Check for updates

The MIT Press

RESEARCH ARTICLE

Google Books, Scopus, Microsoft Academic, and Mendeley for impact assessment of doctoral dissertations: A multidisciplinary analysis of the UK

Kayvan Kousha[iD] and Mike Thelwall[iD]

Statistical Cybermetrics Research Group, School of Mathematics and Computer Science, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1LY, UK

ABSTRACT

A research doctorate normally culminates in publishing a dissertation reporting a substantial body of novel work. In the absence of a suitable citation index, this article explores the relative merits of alternative methods for the large-scale assessment of dissertation impact, using 150,740 UK doctoral dissertations from 2009–2018. Systematic methods for this were designed for Google Books, Scopus, Microsoft Academic, and Mendeley. Fewer than 1 in 8 UK doctoral dissertations had at least one Scopus (12%), Microsoft Academic (11%), or Google Books citation (9%), or at least one Mendeley reader (5%). These percentages varied substantially by subject area and publication year. Google Books citations were more common in the Arts and Humanities (18%), whereas Scopus and Microsoft Academic citations were more numerous in Engineering (24%). In the Social Sciences, Google Books (13%) and Scopus (12%) citations were important and in Medical Sciences, Scopus and Microsoft Academic citations to dissertations were rare (6%). Few dissertations had Mendeley readers (from 3% in Science to 8% in the Social Sciences) and further analysis suggests that Google Scholar finds more citations, but does not report information about all dissertations within a repository and is not a practical tool for large-scale impact assessment.

1. INTRODUCTION

Doctoral dissertations are single-authored outputs usually written by early career researchers summarizing three or more years of academic research. Although sections of some dissertations may also appear in academic publications (Caan & Cole, 2012; Echeverria, Stuart, & Blanke, 2015), including peer-reviewed journals (Evans, Amaro, et al., 2018), they are still considered valuable enough to systematically archive and, increasingly, publish online. A substantial number of doctoral dissertations are produced every year, but no practical method has been developed to assess their scholarly impacts, especially for large research evaluation exercises. This is a problem because universities, departments and research funders could reasonably wish to assess the research-based value of the doctoral training that they support, going beyond simple completion rates or employment statistics. Although many studies have analyzed the scientific productivity or impact of publications resulting from doctoral dissertations (e.g., Breimer, 1996; Echeverria et al., 2015; Hagen, 2010; Lee, 2000; Stewart, Roberts &

Roy, 2007), dissertations could be the only research outputs for many students in the arts, humanities and social sciences (Larivière, 2012) and it would be difficult to accurately identify publications derived from dissertations.

Dissertations are not directly indexed by traditional citation indexes, such as the Web of Science (WoS) and Scopus, and hence their citation counts are not readily available from these sources. Some alternative methods have been used to assess the usage or impact of dissertations, such as statistics about downloads or views of electronic dissertations (e.g., Ferreras-Fernández, García-Peñalvo, & Merlo-Vega, 2015; Zhang, Lee, & You, 2001), cited reference facilities in traditional citation indexes (Larivière, Zuccala, & Archambault, 2008; Rasuli, Schöpfel, & Prost, 2018) and manual Google scholar searches (Bangani, 2018; Wolhuter, 2015). Nevertheless, these methods may not be practical for systematically identifying the scholarly impacts of dissertations for large-scale evaluations. For instance, download statistics for electronic dissertations are commonly limited to local digital libraries of theses and could easily be manipulated. Cited references searches in conventional citation databases (e.g., searching for the terms "thesis*" or "dissertation*") could be more useful to estimate the number of citations to all dissertations from different subjects, universities or years, but may miss many relevant results and retrieve false matches. In the absence of automatic API searches, Google Scholar manual searches could be too time consuming for large sets of dissertations.

Over 185,700 doctoral theses were published by UK universities during 2009–2018 (Figure 1) and in some arts and humanities and social sciences fields, dissertations are usually the only research output of junior researchers. For instance, a large bibliometric analysis of Quebec PhD students publishing during 2000–2007 ($n = 27,393$) found that most students in the arts and humanities (96%) and social sciences (90%) had no academic publications (Larivière, 2012). Hence, evidence of dissertation impact may help early career researchers to demonstrate their research value and may enable universities to monitor the research performance of their doctoral programs. For example, the PhD dissertation, "Mobile sound: media art in hybrid spaces," awarded in 2010 by the School of Media, Film and Music, University of Sussex had 24 Google Scholar, nine Microsoft Academic, seven Scopus, and two Google Books citations by 2019 as well as 14 readers in Mendeley, giving evidence of impactful
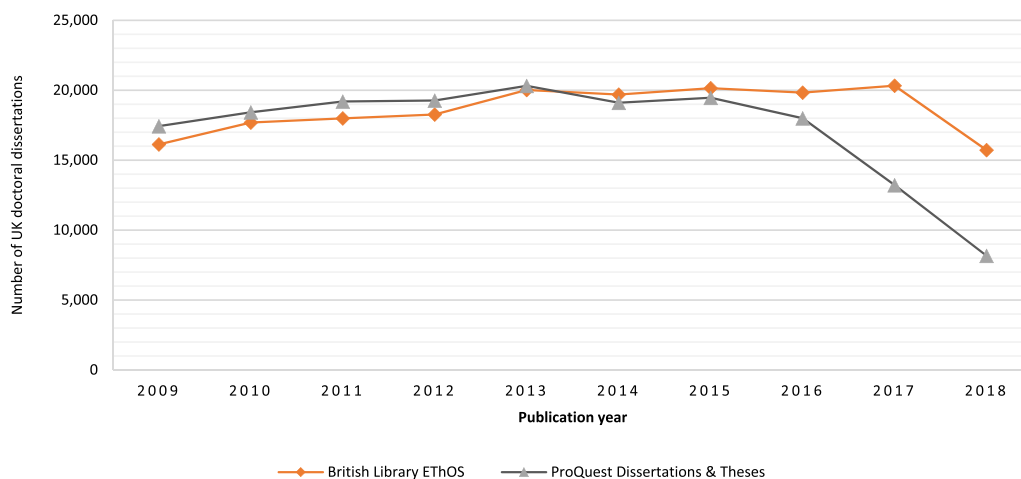

Figure 1. The number of UK doctoral dissertations indexed in the British Library EThOS and ProQuest Dissertations & Theses databases 2009–2018, as of May 2019 (total doctoral dissertations indexed by EThOS: 185,788; and ProQuest Dissertations & Theses: 172,576).

doctoral research. Universities and funders may also wish to assess the impact of traditional dissertations given that there are alternative models for PhDs that they may consider switching to, such as PhDs by publication or "light" dissertations that consist of a collection of published articles with an introductory chapter. Conversely, if dissertations are widely cited then it may be worth encouraging students to publish them formally as books (as is common, for example, in The Netherlands).

This paper investigates whether it is possible to systematically extract scholarly impact evidence for individual doctoral dissertations from Google Books, Scopus, Microsoft Academic, and Mendeley. It goes further than a previous method that found nonexhaustive lists from Google Scholar and exhaustive lists from Mendeley (Kousha & Thelwall, 2019) by finding exhaustive lists for all four sources assessed. It is based on 150,740 UK doctoral dissertations 2009–2018 across 14 fields. The terms "thesis" and "dissertation" are used as synonyms here, although some countries and communities use this terminology in different ways.

2. SCHOLARLY IMPACT ASSESSMENT FOR DISSERTATIONS

A range of alternative sources have been investigated for dissertation impact assessment.

2.1. WoS or Scopus Cited References

Although WoS and Scopus do not directly index dissertations, it is sometimes possible to identify citations to them and other nonindexed items from text searches of indexed reference lists (Bar-Ilan, 2010; Butler & Visser, 2006). Two studies have used the cited reference search options in conventional citation indexes to estimate the number of citations to dissertations from the references of other academic publications.

A WoS cited references search (i.e., querying the metadata of documents extracted by WoS from the reference lists of indexed documents) was used in one study to count citations to dissertations from academic publications between 1900 and 2004 by querying with the term "thesis*" and automatically filtering out most false matches (those where the Cited Work field did not start with "thesis*"). The average number of citations to theses was found to have decreased over time, perhaps because researchers increasingly cited standard published articles and books instead of the dissertations (Larivière et al., 2008). This trend may also be influenced by the increasing ease of access to electronic journal articles. Although this method had high precision, it may miss relevant results when terms such as "doctoral dissertation" or "PhD dissertation" are used instead of "thesis" or none are mentioned in cited references. It is not known whether many citations to dissertations were overlooked because WoS had not identified that the citations were to dissertations rather than to other types of document. Moreover, WoS does not index the publishing university, so the search strategy cannot be used to narrow down the results by university or country.

Another study used a Scopus cited reference search and a long list of related keywords (e.g., "Doctoral dissertation", "Doctoral thesis", or "M.Sc. Thesis" OR "MA thesis") to report the number of Scopus publications with at least one citation to a doctoral or master's dissertation. In contrast to the above investigations, this study found that the proportion of documents citing at least one dissertation had increased in 1996–2018 across four broad fields and was more common in the arts and humanities and social sciences than in science, technology, and medical sciences. It was below 4% in all cases (Rasuli, Schöpfel, & Prost, 2018). This method does not identify the name or any other information about the cited dissertation, however, so it cannot be used for dissertation impact assessment.

2.2. Google Scholar Citations

Google Scholar seems to be most comprehensive source of citations for gray literature impact assessment (Orduna-Malea, Martín-Martín, & López-Cózar, 2017). It can collect citations from journal articles and many (48%-65%) nonjournal scholarly materials, such as books, theses, and unpublished documents (Martín-Martín, Orduna-Malea, Thelwall, & López-Cózar, 2018). Because it disallows automatic querying, with the partial exception of Publish or Perish, it is not a practical tool for large-scale impact assessment.

Several small-scale studies have used Google Scholar to assess the citation impact of dissertations. A study of 16 digitized dissertations awarded by the London School of Economics found no correlation between Google Scholar citations and download counts for these dissertations, although the data set was too small for a meaningful statistical analysis. The most downloaded theses did not necessarily have the most citations from Google Scholar, suggesting that some dissertations might be widely read without being formally cited (Bennett & Flanagan, 2016). Out of 97 South African educational science doctoral dissertations from 2008, a quarter (24%) had at least one citation in Google Scholar (Wolhuter, 2015). An analysis of 125 doctoral dissertations in five broad fields from a Spanish university reported disciplinary differences in their citation impact. Google Scholar citations were more common in Experimental Sciences (32%), Social Sciences (20%), and the Humanities (20%) than in Life Sciences (16%) and Technological Sciences (4%) (Ferreras-Fernández et al., 2015). An investigation of a larger data set of 612 electronic engineering and technology dissertations (2002–2014) from a South African university found that 41% had at least one Google Scholar citation (Bangani, 2018).

