

# JLSC

ISSN 2162-3309 | JLSC is published by the Iowa State University Digital Press | <http://jpsc-pub.org>

Volume 12, 1 (2024)

## Data Ethics in Library Learning Analytics

Dorothea Salo

Salo, D. (2024). Data Ethics in Library Learning Analytics. *Journal of Librarianship and Scholarly Communication*, 12(1), eP16245. <https://doi.org/10.31274/jpsc.16245>

This article underwent fully anonymous peer review in accordance with JLSC's peer review policy.



© 2024 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

## RESEARCH ARTICLE

# Data Ethics in Library Learning Analytics

Dorothea Salo

*University of Wisconsin at Madison*

## ABSTRACT

**Introduction:** This study examines data handling and data ethics in library learning analytics research projects involving the use of data about students as library patrons, alongside a baseline evaluation of the benefits of library learning analytics to libraries practicing it.

**Methods:** Citations were gathered via citation chain aggregation from the original Value of Academic Libraries report, then winnowed to projects with English-language artifacts documenting them, collection and/or analysis of library data about students, and a research question about the contribution of student library use to student success.

**Results and Discussion:** Much of this research is reaching publication despite not employing best practices nor documenting respect for human-subjects research ethics, library-specific privacy and confidentiality ethics, and student data-privacy expectations. Very few projects create direct benefits to libraries. This result would not be possible without gaps or lapses in editorial processes, peer review, and upstream research guidance and ethics reviews.

**Conclusion:** Ethics reforms are required at all stages of research and publication to prevent further unethical exploitation of patron data.

**Keywords:** data ethics, library learning analytics, library-specific privacy, library confidentiality ethics, student data privacy

Received: 04/06/2023 Accepted: 04/06/2024



© 2024 The Author(s). This is an open access article distributed under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)

## IMPLICATIONS FOR PRACTICE

1. Published library learning analytics work has not to date demonstrated significant benefit to libraries engaging in it.
2. Neither institutional review processes nor editorial processes currently prevent library learning analytics work containing ethics lapses and violation of student data-privacy expectations from being published in journals and conferences.
3. Journal editorial boards and conference-planning panels must therefore develop clearer data-ethics guidelines, rejecting work that does not meet them. Some already public work may require expressions of concern or even retractions.
4. Library privacy policies and LIS professional and research organizations must also better govern the use of patron data in research and assessment to prevent ethically unpublishable work from reaching journals and conferences.

## INTRODUCTION AND LITERATURE REVIEW

Publishing processes such as editorial selection and peer review for journals and conferences are nearly the last line of defense against unethical research practices, taking place well after most institutional data-governance and ethics-review processes. Once an article or conference presentation passes editorial and peer review to reach publication, only the risky and logistically challenging process of requesting retraction can remove it from the scholarly and professional literature. Published work documenting unethical decisions and methods therefore becomes an information hazard (Yoose, 2021), falsely legitimizing unethical research practices by misleading potentially everyone in the research process—researchers, data-governance professionals, ethics reviewers, editors, and peer reviewers—into believing those practices acceptable because they achieved publication. Worse yet, published research that contradicts librarianship’s stated ethical values invites patrons to question librarianship’s actual commitment to those values, as happened in Santa Cruz over library use of Gale Analytics’s customer-relationship management tools (*Patron Privacy at Santa Cruz Public Libraries*, 2019).

As an illustration of the hazards of publishing research employing dubiously ethical methods, this piece explores library and information science (LIS) research using higher-education student behavioral-exhaust data, commonly known as “library learning analytics.” This set of methods became prominent after the Association for College and Research Libraries (ACRL) published the Value Agenda for Libraries (VAL) report (Oakleaf, 2010), which advocated for analysis of patron information-use and library-use data, purportedly to demonstrate return on investment in libraries to academic administrators.

An extensive discussion of library learning analytics alongside a discussion of its privacy and ethics implications can be found in the literature review by Jones et al. (2020b). Several recent pieces (Asher, 2017; Asher et al., 2018; Jones & Salo, 2018; Oliphant & Brundin, 2019; Doty, 2020; Huang et al., 2021) raise alarms at length over conflicts between library learning analytics practices and librarianship's stated values and ethics. Researcher librarians self-report considerable awareness of human-subjects research ethics issues, but feel inadequately prepared to address research ethics issues specific to library learning analytics (Jones & Hinchliffe, 2020). Fisher et al. (2019) consider library learning analytics that measure what students *do* inferior to assessment methods that measure what students actually *learn*, especially in a context of unequal power relations between researchers and students. Hathcock (2018) asserts that learning analytics wrongly deprives students of agency they require for successful learning. These discussions parallel concern in the higher-education literature about the ethics of learning analytics generally, as captured in a literature review by Viberg et al. (2022).

One serious library-specific ethical concern is the impact of patron data collection and use on information privacy. The International Federation of Library Associations and Institutions (IFLA)'s Statement on Privacy in the Library Environment (2015) is uncompromising regarding digital surveillance of patrons by libraries:

Library and information services should reject electronic surveillance and any type of illegitimate monitoring or collection of users' personal data or information behaviour that would compromise their privacy and affect their rights to seek, receive and impart information. They should take measures to limit collection of personal information about their users and the services that they use.

The Canadian Federation of Academic Libraries (CFLA)'s Statement on Intellectual Freedom and Libraries (2016) takes the strong stance that "libraries protect the identities and activities of library users except when required by the courts to cede them" (para. 8). The ACRL's Intellectual Freedom Principles for Academic Libraries (American Library Association, 2006) also discusses information confidentiality:

Article III of the Code of Ethics of the American Library Association states that confidentiality extends to "information sought or received and resources consulted, borrowed, acquired or transmitted," including, but not limited to, reference questions and interviews, circulation records, digital transactions and queries, as well as records regarding the use of library resources, services, programs, or facilities.

Several transactions listed above as confidential, of course, leave behavioral data exhaust used in library learning analytics research and assessment. Such data exhaust use may contravene

the IFLA (2015) Statement on Privacy, which notes that “[e]xcessive data collection and use threatens individual users’ privacy and has other social and legal consequences” (para. 2). Similarly, a recent American Library Association (ALA) Council resolution opposes “behavioral tracking,” defined as “the collection of data about an individual’s engagement with the library that, alone or with other data, can identify the user, for purposes of monitoring, tracking, or profiling an individual’s library use beyond operational needs” (ALA Council, 2021, para. 8).

Unfortunately, the exact limits of confidentiality and what counts as “excessive” data collection and use, especially with respect to whether published research or internal assessment count as “operational needs,” have yet to be clarified by the IFLA, ACRL, or ALA (Salo, 2021). The CFLA, by contrast, is quite clear that the only acceptable reason to break patron information-use confidentiality is governmental duress. Professional bodies have also not yet reckoned with increased reidentifiability of data, information security threats, or the temptation that library learning analytics creates to retain patron data longer than otherwise necessary (Salo, 2021; Yoose, 2021), which makes student data unnecessarily vulnerable to data exfiltration (e.g., Collier, 2022) and other inappropriate or dangerous access, including from institutional insiders.

The IFLA, CFLA, and ALA library ethics statements also posit that students, like all library patrons, can expect their information and service use to remain as private and confidential as possible *from librarians themselves*. This prerogative dates back to the establishment of the ALA Code of Ethics, whose privacy and confidentiality stipulations stem in part from early ALA president Arthur Boston’s awareness that librarian surveillance of patrons’ information use would mean fewer patrons (Witt, 2017). In part, also, those stipulations seek to prevent librarians using patrons’ information use against them, as when a librarian detailed immigrant Henry Melnek’s reading habits in a court proceeding against Melnek (Witt, 2017). To be in harmony with the aforementioned ethics codes, given this history and the present day’s highly invasive and punitive surveillance environment (Lamdan, 2022), librarianship would ideally assume *a priori* that identified or reidentifiable patron data is ineligible for reuse outside the specific operational context of its collection and use. Librarians do not enjoy a blanket entitlement to use patron data in research and assessment merely because librarians or library-internal technology collected it (Asher, 2017). Unfortunately, these professional bodies lack enforcement mechanisms for their privacy and confidentiality codes, leaving publishers to prevent lapses in patron privacy and confidentiality from reaching the LIS literature, or should that fail, to use expression-of-concern and retraction processes to flag work containing lapses.

Publishers cannot rely on library privacy policies to govern the ethics of library learning analytics work, since many academic libraries do not have a privacy policy or rely uncritically on that of their institution (Valentine & Barron, 2022; Mann et al., 2023). Moreover,

Institutional Review Boards (IRBs) and analogous ethics-review bodies outside the United States often refuse to review “assessment” or “quality assurance/control” initiatives, even when those may well do harm and even when the data or results will be made public (Metcalf & Crawford, 2016; Oliphant & Brundin, 2019). This can mean no institutional ethical review at all for library learning analytics projects considered by researcher and/or ethics-review body to be assessment or quality control. In the United States, IRBs also focus near exclusively on harm to study subjects from *data collection* specifically. They typically ignore not only harmful reuses of existing data but also other sources of harm to research subjects and communities at large, such as surveillance creep (Hope, 2018), uncontrolled data access and reuse proliferation (Metcalf, 2017), inference harms (Solow-Niederman, 2022), information hazards (Yoose, 2021), and harms to anyone not a study subject from the research or the data collection (Metcalf & Crawford, 2016; Metcalf, 2017). Finally, IRBs do not assess research plans for information privacy and confidentiality, which is not within their purview (Asher et al., 2018). In short, publishers cannot trust that learning analytics manuscripts or proposals have had adequate ethics scrutiny prior to submission.

The LIS literature has documented troubling indications of flawed methods and data-management practices in library learning analytics research to date, though admittedly on small article samples. Research methods employed can carry low statistical validity, and effect sizes for even statistically significant interventions can be quite small (Robertshaw & Asher, 2019). Data-management practices in library learning analytics research, when explained in detail at all, indicate low levels of data security as well as “incomplete anonymization, prolonged data retention, collection of a broad scope of sensitive information, lack of informed consent, and sharing of patron-identified information” (Briney, 2019, p. 1). Such faulty practices risk data leaks and breaches, as well as loss of trust and additional library anxiety from students. They can also make research subjects trivially reidentifiable, as can publishing quantitative data applying to a very small number of research subjects (“low n”), which has been directly observed in the library learning analytics literature with editors from the publishing journal refusing to address it (Briney, 2021).

Under an ethics-of-care (D’Ignazio & Klein, 2020) or contextual-integrity (Nissenbaum, 2010) approach to resolving data-related ethical dilemmas, student expectations around data would be sought prior to conducting research and assessment and broadly respected in relevant processes. Until fairly recently, however, students had not even been systematically consulted for their expectations regarding institutional and library learning analytics work, which likely explains much of the considerable variation in treatment of student research subjects in library learning analytics work that this study documents. Fortunately, the dearth of research into student perspectives has been somewhat rectified. Most consistently across studies, students expect to be told detailed specifics of the data collected from and about them

and how that data will be used (Pardo & Siemens, 2014; Schumacher & Ifenthaler, 2018; Jones et al., 2023). They also expect the opportunity to consent (or refuse consent) to specific collection and use practices, not considering blanket consent acceptable (Jones et al., 2023). Students have expressed that they felt their privacy would be violated if librarians had access to their personally identifiable information, especially if librarians then shared that data with the rest of the institution (Asher et al., 2022). They do not hold high hopes that library learning analytics will be of benefit to them, distrusting the reliability of a correlation between library use and academic achievement and disliking the paternalism of the approach (Gariepy, 2021). A focus-group study by Gray et al. (2022) starkly demonstrates how little institutional staff respect student beliefs and expectations: student subjects mentioned “consent; monitoring [of data access] and transparency” 202 times, staff subjects not even once; “access control” to the resulting data was mentioned by students 57 times, by staff only 16.

