A Scoping Review of Text-Matching Software Used for Student Academic Integrity in Higher Education

Hayden, K. Alix; Eaton, Sarah Elaine; Pethrick, Helen; Crossman, Katherine; Lenart, Bartlomiej A.; Penaluna, Lee-Ann

http://dx.doi.org/10.1155/2021/4834860

Journal Article

Downloaded from PRISM: https://prism.ucalgary.ca
Review Article

A Scoping Review of Text-Matching Software Used for Student Academic Integrity in Higher Education

K. Alix Hayden, Sarah Elaine Eaton, Helen Pethrick, Katherine Crossman, Bartlomiej A. Lenart, and Lee-Ann Penaluna

University of Calgary, 2500 University Dr. NW, Calgary, Alberta T2N 1N4, Canada

Correspondence should be addressed to Sarah Elaine Eaton; seaton@ucalgary.ca

Received 13 July 2021; Revised 5 August 2021; Accepted 11 August 2021; Published 21 September 2021

Academic Editor: Balamuralithara Balakrishnan

Copyright © 2021 K. Alix Hayden et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Text-matching software has been used widely in higher education to reduce student plagiarism and support the development of students’ writing skills. This scoping review provides insights into the extant literature relating to commercial text-matching software (TMS) (e.g., Turnitin) use in postsecondary institutions. Our primary research question was “How is text-matching software used in postsecondary contexts?” Using a scoping review method, we searched 14 databases to find peer-reviewed literature about the use of TMS among postsecondary students. In total, 129 articles were included in the final synthesis, which comprised of data extraction, quality appraisal, and the identification of exemplar articles. We highlight evidence about how TMS is used for teaching and learning purposes to support student success at the undergraduate and graduate levels.

1. Introduction

Plagiarism remains one of the most prominent types of academic misconduct across higher education. Commercially available text-matching software (TMS) has been adopted by many institutions in an effort to reduce violations and help students develop academic writing skills. Understanding the empirical evidence about how TMS is used to prevent violations of academic integrity and promote student learning can help institutions, educators, and policy makers make more informed decisions about how or if to use it.

2. Background: Text-Matching Software in Postsecondary Contexts

Although TMS is used in a variety of contexts including K-12 education, higher education, and scholarly publication, we delimited our study to examining the use of TMS in postsecondary contexts. We defined “postsecondary” as including universities and colleges, regardless of funding type (e.g., private or public) or the types of credentials they awarded (e.g., degrees, diplomas, and certificates). In the subsections that follow, we elaborate on specific topics related to text-matching software in postsecondary contexts including (Section 2.1) plagiarism rates and the influence of the Internet; (Section 2.2) emergence of text-matching software; and (Section 2.3) previous reviews of TMS. This background shows gaps in the extant literature that led us to develop our research questions.

2.1. Plagiarism Rates and the Influence of the Internet

Plagiarism continues to be a pressing issue in higher education [1–3]. One particular aspect of academic integrity scholarship has centered around whether the Internet has resulted in an increase in plagiarism. Scholars are divided on the issue [4] with some arguing that there is a cause-and-effect relationship between advances in digital technologies and increases in academic misconduct [5, 6]. Others have found no empirical evidence to substantiate this claim [7, 8], instead showing that overall rates of plagiarism have decreased during the twenty-first century [2, 9]. Scholars agree that the emergence of sharing culture in the Internet era has led to changes in the ways in which information is shared, including how text can easily be copied and pasted from a
variety of online sources and file sharing is now commonplace among students [4, 10–12]. Hence, there is debate in the field about the ways in which the Internet has influenced plagiarism and other breaches of academic integrity, as well as interventions to address them. One such intervention is TMS.

2.2. Emergence of Text-Matching Software. TMS (also erroneously referred to as “anticheating,” “antiplagiarism,” “plagiarism detection,” or “plagiarism prevention” software) has existed in various forms for decades, though it started to become commercially available in the late 1990s with use expanding into the twenty-first century [5, 13]. The earliest mentions we found of TMS dated back to 1998 [14–16]. Since then, the use of TMS has proliferated and it is now used widely in some countries such as the UK and Australia, with use in other countries, such as Canada, being more limited [4, 17, 18]. TMS has garnered much attention in mainstream media, though stories in news media can sometimes demonstrate a particular viewpoint, either in favour or against the software [15, 16, 19–21]. In our study, we were concerned with what the scholarly literature presented.