One large-scale study of 77,884 American doctoral dissertations from 2013–2017 found that a fifth had at least one Google Scholar citation. Google Scholar citations were more numerous for older dissertations and for the social sciences, arts, and humanities (Kousha & Thelwall, 2019). This used an advanced Google Scholar search for dissertations in the ProQuest website indexed by Google Scholar and Publish or Perish to partly automate the queries for these dissertations in batches, followed by a program to match the Google Scholar records with the original list of dissertations from ProQuest. This method is not easily generalizable to dissertations not indexed by a single repository, however. The method also does not give exhaustive coverage of dissertations from indexed repositories because dissertations with duplicate copies available elsewhere on the web may be assigned to any of the source websites by Google Scholar, rendering them invisible to Google Scholar searches of their other sources.

2.3. Google Books Citations

Google Books is not a citation index and cannot therefore report citation counts. Nevertheless, automatic methods can be used to identify citations from digitized books with high accuracy and coverage via the Google Books API (Kousha & Thelwall, 2015). Google Books citations are up to twice as numerous as WoS citations for journal articles (Kousha & Thelwall, 2009) and up to four times as numerous as Scopus citations to humanities books (Kousha, Thelwall, & Rezaie, 2011). Although it seems that no study has investigated citations from Google Books to dissertations, it is a useful source of impact assessment for gray literature and queries can be automated (Bickley, Kousha, & Thelwall, 2019). Hence, Google Books might also provide an automatic source of citations to doctoral dissertations, and this may be especially valuable in book-based fields where edited books and monographs are important.

2.4. Microsoft Academic

Microsoft Academic claims to have indexed over 220 million scholarly publications, mainly from journal articles and conference papers (https://academic.microsoft.com/ as of August 2019). Microsoft Academic has found slightly more citations to journal articles than conventional citation databases (Harzing & Alakangas, 2017a, 2017b; Thelwall, 2017) and locates two to five times as many citations as Scopus for recently published or in press articles (Kousha, Thelwall, & Abdoli, 2018). Because citations to dissertations are not straightforward to collect systematically from conventional citation databases, Microsoft Academic could be a useful source for this purpose. Microsoft Academic can automatically report citation counts for articles (Hug, Ochsner, & Brändle, 2017) and books (Hug & Brändle, 2017; Kousha & Thelwall, 2018), but it is not known whether Microsoft Academic's coverage of doctoral dissertations and their citation counts are substantial enough for a large-scale impact assessment exercise. However, a study of 3,964 dissertations deposited in the University of Zurich Open Archive and Repository found that only 13% were indexed by Microsoft Academic (Hug & Brändle, 2017).

2.5. Mendeley

The social reference sharing site Mendeley allows users to save document information and subsequently use it to generate reference lists. As a side-effect of this service, it can also report the number of users that have added a document to their Mendeley library. This number can be used as an alternative impact indicator of readership.

There is now much evidence that Mendeley readership counts correlate positively with citation counts for published journal articles (e.g., Costas, Zahedi, & Wouters, 2015; Thelwall, Haustein, et al., 2013; Zahedi, Costas, & Wouters, 2017) and conference papers (Aduku, Thelwall, & Kousha, 2017; Thelwall, 2020). Mendeley readership data is a particularly useful indicator for assessing the early impact of scholarly publications (Thelwall, 2018). Mendeley users may register useful publications that are rarely cited, such as editorials, letters, news, or meeting abstracts (Zahedi & Haustein, 2018). Previous studies have shown that doctoral and PhD students are a majority of Mendeley readers. For instance, a third (33%) of the readers of 1.2 million biomedical journal articles were PhD students (Haustein & Larivière, 2014) and a later investigation found that PhD students were the main Mendeley readers of articles in Clinical Medicine, Engineering and Technology (55.4%), Social Science (54.8%), Physics (51.7%), Chemistry (50.3%), and Clinical Medicine (39.1%) (Mohammadi, Thelwall, et al., 2015).

There has also been one investigation of Mendeley readers of doctoral dissertations. A study of 77,884 American doctoral dissertations showed that 16% had at least one Mendeley reader and PhD students were the main Mendeley readers of doctoral dissertations, ranging from 33% in Agricultural Sciences to 49% in Chemical Sciences. Nevertheless, this study found low or insignificant correlations between Google Scholar citation counts and Mendeley reader counts in many subject areas, suggesting that citations and readers may only loosely reflect similar types of impact for dissertations (Kousha & Thelwall, 2019).

3. RESEARCH QUESTIONS

The following research questions drive this study. The underlying assumption is that citations or evidence of readers are, in general, similar in value, so that higher numbers represent more substantial evidence of impact. In addition, since most dissertations have no citations or readers, other sources of scholarly impact (e.g., citations from patents, academic syllabi,

encyclopedias, or clinical practices) may help to differentiate between dissertations with some impact evidence and dissertations with none. Google Scholar was not included for the main data set because it was not practical to systematically capture the citation impact of all 150,740 UK doctoral dissertations examined. Instead, a manageable method based on the British Library EThOS domain search was used to estimate whether Google Scholar might identify more citations to dissertations than Scopus and Google Books (see section 6).

1. Can useful evidence of dissertation impact be extracted from Google Books and Microsoft Academic for large-scale analyses, and are these more useful than Scopus and Mendeley?
2. Which of the above four sources provides the most useful impact evidence (citations and readership counts) for different fields and years?
3. Do citation and readership indictors for doctoral dissertations reflect a similar type of impact in different subject areas and years?

4. METHODS

The ProQuest Dissertations & Theses database contains a large number of documents uploaded by universities and seems to be the largest international repository of dissertation information. In this study, 150,740 UK doctoral dissertations from 2009–2018 were identified in the ProQuest database and their scholarly impacts were assessed using Google Books, Scopus, Microsoft Academic, and Mendeley in June 2019. Averages were calculated for each field and year to estimate which sources of impact provided the most evidence for each broad subject. Correlation analyses were used to provide indirect statistical evidence about whether the indicators might reflect similar types of impact.

4.1. The ProQuest Data Set

The ProQuest Dissertations & Theses database has been widely used as the data source to investigate aspects of doctoral dissertations and publications resulting from them in different fields (e.g., Andersen & Hammarfelt, 2011; Kim, Hansen, & Helps, 2018; Slagle & Williams, 2019; Truong, Garry, & Hall, 2014). ProQuest Dissertations & Theses indexes a large number of dissertations and theses from universities "in nearly 100 countries."[1] From 2009 to 2016 it indexed a similar number of UK theses to the British Library EThOS system (Figure 1), which is the UK's doctoral research theses archive. It aims to eventually be comprehensive for the UK, but "we do not (yet) hold all records for all institutions."[2] The coverage of EThOS is better for recent British doctoral theses (2017–2018), either due to delayed ProQuest indexing or fewer dissertations being submitted to ProQuest. ProQuest Dissertations & Theses was used as the main data source in this study because EThOS neither supports large-scale metadata collection nor includes subject classifications.

Bibliographic information about all UK doctoral dissertations was manually downloaded with permission from ProQuest. The advanced search option in ProQuest Dissertations & Theses was used and results were limited to "Doctoral Dissertations" with institution location "England" OR "Scotland" OR "Wales" OR "Northern Ireland" OR "United Kingdom" for each year separately. Metadata was saved for all 172,576 UK doctoral dissertations 2009–2018, including Author, Title, Degree, Publication year, Subject, and University/institution. These

---

[1] https://www.proquest.com/products-services/pqdtglobal.html
[2] https://ethos.bl.uk/About.do

years were selected to assess the impact of time on citation counts. Publication years were combined into sets of two for most analyses, giving five 2-year periods 2009–2018, because in some years and subject areas there were too few dissertations for meaningful statistical analyses. Although recently published dissertations 2017–2018 need more time to receive enough citations for a reasonable assessment, these years were included to examine whether Mendeley reader counts could identify early scholarly impacts for doctoral dissertations. Dissertations without subject classifications (12.7%: 21,836 out of 172,576) were removed from the main data set because ProQuest subjects were needed for disciplinary analyses, giving a final data set of 150,740 UK doctoral dissertations.

Each dissertation was searched for separately in Google Books, Scopus, Microsoft Academic, and Mendeley (for methods see below). The 440 ProQuest subjects allocated to the dissertations were too narrow for statistical analyses. For instance, there was only one dissertation for "Performing arts education," "Native American studies," "Plate tectonics," "Osteopathic medicine," "Hydraulic engineering," and "Patent law." The wider ProQuest classification scheme was therefore used to recategorize the dissertations into 20 broad areas[3]. Nevertheless, there were still few dissertations in some broad subjects and years, such as 56 in "Communications and Information Sciences" and 64 in "Architecture." Some small ProQuest broad subjects were therefore merged into broader subjects, resulting in 14 subject areas. The subjects "Architecture" and "Fine and Performing Arts" were combined to form a new category, "Arts and Architecture." Similarly, "Environmental Sciences," "Agriculture," and "Ecosystem Sciences" were combined to form "Environmental and Agricultural Sciences" and dissertations from "Communications and Information Sciences" were added to "Social Sciences."

4.2. Google Books Citation Searches

Google Books API citation searches in the free software Webometric Analyst (http://lexiurl. wlv.ac.uk) were used to automatically generate queries for all 150,740 UK doctoral dissertations. This software uses full-text search heuristics to identify mentions of publications from digitized books indexed by Google Books (see the "Books" tab in Webometric Analyst). The software automatically removes false matches (Kousha & Thelwall, 2015), but the pilot study showed that the filtered results incorrectly included citations to articles based on dissertations with identical or similar titles by the same author (sometimes with collaborators).

New queries were therefore designed to extract citations from Google Books to dissertations but not to other types of publications (e.g., journal articles or conference papers). For this, Google Books queries were generated for all dissertations using the last name of the author (student), a phrase search for the dissertation title, the publication year, and the awarding university. Examples are

> Zarate "Subtitling for deaf children: Granting accessibility to audiovisual programmes in an educational way" 2014 "University College London" [Three Google Books citations]
> Sneath "Consumption, wealth, indebtedness and social structure in early modern England" 2009 "University of Cambridge" [Six Google Books citations]

University names were included to the queries to exclude citations to nondissertation publications with the same author and title. For instance, the query *Morgan* "How do chemotherapeutic

---

[3] http://corpweb.proquest.com/assets/etd/umi_subjectcategoriesguide.pdf

agents damage the ovary?" *2014* found six Google Books citations to a coauthored article published in the journal *Human Reproduction Update* from the dissertation with the same title and author. However, adding the name of awarding university for the dissertation ("University of Edinburgh") excluded the false matches (Morgan "How do chemotherapeutic agents damage the ovary?" "University of Edinburgh" 2014). Hence, university names were necessary for searching and filtering the Google Books citations to dissertations. It is standard practice to cite the awarding university for dissertations, so the new query format should not lose many citations.