While students generally indicate a high level of trust in their institutions to use data for their benefit (Korir et al., 2022), this trust exists among substantial student ignorance of actual institutional, library, and vendor data practices (Jones et al., 2020a). Trust is also commonly modulated by the belief that fellow students with stricter privacy boundaries should have those boundaries respected (Jones et al., 2020a; Gariepy, 2021). Students express strong preferences for their data to be deidentified before use in research and assessment (Gariepy, 2021; Jones et al., 2023), though their near universal ignorance of reidentification attacks and inference harms suggests that they are taking refuge in a false sense of safety-in-anonymity.

Such ethics guidance as exists for learning analytics aligns fairly closely with the student expectations just discussed. In their guidelines, Slade & Tait (2019) suggest the following core ethical issues that student research subjects have also mentioned: transparency (including to students), data ownership and control (which should rest with students), communication, cultural values, inclusion, consent, and student agency and responsibility.

Ethics guidelines specific to publishers, such as those from the Committee on Publication Ethics (COPE; <https://publicationethics.org/>), focus on ethics specific to *publishing* processes—such as authorship assignment, data fabrication, conflicts of interest, plagiarism, and citation farming—rather than research and assessment methods and processes upstream from manuscript submission. They do not address information-privacy ethics at all. The most that COPE has to say about upstream ethics is that “[e]ditors should consider retracting a publication if... [i]t reports unethical research” (COPE Council, 2019). COPE has been working with FORCE11—an organization of scholars and information professionals promoting data and software curation and publication—on ethics guidelines specific to data publication, however (Lowenberg & Puebla, 2022); results so far could be tweaked into applying to

actual publications as well. More general data ethics guidelines such as those from the US General Services Administration (2019) suggest (emphasis added) “uphold[ing] applicable statutes, regulations, *professional practices*, and *ethical standards*” with privacy, confidentiality, and transparency specifically mentioned among those standards.

Once again, all this leaves publishers as the only stakeholders broadly capable of removing work that does not respect library privacy and confidentiality ethics, human-subjects ethics, data ethics, or students’ expectations regarding their data from the LIS literature, and disincentivizing future such work. To assist publishers in this endeavor, this study builds on and adds detail to the prior work of Briney (2019) and Robertshaw & Asher (2019), addressing a larger sample of the journal literature and expanding that sample to conference presentations, theses, and book sections. Its research questions are as follows:

RQ1. What types of (non-self-reported) data are being collected for analysis on student library patrons in library learning analytics publications and conference presentations?

RQ2. How much of this work documents human-subjects ethics review?

RQ3. How much of this work creates privacy risks for student patrons, particularly reidentifiability risks? How much of this work respects library information-privacy ethics?

RQ4. How much of this work respects what is currently known about student data-privacy expectations, specifically notice, consent, and use of only deidentified data?

RQ5. How much library learning analytics work has created concrete benefits for libraries performing it?

## METHODS

### Ethics statement

The author used the University of Wisconsin Human Research Determination worksheet to assess whether this study required IRB review. Because no private information about or bio-specimens from living persons would be collected in the course of the study, no review was required or sought.

The author did not search for, request, or examine data underlying any of the research projects considered for or actually included in this study, so as not to violate the privacy of student research subjects or the privacy and confidentiality obligations of the researchers and their organizations. In the case of one repositied dataset (Pattern, 2011) related to a project deemed study-eligible, the author examined only the metadata for the dataset, not the

dataset itself. Similarly, the author made no attempt to reidentify research subjects from any project considered or included, despite several instances in which reidentification was likely possible.

### Locating artifacts

An “artifact” for purposes of this study is a published, reposted, or openly Web-available document. Journal articles, book chapters or sections, theses, dissertations, reposted pre/post-prints, reposted gray literature, reposted datasets, and conference papers or slidedecks were all considered for inclusion. Videorecorded conference sessions were not, partly due to the difficulty of systematically locating them—they are not included in any of the databases searched for artifacts, and are rarely if ever cited in the published literature—and partly because watching them (in the absence of reliable transcripts) to assess the work against both inclusion criteria and the study’s phenomena of interest would have consumed infeasible amounts of time.

Artifact collection took place via citation chain aggregation (Cribbin, 2011). It began with searches of LISTA, Google Scholar, Web of Science, and Scopus for artifacts citing the original VAL report (Oakleaf, 2010). The resulting citation lists were imported into the Zotero citation manager and deduplicated. Citations to non-English artifacts were put in a separate folder for data publication. These searches turned out, incidentally, to include all journal articles studied in Briney (2019) and Robertshaw & Asher (2019). Completing the chain involved backward citation chasing (Haddaway et al., 2022), scanning literature reviews and works cited sections of especially citation-heavy documents to add more artifacts, including a few actually pre-dating the VAL report. Finally, to remediate gappy indexing coverage of recent work, proceedings of the Library Assessment Conference and the following journals rich in learning analytics work were examined for suitable artifacts:

- *Evidence-Based Library and Information Practice* (full archives, including research summaries pointing to articles published elsewhere; this journal seemed incompletely indexed, which suggested a full archive sweep would be wise)
- *portal: Libraries and the Academy* (2018–22, to catch articles possibly not yet indexed)
- *Journal of Academic Librarianship* (2018–22, to catch articles possibly not yet indexed)
- *Performance Management and Metrics* (2018–22, to catch articles possibly not yet indexed)

## Winnowing artifacts

This study excluded advocacy papers, methods papers, and similar, since such work does not directly use or report on any patron data and therefore cannot cause direct harm to patrons. (Whenever such papers advocate unethical methods or do not explain how to work ethically, they cause indirect harms; this study, however, was only designed to assess direct data-fueled damage to privacy, confidentiality, and student expectations.) Research projects relying exclusively on direct input from student research subjects, as with surveys, interviews, or focus groups, were also excluded, as these methods do not use behavioral-exhaust data—data collected *about*, rather than directly from, student library patrons—and are therefore, per VAL, not actually library learning analytics. Library learning analytics projects augmented with surveys, interviews, or focus groups, however, were included. In summary, to be eligible for this study, a research project must meet the following:

- Documented in at least one English-language non-video artifact,
- Involved analysis of new or stored data about students, where the data was *non-self-reported* (that is, not collected through direct researcher interaction with the student via survey, focus group, or interview), with that data held by the library, the institution, or both, and
- Involved a research question about the contribution of student library use to one or more measures of student success.

Projects that met all these criteria while also employing additional research methods (qualitative or quantitative) or asking additional research questions were included. Projects that failed to meet any criterion above were excluded; this did involve a few possibly incorrect judgment calls about what constituted “stored, non-self-reported data” or whether a project’s research question was aimed at measuring student success.

English-language artifacts from the initial set of gathered citations were first winnowed for eligibility by scanning abstracts and methods sections. Ineligible artifacts had their citations parked in separate Zotero folders (one for non-English artifacts, one for artifacts failing project-eligibility criteria) for data publication. To avoid overstating the prevalence of phenomena based on the same research appearing in multiple artifacts, the remaining artifacts were then grouped into projects based on whether they described substantially the same group of research subjects. In one case, one artifact described its dataset, whereas a later artifact used the same dataset augmented with new data; these artifacts were grouped into a single project.

Most projects were described in only one artifact, but some projects fueled multiple artifacts, as shown in [Table 1](#) below.

Artifact count	Number of projects with this count of associated artifacts
2	9
3	2
4	1
9	1
10	1
11	1

**Table 1.** Projects discussed in more than one artifact

Research projects and citations for artifacts associated with them were then entered into a purposebuilt SQLite database for coding. Eligible artifacts were closely read and coded project-by-project to characterize the project and the institution(s) where the project took place. A few times, this closer read determined that a given project did not actually meet inclusion criteria, so it was removed from the database along with its related artifacts, and its citation(s) were moved to the ineligible-project Zotero folder for data publication. The total number of projects deemed eligible for this study was 62, represented by 102 artifacts made available in or before October 2022, which are listed in the [Appendix](#). The types of literature represented by these 102 artifacts are shown in [Table 2](#).

Journal article	73
Conference paper or slidedeck	13
Reposited gray literature	7
Book, book chapter, or other book segment	4
Magazine (non-peer-reviewed) article	3
Dissertation or thesis	2

**Table 2.** Type of literature for artifacts eligible for this study

Sixteen journals published one eligible article each. Journals that published more than one eligible article are given in [Table 3](#).

College & Research Libraries	18
Journal of Academic Librarianship	10
Evidence Based Library and Information Practice	9
portal: Libraries and the Academy	8
Performance Measurement and Metrics	5
Journal of Library Administration	3
Australian Academic & Research Libraries	2
Library and Information Science Research	2

**Table 3.** Journals publishing artifacts eligible for this study

Eight eligible conference papers and slidedecks were presented at the Library Assessment Conference, with five other conferences hosting one artifact each. The magazine *Against the Grain* published two eligible non-peer-reviewed articles; *Educause Review* published the remaining one.

Of the 62 eligible projects, 56 took place in baccalaureate institutions, five in community/technical colleges, and one in K-12 schools. Five projects were undertaken with library-external funding, three of those from institutional or consortial sources and the other two from national grant programs. [Table 4](#) lists the number of eligible projects performed in specific countries; this represents a substantial undercount of international library learning analytics work because of the restriction of eligible projects to those with available English-language artifacts.

Country	Number of eligible projects
Australia	4
Chile	1
Hong Kong	1
Mongolia	1
South Africa	1
Spain	2
Turkey	2
United Kingdom	4
United States	46

**Table 4.** Countries in which eligible research projects were performed

## Coding

Coding involved enumerating the types of data collected and analyzed for each project as well as assessing whether each project employed a select group of potentially ethically challenging or privacy-damaging practices. Data types were each given a column in their own table. For most data types, a binary decision was recorded according to whether or not the project collected this type of data. A few data types were broken down in slightly more granular fashion; age, for example, could be coded as “date of birth,” “age [in years],” or “bracket” (meaning age was broken into brackets and each subject’s membership in a given age bracket recorded without recording their exact age). The data dictionary published at <https://osf.io/g6y3z/> contains all permitted values for all data types.

The following ethically challenging practices were coded in a separate table:

- Level of subject identification: Deidentified, aggregated, or fully identified
- Notice: Whether the project documented that subjects were notified of the research undertaken; possibilities were “yes,” “no,” “subjects were notified as part of a larger research project,” or “subjects were notified on receipt of their campus ID.”
- Informed consent: Whether the project documented that subjects were given the opportunity to consent or refuse the research undertaken
- Low n: Whether the project reported any results applying to 30 subjects or fewer, a practice that raises reidentifiability concerns
- Mixing library with nonlibrary data: Whether the project mixed data originating in student engagement with the library with data about students from other sources
- Library data outside library: Whether anyone not employed by the library had access to data originating in the library, which raises confidentiality concerns
- Subject of inquiry revealed: Whether any researcher had access to data indicating the subject(s) of student inquiry

## RESULTS AND DISCUSSION

### Data collected and analyzed

Table 5 answers RQ1 with counts of how many projects collected or analyzed specific categories of data. The sum of projects in the table is greater than 62 because most projects

analyzed several categories of data. Full descriptions of variables are available in the data dictionary at <https://osf.io/g6y3z/>. Race and ethnicity were lumped together because many projects conflated them.