2.3. Previous Reviews of TMS. Previous reviews of TMS include evaluations of the tools themselves [22, 23] and scholarly and systematic reviews [24]. Our study builds on previously published reviews by including studies published up to and including February 2021, thus offering the most current findings.

2.4. Review Questions (RQs). We began with a broad general research question, followed by two subquestions, both related to different aspects of the effectiveness of TMS software for (a) reducing academic misconduct and (b) increasing student learning as an educational intervention. The research team developed the research questions through collaborative and iterative team dialogue. The research questions were published in our protocol and remained constant through the duration of the project [25].

Primary RQ: how is text-matching software used in postsecondary contexts?

(i) Sub-RQ1: what is the effectiveness of such software in reducing incidences of plagiarism?

(ii) Sub-RQ2: what is the effectiveness of such software as an educational intervention?

We were particularly concerned with understanding the quality of the extant literature, as well as the contents of the studies themselves. This focus on quality is reflected in Section 3.

3. Methods

3.1. Design. Prior to commencing our review, we developed a detailed protocol, which was subsequently published [25]. The protocol guided our research, and any deviations are reported. The review is reported in accordance with the PRISMA 2020 reporting guidelines [26].

We initially planned to conduct a systematic review; however, we discovered that the literature on TMS is diversified and heterogeneous. Our research subquestions which focused on effectiveness could not be addressed because of this heterogeneity. Therefore, we decided to pivot to a scoping review approach. Scoping reviews are appropriate when a research team wants to “determine the scope or coverage of a body of literature on a given topic and give clear indication of the volume of literature and studies available as well as an overview (broad or detailed) of its focus” ([27], p. 2). Furthermore, the aim of scoping reviews is to map the literature, which responds to our primary research question.

The research team comprised six members including two academic librarians, one of whom had extensive experience with systematic and scoping reviews. Four academic integrity subject-matter experts collaborated with the librarians. The review took place from October 2018 to May 2021, with the protocol published in the first year of the project [25].

3.2. Eligibility: Inclusion and Exclusion Criteria. Our eligibility (i.e., inclusion and exclusion) criteria were informed by the Johanna Briggs Institute’s mnemonic PICO: Population, Phenomenon of Interest, Context, Outcomes [28]. The research team engaged in extensive and detailed dialogue to meticulously consider each and every inclusion and exclusion criterion.

We delimited our review to the population focused on postsecondary students including undergraduates and graduates (master’s and doctoral). Studies that focused only on faculty perspectives or experiences were excluded; however, articles that incorporated faculty activities to educate students or reduce plagiarism were included. Our rationale for this is that were inherently interested in how TMS is used with respect to students. Therefore, we determined that studies would need to have at least some indication that the student experience was considered.

Studies focused on the K-12 population were excluded. The context was postsecondary education, which we defined as “universities, community college, trade, and vocational training centres” [29]. Due to the exploratory nature of scoping reviews, we did not identify specific outcomes associated with using TMS in a higher education environment.

The phenomenon of interest included commercially available TMS. As such, we excluded studies that focused on independent or individual coders who wrote their own software as an experiment or open access software, as these are not generally licensed at scale by educational institutions.

We did not exclude studies based on study or publication design. We included quantitative and qualitative studies, as well as text and opinion articles, legal overviews, and theoretical papers. However, we excluded popular media including blogs and social media. We also excluded promotional materials and conference abstracts. There were no restrictions on publication date or geographical location. All included articles needed to be available in English.
3.3. Sources and Search Strategy. Our search and information sources were exhaustive. The search was developed by two librarians (KAH and BL) and pilot tested against known relevant studies. For the education-related databases, only the concept of text matching software was searched. For all the other databases, the search focused on two main concepts: text matching software and postsecondary education. Each concept included both keywords and subject headings. Keywords were constant across all databases. Subject headings were determined by a database’s controlled vocabulary. The search was developed in ERIC and then translated for other databases (see Table 1). All search strategies were completed as per PRISMA 2020 guidelines and saved on an open access database [30]. Searches are available in September 2019 and rerun again in February 2021 to update the results.

Both discipline-focused and interdisciplinary databases were searched. Databases to search were first identified by running a quick search of “text matching software” or “Turnitin” to determine whether the topic was covered by the database. We searched 14 relevant databases (Table 2). All were searched from the database inception until September 13/14, 2019, and updated on February 10, 2021.

Given the large number of studies included and the heterogeneity of the included studies, we did not search the grey literature, as originally planned nor did we complete a reference list scan or forward citation tracking because of feasibility concerns. These are considered limitations of our review.

