwert, R., Heck, T., Bluemel, C., ... & Skupien, S. (2022). MapOSR: A mapping review dataset of empirical studies on Open Science. F1000Research, 11, 535.

Makel, M. C., Caroleo, S., Meyer, M., Pei, M., Fleming, J., Hodges, J., ... & Plucker, J. (2026). "Don't hate the players, hate the game": Qualitative insights from education researchers on questionable and open research practices. Royal Society Open Science, 13(1), 251888.

Makel, M. C., Hodges, J., Cook, B. G., & Plucker, J. A. (2021). Both questionable and open research practices are prevalent in education research. Educational Researcher, 50(8), 493–504.

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. Educational Researcher, 43(6), 304–316.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., ... & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. Annual Review of Psychology, 73, 719–748.

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

Priem, J., Piwowar, H., & Orr, R. (2022). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts (arXiv:2205.01833). arXiv. https://doi.org/10.48550/arXiv.2205.01833

Prosser, A. M. B., Hamshaw, R. J. T., Meyer, J., Bagnall, R., Blackwood, L., Huysamen, M. , Jordan, A., Vasileiou, K., & Walter, Z. (2022). When open data closes the door: A critical examination of the past, present and the potential future for open data guidelines in journals. British Journal of Social Psychology, 00, 1–17. https://doi.org/10.1111/bjso.12576

Rowhani-Farid, A., Aldcroft, A., & Barnett, A. G. (2020). Did awarding badges increase data sharing in BMJ Open? A randomized controlled trial. Royal Society Open Science, 7(3), 191818.

Schmidt-Hertha, B., Rittberger, M., & König, A. (2024). Forschung und Publikationskulturen. In B. Schmidt-Hertha, A. Tervooren, R. Martini, & I. Züchner (Eds.), Datenreport Erziehungswissenschaft 2024: Erstellt im Auftrag der Deutschen Gesellschaft für Erziehungswissenschaft (DGfE) (1st ed., pp. 173–200). Verlag Barbara Budrich. https://doi.org/10.2307/jj.11786270.11

Schneider, J. (2024). Sorry we're open, come in we're closed: Different profiles in the perceived applicability of open science practices to completed research projects. Royal Society Open Science, 11(1), 230595.

Schneider, J., Rosman, T., Kelava, A., & Merk, S. (2022). Do open-science badges increase trust in scientists among undergraduates, scientists, and the public? Psychological Science, 33(9), 1588–1604.

Tedersoo, L., Küngas, R., Oras, E., Köster, K., Eenmaa, H., Leijen, Ä., ... & Sepp, T. (2021). Data sharing practices and data availability upon request differ across scientific disciplines. Scientific Data, 8(1), 192.

Van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open science in education sciences. Journal of Learning Disabilities, 54(2), 139–152.

Vanpaemel, W., Vermorgen, M., Deriemaecker, L., & Storms, G. (2015). Are we wasting a good

crisis? The availability of psychological research data after the storm. Collabra, 1(1), 3.

Watson, M. (2015). When will “open science” become simply “science”? Genome Biology, 16(1), 101.

Zong, Q., Huang, Z., & Huang, J. (2023). Do open science badges work? Estimating the effects of open science badges on an article’s social media attention and research impacts. Scientometrics, 128(6), 3627–3648. https://doi.org/10.1007/s11192-023-04720-7