Data category	Number of projects including it (n = 62)
Major or program undertaken (any granularity)	41 (66%)
Year within program	38 (61%)
Gender	34 (55%)
Race and/or ethnicity	27 (44%)
Library-instruction attendance (in any form: optional or required workshops, one-shot sessions, full library-taught courses)	26 (42%)
Circulation of physical library materials to specific students (includes circulation counts, interlibrary loan of physical items)	22 (35%)
Age (date of birth, age at time of data collection, or age bracket)	19 (31%)
Proxy-server data about electronic-resource use (includes login counts)	17 (27%)
Enrollment in a specific course	17 (27%)
Use of a library service not covered under circulation, proxy, or instruction (e.g., reference, study rooms, workstations)	17 (27%)
Socioeconomic status (any measure or proxy measure, student and/or family)	13 (21%)
High-school performance (grade point average, class standing, ACT/SAT or other standardized test performance, Advanced Placement credits, remedial coursework)	13 (21%)
Geographic data about students and/or their families (includes swipe-card data on library entry, student/family addresses, on/off-campus residency, use of workstations in a specific library)	7 (11%)
First-generation student status	6 (10%)
National origin and/or citizenship status of student and/or family	4 (6%)
Military and/or veteran status	3 (5%)
Disability status	1 (2%)

**Table 5.** Data categories collected and analyzed

**Table 6** counts how many of the projects aimed to assess data against specific student-success outcomes. The sum in this table is greater than 62 because several projects assessed more than one outcome.

Outcome	Number of projects including it (n = 62)
GPA or analogue (overall or for a limited time such as a term or a year)	47 (76%)
Retention (any duration)	23 (37%)
Grade in a specific for-credit course	15 (24%)
Graduation	11 (18%)
Performance on a specific test, exam, assignment, or other assessment	11 (18%)

**Table 6.** Outcomes assessed

### Ethics review

Analysis for RQ2 demonstrated that significant swathes of library learning analytics research do not document ethics review. Because ethics-review regimes differ significantly globally, this paper considers projects performed inside and outside the United States separately. Of the 46 eligible projects performed in the United States, 11 passed IRB review, 4 were declared exempt by their IRB, and the remaining 31 did not mention any institutional ethics review, though it is possible (Jess Schomberg, personal communication, August 11, 2022) that such mention was either not requested or indeed actively removed by publication outlets. Of the 16 projects performed outside the United States, only one mentioned an institutional ethics review. Of course, as discussed earlier, IRB and analogous institutional research ethics review does not consider library-specific risks to subjects, such as information-privacy risks. Given that no person or group systematically assesses library-specific ethical questions in research, it is fair to say that no project eligible for this study had a full and complete review of its ethics.

Even so, a few researchers thoroughly and thoughtfully addressed the ethics of their processes and analyses in their writing. One artifact of an eligible project (Wittkower et al., 2022), for example, contains a substantial ethics statement that is an exemplary model of careful research design and meticulous methods documentation. Another (LeMaistre et al., 2018) describes data fuzzing, deidentification, aggregation, and security practices in laudable detail.

### Privacy risks

Analysis for RQ3 showed that significant swathes of eligible projects create privacy risks for student patrons. In all but three projects, data originating from library use was mixed with data obtained from sources outside the library, a practice that deserves reconsideration in light of library confidentiality ethics. In 24 projects, data about patron library use was shared outside

the library, typically with a co-investigator outside the library, a graduate student, a student affairs office, or an institutional data office. Such sharing can fairly be considered a breach of library confidentiality.

Globally, 11 projects, eight of which fully identified student subjects, contained data documenting one or more subjects of patron inquiry through specific items circulated or specific databases or electronic materials consulted. In six of those projects, five fully identifying student subjects, such library data was shared outside the library. This practice, while it certainly contradicts the IFLA's injunction to avoid compromising patron privacy, is especially concerning in the United States because it violates a bright-line stipulation in the ALA's Privacy Interpretation (ALA, 2006, para. 3 [emphasis added]): "The right to privacy includes the right to open inquiry without having *the subject of one's interest* examined or scrutinized by others, in person or online." Among the 46 projects performed in the United States, eight documented subjects of patron inquiry, six of those on fully identified student subjects. Four of the eight shared data outside the library, three of those with fully identified student subjects. A softer version of the same problem occurs in the 41 projects recording student majors/programs and the 17 projects noting student enrollment in a specific course; these, too, are substantial hints to the subjects of student inquiry, and as such, library ethics codes indicate that they deserve full confidentiality.

### Reidentifiability of student patrons

As with transparency and consent, deidentification before data analysis has been demonstrated to be a baseline student expectation of library learning analytics, such that flouting it invites ethical challenge and student distrust. Ease of subject reidentification must also be taken into account, since in the event of a breach or leak, concerned students are unlikely to care whether data were not deidentified at all or merely *inadequately* deidentified.

**Deidentification and its documentation.** This study lumped together deidentification and anonymization claims from eligible projects for two reasons:

- Full anonymization—defined as “no research subject can be identified as being represented in the research data”—of reasonably high-quality and high-dimensional data is thought to be impossible (Narayanan & Felten, 2014; Rocher et al., 2019), short of sophisticated techniques such as differential privacy that no project in this study attempted.
- Many artifacts claiming deidentified or anonymized data contained only the bare claim (e.g., Castillo-Manzano et al., 2020; Wright, 2021), leaving it impossible to assess the quality, thoroughness, or efficacy of the deidentification process.

One project (Jara et al., 2017) performed deidentification through hashing subjects’ identifiers. This was known at the time to be inadequate to protect research subjects (Narayanan & Felten, 2014) because any attacker with dataset access who knows a research subject’s identifier can hash it to determine whether that subject is represented in the dataset. If the subject is indeed present, the attacker can easily learn the values of all other data variables pertaining to them. Best practice is to remove all direct identifiers and use an opaque subject identifier with no real-world connection to any research subject.

Table 7 counts claims about data deidentification from the 62 projects eligible for the study. Projects whose artifacts made no statements about data deidentification or aggregation were assumed not to have attempted either.

Claim	Eligible projects making this claim (n = 62); eligible US projects making this claim (n = 46)
Aggregate data only	2 (3%); 2 (4%)
No deidentification; fully identified data analyzed	35 (56%); 27 (59%)
Some form of deidentification employed	25 (40%); 17 (37%)

**Table 7.** Data deidentification claims

**Low-n and outlier groups.** Fifteen projects, 14 performed in the United States, published a data variable applying to 30 or fewer research subjects. This “low n” reporting in quantitative research greatly improves the odds that an individual research subject or group of subjects can be reidentified, and this or other data variables (possibly damaging ones) associated with them. Populations this small should either be merged with other populations if appropriate or omitted altogether from publication (Briney, 2021). The number 30 is admittedly somewhat arbitrary; this result should be taken more as an indicator of likely researcher and publisher carelessness with data and reporting than an absolute statement about subject reidentifiability. In contrast, a few projects (e.g., those represented by the artifacts Goodall & Pattern, 2011; Taylor, 2013; Wong & Cmor, 2017) documented their decisions to eliminate low-n variables from publication.

**Dragnet project enrollment.** In projects such as the University of Wollongong’s Library Cube (represented in the artifacts Jantti & Cox, 2010a, 2010b, 2013; Cox & Jantti, 2012a, 2012b, 2015; Jantti, 2014, 2016; Jantti & Heath, 2016) and the University of Minnesota cohort project (represented in the artifacts Nackerud et al., 2012, 2013, 2015; Oakleaf et al., 2013; Soria, 2013; Soria et al., 2013, 2014, 2017a, 2017b; Fransen & Peterson, 2016), subjects are easily reidentifiable as having taken part in the project because the project employed neither sampling methods nor consent practices, instead silently enrolling every

student in a substantial fully described cohort such as an entering class. This style of dragnet enrollment, especially without any attempt at deidentification or aggregation, may be ethically acceptable in strictly internal, fully confidential assessment—though even that deserves considered discussion in light of students expecting consent processes—but is likely inappropriate for publication, as it violates student expectations of deidentification and consent, and may (depending on the variables collected, analyzed, and published) expose students to inference harms (Solow-Niederman, 2022).

**The Sweeney test and similar reidentifiability tests.** Sweeney (2000) famously determined that nearly nine in ten Americans were uniquely identifiable based on three indirect identifiers: date of birth, residential ZIP code, and gender (binary assumed); with date of birth fuzzed to birth year, over seven in ten remained uniquely identifiable. Six projects in this study contained the Sweeney trifecta—age in years or exact date of birth, gender, and some form of geolocation. Ten more contained age/date of birth and gender, which—given that residential college students usually live near the college—may suffice to complete the Sweeney test. Two projects contained subject IP addresses, which can often be used to geolocate the network-connected device, especially combined with a timestamp.

As has been noted with respect to academic library patron data (Briney, 2021), the Sweeney trifecta is far from the only combination of variables likely to lead to reidentification of small, often vulnerable, groups of students or even individual students. Twelve projects in this study contained data on students' race/ethnicity, gender, year of study, and major/program; six more contained race/ethnicity, gender, and major/program. The exact reidentification risk of these combinations of variables in student populations is not presently known, but in the United States, numerical-minority status makes the following groups of students more reidentifiable when relevant variables are included in the data:

- students of color at predominantly white institutions;
- international students;
- first-generation students;
- non-traditional-aged undergraduates;
- transgender, non-binary, or agender students; and
- students in highly gendered fields of study who are not of the most common gender in that field.

One artifact representing a project eligible for this study (Jantti & Cox, 2010a) calls the likelihood of reidentification by variable combination “highly unlikely” despite reporting no effort to

assess its possibility in the dataset. Another artifact documenting the same project (Jantti & Cox, 2010b) exhibits a data-query interface (on fully identified data, to boot) that appears to make combined-variable queries quite simple to perform, casting considerable further doubt on the unlikelihood of singling out vulnerable groups or individual subjects via indirect identifiers.

### Respecting student expectations

**Transparency and consent.** Multiple studies discussed in the literature review illustrate students' clear expectation for researcher and assessor transparency about data collection and analysis and for student opportunity to issue, refuse, and withdraw consent. To those following ethics-of-care tenets, notification and consent practices are the only courteous way to proceed while respecting student agency and autonomy and reducing power imbalances between researchers and students. They may also help allay student concerns and even encourage research participation (Michalak & Rysavy, 2019). Allowing research and assessment reporting that violates these expectations to remain in the LIS literature casts LIS publishers in a poor ethical light, even considering that much of this research was performed before these expectations were known, as expressions of concern and retractions are viable publisher tools for retroactive literature correction.

Analysis for RQ4 indicates that, of the 62 eligible projects, only 11 informed students of the specific research being undertaken using their data; one more stated that students were informed that research would take place using their data when they signed up for their campus identification cards. Only five projects obtained informed consent from student subjects.

Of the 12 projects that underwent institutional ethics review, only three both informed student subjects of the specific research project and obtained informed consent from them. None of the projects declared exempt either notified student subjects or obtained informed consent from them. This strongly suggests an unaddressed disconnect between phenomena considered during ethics review and student expectations of notice and consent.