4.3. Microsoft Academic Citation Searches

Microsoft Academic was used as a second automatic citation data source for dissertations following evidence that it also indexes dissertations (Hug & Brändle, 2017; Harzing & Alakangas, 2017b). Pilot testing suggested that Microsoft Academic indexes many UK dissertations from different sources, such as EThOS, university repositories, and other digital libraries, such as arXiv.org. For instance, Microsoft Academic found 36 citations to the doctoral dissertations "Justification based explanation in ontologies" by Matthew Horridge published in 2014, providing external links to EThOS and the University of Manchester repository for the metadata and full text of the dissertation.

The Microsoft Academic API in Webometric Analyst (see "Microsoft Academic" option in the "Citation" menu) was used to generate and run automatic citation searches for all 150,740 UK doctoral dissertations (see examples below). Only dissertation titles were used, excluding author names, publication year and other bibliographic information to maximize search recall, based upon previous experience (Kousha & Thelwall, 2018; Thelwall, 2017, 2018). Webometric Analyst uses lowercase letters and omits some characters in dissertation titles to match Microsoft Academic's indexing policy, as in the examples below.

Ti= 'an investigation of a frequency diverse array'
Ti= 'the europeanisation of contract law'

Additional filtering by authors, title, and publication year was used to remove false matches with a program designed for this purpose in the Webometric Analyst (see "Filter results to remove matches for incorrect documents" option under "Microsoft Academic"). The Microsoft Academic citation searches for dissertations sometimes retrieved records with the same title and author (student) published in journals or conferences. An extra step was therefore necessary to exclude nondissertations from the Microsoft Academic results. For this, all Microsoft Academic search results with any information in the "DOI," "Journal or Conference ID," or "Journal Full Name" fields were removed. For instance, the initial Microsoft Academic search for the doctoral thesis "Non-UK university students stress levels and their coping strategies" by Mark Owusu Amponsah, published in 2009, found nine citations, but all citations were to an article with the same title and author published in the *Educational Research* journal rather than the original University of Manchester doctoral dissertation.

4.4. Scopus Cited Reference Searches

Scopus does not directly index dissertations, but it is possible to search for citations to them within the references of indexed publications. The titles of all 150,740 ProQuest dissertations were searched for in the Scopus Reference field ("REF") using the Scopus advanced search option. Eighty-nine separate searches covering 1,700 dissertation titles were combined with

the OR operator as phrase searches and the results were downloaded from Scopus for analysis. For this, Webometric Analyst was used to automatically generate Scopus cited reference queries based on dissertation titles, as shown in the example below (see "Make Scopus queries from bibliographic information" options in "Citations" menu). The program generates queries preform Scopus cited references searches effectively by using lowercase letters and excluding non-ASCII characters in dissertation titles, as shown in the example below.

> REF("travel and communication in the landscape of early medieval wessex") OR REF ("developing a bim based methodology to support renewable energy assessment of buildings") OR REF("essays on networks and market design") OR REF("the wood boring amphipod chelura terebrans") OR …

Because it was not possible to include other information, such as university names or author last names, in the Scopus cited reference queries, a program was written and added to Webometric Analysist to automatically identify citations to dissertations from the Scopus references by matching (a) the title, (b) the author last name, and (c) the publishing university name (see "Count matches of content of col in file 1 -Scopus reference lists" under "Count frequency of text or words" in "Tab-Sep" menu). This step was necessary to remove false matches, including citations to articles resulting from a dissertation by the same author and with the same title. Below is an example of a cited reference record found in Scopus, indicating that a UK dissertation had been cited 58 times by Scopus-indexed publications.

> Tagg, C. (2009). A corpus linguistics study of SMS text messaging. **Cited 58 times**. PhD thesis, University of Birmingham, http://etheses.bham.ac.uk/253/1/Tagg09PhD.pdf

Matching Scopus cited references using only title and author last name would retrieve many false matches. For example, the doctoral dissertation "Oral prednisolone for preschool children with acute virus-induced wheezing" by Jayachandran Panickar had 219 citations, but all were to a coauthored article with the same title and (first) author in the *New England Journal of Medicine*. For this case, adding the name of the awarding university (University of Leicester) as an additional matching term found no citation matches. However, in some cases both dissertations and articles resulting from dissertations with the same title and author could be cited separately, such as "Primary headteachers: New leadership roles inside and outside the school" by Susan Robinson, with two Scopus citations to the original dissertation awarded by Birmingham City University and eight other Scopus citations to an article with the same title and author derived from the dissertation in the journal *Educational Management Administration & Leadership*; such incorrect matches were removed by the final filtering stage above.

4.5. Mendeley Reader Counts

To assess Mendeley reader counts, all 150,740 dissertations were searched in Mendeley via its API in Webometric Analyst after extra filtering to identify correct Mendeley reader counts for each dissertation (see "Mendeley" option in Webometric Analyst). For this, titles and last names of authors of dissertations were automatically searched and filtering rules were applied to capture Mendeley reader counts to dissertations and to remove reader counts to other publications (e.g., journal articles or conference papers). For the final filtering stage, reader counts

from any records with "Scopus ID," "DOI," or "ISSN" were ignored and the Mendeley output was limited to dissertations or thesis using the "Source" or "Type" fields in the analysis (e.g., Thesis, Dissertation, PhD thesis, ProQuest Dissertations and Theses, PQDT—UK & Ireland, Dissertation Abstracts).

5. PRIMARY RESULTS AND DISCUSSION

Fewer than 1 in 8 UK doctoral dissertations 2009–2018 (n = 150,740) had at least one Scopus (12%: 17,662), Microsoft Academic (11%: 17,206), or Google Books citation (9%: 13,229) or Mendeley reader (5%: 7,405). The most numerous source varies substantially between subjects, however (Figure 2). Google Books is dominant for Arts and Humanities and Scopus is predominant in Engineering, but Scopus and Microsoft Academic are similar for Medical Sciences and Sciences, whereas Google Books, Scopus and Microsoft Academic are similar for Social Sciences. Mendeley is the weakest indicator in almost all cases, with the partial exception of Medical Sciences.

Analyzing the sources separately, a much higher proportion of doctoral dissertations had been cited in Google Books for Arts and Humanities (18%) than for Medical Sciences (3%), Sciences (4%), and Engineering (8%) (Figure 2). In contrast, Scopus and Microsoft Academic citations are more numerous for Engineering, where a quarter (24%) of dissertations had at least one Scopus citation compared with 8% in Arts and Humanities.

At the level of broad subjects, in the Social Sciences, relatively many doctoral dissertations had Google Books (13%) or Scopus (12%) citations, suggesting that citations from both book and article citations could be useful for monitoring the impact of social science doctoral research. In the Sciences, 12% of doctoral dissertations had at least one citation from Microsoft Academic and Scopus. In Medical Sciences, citations to dissertations were rare overall, with only 6% having at least one citation from Scopus and Microsoft Academic and only 3% having citations from Google Books. The relatively extensive rate of citing dissertations is not surprising in the book-based Arts and Humanities, but it is not clear why Engineering theses should be the most cited.
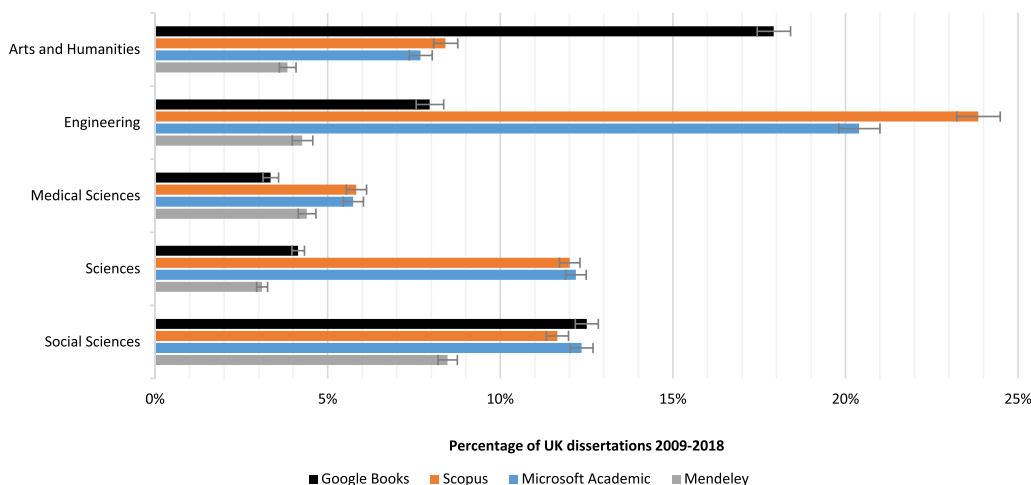

Figure 2.   The proportion of UK doctoral dissertations 2009–2018 with at least one citation in Google Books, Microsoft Academic, and Scopus or at least one Mendeley reader (Arts and Humanities: n = 24,255; Engineering: n = 17, 602; Medical Sciences: n = 24, 528; Sciences: n = 46,470; Social Sciences: n = 37,885).

5.1. Average Dissertation Impact Across 14 Fields

The average numbers of Google Books, Scopus, and Microsoft Academic citations and Mendeley readers were compared between 14 fields. Geometric means were used instead of arithmetic means because they are a more effective indicator of central tendency for highly skewed citation and altmetric data (Thelwall, 2016).

5.1.1. Google Books citations

The average (geometric mean) number of Google Books citations to doctoral dissertations is generally highest in the humanities (Figure 3). For instance, Google Books citations to doctoral dissertations published during 2009–2016 (giving a minimum of 2 years of citations for all documents) in History are four to 15 times higher, on average, than in science, technology, and biomedical fields and 1.4 to 4.7 times higher than in social science fields. In Philosophy, Religion, and Ethnic Studies, and Language and Literature, Google Books citations average three to 11 and two to eight times higher than science, technology, and biomedical disciplines. Thus, Google Books citations are numerically the most common, and therefore probably most useful, for humanities dissertations, reflecting the importance of humanities books for research communication. In Arts and Architecture, Social Sciences, and Education, dissertations published during 2009–2016 tend to be more cited by other books than in science, technology, and biomedical fields (Figure 3).