3.4. Source Selection. Search results were imported into Covidence for screening and were automatically deduplicated. Covidence deduplicates based on title, year, and volume which need to be exact matches, and author names are similar [31].

Prior to screening, an interrater calibration exercise was conducted with all screeners. Randomly generated 100 titles/abstracts were screened independently by each reviewer, noting reasons for “No” and “Maybe” decisions. The screeners then met to clarify the inclusion/exclusion criteria and came to consensus on any disagreements. Screening occurred in two phases. The first phase included title/abstract screening. Records selected as potentially relevant were moved to the second phase, which involved reviewing the full-text articles. All screening was conducted independently, and all records and full-text articles were screened by two reviewers. Disagreements were resolved through discussion.

3.5. Data Extraction. All authors contributed to the data extraction process. Prior to data extraction, a calibration exercise was conducted with three studies to ensure that the data extraction template included all required elements and that each researcher was extracting data in the same way and understood the process. We extracted data from each included article under the following categories: purpose/research question, country where the study took place (if applicable), population (if applicable), text-matching software used, study design, methodology details, key findings, themes (for qualitative study designs), and limitations. Data were also extracted with regard to the focus of the use of the text-matching software in the article (i.e., whether it was an educational intervention or used for punitive purposes).

3.6. Quality Appraisal. Four authors (BL, HP, KC, and LAP) completed quality appraisal. Due to the diversity of study designs in the field of educational research, we used two critical appraisal tools to assess the quality of the included studies. The Mixed Methods Appraisal Tool (MMAT) provided assessment criteria for empirical mixed methods and qualitative and quantitative study designs [32]. We used the Joanna Briggs Institute (JBI) Checklist for Text and Opinion to assess articles that were theoretical, program descriptions, or opinion pieces [33]. The MMAT produced a score out of seven items for each empirical article, and the JBI checklist produced a score out of six. Once scored, the articles were categorized as high, medium, or low quality. The MMAT guidelines do not specify a particular ranking system for article quality, but rather request that authors be transparent about how the scores were interpreted [32]. High-quality articles were scored five–seven with the MMAT tool or six–seven with the JBI checklist; medium-quality articles were scored three–four (both MMAT and JBI); and low-quality articles were scored zero–two (both MMAT and JBI). Articles were not excluded based on their quality scores.

3.7. Data Synthesis. Two authors (KAH and HP) compiled the data extraction and quality appraisal from the rest of the research team for data synthesis. We summarized the data in a series of descriptive tables and an accompanying narrative synthesis. Due to the large number of included articles and the heterogeneity of these studies, it was not possible to include every article in the narrative synthesis. Through group consensus, the authors identified a series of exemplar articles that scored high in the quality appraisal process and had notable methodologies or findings. These were the articles that were emphasized in the narrative synthesis.

4. Results

Figure 1 shows the flow of articles through the searching and screening process. The original search (September 2019) resulted in 7,352 records, of which 3,403 were duplicates. Titles and abstracts of 3,949 were screened, resulting in 363 full-text articles reviewed for eligibility, of which 116 met the inclusion criteria. The updated search (February 2021) identified an additional 1050 records, with 391 duplicates. After screening 659 titles/abstracts, 37 studies’ full texts were reviewed and an additional 13 articles met the inclusion criteria. In total, 129 articles were included in the final analysis [5, 17, 18, 22, 33–155].

4.1. Study Characteristics. As shown in Table 3, most of the articles were from English-speaking countries, with 33 countries representing overall. The most common country where empirical studies took place was the United States of America (n = 32, 24.8%), followed by the United Kingdom.
Turnitin was the most frequently used text-matching software (n = 111, 86.0%) by far; the next most frequent software was SafeAssign (n = 9, 7.0%) and Urkund (n = 3, 2.3%). Other software used was Ephorus, EVE2, Grademark, JISC Plagiarism Detection Service, Scriptum, WCopyfind, iThenticate, MyDropBox, and Veriguide (all present once or twice in the included articles). Generic software or other programs accounted for 14 (10.9%) uses across the included articles. A total of 11 (8.5%) articles featured more than one type of software.

The study population is shown in Table 6. Some included articles (n = 24, 18.6%) did not have a study population because they were opinion articles. Of those articles with a study population, the majority included only postsecondary students (n = 84, 65.1%). An additional 21 studies (16.3%) included both students and faculty/instructors. We were unable to report on students’ year of study or status (undergraduate or graduate), as well as discipline, because many articles included multiple populations or the reporting was unclear.