**Collection and use of sensitive data.** Exactly which student data variables count as “sensitive” differs across sources of guidance. The US educational-records law Family Educational Rights and Privacy Act (FERPA), of course, considers grades and other records of achievement to be sensitive. The National Information Standards Organization (NISO) mentions “race, gender, socioeconomic class, [and] ability” in its Privacy Principles (NISO, 2015). As yet, students have not been broadly and systematically canvassed for their perceptions of data sensitivity, particularly around campus-specific data, though one survey-based study (Asher et al., 2022) showed students placing financial information at the top of their privacy expectations, with physical and

mental health just below it. While use of library materials (print or electronic) was at the bottom of the list, a full quarter of students surveyed still considered it “completely private.”

Table 8 counts the projects containing variables mentioned above as potentially sensitive.

Type of sensitive data	Number of projects analyzing it (n = 62)
Grades (including GPA or analogue)	52 (84%)
Race and/or ethnicity	27 (44%)
Gender	34 (55%)
Socioeconomic status/financial information	13 (21%)
Disability	1 (2%)

**Table 8.** Sensitive data present in eligible projects

Table 9 counts the projects containing the above variables in data for which no attempt at deidentification or aggregation was documented.

Type of sensitive data	Number of projects not documenting deidentification methods that analyze it (n = 35)
Grades (including GPA or analogue)	29 (83%)
Race and/or ethnicity	15 (43%)
Gender	21 (60%)
Socioeconomic status/financial information	6 (17%)
Disability	0

**Table 9.** Sensitive data present without any documentation of deidentification in eligible projects

Also worth noting is the presence in the projects studied of additional data variables at least some students are likely to consider sensitive: citizenship or international status, first-generation status, veteran/military status, and high-school performance measures.

**Efficacy claims.** The stated *raison d’être* for VAL (Oakleaf, 2018) is earning continued and even additional support for libraries from institutions and their administrators. This already shades the ethics of the endeavor, because under present-day formulations of research ethics, ethical research must operate from a principle of benevolence toward research subjects rather than the self-interest of researchers or their organizations (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research, 1979).

It is nonetheless striking that the answer to RQ5 is that only five projects (8%) of the 62 in this study claimed any concrete present or past benefit to the library from undertaking the research.

No library claiming a benefit stated that the benefit was additional monetary or staff resources. Remaining claims were all couched in the future tense or the conditional mood—benefits “will” or “could” or “may” or “might” accrue, but none actually had at time of publication. These findings at minimum suggest that retracting articles and conference presentations found to contain ethical lapses will not seriously damage librarianship’s ability to advocate for itself. They also indicate a need for evidence-based reconsideration of VAL as a principled and efficacious method of library advocacy, especially in light of the ethics challenges pointed out in this study.

### **Suggestions for related work**

**Learning analytics ethics.** Inclusion criteria for this study were quite narrow; expanded studies would be welcome, as would multilingual and international studies. Qualitative studies of how authors, editors, and peer reviewers approach ethics questions during research and publication could elucidate how best to guarantee an ethics-minded LIS literature. This study also occludes the role of library and institutional consortia, as well as library-technology, edtech, and content vendors, in learning analytics. Research into student perceptions of learning analytics (with or without library involvement) once they are told how vendors and consortia are involved, and what difference that makes to how far their data travels and how many people have access to it, is urgently needed if librarians are to negotiate contracts and perform research and assessment while protecting student interests and honoring their expectations. Researchers interested in additional investigation into student expectations for learning analytics, inside or outside libraries, now have a concrete list of data variables and practices known to exist in the literature to consult with students about; assessing the perceived sensitivity of the data variables identified in this study would be worthwhile.

Research into the reidentifiability of deidentified patron data (after [Sweeney, 2000](#) and [Narayanan & Shmatikov, 2008](#)) would be valuable in helping libraries, researchers, and publishers contend with present-day threats to information privacy. Addressing this question in the academic library context must take into account other campus information stores, given the spreading availability of single-sign-on authentication, campus and third-party data warehouses, learning analytics, data dashboards, website tracker use, and similar data-intensive practices in higher education. Attention to the applicability of more sophisticated methods of deidentification and reidentifiability assessment such as differential privacy and *k*-anonymity to library patron data would be useful as well, especially to publishers wishing to improve author guidance.

A more interventionist research agenda could couple education on risks and harms of data collection and analysis in general and learning analytics in particular with assessment of student attitudinal change toward learning analytics practices; a pre-test/post-test methodology could elicit important insights. Exploiting student ignorance of data-related risks and harms is of course

questionably ethical at best, so the more that is known about how students' beliefs and expectations change once they understand more about data-related risks and harms, the better.

**Patron data in LIS research publications.** Library learning analytics is not the only LIS research and assessment modality presently relying on analysis of patron information and service-use data. Future content-analysis studies could examine the LIS literature for the ethics and practices of research and assessment use of library website trackers, proxy-server data, reference chat logs, integrated library system analytics suites such as (but not limited to) Alma Analytics, or customer-relationship management software such as (but not limited to) OCLC Wise or OrangeBoy. Papers that only described methods without actually implementing them were ineligible for this study; a future study could assess attention to information privacy and ethics in data-centric methods papers.

LIS publishing professionals, many of whom are librarians themselves, can reduce the flow of submissions containing unethical exploitation of patron data by advocating for change in academic libraries, master's and doctoral programs in LIS, research funders, and professional organizations. Academic libraries without privacy policies can follow the ACRL's clear injunction to create them (ALA, 2006); privacy policies can also be communicated to students and institutional data stakeholders. Academic libraries whose privacy policies do not address patron data use in assessment and research can update them to do so. LIS education at the master's and especially doctoral levels can review how it addresses patron data in research methods courses, inculcating appropriate confidentiality and privacy practices based on library ethics codes. LIS professional and research organizations can create clear guidance on best practices and unacceptable practices for collection, storage, deidentification, security, and destruction of patron data in research and assessment. LIS research funders can ensure through their review processes that research out of line with such best practices is not funded.

Publishers of LIS journals and conference proceedings face additional complex challenges, however (Luna, 1997), especially given the dearth of ethics guidance and processes upstream from them. Because their research-dissemination role is highly visible and strongly associated with quality control, they risk more public blame for promulgating ethically dubious research than do libraries, LIS programs, or funders. Unfortunately, even with the advocacy suggested earlier, publishers cannot prevent research lacking due attention to information privacy and ethics from being submitted to them, nor can they retroactively instill best practices into a project lacking them. They can, however, concentrate on preventing ethically dubious work from achieving publication and, should this fail, removing it via retraction afterward. Outlets employing developmental editing and similar mentoring processes can add addressing ethical challenges to their research-design assistance work, and every journal and conference

promulgating research using patron data could benefit from well-communicated expectations. The COPE/FORCE11 work on the ethics of data publication (Lowenberg & Puebla, 2022) could be a fruitful source of viable expectations. Another example of progress is the guidance revision for library learning analytics work by the editorial board of *portal: Libraries and the Academy* (Jones, 2021), which requests that learning analytics researchers explicitly examine the ethics of their methods and processes in their manuscripts.

Other possibilities for publishers wishing to improve publication ethics include the following:

- Requiring author manuscripts for any work involving human subjects to include an ethics statement documenting all ethical reviews the project received and attesting clearance for all necessary ethics reviews and data-governance processes required by their library, institution, and any involved funders. Should the statement prove false, retraction becomes a much simpler decision for the editors.
- Establishing specific and binding policy for the journal or conference with respect to research and assessment use of patron data in general and data from vulnerable patron populations such as students in particular. Any submission not in line with this policy must receive a desk rejection.
- Adding ethics and data-handling concerns to author guidance and peer-review rubrics. Helpful guidance could include acceptable sampling methods (disallowing dragnet enrollment), notification and consent, data minimization, data collection, secure storage, data deidentification, data destruction, low-n variable publication, combination of patron data with library-external data sources, and when it is acceptable (or not) to share patron data beyond the library.
- Establishing a process separate from peer review for ethics review of submissions, particularly submissions with associated datasets.
- Reviewing policies and processes for expressions of concern and retractions to ensure that ethics and data-handling lapses are grounds for editorial action on already published materials. Reviewing procedures around reporting submissions to institutional and funder research-integrity and data-governance offices may also be warranted.
- Adding an ethics watchdog and dataset reviewers to editorial boards, especially important in light of the patron-data breach at Code4Lib Journal (2023). Procedures for choosing people to serve in this capacity should guard against subversion of the venue's commitment to ethics and security.

What LIS publishers should do about the parlous state of the existing literature as documented in this study is a fraught question. Ideally, authors who recognize that their work was careless with

data or did not treat student patrons with due respect would request retractions themselves. Author-initiated retractions, especially of work that took place before research was published documenting student expectations for their behavioral data exhaust, are highly praiseworthy and should absolutely *not* count against authors in tenure and promotion processes, as they document those authors' strong commitment to ethical responsibility. Lacking that, journal and conference editors should consider placing expressions of concern on, or even retracting, articles and conference papers documenting inadequate or unethical processes. Although this will cause regrettable career damage to authors, leaving the work alone creates an information hazard (Yoose, 2021) inviting more researchers to treat students and their information privacy carelessly. Even more important is fully retracting methods and advocacy papers that recommend research or assessment using student data without sufficient (or often any) attention to ethics, privacy and confidentiality, data security, and student expectations (e.g., Dillalogue & Koehn, 2020; Hart & Amos, 2018; and especially Zucca, 2013). For these, an expression of concern cannot mend the harm of misleading researchers into believing that suboptimal practices are acceptable.

Leaving individual researchers to work through ethical questions for themselves has not as yet succeeded, despite recently published resources for ethical reflection (Jones, 2021; Young, 2022). Without additional interventions, it seems unlikely to fare better in the future, given the gaps this study exposes in library privacy policies, human-subjects ethics review, institutional and library data-governance processes, and journal and conference review. These gaps leave authors exposed to significant risk of retractions, expressions of concern, and other career-damaging negative actions on their publications. To protect them best, systems throughout LIS can institute the systemic reforms listed in detail earlier to respond appropriately to ethically questionable practices and, ideally, prevent ethics problems altogether. Until that happens, however, publishers have little choice but to shore up their ethics-related editing, review, and retraction processes as discussed earlier if they are to maintain their reputation for research integrity and librarianship's commitments to patron privacy and confidentiality.

## CONCLUSION

It is clear from this study that much library learning analytics research to date has not been performed in ways that respect student autonomy, dignity, and agency while adhering to human-subjects research ethics processes, library privacy and confidentiality ethics, and best practices in data management and data security. A few projects studied, however, equally clearly demonstrate that it is at least *possible* for learning analytics research to do so.

Unethical, risky, insecure, and/or exploitative use of student data in LIS research and assessment publications can only continue at a significant cost to librarianship's trust in its publishers and to patron trust in librarianship's stated privacy and confidentiality commitments. Since

the ACRL is broadly responsible for substantial growth in library learning analytics research due to its publication and promotion of VAL, it seems fitting that the ACRL undertake the necessary work to rein in the associated problems documented in this study, especially considering this study's finding that VAL does not appear to have created significant benefits for libraries employing it.

## ACKNOWLEDGMENTS

The author is grateful to David Fiander for information on ethics-review processes in Canada, to beta readers who do not wish to be named for valuable feedback, and to anonymous peer reviewers for a wealth of well-considered feedback and commentary that improved this article immensely. Remaining errors and infelicities are of course the author's.

**Data statement:** The dataset underlying this study has been made available at <https://osf.io/g6y3z/>. This includes RIS files exported from Zotero for artifacts studied as well as artifacts representing projects that did not meet inclusion criteria. The SQLite database used for coding and analysis is there with associated data dictionary. For data reusers' convenience and some protection from file-format obsolescence, CSV files were also exported for each table and view in the database.