5.1.2. Scopus citations

The average number of Scopus citations per dissertation is highest in Engineering and Technology, Environmental and Agricultural Sciences, and Mathematical and Physical Sciences, where journal articles and conference papers dominate research communication (Figure 4). For instance, in Engineering and Technology the geometric mean number of Scopus citations to dissertations published during 2013–2014 (with at least 4 years to attract citations) is 4.5 to 3.0 and 3.7 to 2.1 times higher than in the arts and humanities and social
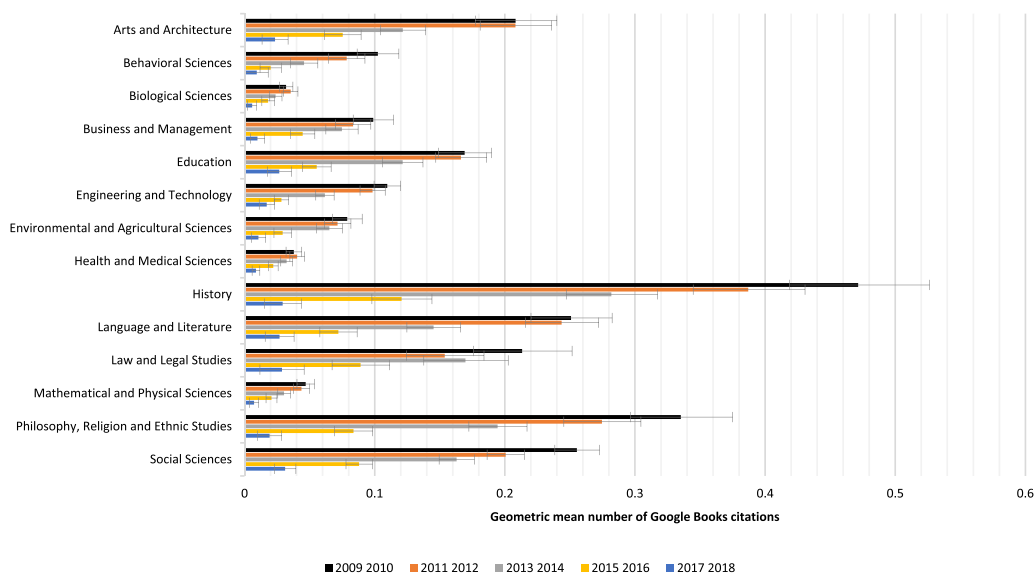

Figure 3. Geometric mean number of Google Books citations to UK doctoral dissertations 2011–2018 across 14 fields.
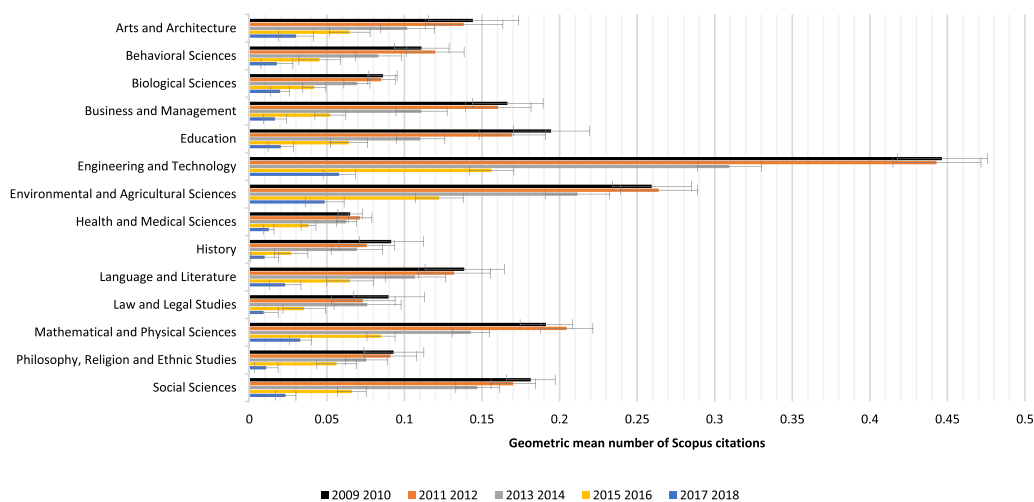
Figure 4. Geometric mean number of Scopus citations to UK doctoral dissertations 2011–2018 across 14 fields.

science subject areas, respectively. This suggests that Scopus citations are useful for assessing the scholarly impact of scientific dissertations. Surprisingly, in Health and Medical Sciences and Biological Sciences the average number of Scopus citations per dissertation is low for nearly all fields and years, despite the importance of journal articles in these areas. This suggests that dissertations are comparatively unimportant for scholarly communication in these areas.

5.1.3. Microsoft Academic citations

Microsoft Academic finds slightly more citations (6%) than Scopus to journal articles (Thelwall, 2017). For dissertations, the average number of Microsoft Academic citations is almost the same as for Scopus and is higher in Engineering and Technology, Environmental and Agricultural Sciences, and Mathematical and Physical Sciences (Figure 5). The similar average citation counts from Microsoft Academic and Scopus suggest that these databases cover similar citing publications for doctoral dissertations. In particular, the potentially wider coverage
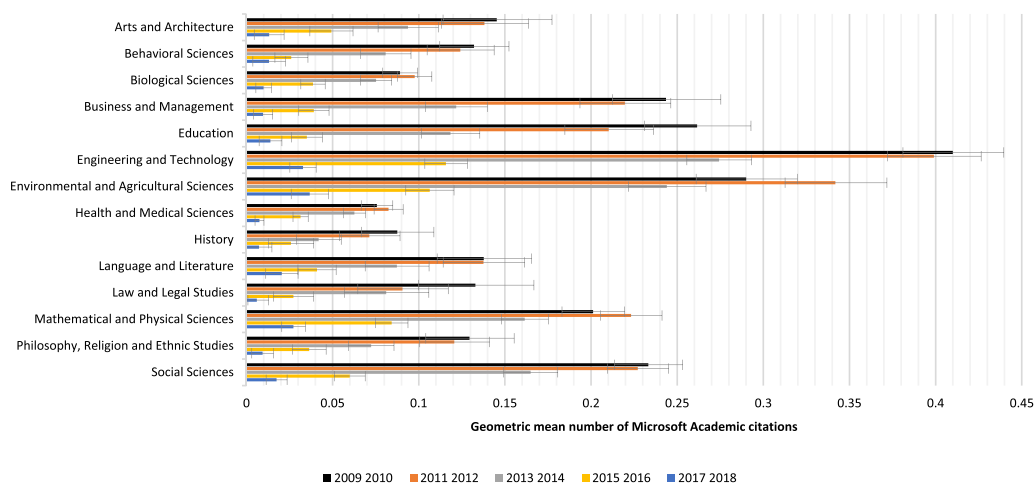

Figure 5. Geometric mean number of Microsoft Academic citations to UK doctoral dissertations 2011–2018 across 14 fields.

of Microsoft Academic through its web searches (including dissertations in arXiv and university repositories, for example), has not translated into additional citations overall.

5.1.4. Mendeley readers

Mendeley reader counts are able to identify the impact of articles earlier than citations (e.g., Thelwall, 2018), and are especially useful in the first few years after publication. This is also evident in Education, Behavioral Sciences, and Business fields for recently published dissertations (2017–2018) (Figure 6). For example, in Education the average number of Mendeley readers is 2.3 to 4.1 times higher than the arts and humanities disciplines and 2.3 to 4.2 times higher than in science, technology, and biomedical subjects. One reason could be that social sciences scholars use Mendeley more for their research or teaching. For instance, a survey showed that Mendeley users in the social sciences more often bookmark publications for future citation (91%) and teaching (35%) than users in other fields (Mohammadi, Thelwall & Kousha, 2015).

5.2. Comparison Between Impact Indicators

Figures 7–11 compare the average (geometric mean) number of Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for each 2-year period. In the five arts and humanities fields, Google Books citations to doctoral dissertations were much more common than Scopus and Microsoft Academic citations and Mendeley readership counts. This difference is statistically significant at the 95% level for all dissertations published during 2009–2014 except for Arts and Architecture for 2013–2014 (Figures 7–9). Most notably, in History the average number of Google Books citations to doctoral theses was up to 5.2, 6.7, and 15.5 times higher than Scopus and Microsoft Academic and Mendeley readership counts, respectively, and this difference is statistically significant for dissertations published during 2009–2016 (Figures 7–10). Thus, the result suggests that arts and humanities doctoral dissertations tend to be most cited by books and Google Books citations seem likely to be the most useful indicator for assessing the intellectual impact of these dissertations.

In contrast, in engineering, science and biomedical fields the average numbers of citations to doctoral dissertations (2009–2016) from Scopus were up to 5.5 times higher than from Google Books. This difference was statistically significant at the 95% level (Figures 7–10).
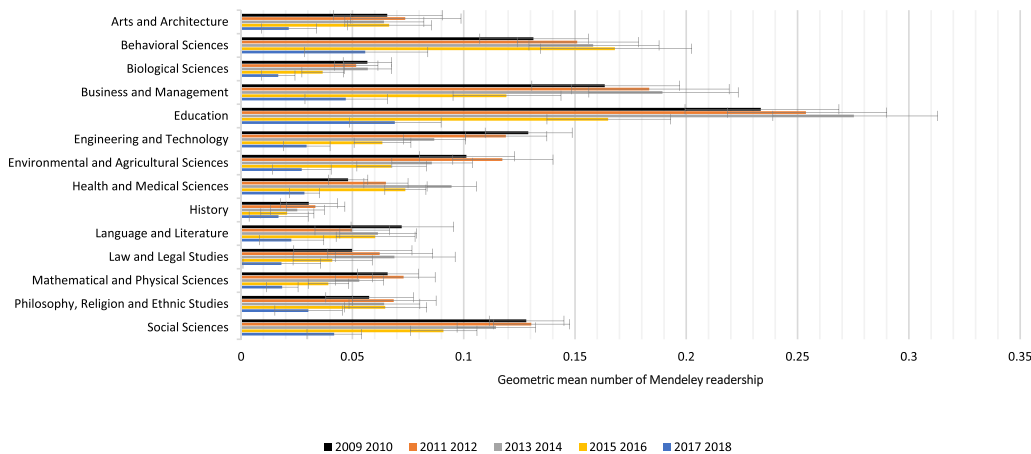

Figure 6. Geometric mean number of Mendeley readers to UK doctoral dissertations 2011–2018 across 14 fields.
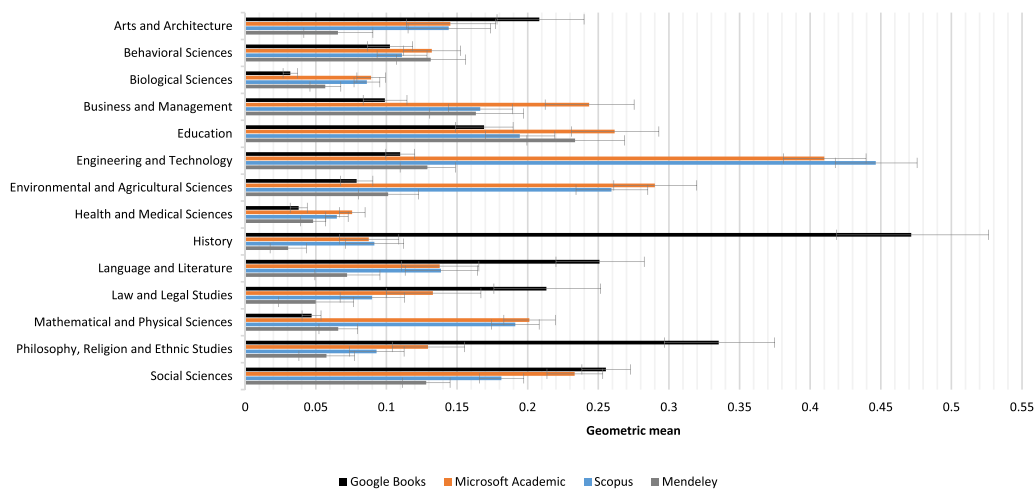
Figure 7. Geometric mean number of Google Books, Microsoft Academic, and Scopus citations and Mendeley readers for UK doctoral dissertations 2009–2010 across 14 fields.