Not surprisingly, articles focused specifically on student work were most frequent (n = 47, 36.4%). These articles utilized student papers, essays, or written assignments as the focus of their study. Articles in this category investigated ways to detect, compare, measure, deter, or minimize plagiarism; influence of TMS on student work; prevalence or incidence of plagiarism in student essays; positive and negative aspects of using TMS on student work; using TMS to improve student writing; and reliability, functionality, and accuracy of TMS.

Some studies (n = 17, 13.2%) investigated student perspectives of text-matching software. These articles explored reasons why students plagiarize, as well as students’ understanding and awareness of, and attitudes towards TMS and plagiarism. Articles in this category also addressed

---

**Table 1: ERIC (Ebsco) search strategy.**

<table>
<thead>
<tr>
<th>#</th>
<th>Query</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>SU “Computer Software”</td>
</tr>
<tr>
<td>S2</td>
<td>SU “Computer Uses in Education”</td>
</tr>
<tr>
<td>S3</td>
<td>SU “Information Technology”</td>
</tr>
<tr>
<td>S4</td>
<td>S1 OR S2 OR S3</td>
</tr>
<tr>
<td>S5</td>
<td>SU “Plagiarism”</td>
</tr>
<tr>
<td>S6</td>
<td>SU “Cheating”</td>
</tr>
<tr>
<td>S7</td>
<td>S5 OR S6</td>
</tr>
<tr>
<td>S8</td>
<td>S4 AND S7</td>
</tr>
<tr>
<td>S9</td>
<td>(plagiarism or cheating) N3 detect*</td>
</tr>
<tr>
<td>S10</td>
<td>(plagiarism or cheating) N3 (software or tool* or program* or computer* or online or Internet or product* )</td>
</tr>
<tr>
<td>S11</td>
<td>anti-plagiarism N3 (software or tool* or program* or computer* or online or Internet or product* )</td>
</tr>
<tr>
<td>S12</td>
<td>anti-plagiarism N3 (software or tool* or program* or computer* or online or Internet or product* )</td>
</tr>
<tr>
<td>S13</td>
<td>“text match” N5 (software or tool* or program* or computer* or online or Internet or product* )</td>
</tr>
<tr>
<td>S14</td>
<td>(Turnitin* or iThenticate or SafeAssign or CrossCheck or Copyscape or CopyCatch or Urkund)</td>
</tr>
<tr>
<td>S15</td>
<td>S9 OR S10 OR S11 OR S12 OR S13 OR S14</td>
</tr>
<tr>
<td>S16</td>
<td>S8 OR S15</td>
</tr>
</tbody>
</table>

**Table 2: Databases searched.**

<table>
<thead>
<tr>
<th>Database</th>
<th>Interface</th>
<th>Discipline</th>
</tr>
</thead>
<tbody>
<tr>
<td>ERIC</td>
<td>EBSCO</td>
<td>Education</td>
</tr>
<tr>
<td>Education Research Complete</td>
<td>EBSCO</td>
<td>Education</td>
</tr>
<tr>
<td>Library and Information Science Source</td>
<td>EBSCO</td>
<td>Library science</td>
</tr>
<tr>
<td>Academic Search Complete</td>
<td>EBSCO</td>
<td>Interdisciplinary</td>
</tr>
<tr>
<td>CINAHL Plus with Full Text</td>
<td>EBSCO</td>
<td>Nursing and allied health</td>
</tr>
<tr>
<td>Business Source Complete</td>
<td>EBSCO</td>
<td>Business</td>
</tr>
<tr>
<td>International Bibliography of Social Sciences</td>
<td>ProQuest</td>
<td>Social sciences</td>
</tr>
<tr>
<td>Library and Information Science Abstracts</td>
<td>ProQuest</td>
<td>Library science</td>
</tr>
<tr>
<td>ABI/Business Premium Collection</td>
<td>ProQuest</td>
<td>Business</td>
</tr>
<tr>
<td>Compendex (Engineering Village)</td>
<td>Elsevier</td>
<td>Engineering/computer science</td>
</tr>
<tr>
<td>Scopus</td>
<td>Elsevier</td>
<td>Interdisciplinary</td>
</tr>
<tr>
<td>MEDLINE(R) and Epub Ahead of Print, In-Process, In-Data-review and Other Nonindexed Citations and Daily</td>
<td>OVID</td>
<td>Medicine/health sciences</td>
</tr>
<tr>
<td>APA PsycInfo</td>
<td>OVID</td>
<td>Psychology</td>
</tr>
<tr>
<td>Web of Science Core Collection</td>
<td>Clarivate</td>
<td>Interdisciplinary</td>
</tr>
</tbody>
</table>
Table 3: Study characteristics: country of study.