## REFERENCES

ALA (American Library Association) Council. (2021, February 25). *Resolution on the misuse of behavioral data surveillance in libraries*. <https://www.ala.org/advocacy/intfreedom/datasurveillanceresolution>

ALA (American Library Association). (2006, July 26). *Intellectual freedom principles for academic libraries: An interpretation of the Library Bill of Rights*. <http://www.ala.org/advocacy/intfreedom/librarybill/interpretations/intellectual>

Asher, A. (2017). Risk, benefits, and user privacy: Evaluating the ethics of library data. In B. Newman & B. Tijerina (Eds.), *Protecting patron privacy: A LITA guide* (pp. 43–56). Rowman & Littlefield. [https://scholarworks.iu.edu/dspace/bitstream/handle/2022/22035/Asher-Risk\\_Benefits\\_User\\_Privacy\\_Final.pdf;sequence=1](https://scholarworks.iu.edu/dspace/bitstream/handle/2022/22035/Asher-Risk_Benefits_User_Privacy_Final.pdf;sequence=1)

Asher, A., Briney, K. A., Gardner, G. J., Hinchliffe, L. J., Nowviskie, B., Salo, D., & Shorish, Y. (2018). *Ethics in research use of library patron data: Glossary and explainer*. <https://doi.org/10.17605/OSF.IO/XFKZ6>

Asher, A., Briney, K. A., Jones, K. M. L., Regalado, M., Perry, M. R., Goben, A., Smale, M. A., & Salo, D. (2022). Questions of trust: A survey of student expectations and perspectives on library learning analytics. *Library Quarterly*, 92(2), 151–171. <https://doi.org/10.1086/718605>

Briney, K. A. (2019). Data management practices in academic library learning analytics: A critical review. *Journal of Librarianship and Scholarly Communication*, 7(1), Article 1. <https://doi.org/10.7710/2162-3309.2268>

Briney, K. A. (2021, April 1). *Citation omitted: A story of re-identification*. Data Ab Initio. <http://dataabinitio.com/?p=1012>

- Canadian Federation of Library Associations. (2016, November 15). *Statement on intellectual freedom and libraries*. <http://cfla-fcab.ca/en/guidelines-and-position-papers/statement-on-intellectual-freedom-and-libraries/>
- Castillo-Manzano, J. I., González-Fernández-Villavicencio, N., Lopez-Valpuesta, L., & Pozo-Barajas, R. (2020). When the road to academic success goes through the library: A case study at the University of Seville. *Journal of Academic Librarianship*, 46(4), 102162. <https://doi.org/10.1016/j.acalib.2020.102162>
- Code4Lib Journal. (2023). Extra Editorial: On the release of patron data in Issue 58 of Code4Lib Journal. *code4lib*, 58. <https://journal.code4lib.org/articles/18040>
- Collier, K. (2022, December 20). Ransomware hackers take demands directly to college students: ‘For you, it’s a sad day.’ *NBC News*. <https://www.nbcnews.com/tech/security/ransomware-hackers-take-demands-directly-college-students-s-sad-day-rcna61253>
- COPE Council. (2019). *COPE guidelines: Retraction guidelines*. <https://doi.org/10.24318/cope.2019.1.4>
- Cox, B., & Jantti, M. (2012a). Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement. *Library and Information Science Research*, 34(4), 308–316. <https://doi.org/10.1016/j.lisr.2012.06.002>
- Cox, B., & Jantti, M. (2012b, July 17). *Discovering the impact of library use and student performance*. Educause Review. <https://er.educause.edu/articles/2012/7/discovering-the-impact-of-library-use-and-student-performance>
- Cox, B., & Jantti, M. (2015). The Library Cube: Revealing the impact of library use on student performance (University of Wollongong). In B. Showers (Ed.), *Library analytics and metrics: Using data to drive decisions and services* (pp. 66–74). Facet Publishing.
- Cribbin, T. F. (2011). Citation chain aggregation: An interaction model to support citation cycling. *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 2149–2152. <https://doi.org/10.1145/2063576.2063913>
- D’Ignazio, C., & Klein, L. (2020, March 16). Our values and our metrics for holding ourselves accountable. In *Data feminism*. MIT Press. <https://data-feminism.mitpress.mit.edu/pub/3hxx4l8o/release/2>
- Dillalogue, E., & Koehn, M. (2020). Assessment on a dime: Low cost user data collection for assessment. *Evidence Based Library and Information Practice*, 15(3), Article 3. <https://doi.org/10.18438/ebliip29582>
- Doty, P. (2020). Library analytics as moral dilemmas for academic librarians. *The Journal of Academic Librarianship*, 46(4), 102141. <https://doi.org/10.1016/j.acalib.2020.102141>
- Fisher, Z., Magnus, E., & Branch, N. (2019, March 13). *Critical assessment practices: A discussion on when and how to use student learning data without doing harm* [Presentation]. Santa Clara University Scholar Commons. <https://scholarcommons.scu.edu/library/205>
- Fransen, J., & Peterson, K. (2016). Graduate in four years? Yes, the library can help with that! *Proceedings of the 2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*, 474–484. <http://old.libraryassessment.org/bm~doc/77-fransen-2016.pdf>

- Gariepy, L. W. (2021). Acceptable and unacceptable uses of academic library search data: An interpretive description of undergraduate student perspectives. *Evidence Based Library and Information Practice*, 16(2), 22–44. <https://doi.org/10.18438/eblip29923>
- Goodall, D., & Pattern, D. (2011). Academic library non/low use and undergraduate student achievement: A preliminary report of research in progress. *Library Management*, 32(3), 159–170. <https://doi.org/10.1108/01435121111112871>
- Gray, G., Schalk, A. E., Cooke, G., Murnion, P., Rooney, P., & O'Rourke, K. C. (2022). Stakeholders' insights on learning analytics: Perspectives of students and staff. *Computers & Education*, 187, 104550. <https://doi.org/10.1016/j.compedu.2022.104550>
- Haddaway, N. R., Grainger, M. J., & Gray, C. T. (2022). Citationchaser: A tool for transparent and efficient forward and backward citation chasing in systematic searching. *Research Synthesis Methods*, 13(4), 533–545. <https://doi.org/10.1002/jrsm.1563>
- Hart, S., & Amos, H. (2018). The library assessment capability maturity model: A means of optimizing how libraries measure effectiveness. *Evidence Based Library and Information Practice*, 13(4), 31–49. <https://doi.org/10.18438/eblip29471>
- Hathcock, A. (2018, January 24). *Learning agency, not analytics*. At The Intersection. <https://aprilhathcock.wordpress.com/2018/01/24/learning-agency-not-analytics/>
- Hope, A. (2018). Creep: The growing surveillance of students' online activities. *Education and Society*, 36(1), 55–72. <https://doi.org/10.7459/es/36.1.05>
- Huang, C., Samek, T., & Shiri, A. (2021). AI and ethics: Ethical and educational perspectives for LIS. *Journal of Education for Library and Information Science*, 62(4), 351–365. <https://doi.org/10.3138/jelis-62-4-2020-0106>
- IFLA (International Federation of Library Associations and Institutions). (2015, August 14). *IFLA statement on privacy in the library environment*. International Federation of Library Associations and Institutions. <https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/>
- Jantti, M. (2014). *Aspiring to excellence: Maximising data to sustain, shift, and reshape a library for the future* [Presentation]. 2014 Library Assessment Conference, Association of Research Libraries, Seattle, WA. <https://www.libraryassessment.org/wp-content/uploads/bm~doc/1keynotejanti.pdf>
- Jantti, M. (2016). Libraries and big data: A new view on impact and affect. In J. Atkinson (Ed.), *Quality and the academic library* (pp. 267–273). Chandos Publishing. <https://doi.org/10.1016/B978-0-12-802105-7.00026-9>
- Jantti, M., & Cox, B. (2010a). Measuring the value of library resources and student academic performance through relational datasets. In S. Hiller, K. Justh, M. Kyrillidou, & J. Self (Eds.), *Proceedings of the 2010 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment* (pp. 525–532). <http://old.libraryassessment.org/bm~doc/proceedings-lac-2010.pdf>
- Jantti, M., & Cox, B. (2010b). *The Library Cube* [Conference session]. Library Assessment Conference, Baltimore, MD. [https://www.libraryassessment.org/wp-content/uploads/bm~doc/jantti\\_margie.pdf](https://www.libraryassessment.org/wp-content/uploads/bm~doc/jantti_margie.pdf)

- Jantti, M., & Cox, B. (2013). Measuring the value of library resources and student academic performance through relational datasets. *Evidence Based Library and Information Practice*, 8(2), Article 2. <https://doi.org/10.18438/B8Q89F>
- Jantti, M., & Heath, J. (2016). What role for libraries in learning analytics? *Performance Measurement and Metrics*, 17(2), 203–210. <https://doi.org/10.1108/PMM-04-2016-0020>
- Jara, M., Clasing, P., González, C., Montenegro, M., Kelly, N., Alarcón, R., Sandoval, A., & Saurina, E. (2017). Patterns of library use by undergraduate students in a Chilean university. *Portal*, 17(3), 595–615. <https://doi.org/10.1353/pla.2017.0036>
- Jones, K. M. L. (2021). Library learning analytics: Addressing the relationship between professional, research, and publication ethics. *Portal*, 21(3), 417–423. <https://doi.org/10.1353/pla.2021.0022>
- Jones, K. M. L., Asher, A., Goben, A., Perry, M. R., Salo, D., Briney, K. A., & Robertshaw, M. B. (2020a). “We’re being tracked at all times”: Student perspectives of their privacy in relation to learning analytics in higher education. *Journal of the Association for Information Science and Technology*, 71(9), 1044–1059. <https://doi.org/10.1002/asi.24358>
- Jones, K. M. L., Briney, K. A., Goben, A., Salo, D., Asher, A., & Perry, M. R. (2020b). A comprehensive primer to library learning analytics practices, initiatives, and privacy issues. *College & Research Libraries*, 81(3). <https://doi.org/10.5860/crl.81.3.570>
- Jones, K. M. L., Goben, A., Perry, M. R., Regalado, M., Salo, D., Asher, A., Smale, M. A., & Briney, K. A. (2023). Transparency and consent: Student perspectives on educational data analytics scenarios. *portal: Libraries and the Academy*, 23, 485–515.
- Jones, K. M. L., & Hinchliffe, L. J. (2020, October). *New methods, new needs: Preparing academic library practitioners to address ethical issues associated with learning analytics* [Conference session]. Annual Meeting of the Association for Library and Information Science Education (ALISE), online.
- Jones, K. M. L., & Salo, D. (2018). Learning analytics and the academic library: Professional ethics commitments at a crossroads. *College & Research Libraries*, 79(3), 304–323. <https://doi.org/10.5860/crl.79.3.304>
- Korir, M., Slade, S., Holmes, W., Héliot, Y., & Rienties, B. (2022). Investigating the dimensions of students’ privacy concern in the collection, use, and sharing of data for learning analytics. *Computers in Human Behavior Reports*, 100262. <https://doi.org/10.1016/j.chbr.2022.100262>
- Lamdan, S. (2022). *Data cartels: The companies that control and monopolize our information*. Stanford University Press.
- LeMaistre, T., Shi, Q., & Thanki, S. (2018). Connecting library use to student success. *Portal*, 18(1), 117–140. <https://doi.org/10.1353/pla.2018.0006>
- Lowenberg, D., & Puebla, I. (2022, April 7). *Joint FORCE11 & COPE Research Data Publishing Ethics Working Group Policy Templates*. <https://zenodo.org/records/6422102>