This suggests that in article-based fields Scopus is more useful than Google Books for monitoring the scholarly impacts of doctoral dissertations. Nevertheless, Scopus has no obvious citation advantage over Microsoft Academic because it found 0.8 to 1.2 times more citations to engineering, science, or biomedical dissertations 2009–2016, and the confidence intervals overlap, except for engineering in 2015–2016 (Figures 7–10). Microsoft Academic only found statistically significantly more citations than Scopus for older dissertations published during 2009–2012 in Social Sciences, Business and Management, and Education, suggesting that it may have better coverage of social science publications than Scopus (Figures 6–7). This would reflect the incomplete coverage of social sciences journals by Scopus (Mongeon & Paul-Hus, 2016).

In three social science disciplines (Behavioral Sciences, Business, and Management and Education) and Health and Medical Sciences, the average Mendeley reader counts for UK
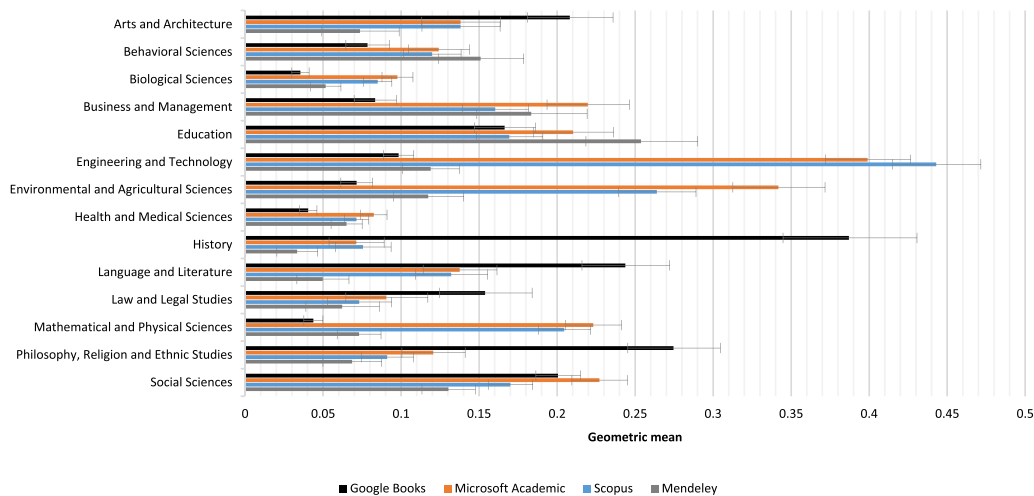
Figure 8. Geometric mean number of Google Books, Microsoft Academic, and Scopus citations and Mendeley readers for UK doctoral dissertations 2011–2012 across 14 fields.
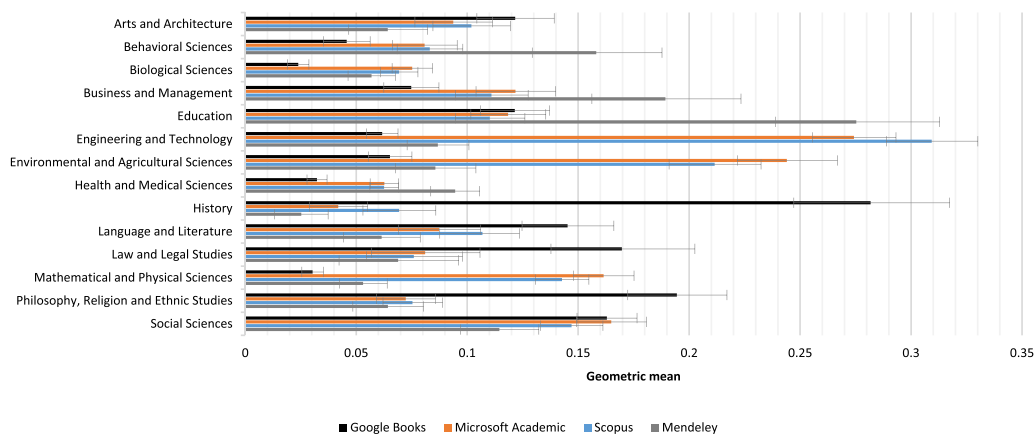
Figure 9. Geometric mean number of Google Books, Microsoft Academic, and Scopus citations or Mendeley readers for UK doctoral dissertations 2013–2014 across 14 fields.

doctoral dissertations 2012–2018 were higher than Scopus, Microsoft Academic, and Google Books citations, and this difference was statistically significant and larger for more recently published dissertations (Figures 9–11). For example, in Education the average numbers of Mendeley readers for doctoral theses 2017–2018 were up to 3.4, 5.0, and 2.6 times higher than Scopus, Microsoft Academic, and Google Books citations, respectively, which are statistically significant differences. Similarly, in Health and Medical Sciences the average numbers of Mendeley readers for 2017–2018 dissertations were 2.3, 3.8, and 3.2 times higher than citation counts from Scopus, Microsoft Academic, and Google Books, respectively. This supports the previous result that Mendeley had an advantage over Google Scholar citations by finding more readers for recently published American doctoral dissertations in Medical Sciences, Social Sciences, Economics and Business, Psychology, and Educational Sciences (Kousha & Thelwall, 2019). This suggests that in these fields a greater number of doctoral dissertations might be read by students and academics than in most other fields.
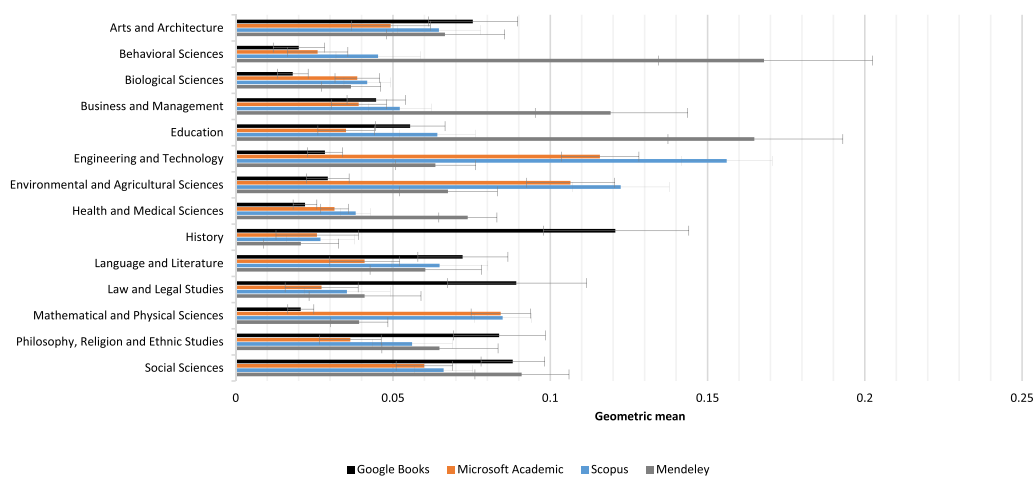

Figure 10. Geometric mean number of Google Books, Microsoft Academic, and Scopus citations or Mendeley readers for UK doctoral dissertations 2015–2016 across 14 fields.
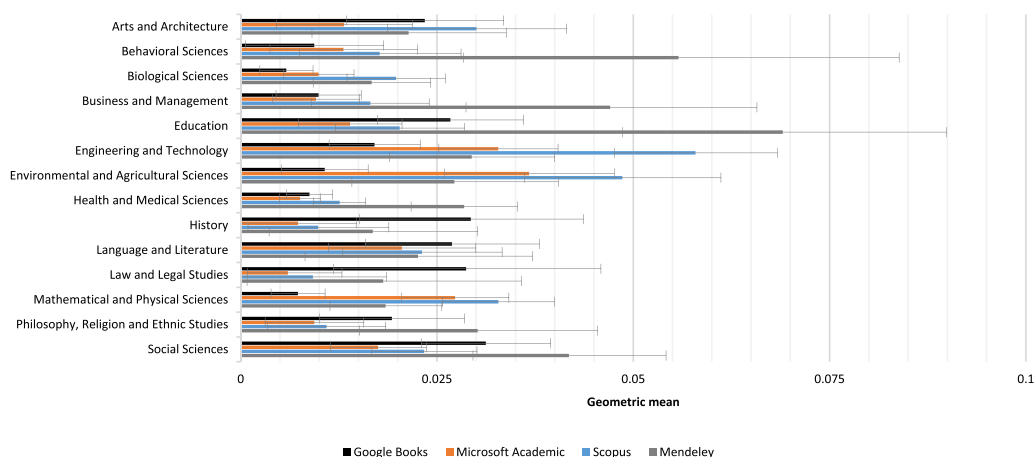
Figure 11. Geometric mean number of Google Books, Microsoft Academic, and Scopus citations or Mendeley readers for UK doctoral dissertations 2017–2018 across 14 fields.

5.3. Correlations Between Indicators

Spearman correlation tests were calculated separately for each of the 14 fields and each set of 2 years to assess the degree of similarity between the indicators for UK doctoral dissertations. The highest (and statistically significant at the $p = 0.01$ level) positive Spearman correlations are between the Scopus and Microsoft Academic citations in all subjects areas and years (Figures 12–16). This confirms (from previous journal article comparisons of the two) that these citation databases reflect similar types of intellectual impact and probably have broadly similar coverage of scientific publications (see Harzing & Alakangas, 2017b; Thelwall, 2017). The correlation is usually highest for science and technological fields, such as Engineering and Technology (ranging from .509 to .634), Mathematical and Physical Sciences (.467 to .634), and Environmental and Agricultural Sciences (.465 to .563).