<table>
<thead>
<tr>
<th>Country</th>
<th>n</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>11</td>
<td>8.5</td>
</tr>
<tr>
<td>Botswana</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Canada</td>
<td>4</td>
<td>3.1</td>
</tr>
<tr>
<td>China</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Croatia</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Cyprus</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Finland</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Germany</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>India</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Indonesia</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Iraq</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Ireland</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Japan</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Latvia</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Malaysia</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>Mexico</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>New Zealand</td>
<td>5</td>
<td>3.9</td>
</tr>
<tr>
<td>Oman</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Pakistan</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Peru</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Philippines</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Qatar</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Saudi Arabia</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>South Africa</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Spain</td>
<td>2</td>
<td>1.6</td>
</tr>
<tr>
<td>Taiwan</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>Turkey</td>
<td>3</td>
<td>2.3</td>
</tr>
<tr>
<td>United Arab Emirates</td>
<td>1</td>
<td>0.8</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>19</td>
<td>14.7</td>
</tr>
<tr>
<td>United States of America</td>
<td>32</td>
<td>24.8</td>
</tr>
</tbody>
</table>

Figure 1: PRISMA Diagram. The flow of articles through the screening and searching process is shown, using the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines [26].
students’ perspectives of learning when taught about plagiarism and text-matching programs. Interestingly, some studies \((n = 20, 15.5\%)\) included both student work and student perspectives. Most often, these articles provided instructional sessions on academic integrity or TMS, assessed students’ work with TMS, and additionally included student perspectives of TMS. As mentioned, 21 studies included both students and instructors. The majority focused on the perspectives (detailed above) of both students and instructors with respect to TMS \((n = 15, 11.6\%)\). Instructor perspectives of student work assessed by TMS were addressed by a small number of studies \((n = 5, 3.9\%)\). These articles focused on problems and benefits of using TMS on student work, reducing instructor
workload/marketing with TMS, and using TMS to enhance learning. One article investigated both student and instructor perspectives on student work which was analyzed by a TMS program.

Table 7 shows the uses of the text-matching software included as an educational intervention \((n = 50, 38.8\%)\), a punitive measure \((n = 51, 39.5\%)\), both or other/legal/ethical \((n = 17, 13.2\%)\), or was unclear or not relevant \((n = 11, 8.5\%)\). Educational interventions included studies focused on teaching students about plagiarism and/or text-matching software. Student work was often assessed by TMS, and then, feedback was provided to students on how to avoid plagiarism or strategies to write more effectively. Often, the TMS was presented in a positive light and as a support for student learning, and students were provided with the opportunity to revise work after submission. Punitive measures included those articles where no remediation or learning opportunities were provided to students. Often, these studies focused on detection of plagiarism through the use of TMS. Examples of articles where TMS was coded as “Other” included those that investigated advantages, disadvantages, or various uses of TMS, discussed legal or ethical implications of TMS use in postsecondary education, or questioned if TMS was punitive or educational.

4.2. Quality Appraisal. Table 8 shows the quality scores of the included articles, which varied from high \((n = 72, 55.8\%)\) to medium \((n = 30, 23.3\%)\) to low \((n = 27, 20.9\%)\). Of the different study designs, opinion articles had the highest proportion of high-quality scores \((n = 13, 72.2\%)\). There was a large number and proportion of mixed-method articles \((n = 11, 28.9\%)\) and quantitative descriptive articles \((n = 8, 29.6\%)\) with low-quality scores. The most common cause of low-quality scores was the lack of a clear research question, which was scored as “No” or “Can’t Tell” in 57 of the 106 empirical articles (i.e., mixed-method, qualitative, and quantitative articles).

4.3. Exemplars of Excellence. Using our quality score as a guide, we selected four studies, each using a different methodology, to highlight as an exemplar of excellence. The selection of the studies was led by one author (SEE) and confirmed by the others. None of the exemplars included all of the high-quality indicators for their study type, but nevertheless provide useful models for researchers with particular methodological expertise.