Luna, F. (1997). Vulnerable populations and morally tainted experiments. *Bioethics*, 11(3–4), 256–264. <https://doi.org/10.1111/1467-8519.00064>

Mann, E. Z., Jacobs, S. A., Kinsley, K. M., & Spears, L. I. (2023). Tracking transparency: An exploratory review of Florida academic library privacy policies. *Information and Learning Sciences*, 124(9/10), 285–305. <https://doi.org/10.1108/ILS-04-2023-0038>

Metcalf, J. (2017, November 30). “The study has been approved by the IRB”: Gayface AI, research hype and the pervasive data ethics gap. Medium. <https://medium.com/pervade-team/the-study-has-been-approved-by-the-irb-gayface-ai-research-hype-and-the-pervasive-data-ethics-ed76171b882c>

Metcalf, J., & Crawford, K. (2016). Where are human subjects in Big Data research? The emerging ethics divide. *Big Data & Society*, 3(1), 205395171665021. <https://doi.org/10.1177/2053951716650211>

Michalak, R., & Rysavy, M. D. T. (2019). Data privacy and academic libraries: Non-PII, PII, and librarians’ reflections (part 2). *Journal of Library Administration*, 59(7), 768–785. <https://doi.org/10.1080/01930826.2019.1649969>

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2013). Analyzing demographics: Assessing library use across the institution. *Portal*, 13(2), 131–145. <https://doi.org/10.1353/pla.2013.0017>

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2015). Retention, student success and academic engagement at Minnesota (University of Minnesota). In B. Showers (Ed.), *Library analytics and metrics: Using data to drive decisions and services*. Facet Publishing.

Nackerud, S., Fransen, J., Peterson, K., Mastel, K., Soria, K., & Peterson, D. (2012, March 14). *Library data and student success* [Presentation]. Library Technology Conference, Minneapolis, MN. [https://digitalcommons.maclester.edu/libtech\\_conf/2012/sessions/28](https://digitalcommons.maclester.edu/libtech_conf/2012/sessions/28)

Narayanan, A., & Felten, E. W. (2014, July 9). No silver bullet: De-identification still doesn’t work. <http://randomwalker.info/publications/no-silver-bullet-de-identification.pdf>

Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. *2008 IEEE Symposium on Security and Privacy (Sp 2008)*, 111–125. <https://doi.org/10.1109/SP.2008.33>

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *Ethical principles and guidelines for the protection of human subjects of research*. <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/read-the-belmont-report/index.html>

NISO. (2015, December 10). *NISO consensus principles on users’ digital privacy in library, publisher, and software-provider systems (NISO Privacy Principles)*. National Information Standards Organization. <https://www.niso.org/publications/privacy-principles>

Nissenbaum, H. F. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford Law Books.

Oakleaf, M. (2010). *Value of academic libraries: A comprehensive research review and report*. Association of College and Research Libraries. <http://hdl.handle.net/11213/17187>

Oakleaf, M. (2018). The problems and promise of learning analytics for increasing and demonstrating library value and impact. *Information and Learning Science*, 119(1–2), 16–24. <https://doi.org/10.1108/ILS-08-2017-0080>

Oakleaf, M., Stone, G., Pattern, D., Bowles-Terry, M., Peterson, K., Nackerud, S., & Fransen, J. (2013, April 13). *Do or do not...there is no try: The quest for library value* [Conference]. Association of College & Research Libraries Conference, Indianapolis, Indiana. <https://eprints.hud.ac.uk/id/eprint/16421/>

Oliphant, T., & Brundin, M. R. (2019). Conflicting values: An exploration of the tensions between learning analytics and academic librarianship. *Library Trends*, 68(1), 5–23. <https://doi.org/10.1353/lib.2019.0028>

Pardo, A., & Siemens, G. (2014). Ethical and privacy principles for learning analytics. *British Journal of Educational Technology*, 45(3), 438–450. <https://doi.org/10.1111/bjjet.12152>

Pattern, D. (2011). *Library Impact Data Project (LIDP) data* [Data set]. *Library Impact Data Project*. <https://library.hud.ac.uk/archive/projects/lidp/>

*Patron privacy at Santa Cruz Public Libraries: Trust and transparency in the age of data analytics*. (2019, June 24). [http://www.co.santa-cruz.ca.us/Portals/0/County/GrandJury/GJ2019\\_final/SantaCruzPublicLibrariesReport.pdf](http://www.co.santa-cruz.ca.us/Portals/0/County/GrandJury/GJ2019_final/SantaCruzPublicLibrariesReport.pdf)

Robertshaw, M. B., & Asher, A. (2019). Unethical numbers? A meta-analysis of library learning analytics studies. *Library Trends*, 68(1), 76–101. <https://doi.org/10.1353/lib.2019.0031>

Rocher, L., Hendrickx, J. M., & de Montjoye, Y.-A. (2019). Estimating the success of re-identifications in incomplete datasets using generative models. *Nature Communications*, 10(1), Article 1. <https://doi.org/10.1038/s41467-019-10933-3>

Salo, D. (2021). Physical-equivalent privacy. *The Serials Librarian*, 1–15. <https://doi.org/10.1080/0361526X.2021.1875962>

Schumacher, C., & Ifenthaler, D. (2018). Features students really expect from learning analytics. *Computers in Human Behavior*, 78(January), 397–407. <https://doi.org/10.1016/j.chb.2017.06.030>

Slade, S., & Tait, A. (2019, March). *Global guidelines: Ethics in learning analytics* (pp. 1–16). International Council for Open and Distance Education. <https://static1.squarespace.com/static/5b99664675f9eea7a3ecce82/t/5ca37c2a24a694a94e0e515c/1554218087775/Global+guidelines+for+Ethics+in+Learning+Analytics+Web+ready+March+2019.pdf>

Solow-Niederman, A. (2022). Information privacy and the inference economy. *Northwestern University Law Review*, 117(2), 357–424. <https://scholarlycommons.law.northwestern.edu/nulr/vol117/iss2/1>

Soria, K. M. (2013). Factors predicting the importance of libraries and research activities for undergraduates. *Journal of Academic Librarianship*, 39(6), 464–470. <https://doi.org/10.1016/j.acalib.2013.08.017>

Soria, K. M., Fransen, J., & Nackerud, S. (2013). Library use and undergraduate student outcomes: New evidence for students' retention and academic success. *Portal*, 13(2), 147–164. <https://doi.org/10.1353/pla.2013.0010>

Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *Journal of Academic Librarianship*, 40(1), 84–91. <https://doi.org/10.1016/j.acalib.2013.12.002>

Soria, K. M., Fransen, J., & Nackerud, S. (2017a). Beyond books: The extended academic benefits of library use for first-year college students. *College & Research Libraries*, 78, No. 1. <https://doi.org/10.5860/crl.78.1.8>

Soria, K. M., Fransen, J., & Nackerud, S. (2017b). The impact of academic library resources on undergraduates' degree completion. *College & Research Libraries*, 78(6), 812–823. <https://doi.org/10.5860/crl.78.6.812>

Sweeney, L. (2000). *Simple demographics often identify people uniquely*. <http://dataprivacylab.org/projects/identifiability/paper1.pdf>

Taylor, M. A. (2013, May). *Academic library use and undergraduate engagement and persistence* [Doctoral dissertation, University of Texas]. <https://repositories.lib.utexas.edu/handle/2152/21864>

US General Services Administration. (2019). *Federal data strategy: Data ethics framework* (pp. 1–28). <https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf>

Valentine, G., & Barron, K. (2022). An examination of academic library privacy policy compliance with professional guidelines. *Evidence Based Library and Information Practice*, 17(3), Article 3. <https://doi.org/10.18438/eblip30122>

Viberg, O., Mutimukwe, C., & Grönlund, Å. (2022). Privacy in LA research: Understanding the field to improve the practice. *Journal of Learning Analytics*, 9(3), Article 3. <https://doi.org/10.18608/jla.2022.7751>

Witt, S. (2017). The evolution of privacy within the American Library Association, 1906–2002. *Library Trends*, 65(4), 639–658. <https://doi.org/10.1353/lib.2017.0022>

Wittkower, L. R., McInnis, J. W., & Pope, D. R. (2022). An examination of relationships between library instruction and student academic achievement. *Journal of Library Administration*, 1–12. <https://doi.org/10.1080/01930826.2022.2117953>

Wong, S. H. R., & Cmor, D. (2017). Measuring association between library instruction and graduation GPA. *College & Research Libraries*, 72, No 5. <https://doi.org/10.5860/crl-151>

Wright, L. B. (2021). Assessing library instruction: A study of the relationship between attendance, retention, and student success. *Journal of Academic Librarianship*, 47(5). <https://doi.org/10.1016/j.acalib.2021.102431>

Yoose, B. (2021, September 27). *Is library scholarship a privacy information hazard?* LDH Consulting Services. <https://ldhconsultingservices.com/is-library-scholarship-a-privacy-information-hazard/>

Young, S. (2022, October 25). *A theoretical framework and practical toolkit for ethical library assessment* [Doctoral dissertation, Humboldt-Universität zu Berlin]. <https://doi.org/10.18452/25372>

Zucca, J. (2013). Business intelligence infrastructure for academic libraries. *Evidence Based Library and Information Practice*, 8(2), Article 2. <https://doi.org/10.18438/B83G75>

## APPENDIX: CITATIONS FOR ELIGIBLE STUDIES

- Albelda, B., & de la Mano, M. (2017). Libraries contribution to the acquisition of reading literacy at primary and secondary school in Spain: An approach based on international assessment studies. *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*, 1–15.
- Allison, D. (2015). Measuring the academic impact of libraries. *Portal: Libraries and the Academy*, 15(1), 29–40. <https://doi.org/10.1353/pla.2015.0001>
- Asher, A. (2017). *Evaluating the effect of course-specific library instruction on student success*. <https://scholarworks.iu.edu/dspace/handle/2022/21277>
- Barkey, P. (1965). Patterns of student use of a college library. *College & Research Libraries*, 26(2). [https://doi.org/10.5860/crl\\_26\\_02\\_115](https://doi.org/10.5860/crl_26_02_115)
- Beile, P., Choudhury, K., Mulvihill, R., & Wang, M. (2020). Aligning library assessment with institutional priorities: A study of student academic performance and use of five library services. *College & Research Libraries*, 81(3), 435–458. <https://doi.org/10.5860/crl.81.3.435>
- Beile, P., Choudhury, K., & Wang, M. C. (2017). Hidden treasure on the road to Xanadu: What connecting library service usage data to unique student ids can reveal. *Journal of Library Administration*, 57(2), 151–173. <https://doi.org/10.1080/01930826.2016.1235899>
- Black, E. L., & Murphy, S. A. (2017). The out loud assignment: Articulating library contributions to first-year student success. *Journal of Academic Librarianship*, 43(5), 409–416. <https://doi.org/10.1016/j.acalib.2017.06.008>
- Bowles-Terry, M. (2012). Library instruction and academic success: A mixed-methods assessment of a library instruction program. *Evidence Based Library and Information Practice*, 7(1), 82–95. <https://doi.org/10.18438/b8ps4d>
- Castillo-Manzano, J. I., González-Fernández-Villavicencio, N., Lopez-Valpuesta, L., & Pozo-Barajas, R. (2020). When the road to academic success goes through the library: A case study at the University of Seville. *Journal of Academic Librarianship*, 46(4), 102162. <https://doi.org/10.1016/j.acalib.2020.102162>