Although the correlations between Google Books citations and both Scopus and Microsoft Academic citations are mostly statistically significant and positive across all fields and years (except in History for 2015–2016 and 2017–2018 and Philosophy, Religion, and Ethnic
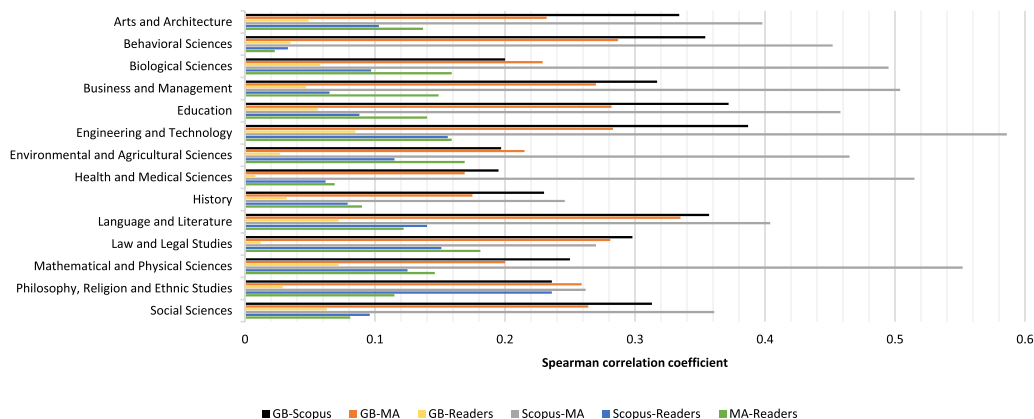

Figure 12. Spearman correlations between Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for 2009–2010 UK doctoral dissertations across 14 fields.
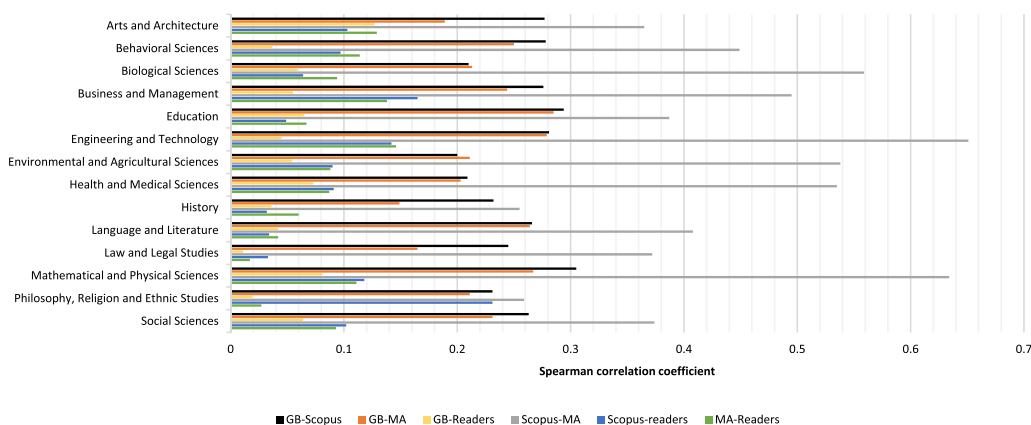
Figure 13. Spearman correlations between Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for 2011–2012 UK doctoral dissertations across 14 fields.

Studies for 2017–2018), this association is very low, suggesting that Google Books citations reflect different types of dissertation impacts compared with mainly article-based citations from Scopus and Microsoft Academic. For instance, the lowest Spearman correlations between Google Books with Scopus and Microsoft Academic citations were in History, although the average numbers of Google Books citations to doctoral theses were up to 5.2 and 6.7 times higher than Scopus and Microsoft Academic respectively (see Figures 7–10). Google Books citations might reflect the impact from (book-based) humanities areas, for example.

There were very low significant or insignificant correlations between Mendeley reader counts and citation indicators for doctoral dissertations in all fields and years, suggesting that reader counts for dissertations rarely translate into the citations found by Google Books, Scopus, or Microsoft Academic. For instance, there were very low correlations between Mendeley readers and Scopus citations for older dissertations published 2009–2010 (ranging from .062 to .152) and no significant correlations were found between them for the recently published dissertations published 2017–2018 in 11 out of 14 subject areas. This result is in contrast to much prior evidence that Mendeley readership counts highly or moderately correlate with citation counts for journal articles (e.g., Costas et al., 2015; Zahedi et al., 2017), but
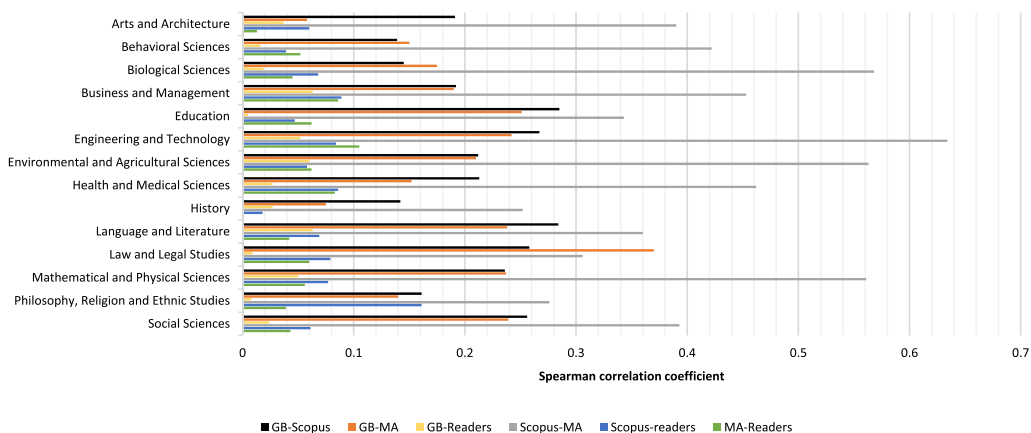

Figure 14. Spearman correlations between Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for 2013–2014 UK doctoral dissertations across 14 fields.
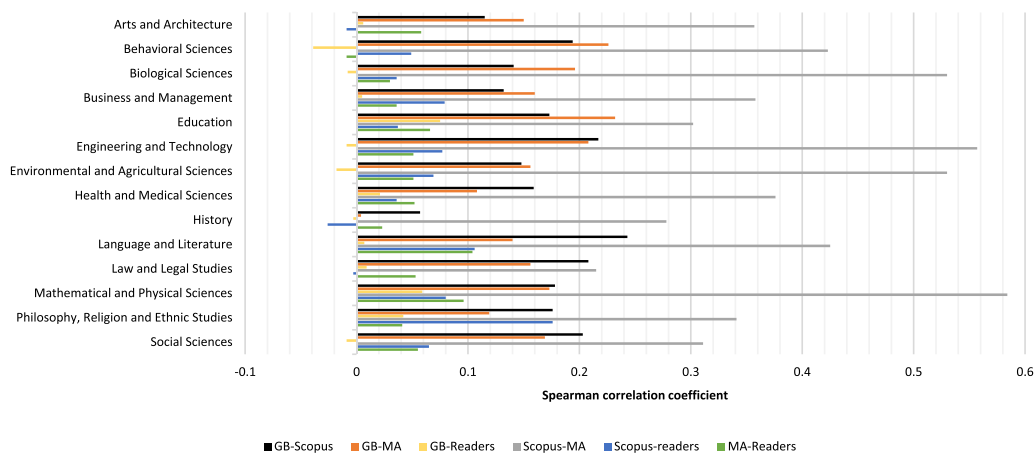
Figure 15. Spearman correlations between Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for 2015–2016 UK doctoral dissertations across 14 fields.

supports a previous investigation of American doctoral dissertations that Mendeley reader counts and Google Scholar counts loosely reflect similar types of impact (Kousha & Thelwall, 2019). Further analysis of the readership status of doctoral dissertations in Mendeley showed that overall 71% of readers were students (PhD or doctoral: 43%; Master or Postgraduate: 21%; and Bachelor: 7%), whereas 18% were academics or researchers (Researcher: 10%; Professor: 2%; Associate Professor: 2%; Lecturer: 3%; and Senior Lecturer: 1%) and 11% were other readers. Hence, it is likely that many doctoral dissertations are read by students for their research without being cited in the scholarly publications and vice versa, depending on the information seeking and referencing behavior of doctoral students in different fields (Larivière, Sugimoto, & Bergeron, 2013).

6. GOOGLE SCHOLAR DISSERTATION SEARCHES

Google Scholar covers a wider range of international scholarly publications than the other sources used in this study, so further investigations were conducted to estimate whether it could find more citations to UK doctoral dissertations.
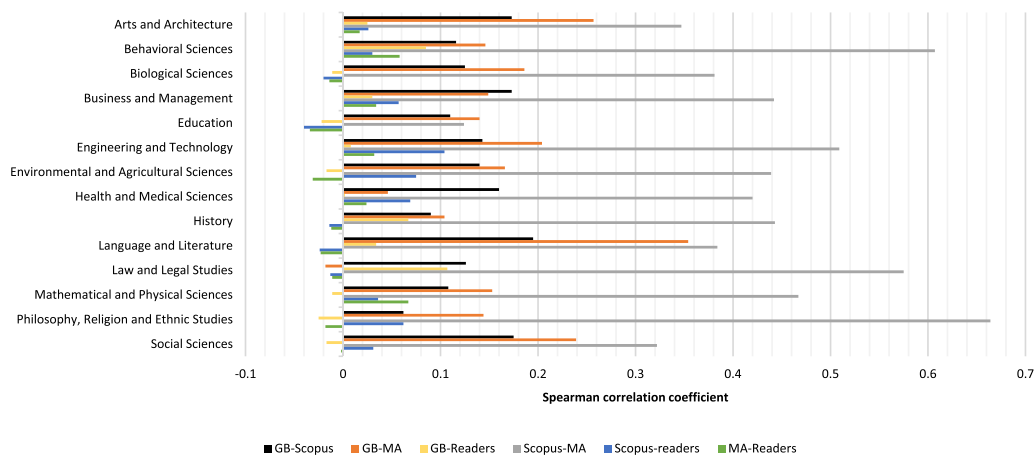

Figure 16. Spearman correlations between Google Books, Scopus, and Microsoft Academic citations and Mendeley readers for 2017–2018 UK doctoral dissertations across 14 fields.