4.3.1. Exemplar #1: Quantitative Descriptive Study. The exemplar we selected in the category of quantitative descriptive studies was Dawson et al.’s study, “Can Software Improve Marker Accuracy at Detecting Contract Cheating? A Pilot Study of the Turnitin Authorship Investigate Alpha” [67]. In this study, “twenty-four experienced markers from five units of study were asked to make decisions about the presence of contract cheating in bundles of 20 student assignments, which included 14 legitimate assignments and six purchased from contract cheating sites” ([67], p. 473). The markers were representative of the disciplines of world history \((n = 4)\); cells and genes \((n = 5)\); personality \((n = 5)\); psychopathology \((n = 5)\); and nutritional physiology \((n = 5)\). Details were also provided about the number of students enrolled in each course, along with the number who participated in the study.

The method and analysis were explained in sufficient detail that we determined the study could be replicated either exactly or closely based on the published article. Although the paper did not include a section explicitly addressing the limitations of the study, there was some indirect discussion related to general limitations of detection of contract cheating.

4.3.2. Exemplar #2: Qualitative Study. We selected Rees and Emerson’s study “The impact that Turnitin® has had on text-based assessment practice” as our exemplar of excellence of a qualitative study [131]. Rees and Emerson conducted structured interviews with staff \((n = 9)\) who had been trained to use Turnitin®, understood its capabilities, and were “high-volume” ([131], p. 4) users of the software. The authors analyzed case studies of courses at Massey University in New Zealand, where Turnitin® was available to all staff (including faculty) and its use was mandated in some programmes.

Rees and Emerson included a copy of their structured interview questions in their article and included participant quotations in their results. With regards to the two case studies of programmes, Rees and Emerson selected one program in nursing and another in communication in the sciences, presenting a robust discussion of the use of text-matching software in the selected programs.

The authors concluded by noting that although there were opportunities for faculty to adjust their assessment approaches using text-matching software to do so, they remained more likely to view the software as a tool for detecting plagiarism. The authors contemplated why faculty change regarding assessment has not occurred and called for further research.

Our evaluation of this study was that it presented a comprehensive institutional case study that others might be able to replicate. This was due, in part, to the level of detail provided in the methods and analysis sections of the study.

4.3.3. Exemplar #3: Mixed-Method Study. The exemplar of excellence we select for this category was the study of Kaktiņš, “Does Turnitin support the development of international students’ academic integrity?” [96]. Kaktiņš conducted a case study at a commercial educational institution offering higher education pathway programs for international students. Quantitative data were collected using a student survey, supplemented by two types of qualitative data collected using (a) a focus group with students and (b) interviews with teachers. Survey respondents \((n = 260)\) represented an 89% response rate. The focus group included students \((n = 5)\) from different countries to allow for representation of the student body at the school. Teachers \((n = 12)\) were interviewed, with details included about their
length of time teaching at the school and other demographic specifications.

The quality and depth of analysis, as well as the clear writing style of the author, allowed the reader to easily understand how the study was conducted and how results were analyzed. The discussion was balanced in that it explored both positive and negative aspects of using text-matching software. In addition, the author discussed limitations of the study and offered directions for future research. Kaktiŋał concluded with a call for higher-education institutions:

To recognise such software as part of a suite of strategies within a broader context of developing academic integrity, in spirit as well as practice. To do otherwise would encourage a very narrow and potentially counterproductive attitude that could stunt the development of both students’ critical thinking and their in-depth understanding of academic integrity ([96], p. 445).

Overall, this article presented an evidence-informed and balanced study that demonstrated methodological rigour.

### 4.3.4. Exemplar #4: Text and Opinion Study.
Studies falling into this category include “text- and opinion-based evidence (which may also be referred to as nonresearch evidence) which is drawn from expert opinions, consensus, current discourse, comments, assumptions, or assertions that appear in various journals, magazines, monographs, and reports” [28]. We note that the JBI critical appraisal checklists were originally developed for use in medical and health sciences and as such, this definition is an imperfect fit with studies conducted in other disciplines such as the humanities, social sciences, or law. However, owing to the absence of any other protocol designed for use in other disciplines, the research team used the JBI definition to guide our work. Studies that fell into this category generally resulted in close scrutiny by at least two researchers, often with extensive discussion regarding how to extract data and assess quality.

We selected “When Students Won’t Turnitin: An Examination of the Use of Plagiarism Prevention Services in Canada” [141] as our exemplar of excellence in this category. Though dated, this study presented a comprehensive analysis of a legal considerations related to the use of a commercial text-matching software. Strawczynski documented and analyzed a precedent-setting legal case in Canada (Rosenfeld v. McGill University) in which a student took the university to court over the use of text-matching software in the 2003-2004 academic year. Strawczynski provided an in-depth analysis of copyright law in Canada and considerations of the Canadian Charter of Rights and Freedoms that led the courts to find in favour of the student.