- Catalano, A. J., & Phillips, S. R. (2016). Information literacy and retention: A case study of the value of the library. *Evidence Based Library and Information Practice*, 11(4), Article 4. <https://doi.org/10.18438/B82K7W>
- Çetin, Y., & Howard, V. (2016). An exploration of the relationship between undergraduate students' library book borrowing and academic achievement. *Journal of Librarianship and Information Science*, 48(4), 382–388. <https://doi.org/10.1177/0961000615572404>
- Çetin, Y., & Kinay, H. (2011). A causal relationship between Turkish EFL students' academic achievement and borrowed library books. *Gaziantep University Journal of Social Sciences*, 10(4), Article 4. <https://dergipark.org.tr/en/pub/jss/issue/24241/256987>
- Cherry, E., Rollins, S. H., & Evans, T. (2013). Proving our worth: The impact of electronic resource usage on academic achievement. *College & Undergraduate Libraries*, 20(3–4), 386–398. <https://doi.org/10.1080/10691316.2013.829378>
- Chiteng Kot, F., & Jones, J. L. (2015). The impact of library resource utilization on undergraduate students' academic performance: A propensity score matching design. *College & Research Libraries*, 76(5), 566–586. <https://doi.org/10.5860/crl.76.5.566>
- Collins, E., & Stone, G. (2014). Understanding patterns of library use among undergraduate students from different disciplines. *Evidence Based Library and Information Practice*, 9(3), 51–67. <https://doi.org/10.18438/b8930k>
- Cook, J. M. (2014). A library credit course and student success rates: A longitudinal study. *College & Research Libraries*, 75(3), 272–283. <https://doi.org/10.5860/crl12-424>
- Corlett, D. (1974). Library skills, study habits and attitudes, and sex as related to academic achievement. *Educational and Psychological Measurement*, 34(4), 967–969. <https://doi.org/10.1177/001316447403400430>
- Coulter, P., Clarke, S., & Scamman, C. (2007). Course grade as a measure of the effectiveness of one-shot information literacy instruction. *Public Services Quarterly*, 3(1–2), 147–163. [https://doi.org/10.1300/J295v03n01\\_08](https://doi.org/10.1300/J295v03n01_08)
- Cox, A., Gruber, A. M., & Neuhaus, C. (2019). Complexities of demonstrating library value: An exploratory study of research consultations. *Portal: Libraries and the Academy*, 19(4), 577–590. <https://doi.org/10.1353/pla.2019.0036>

- Cox, B., & Jantti, M. (2012a). Capturing business intelligence required for targeted marketing, demonstrating value, and driving process improvement. *Library and Information Science Research*, 34(4), 308–316. <https://doi.org/10.1016/j.lisr.2012.06.002>
- Cox, B., & Jantti, M. (2012b, July 17). Discovering the impact of library use and student performance. *Educause Review*. <https://er.educause.edu/articles/2012/7/discovering-the-impact-of-library-use-and-student-performance>
- Cox, B., & Jantti, M. (2015). The Library Cube: Revealing the impact of library use on student performance (university of wollongong). In B. Showers (Ed.), *Library Analytics and Metrics: Using data to drive decisions and services* (pp. 66–74). Facet Publishing.
- Davidson, K. S., Rollins, S. H., & Cherry, E. (2013). Demonstrating our value: Tying use of electronic resources to academic success. *The Serials Librarian*, 65(1), 74–79. <https://doi.org/10.1080/0361526X.2013.800630>
- De Groote, S. L., & Scoulas, J. M. (2022). The impact of the academic library on students' success, in their own words. *Portal: Libraries and the Academy*, 22(2), 355–374. <https://doi.org/10.1353/pla.2022.0021>
- de Jager, K. (2001). Impacts & outcomes: Searching for the most elusive indicators of academic library performance. *Proceedings of the 4th Northumbria International Conference on Performance Measurement in Libraries and Information Services*, 291–298. [https://www.libqual.org/documents/admin/4np\\_secure.pdf](https://www.libqual.org/documents/admin/4np_secure.pdf)
- de Jager, K. (2002). Successful students: Does the library make a difference? *Performance Measurement and Metrics*, 3(3), 140–144. <https://doi.org/10.1108/14678040210453564>
- de Jager, K., Nassimbeni, M., Daniels, W., & D'Angelo, A. (2018). The use of academic libraries in turbulent times: Student library behaviour and academic performance at the University of Cape Town. *Performance Measurement and Metrics*, 19(1), 40–52. <https://doi.org/10.1108/PMM-09-2017-0037>
- Fransen, J., & Peterson, K. (2016). Graduate in four years? Yes, the library can help with that! *Proceedings of the 2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*, 474–484. <http://old.libraryassessment.org/bm~doc/77-fransen-2016.pdf>

Gaha, U., Hinnefeld, S., & Pellegrino, C. (2018). The academic library's contribution to student success: Library instruction and GPA. *College & Research Libraries*, 79(6), 737–746. <https://doi.org/10.5860/crl.79.6.737>

Goodall, D., & Pattern, D. (2011). Academic library non/low use and undergraduate student achievement: A preliminary report of research in progress. *Library Management*, 32(3), 159–170. <https://doi.org/10.1108/01435121111112871>

Gores, J., Springfield, C., & Coan, M. (2017). How data from an unlikely source pointed us in a new direction for information literacy assessment and charted a course to student success. *At the Helm: Leading Transformation*, 283–291.

Haddow, G. (2013). Academic library use and student retention: A quantitative analysis. *Library & Information Science Research*, 35(2), 127–136. <https://doi.org/10.1016/j.lisr.2012.12.002>

Haddow, G., & Joseph, J. (2010). Loans, logins, and lasting the course: Academic library use and student retention. *Australian Academic & Research Libraries*, 41(4), 233–244. <https://doi.org/10.1080/00048623.2010.10721478>

Heady, C., Morrison, M. M., & Vossler, J. (2018). Ecological study of graduation rates and GPA in a library credit course. *Journal of Academic Librarianship*, 44(5), 642–649. <https://doi.org/10.1016/j.acalib.2018.07.010>

Hill, L., Maier-Katkin, D., Ladny, R., & Kinsley, K. (2018). When in doubt, go to the library: The effect of a library-intensive freshman research and writing seminar on academic success. *Journal of Criminal Justice Education*, 29(1), 116–136. <https://doi.org/10.1080/10511253.2017.1372498>

International Federation of Library Associations and Institutions. (2015, August 14). *IFLA Statement on Privacy in the Library Environment*. International Federation of Library Associations and Institutions. <https://www.ifla.org/publications/ifla-statement-on-privacy-in-the-library-environment/>

Jantti, M. (2014). Aspiring to excellence: Maximising data to sustain, shift, and reshape a library for the future. *Proceedings of the 2014 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*, 15–22. <https://www.libraryassessment.org/wp-content/uploads/bm~doc/1keynotejanti.pdf>

Jantti, M. (2016). Libraries and big data: A new view on impact and affect. In J. Atkinson (Ed.), *Quality and the Academic Library* (pp. 267–273). Chandos Publishing. <https://doi.org/10.1016/B978-0-12-802105-7.00026-9>

Jantti, M., & Cox, B. (2010a). Measuring the value of library resources and student academic performance through relational datasets. *Proceedings of the 2010 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*, 525–532. <http://old.libraryassessment.org/bm~doc/proceedings-lac-2010.pdf>

Jantti, M., & Cox, B. (2010b). *The Library Cube*. Library Assessment Conference, Baltimore, MD. [https://www.libraryassessment.org/wp-content/uploads/bm~doc/jantti\\_margie.pdf](https://www.libraryassessment.org/wp-content/uploads/bm~doc/jantti_margie.pdf)

Jantti, M., & Cox, B. (2013). Measuring the value of library resources and student academic performance through relational datasets. *Evidence Based Library and Information Practice*, 8(2), Article 2. <https://doi.org/10.18438/B8Q89F>

Jantti, M., & Heath, J. (2016). What role for libraries in learning analytics? *Performance Measurement and Metrics*, 17(2), 203–210. <https://doi.org/10.1108/PMM-04-2016-0020>

Jara, M., Clasing, P., González, C., Montenegro, M., Kelly, N., Alarcón, R., Sandoval, A., & Saurina, E. (2017). Patterns of library use by undergraduate students in a Chilean university. *Portal*, 17(3), 595–615. <https://doi.org/10.1353/pla.2017.0036>

Jones, W. L., & Mastroilli, T. (2022). Assessing the impact of an information literacy course on students' academic achievement: A mixed-methods study. *Evidence Based Library and Information Practice*, 17(2), Article 2. <https://doi.org/10.18438/eblip30090>

Kelly, N., Montenegro, M., González, C., Jara, M., Alarcón, R., Saurina, E., & Cano, F. (2014). Exploring the relationship between undergraduate students' use of library resources and learning outcomes. *Proceedings of Library Assessment Conference. Building Effective, Sustainable, Practical Assessment. Oregon, Washington 4th-6th August*. Library Assessment Conference, Oregon. <https://www.libraryassessment.org/wp-content/uploads/bm~doc/13saurinadiscussionession.pdf>

Killick, S., Nurse, R., & Clough, H. (2019). The continuing adventures of library learning analytics: Exploring the relationship between library skills training and student success. *Proceedings of the 2018 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment: December 5–7, 2018, Houston, TX*, 188–199. <https://doi.org/10.29242/lac.2018.15>

Kohler, E., & Stovall, C. (2019). Mining EZProxy data: User demographics and electronic resources. *Proceedings of the 2018 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment: December 5–7, 2018, Houston, TX*, 728–741. <https://doi.org/10.29242/lac.2018.68>

Kramer, L. A., & Kramer, M. B. (1968). The college library and the drop-out. *College & Research Libraries*. <http://crl.acrl.org/crl/article/download/12049/13495>

Krieb, D. (2018). Assessing the impact of reference assistance and library instruction on retention and grades using student tracking technology. *Evidence Based Library and Information Practice*, 13(2), 2–12. <https://doi.org/10.18438/eblip29402>

Laskin, M., & Zoe, L. (2017). *Information literacy and institutional effectiveness: A longitudinal analysis of performance indicators of student success*. CUNY Academic Works. [https://academicworks.cuny.edu/ho\\_pubs/60/](https://academicworks.cuny.edu/ho_pubs/60/)

LeMaistre, T., Shi, Q., & Thanki, S. (2018). Connecting library use to student success. *Portal*, 18(1), 117–140. <https://doi.org/10.1353/pla.2018.0006>

Mao, J., & Kinsley, K. (2017). Embracing the generalized propensity score method: Measuring the effect of library usage on First-Time-In-College student academic success. *Evidence Based Library and Information Practice*, 12(4), 129–157. <https://doi.org/10.18438/B8BH35>

Massengale, L., Piotrowski, P., & Savage, D. (2017). *Identifying and articulating library connections to student success*. <https://doi.org/10.5860/crl.77.2.227>

McCarthy, S. C. (2017). At issue: Exploring library usage by online learners with student success. *The Community College Experience*, 23(2), 27–31.