6.1. Methods

Google Scholar does not support automatic searches for large-scale citation analysis and hence it was impractical to query all 150,740 dissertations individually in Google Scholar. However, Google Scholar indexes many UK doctoral dissertations from the British Library EThOS service. Hence, a combination of site, phrase, and author searches was used to extract Google Scholar records for UK doctoral dissertations from EThOS (see Kousha & Thelwall, 2019). By June 2019, Google Scholar reported indexing only 58,000 doctoral dissertations directly from EThOS 2009–2018[4]. The primary reason for the discrepancy is presumably that many UK doctoral dissertations are also indexed by Google Scholar from university repositories, digital libraries, and commercial publishers, such as ProQuest, and these additional versions would be registered by Google Scholar primarily with the domain first found and hence would not be searchable in Google Scholar via the British Library EThOS domain (advanced query site:ethos.bl.uk). Therefore, the method applied here misses many UK dissertations that were indexed by Google Scholar from other sources first (e.g., from the ProQuest website), although they may be also indexed from EThOS, as is visible by checking the "all versions" link for each Google Scholar result. It is also possible that some EThOS records were ignored by Google Scholar because EThOS did not contain the full text, but no examples of this were found.

The "site:ethos.bl.uk" advanced Google Scholar query was used to match records from the British Library EThOS database. Because Google Scholar displays the first 1,000 results, all searches were restricted to each year 2009–2018 and an additional "author:" search command was used to limit results to an initial of the authors' first names (see Kousha & Thelwall, 2019). For instance, the query "site:ethos.bl.uk author:C" returned 829 results for the year 2009 which was less than the maximum of 1,000 hits. This query finds dissertations from EThOS where the letter C is anywhere in the authors' first name initials, such as "**C** Wilson," "**C**J Jones," "D**C** Corcoran," or "DA**C** Narciso." However, some letters returned more than 1,000 hits (e.g., A, M, J and S), such as the query "site:ethos.bl.uk author:S" for the year 2009, with 1,070 hits. For these, previously searched letters with fewer than 1,000 hits were excluded using the "-author:" command to have fewer than 1,000 results, as shown in the example below returning 628 hits for 2009 from EThOS.

> site:ethos.bl.uk **author:S** -author:C -author:R -author:L -author:E -author:D -author:K -author:P -author:H -author:N -author:T -author:G -author:B -author:F -author:W -author:Y -author:I -author:O -author:V -author:Z -author:X -author:U -author:Q

The Publish or Perish software was used to facilitate systematic data collection for all queries (Harzing, 2007; Harzing & van der Wal, 2008) and duplicate dissertation titles retrieved from different Google Scholar queries were removed. Combining the results, a quarter of all UK doctoral dissertations in the main data set were matched with dissertations found by Google Scholar searches from the EThOS domain (24% or 36,354 out of 150,740) for further analysis. It is not clear why this number is lower than the Google Scholar estimate of 58,000, but search engine hit count estimates are known to be inaccurate (Sánchez, Martínez-Sanahuja, & Batet, 2018).

---

[4] This is an estimation based on the site:ethos.bl.uk command in Google Scholar for 2009–2018.

6.2. Results

The average (geometric mean) number of Google Scholar citations per dissertation is much higher than for both Google Books and Scopus; this is statistically significant for most fields and years (Figures 17–20). The average numbers of Google Scholar citations are up to six and 12 times higher than for Scopus and Google Books citations, respectively, for 2009–2016 (with at least a 2-year time window for citations). Google Scholar finds citations from journal articles, conference papers, books, and scholarly related publications in different languages and countries, but it is not clear why it finds more citations than Microsoft Academic, which also uses web crawling (from Bing). Although it is theoretically possible that the dissertations returned by Google Scholar for the "site:ethos.bl.uk" queries described above would be more likely to be cited than dissertations not returned, the converse seems more likely. This is because dissertations not returned are more likely to be in another online repository, with that record overriding the EthOS record in Google Scholar.

It is possible that Google Scholar is more effective at identifying citations from nonjournal sources (48%–65%) (Martín-Martín, Orduna-Malea, et al., 2018) or academic publications online even when they are not within a publisher website. For instance, the doctoral thesis "Cohabitation and convivencia: Comparing conviviality in Casamance and Catalonia" by Tilmann Heil from the University of Oxford in 2013 had 49 Google Scholar citations, but only three Scopus citations. Manual checks of the Google Scholar citing sources showed that the main citation advantage of Google Scholar over Scopus was in locating citations from books, eprints, ResearchGate publications, working papers, theses, and other nonjournal publications. Hence, in subject areas where both articles and books are important for research communication, such as social sciences fields, Google Scholar could be more useful than Google Books and Scopus individually. For instance, in Education, Google Scholar found 4.2 and 5.3 times more citations than Scopus and Google Books, respectively, to doctoral dissertations 2009–2010. Similarly, in Business and Management, Social Sciences, and Behavioral Sciences Google Scholar found more citations than both Scopus and Google Books. However, in book-based fields such as History and Philosophy, Religion, and Ethnic Studies, where citations from books are important, Google Scholar found only slightly more citations than Google Books. Because Google Scholar does not allow automatic data collection, Google Books citations searches would therefore still be a good choice for large-scale
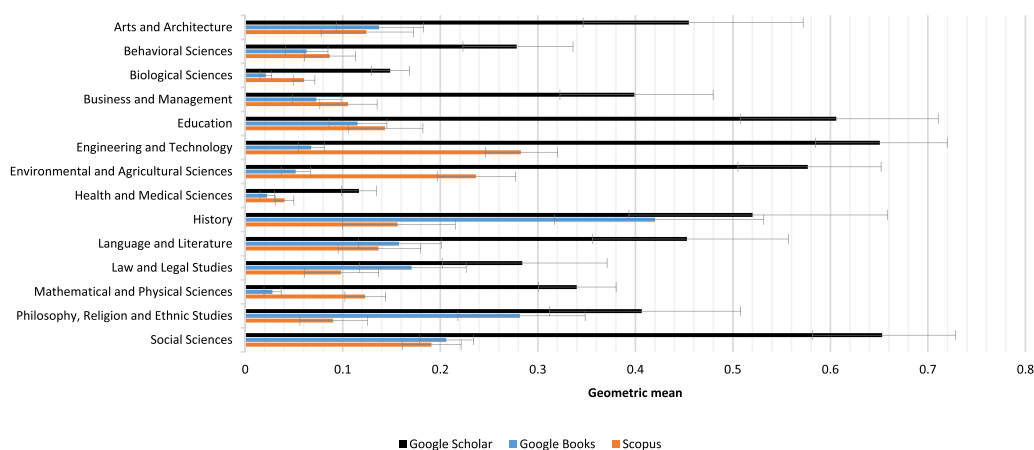

Figure 17. Geometric mean number of Google Scholar, Google Books, and Scopus citations to UK doctoral dissertations 2009–2010 across 14 fields.
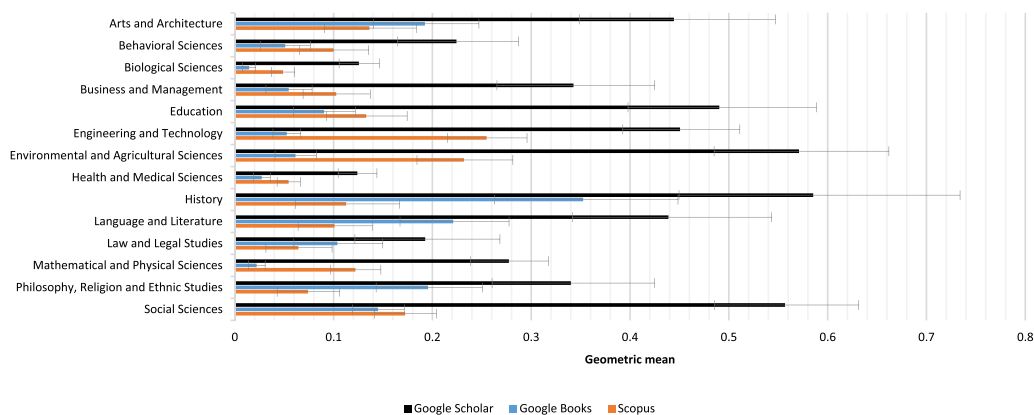
Figure 18. Geometric mean number of Google Scholar, Google Books, and Scopus citations to UK doctoral dissertations 2011–2012 across 14 fields.

citation analyses of dissertations in these areas. Nevertheless, manual checks of some dissertations in History suggested that Google Scholar does not always find citations from Google Books. For example, the doctoral dissertation in oriental studies "Mongol invasions in Southeast Asia and their impact on relations between Dai-Viet and Champa (1226–1326)" from the University of London had seven Google citations from nonbook materials and did not find any of the four unique Google Books citations from other books.

Another reason for finding more citations in Google Scholar than in other sources could be that some authors deposit preprints or postprints of their publications in open access repositories or digital libraries, such as arXiv.org, ResearchGate, and Academia.edu. Google Scholar can index these and may find citations in them to the indexed dissertations, even if the preprints were not subsequently accepted for publication. For example, the 2010 doctoral dissertation "On special functions and composite statistical distributions and their applications in digital communications over fading channels" had 43 Google Scholar and 24 Scopus citations. Manual checks of these results showed that Google Scholar had identified 14 unique citations to the dissertation from arXiv.org (all of which were author self-citations), whereas Scopus had not found any arXiv citations because it does not index the site and because none
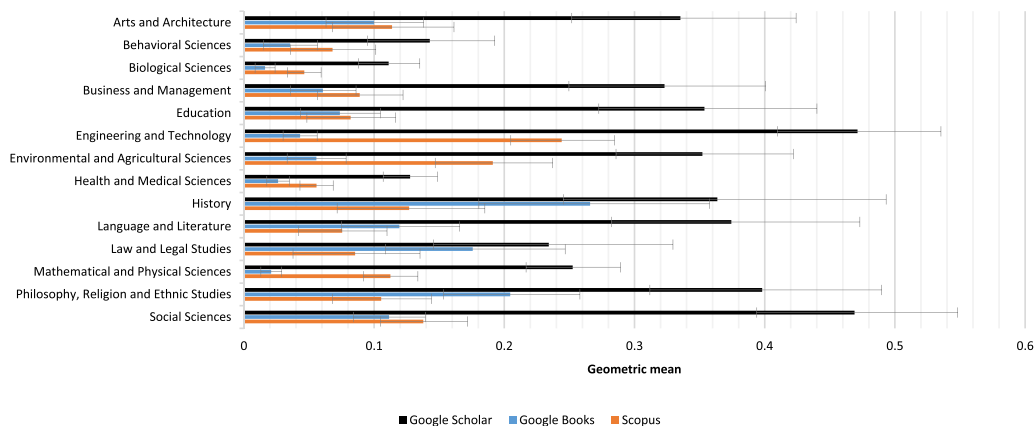