The case not only resulted in McGill University suspending its use of the software but also set legal precedent, ultimately resulting in more limited use of text-matching software across Canada compared with other countries. Strawczynski’s article was comprehensive in its treatment of the case and legal underpinnings that resulted in the court’s decision.

### 5. Discussion

Our study resulted in a large and nonhomogeneous dataset, necessitating the use of multiple tools to analyze the selected studies. It is unusual in a systematic review to use different tools to analyze the data, but the researchers determined that neither the Mixed Methods Appraisal Tool (MMAT) nor the Joanna Briggs Institute (JBI) checklist were appropriate for use on all the studies included. The methodological decision-making process to determine which systematic review tool would be most appropriate for each study was an ongoing process involving at least two researchers.
and iterative process throughout the project. We engaged in this decision-making process using both individual researcher evaluation of each study for possible fit with each tool, supplemented by collaborative dialogue at research team meetings.

5.2. Results As Related to the Review Questions. With regards to our review questions, we were able to answer the primary RQ: how is text-matching software used in post-secondary contexts? However, we were ultimately unable to arrive at satisfactory answers to our two subquestions:

(i) Sub-RQ1: what is the effectiveness of such software in reducing incidences of plagiarism?

Based on the 129 studies we reviewed, the evidence about the effectiveness of text-matching tools to reduce incidences of plagiarism was inconclusive. There was insufficient evidence to support cause-and-effect claims that text-matching software by itself is effective in reducing plagiarism. We noted that several studies [50, 72, 96] advocated for the use of text-matching software as one tool used as part of a comprehensive instructional and institutional academic integrity strategy.

(ii) Sub-RQ2: what is the effectiveness of such software as an educational intervention?

Numerous studies advocated for the use of text-matching software as an educational intervention [34, 35, 40, 42, 47, 48, 55], but with few exceptions [34, 55], and almost none of these studies included both control and experimental groups. As such, it was difficult to determine with certainty to what extent text-matching software is effective as an educational intervention. We can speculate that individual and institutional approaches to teaching and assessment may play a role in this regard, with pedagogical approaches varying among individual educators, in particular in jurisdictions were instructional staff have high levels of pedagogical autonomy and academic freedom with regards to teaching and assessment.

Another factor we considered was the sample size of the population under study. We noted that some studies that focused on using text-matching software as an educational intervention included author-identified limitations that included a small sample size [42, 54, 64, 75, 102, 119, 142]. The nature of the studies involving teaching and learning interventions varied in their scope and approach, and we found no evidence of educational intervention studies that had replicated previous studies, further limiting our ability to definitively determine the effectiveness of TMS as an educational intervention.

The research team consistently referred back to the review questions, as published in our protocol, making every effort to answer them. In the end, we recognized that we were unable to provide conclusive answers to these questions, and in the interest of research integrity, we determined that it was important to address this. We contend that our inability to answer our original research questions was not due to deficiencies in the abilities of the research team, but rather to the nature of the body of literature that exists in the field. As discussed in the methods, we pivoted to a scoping review approach because of the lack of heterogeneity of the included studies. Our scoping review provides a map of the evidence and literature focused on the use of text matching software with the international postsecondary context.

5.3. Plagiarism Prevention versus Detection. Through this review we noted that TMS is used both for teaching and learning purposes (i.e., prevention) and as a plagiarism detection tool. When utilized for teaching and learning, TMS can be used as a mechanism to provide formative feedback to students to help them improve their academic writing skills, along with citing and referencing skills. We were not able to determine through this study if TMS is actually an effective intervention in terms of reducing academic misconduct. One reason for this is that the motivations for academic misconduct can be complex, varying from one individual and one case, to the next. Previous research has shown that it can be more appropriate to consider academic risk factors for students that might result in academic misconduct and consider the possibility of multiple compounded risk factors for some students [18, 157], as opposed to trying to establish a reductionist or linear cause-and-effect relationship between one factor (e.g., the use of TMS) and increases or decreases in academic misconduct. Our study supports the notion that TMS should not be used or viewed as a single solution to the prevention of plagiarism, but instead, it can be used as one aspect of a comprehensive institutional approach to establishing a culture of academic integrity.

5.4. Quality of TMS Research. Our study showed that just over half of the quantitative studies included in our review were determined to be high quality, and fewer than half of qualitative studies were identified as being high quality. This points to the need for more scholarly rigour in studies on TMS in order to better inform the field of academic integrity research, as well as policy and technology licensing decisions at higher education institutions. This is consistent with previous studies that have shown that academic integrity is a less mature field of study than other areas of educational research such as assessment [158].

