Abstract

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

**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 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: Trust and Transparency in the Age of Data Analytics, 2019).

As an illustration of the hazards of publishing research employing dubiously ethical methods, this piece explores 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.

Extensive discussion of library learning analytics alongside 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; Doty, 2020; Huang et al., 2021; Jones & Salo, 2018; Oliphant & Brundin, 2019) 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' (IFLA) 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) Statement on Intellectual Freedom in 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.” 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 IFLA's Statement on Privacy (2015), which notes that “[e]xcessive data collection and use threatens individual users' privacy and has other social and legal consequences.” Similarly, a recent 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).

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 IFLA, ACRL, or ALA (Salo, 2021). 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, nor the temptation 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.

IFLA’s, CFLA’s, and ALA’s 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 abovementioned 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 their institution’s (Valentine & Barron, 2022; Mann et al., 2023). Moreover, Institutional Review Boards (IRBs) and analogous ethics-review bodies outside the US
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 US, 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, 2017; Metcalf & Crawford, 2016). 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). 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 (Jones et al., 2023; Pardo & Siemens, 2014; Schumacher & Ifenthaler, 2018). 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 (Gariepy, 2021; Jones et al., 2020a). 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 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 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:

**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 biospecimens 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 reposed 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 where reidentification was likely possible.

Locating artifacts

An “artifact” for purposes of this study is a published, reposed, or openly Web-available document. Journal articles, book chapters or sections, theses, dissertations, reposed pre/postprints, reposed gray literature, reposed 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 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 have:

- Been 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 the above criteria while also employing additional research methods (qualitative or quantitative) or asking additional research questions were included. Projects that failed 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, while 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.

<table>
<thead>
<tr>
<th>Artifact count</th>
<th>Number of projects with this count of associated artifacts</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
</tr>
</tbody>
</table>

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) moved to the ineligible-project Zotero folder for data publication. The total number of projects deemed eligible for this study is 62, represented by 102 artifacts made available in or before October 2022, which are listed in the appendix.
The 102 artifacts eligible for this study were of the following types:

<table>
<thead>
<tr>
<th>Type of literature</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Journal article</td>
<td>73</td>
</tr>
<tr>
<td>Conference paper or slidedeck</td>
<td>13</td>
</tr>
<tr>
<td>Reposited gray literature</td>
<td>7</td>
</tr>
<tr>
<td>Book, book chapter, or other book segment</td>
<td>4</td>
</tr>
<tr>
<td>Magazine (non-peer-reviewed) article</td>
<td>3</td>
</tr>
<tr>
<td>Dissertation or thesis</td>
<td>2</td>
</tr>
</tbody>
</table>

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

Sixteen journals published one eligible article each. The following journals published more than one eligible article:

<table>
<thead>
<tr>
<th>Journal Name</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>College &amp; Research Libraries</td>
<td>18</td>
</tr>
<tr>
<td>Journal of Academic Librarianship</td>
<td>10</td>
</tr>
<tr>
<td>Evidence Based Library and Information Practice</td>
<td>9</td>
</tr>
<tr>
<td>portal: Libraries and the Academy</td>
<td>8</td>
</tr>
<tr>
<td>Performance Measurement and Metrics</td>
<td>5</td>
</tr>
<tr>
<td>Journal of Library Administration</td>
<td>3</td>
</tr>
<tr>
<td>Australian Academic &amp; Research Libraries</td>
<td>2</td>
</tr>
<tr>
<td>Library and Information Science Research</td>
<td>2</td>
</tr>
</tbody>
</table>

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, 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.

<table>
<thead>
<tr>
<th>Country</th>
<th>Number of eligible projects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>4</td>
</tr>
<tr>
<td>Chile</td>
<td>1</td>
</tr>
<tr>
<td>Hong Kong</td>
<td>1</td>
</tr>
<tr>
<td>Mongolia</td>
<td>1</td>
</tr>
<tr>
<td>South Africa</td>
<td>1</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
</tr>
<tr>
<td>Turkey</td>
<td>2</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>4</td>
</tr>
<tr>
<td>United States</td>
<td>46</td>
</tr>
</tbody>
</table>

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

**Coding**

Coding involved enumerating the types of data collected and analyzed for each project, and 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.

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

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

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, four 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) 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 above, 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 patrons’ 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 IFLA’s injunction to avoid compromising patron privacy, is especially concerning in the United States because it violates a bright-line stipulation in ALA’s Privacy Interpretation (American Library Association, 2006, 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.

<table>
<thead>
<tr>
<th>Claim</th>
<th>Eligible projects making this claim (n=62) / Eligible US projects making this claim (n=46)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate data only</td>
<td>2 (3%) / 2 (4%)</td>
</tr>
<tr>
<td>No deidentification; fully-identified data analyzed</td>
<td>35 (56%) / 27 (59%)</td>
</tr>
<tr>
<td>Some form of deidentification employed</td>
<td>25 (40%) / 17 (37%)</td>
</tr>
</tbody>
</table>

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 Cox & Jantti, 2012a, 2012b, 2015; Jantti, 2014, 2016; Jantti & Cox, 2010a, 2010b, 2013; Jantti & Heath, 2016) and the University of Minnesota cohort project (represented in the artifacts Fransen & Peterson, 2016; Nackerud et al., 2012, 2013, 2015; Oakleaf et al., 2013; Soria, 2013; Soria et al., 2013, 2014, 2017a, 2017b), 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 considerable 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 US, 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 commonest 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 FERPA, of course, considers grades and other records of achievement to be sensitive. 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.
<table>
<thead>
<tr>
<th>Type of sensitive data</th>
<th>Number of projects analyzing it (n=62)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grades (including GPA or analogue)</td>
<td>52 (84%)</td>
</tr>
<tr>
<td>Race/ethnicity</td>
<td>27 (44%)</td>
</tr>
<tr>
<td>Gender</td>
<td>34 (55%)</td>
</tr>
<tr>
<td>Socioeconomic status / financial information</td>
<td>13 (21%)</td>
</tr>
<tr>
<td>Disability</td>
<td>1 (2%)</td>
</tr>
</tbody>
</table>

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.

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

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, and in no case was the benefit claimed 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 Ph.D programs in LIS, research funders, and professional organizations. Academic libraries without privacy policies can follow ACRL’s clear injunction to create them (American Library Association, 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 above, publishers cannot prevent research lacking due attention to information privacy and ethics from being submitted to them, nor can they retroactively install 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 afterwards. 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:

- 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 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 above to respond appropriately to questionably-ethical 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 above 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 ACRL is broadly responsible for substantial growth in library learning analytics research due to its publication and promotion of VAL, it seems fitting that 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.

Works cited


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


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


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

Appendix: Citations for eligible studies


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


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


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