McDonald, J. D., & Levine-Clark, M. (2020). Library learning analytics: A proof of concept study. *Against the Grain*, 32(1). <https://www.charleston-hub.com/2020/04/v321-library-learning-analytics-a-proof-of-concept-study/>

Michalak, R., & Rysavy, M. D. T. (2019). Data privacy and academic libraries: Non-PII, PII, and librarians' reflections (part 2). *Journal of Library Administration*, 59(7), 768–785. <https://doi.org/10.1080/01930826.2019.1649969>

Michalak, R., & Rysavy, M. D. T. (2020). Data privacy and information literacy assessment: A case study. *Against the Grain*, 32(1). <https://www.charleston-hub.com/2020/04/v321-data->

[privacy-and-information-literacy-assessment-a-case-study/](https://www.charleston-hub.com/2020/04/v321-data-privacy-and-information-literacy-assessment-a-case-study/), <https://www.charleston-hub.com/2020/04/v321-data-privacy-and-information-literacy-assessment-a-case-study/>

Mohr, S. M., Ziegler, A., Lehan, T. J., & Wengel, J. (2022). The road less traveled: Assessing the impact of virtual research consultations on online doctoral student persistence. *Internet Reference Services Quarterly*. <https://doi.org/10.1080/10875301.2022.2072043>

Montenegro, M., Clasing, P., Kelly, N., Gonzalez, C., Jara, M., Alarcón, R., Sandoval, A., & Saurina, E. (2016). Library resources and students' learning outcomes: Do all the resources have the same impact on learning? *Journal of Academic Librarianship*, 42(5), 551–556. <https://doi.org/10.1016/j.acalib.2016.06.020>

Moore, D., Brewster, S., Dorroh, C., & Moreau, M. (2002). Information competency instruction in a two-year college: One size does not fit all. *Reference Services Review*, 30(4), 300–306. <https://doi.org/10.1108/00907320210451286>

Murray, A., Ireland, A., & Hackathorn, J. (2016). The value of academic libraries: Library services as a predictor of student retention. *College and Research Libraries*, 77(5), 631–642. <https://doi.org/10.5860/crl.77.5.631>

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2013). Analyzing demographics: Assessing library use across the institution. *Portal*, 13(2), 131–145. <https://doi.org/10.1353/pla.2013.0017>

Nackerud, S., Fransen, J., Peterson, K., & Mastel, K. (2015). Retention, student success and academic engagement at Minnesota (University of Minnesota). In B. Showers (Ed.), *Library Analytics and Metrics: Using data to drive decisions and services*. Facet Publishing.

Nackerud, S., Fransen, J., Peterson, K., Mastel, K., Soria, K., & Peterson, D. (2012, March 14). *Library data and student success*. Library Technology Conference, Minneapolis, MN. [https://digitalcommons.macalester.edu/libtech\\_conf/2012/sessions/28](https://digitalcommons.macalester.edu/libtech_conf/2012/sessions/28)

Nurse, R., Baker, K., & Gambles, A. (2018). Library resources, student success and the distance-learning university. *Information and Learning Science*, 119(1/2), 77–86. <https://doi.org/10.1108/ILS-03-2017-0022>

Oakleaf, M., Stone, G., Pattern, D., Bowles-Terry, M., Peterson, K., Nackerud, S., & Fransen, J. (2013, April 13). *Do or do not...there is no try: The quest for library value*

[Conference]. Association of College & Research Libraries Conference, Indianapolis, Indiana. <https://eprints.hud.ac.uk/id/eprint/16421/>

O’Kelly, M. (2015, November 16). Correlation between library instruction and student retention. *Presentations*. Southeastern Library Assessment Conference, Atlanta, GA. [https://scholarworks.gvsu.edu/library\\_presentations/55](https://scholarworks.gvsu.edu/library_presentations/55)

O’Kelly, M. (2016). Academic libraries and student retention: The implications for higher education. *2016 Library Assessment Conference: Building Effective, Sustainable, Practical Assessment*. Library Assessment Conference, Arlington, VA. <http://libraryassessment.org/archive/conference-proceedings-2016.shtml>

Pattern, D. (2011). *Library Impact Data Project Data* [dataset]. <https://library.hud.ac.uk/archive/projects/lidp/>

Purevsuren, T., Enkhsaruul, M., Adiya, G., & Davaa, A. (2021). Impact of academic library on student learning achievement in engineering field. *Learner Centered Learning 2020*. [https://www.ibs-laubusch.de/wp-content/uploads/2021/03/IBS\\_Band11.pdf#page=132](https://www.ibs-laubusch.de/wp-content/uploads/2021/03/IBS_Band11.pdf#page=132)

Rowe, J., Leuzinger, J., Hargis, C., & Harker, K. R. (2021). The impact of library instruction on undergraduate student success: A four-year study. *College and Research Libraries*, 82(1), 7–18. <https://doi.org/10.5860/crl.82.1.7>

Russell, F., Rawson, C., Freestone, C., Currie, M., & Kelly, B. (2018). Parallel lines: A mixed methods impact analysis of co-curricular digital literacy online modules on student results in first-year nursing. *COLLEGE & RESEARCH LIBRARIES*, 79(7), 948–971. <https://doi.org/10.5860/crl.79.7.948>

Samson, S. (2014). Usage of e-resources: Virtual value of demographics. *Journal of Academic Librarianship*, 40(6), 620–625. <https://doi.org/10.1016/j.acalib.2014.10.005>

Scott, M. (2014). Interlibrary loan article use and user GPA: Findings and implications for library services. *Journal of Access Services*, 11(4), 229–238. <https://doi.org/10.1080/15367967.2014.945116>

Scoulas, J. M. (2021). STEM undergraduate students: Library use, perceptions and GPA. *Performance Measurement and Metrics*, 22(2), 137–148. <https://doi.org/10.1108/PMM-04-2020-0021>

Scoulas, J. M., & De Groote, S. L. (2019). The library's impact on university students' academic success and learning. *Evidence Based Library and Information Practice*, 14(3), 2–27. <https://doi.org/10.18438/eblip29547>

Selegean, J. C., Thomas, M. L., & Richman, M. L. (2017). *Long-range effectiveness of library use instruction (research note)*. [https://doi-org.ezproxy.library.wisc.edu/10.5860/crl\\_44\\_06\\_476](https://doi-org.ezproxy.library.wisc.edu/10.5860/crl_44_06_476)

Shao, X., & Purpur, G. (2016). Effects of information literacy skills on student writing and course performance. *The Journal of Academic Librarianship*, 42(6), 670–678. <https://doi.org/10.1016/j.acalib.2016.08.006>

Showers, B., & Stone, G. (2014). Safety in numbers: Developing a shared analytics service for academic libraries. *Performance Measurement and Metrics*, 15(1/2), 13–22. <https://doi.org/10.1108/PMM-03-2014-0008>

Soria, K. M. (2013). Factors predicting the importance of libraries and research activities for undergraduates. *Journal of Academic Librarianship*, 39(6), 464–470. <https://doi.org/10.1016/j.acalib.2013.08.017>

Soria, K. M., Fransen, J., & Nackerud, S. (2013). Library use and undergraduate student outcomes: New evidence for students' retention and academic success. *Portal*, 13(2), 147–164. <https://doi.org/10.1353/pla.2013.0010>

Soria, K. M., Fransen, J., & Nackerud, S. (2014). Stacks, serials, search engines, and students' success: First-year undergraduate students' library use, academic achievement, and retention. *Journal of Academic Librarianship*, 40(1), 84–91. <https://doi.org/10.1016/j.acalib.2013.12.002>

Soria, K. M., Fransen, J., & Nackerud, S. (2017a). The impact of academic library resources on undergraduates' degree completion. *College and Research Libraries*, 78(6), 812–823. <https://doi.org/10.5860/crl.78.6.812>

Soria, K. M., Fransen, J., & Nackerud, S. (2017b). *Beyond books: The extended academic benefits of library use for first-year college students*. <https://doi.org/10.5860/crl.78.1.8>

Squibb, S. D., & Mikkelsen, S. (2017). *Assessing the value of course-embedded information literacy on student learning and achievement*. <https://doi.org/10.5860/crl.77.2.164>

Stemmer, J. K., & Mahan, D. M. (2015). Assessing the library's influence on freshman and senior level outcomes with user surveys. *Evidence Based Library and Information Practice*, 10(2), 8–20. <https://doi.org/10.18438/b8pg62>

Stemmer, J. K., & Mahan, D. M. (2016). Investigating the relationship of library usage to student outcomes. *College and Research Libraries*, 77(3), 359–375. <https://doi.org/10.5860/crl.77.3.359>

Stone, G. (2015). Library impact data: Investigating library use and student attainment. In *Library Analytics and Metrics: Using data to drive decisions and services*. Facet Publishing.

Stone, G., Pattern, D., & Ramsden, B. (n.d.). *Library Impact Data Project* (p. 4).

Stone, G., Pattern, D., & Ramsden, B. (2011). Does library use affect student attainment? A preliminary report on the library impact data project. *LIBER Quarterly: The Journal of the Association of European Research Libraries*, 21(1), Article 1. <https://doi.org/10.18352/lq.8005>

Stone, G., Pattern, D., & Ramsden, B. (2012). *Library Impact Data Project: Hit, miss or maybe* (I. Hall, S. Thornton, & S. Town, Eds.; pp. 385–390). University of York. <http://www.york.ac.uk/media/abouttheuniversity/supportservices/informationdirectorator/documents/9th%20Northumbria%20Conference%20Proceedings.pdf>

Stone, G., & Ramsden, B. (2017). *Library Impact Data Project: Looking for the link between library usage and student attainment*. <https://doi.org/10.5860/crl12-406>

Stone, G., Sharman, A., Dunn, P., & Woods, L. (2015). Increasing the impact: Building on the library impact data project. *Journal of Academic Librarianship*, 41(4), 517–520. <https://doi.org/10.1016/j.acalib.2015.06.003>

Taylor, M. A. (2013). *Academic library use and undergraduate engagement and persistence* [University of Texas]. <https://repositories.lib.utexas.edu/handle/2152/21864>

Thorpe, A., Lukes, R., Bever, D. J., & He, Y. (2016). The impact of the academic library on student success: Connecting the dots. *Portal*, 16(2), 373–392. <https://doi.org/10.1353/pla.2016.0027>

Vance, J., Kirk, R., & Gardner, J. (2012). Measuring the impact of library instruction on freshmen success and persistence: A quantitative analysis. *Communications in Information Literacy*, 6(1). <https://doi.org/10.15760/comminfolit.2012.6.1.117>

- Wells, J. (1995). The influence of library usage on undergraduate academic success. *Australian Academic & Research Libraries*, 26(2), 121–128. <https://doi.org/10.1080/00048623.1995.10754923>
- White, S., & Stone, G. (2010). Maximizing use of library resources at the University of Huddersfield. *Serials*, 23(2), Article 2. <https://doi.org/10.1629/2383>
- Wittkower, L. R., McInnis, J. W., & Pope, D. R. (2022). An examination of relationships between library instruction and student academic achievement. *Journal of Library Administration*, 1–12. <https://doi.org/10.1080/01930826.2022.2117953>
- Wong, S. H. R., & Cmor, D. (2017). *Measuring association between library instruction and graduation GPA*. <https://doi.org/10.5860/crl-151>
- Wong, S. H. R., & Webb, T. D. (2017). *Uncovering meaningful correlation between student academic performance and library material usage*. <https://doi.org/10.5860/crl-129>
- Wright, L. B. (2019). *Library instruction's relationship to student success* [PhD Thesis]. Valdosta State University.
- Wright, L. B. (2021). Assessing library instruction: A study of the relationship between attendance, retention, and student success. *Journal of Academic Librarianship*, 47(5). <https://doi.org/10.1016/j.acalib.2021.102431>