It would be reasonable to expect that the quality of available research will continue to improve as the field of academic integrity research continues to mature.

5.5. Strengths and Limitations. Although the research team members held varying levels of proficiency in numerous languages other than English, we made the methodological decision to only include studies that were published in English. It is possible that we missed some studies published in other languages.
As with any scoping or systematic review, we tightly controlled our search terms in order to focus our review. We also made the methodological decision to exclude the grey literature. As such, our review, while comprehensive, may not be exhaustive of all available scholarships.

This was a methodologically complex study to undertake, as it spanned multiple disciplines, which necessitated the use of both the Mixed Methods Appraisal Tool (MMAT) and the JBI Checklist. The research team engaged in iterative and ongoing dialogue throughout the project, with quality assurance, internal peer review, and interrater reliability being consistent themes during the research process.

Although we began our project with the intention of undertaking a systematic review, we recognized through our analysis of the results that we were unable to adhere to the accepted standards of systematic reviews. We noted a couple of examples of published studies that have claimed to be systematic reviews of topics related to academic integrity [159, 160] but in our estimation, did not necessarily meet the rigorous criteria to be classified as such. It is not our intention to disparage the scholarly efforts of others who endeavour to contribute to research in this field of academic integrity, but rather to emphasize the need to pursue quality in research, above all else. To that end, and with the goal of upholding scholarly rigour and ensuring integrity in the reporting of our own results, we carefully considered our analysis and determined that, in the end, our study would more accurately be described as a scoping review, rather than a systematic one.

6. Recommendations for Future Research

As a result of this study, we recommend that independent scholarly research into text-matching software continue, as the field continues to evolve. We identified a general lack of longitudinal studies or those that replicated methods of previous studies, pointing to the need for more consistency in the ways in which studies about TMS are conducted.

We have further noted that numerous studies included small sample sizes, with studies often being conducted at a single university or in a single country. There is a need for multi-institutional, multicountry research, as well as longitudinal research and studies that replicate previous research. This would help to contribute to the overall rigour and quality of research related to TMS and academic integrity in general.

In recent years, companies that produce TMS have expanded their offerings to include complementary products, such as those designed to recognise writing that may be a result of contract cheating. It is likely that machine learning and artificial intelligence and, in particular, technologies such as the Generative Pretrained Transformer (GPT) 3 [4, 161, 162] will expand concerns about student writing and authorship, as well as the notion of what constitutes original work on the part of the student. Research into these new technologies will become an important and urgent question for academic integrity scholars in the coming years.

At the beginning of our study, there were few tools or resources available for knowledge synthesis (i.e., systematic or scoping) reviews outside of medical and health sciences. During the course of our project, additional resources became available to support knowledge synthesis reviews in educational contexts [159, 163]. As such, we advocate for more systematic and scoping reviews to be conducted on topics related to academic integrity, with an ongoing emphasis on quality and methodological rigour.

7. Conclusions

Whether text-matching software is used as a tool to educate students, to detect misconduct, or a combination of both, the technology is here to stay. Our review showed methodological breadth in approaches to studying the use of such software, including qualitative, quantitative, mixed methods, and other kinds of research design, with studies of varying quality across all research design types.

Although we were unable to provide conclusive advice regarding the efficacy of text-matching software as a tool to reduce plagiarism or as an educational intervention, we have shown how commercial text-matching software is used in postsecondary education as both a punitive and pedagogical tool at undergraduate and graduate levels.

This study contributes to the body of scholarship regarding text-matching software, offering evidence-informed insights as to methodological possibilities for research and the need to focus on rigour in scholarship to further develop academic integrity as an area of research. This study may also be useful to educational decision makers seeking advice on what to consider prior to licensing a commercial text-matching software and, in particular, the need to understand the limitations of both the software itself and the extant evidence to inform such decisions.

Data Availability

All data sources are cited in the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

The authors are grateful to the University of Calgary for providing funding for this project. The authors extend our appreciation to their research assistant, Claire Rockliff.

References


[40] C. Bailey, R. Challen, C. Bailey, and R. Challen, "Student perceptions of the value of Turnitin text-matching software


A. Ledwith and A. Ríosquez, “Using anti-plagiarism software to promote academic honesty in the context of peer reviewed


[133] V. Rolfe, “Can Turnitin be used to provide instant formative feedback?”, *British Journal of Educational Technology*, vol. 42, no. 4, pp. 701–710, 2011.