Figure 19. Geometric mean number of Google Scholar, Google Books, and Scopus citations to UK doctoral dissertations 2013–2014 across 14 fields.
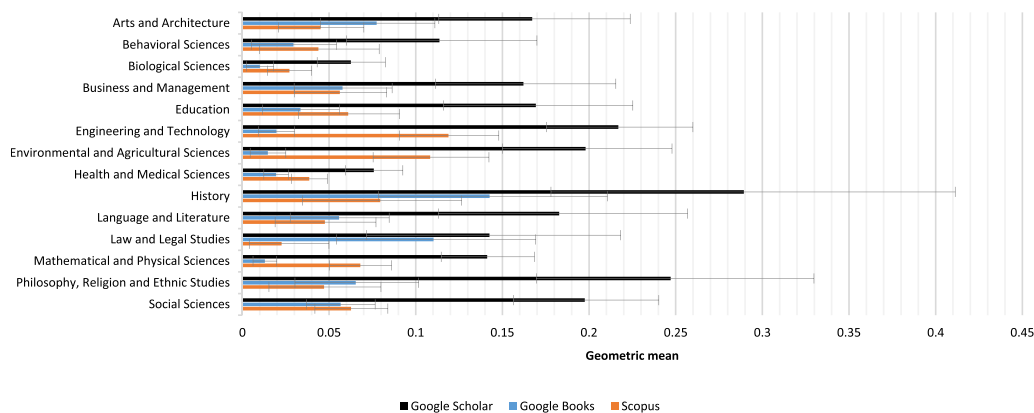
Figure 20. Geometric mean number of Google Scholar, Google Books, and Scopus citations to UK doctoral dissertations 2015–2016 across 14 fields.

of the arXiv.org citations in this case were from preprints of journal articles or conference papers indexed in Scopus (unless their names had been changed). This partly explains the large differences between Google Scholar and Scopus average citation counts, especially in Engineering and Technology and Mathematical and Physical Sciences, which have strong preprint cultures.

7. LIMITATIONS

7.1. ProQuest Coverage

The coverage of the ProQuest Dissertations & Theses database of the UK doctoral dissertations is not complete, especially for dissertations published during 2017–2018 (Figure 1). Hence, fewer recent dissertations were investigated, and these may form a biased sample.

7.2. Mentions of University Names

The method used here captured citations to dissertations using university names as filtering terms and will miss university names translated into other languages, such as in Spanish (e.g., "Cleland, 2000. Building successful brands on the Internet. Tesis doctoral. Universidad de Cambridge" or "Fahey, R.P. Rate effects in speech and nonspeech. (1993) Tesis Doctoral. Universidad de Oxford"). University names could also be mentioned incorrectly in the cited references or abbreviated, such as "Birmingham University" or "Univ. Birmingham" instead of "University of Birmingham." In these cases correct matches would be filtered out (e.g., "Sheldrake, S., 2010. The experiences of being a teenage father: An IPA analysis. Birmingham University, UK" or "Köster, K. Robust Clustering and Image Segmentation, PhD thesis, the Univ. Birmingham, School of Electronic and Electrical Eng.").

7.3. Merging Subjects

Related subject areas were merged into 14 broader fields based on the ProQuest classification scheme for the statistical analysis. For instance, Mathematical and Physical Sciences includes all 32 narrow subjects under this category. However, the scholarly impact of doctoral theses in Materials Science, Mathematics Physics, or Chemistry could differ and the merging strategy hides disciplinary differences. Hence, future research could analyze the impact indicators assessed here using larger data sets across narrow subject areas. For instance, there are 1,428,

1,219, 1,034, and 1,011 ProQuest-indexed American doctoral dissertations from 2017 in Materials Science, Chemistry, Mathematics, and Physics, respectively, which is large enough for a meaningful analysis.

7.4. Other Issues

In many fields articles derived from dissertations could be (highly) cited rather than the theses, and this study has only attempted to identify citations to dissertations, without identifying derivative publications. Hence, the methods here should be interpreted in terms of the impact of dissertation documents rather than the underlying PhD research. Connecting journal articles to dissertations seems to be a difficult task to automate, but this would give a fuller picture of the value of doctoral research. This study also does not capture the impacts of dissertations that do not translate into citations, such as use in education or professional contexts. Such uses seem likely to be rare in comparison to journal articles, however, due to the extra effort required to read dissertations. International and language differences have not been assessed and it is possible that dissertations are more cited in countries with a weaker culture of research publishing or less cited if they are not in English.

8. CONCLUSIONS

This article introduced methods to systematically assess the citation impact of dissertations and compared them on a large set from the UK. The systematic methods can be used to capture scholarly evidence of impacts to the individual dissertations from Google Books, Scopus, Microsoft Academic, and Mendeley for large-scale analyses. The application of methods for universities or departments could be to directly assess the value of publishing dissertations and indirectly (because it excludes the impact of publications derived from dissertations) assess the success of doctoral programs.

There are disciplinary differences in where doctoral dissertations are cited. In the arts and humanities, Google Books citations to dissertations were relatively common (18%), whereas in Engineering a quarter (24%) of theses had at least one citation in Scopus or Microsoft Academic. The methods used here are most useful in fields where scholars rarely publish articles from their doctoral research, such as in many arts and humanities fields. Google Scholar is the best source for the citation impact assessment of dissertations because of its coverage of a wide range of citing document types, but is most useful when a manageably small number of dissertations need to be investigated. If impact comparisons are to be made between sets of dissertations, then the differences found here show that the dissertations must be classified by field and compared only between fields. ProQuest currently seems to be the only large-scale source of subject classifications for dissertations, however, so comparison studies may need to access ProQuest data.

For young researchers, reporting citation counts for their dissertations and comparing them to the benchmark values reported in this article could be helpful for job applications or promotion, especially in fields where journal articles are not mainstream. Since most dissertations were not cited in Scopus (88%), Microsoft Academic (89%), or Google Scholar (91%) or read in Mendeley (95%), even the presence of a single citation to a dissertation from any of these sources is evidence of above average impact (unless other countries' dissertations are more frequently cited).

From a policy perspective, the results show that it is now possible to use a range of sources to automatically generate impact data for large numbers of dissertations in order to help investigate their value, although the best source of evidence depends on the subject area. Possible

applications include assessing whether particular repositories help promote the impact of dissertations as well as comparing the relative impact of dissertations between funding schemes or comparable universities to identify areas of good practice. Individual universities might also want to benchmark their PhD programs against those of similar institutions. Alternatively, the individual citing documents (except Mendeley) might be analyzed qualitatively to investigate at a deeper level how PhD research generates impact, other than through derivative publications.

AUTHOR CONTRIBUTIONS

Kayvan Kousha: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing—original draft, Writing—review & editing. Mike Thelwall: Software, Writing—review & editing.

COMPETING INTERESTS

FUNDING INFORMATION

DATA AVAILABILITY

Based on the terms and conditions agreed with ProQuest for providing information of the UK doctoral dissertations, we agreed that "the ProQuest bibliographic information of dissertations will be used for research purposes only and will be deleted permanently after analysis without sharing the collected data with third parties."

REFERENCES

Aduku, J., Thelwall, M., & Kousha, K. (2017). Do Mendeley reader counts reflect the scholarly impact of conference papers? An investigation of Computer Science and Engineering. *Scientometrics*, *112*(1), 573–581. https://doi.org/10.1007/s11192-017-2367-1

Andersen, J. P., & Hammarfelt, B. (2011). Price revisited: On the growth of dissertations in eight research fields. *Scientometrics*, *88*(2), 371–383. https://doi.org/10.1007/s11192-011-0408-8

Bangani, S. (2018). The impact of electronic theses and dissertations: A study of the institutional repository of a university in South Africa. *Scientometrics*, *115*(1), 131–151. https://doi.org/10.1007/s11192-018-2657-2

Bar-Ilan, J. (2010). Citations to the "Introduction to informetrics" indexed by WOS, Scopus and Google Scholar. *Scientometrics*, *82*(3), 495–506. https://doi.org/10.1007/s11192-010-0185-9

Bennett, L., & Flanagan, D. (2016). Measuring the impact of digitized theses: A case study from the London School of Economics. *Insights: The UKSG Journal*, *29*(2), 111–119. https://doi.org/10.1629/uksg.300

Bickley, M., Kousha, K., & Thelwall, M. (2019). Can the impact of grey literature be assessed? An investigation of UK government publications cited by articles and books. In *ISSI2019 Proceedings*. Sapienza University, Rome, Italy. https://wlv.openrepository.com/handle/2436/622832

Breimer, L. H. (1996). Age, sex and standards of current doctoral theses by Swedish medical graduates. *Scientometrics*, *37*(1), 171–176. https://doi.org/10.1007/BF02093493

Butler, L., & Visser, M. S. (2006). Extending citation analysis to non-source items. *Scientometrics*, *66*(2), 327–343. https://doi.org/10.1007/s11192-006-0024-1

Caan, W., & Cole, M. (2012). How much doctoral research on clinical topics is published? *Evidence Based Medicine*, *17*(3), 71–74. https://doi.org/10.1136/ebmed-2011-100227

Costas, R., Zahedi, Z., & Wouters, P. (2015). Do "altmetrics" correlate with citations? Extensive comparison of altmetric indicators with citations from a multidisciplinary perspective. *Journal of the Association for Information Science and Technology*, *66*(10), 2003–2019. https://doi.org/10.1002/asi.23309

Echeverria, M., Stuart, D., & Blanke, T. (2015). Medical theses and derivative articles: Dissemination of contents and publication patterns. *Scientometrics*, *102*(1), 559–586. https://doi.org/10.1007/s11192-014-1442-0

Evans, S. C., Amaro, C. M., Herbert, R., Blossom, J. B., & Roberts, M. C. (2018). "Are you gonna publish that?" Peer-reviewed publication outcomes of doctoral dissertations in psychology. *PLOS ONE*, *13*(2), e0192219. https://doi.org/10.1371/journal.pone.0192219

Ferreras-Fernández, T., García-Peñalvo, F. J., & Merlo-Vega, J. A. (2015). Open access repositories as channel of publication scientific grey literature. In *TEEM '15: Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality* (pp. 419–426). https://doi.org/10.1145/2808580.2808643

Hagen, N. T. (2010). Deconstructing doctoral dissertations: How many papers does it take to make a PhD? *Scientometrics*, *85*(2), 567–579. https://doi.org/10.1007/s11192-010-0214-8

Harzing, A. (2007). Publish or Perish, available from https://harzing.com/resources/publish-or-perish

Harzing, A., & van der Wal, R. (2008). Google Scholar as a new source for citation analysis. *Ethics in Science and Environmental Politics*, *8*(1), 61–73. https://doi.org/10.3354/esep00076

Harzing, A., & Alakangas, S. (2017a). Microsoft Academic: Is the phoenix getting wings? *Scientometrics*, *110*(1), 371–383. https://doi.org/10.1007/s11192-016-2185-x

Harzing, A., & Alakangas, S. (2017b). Microsoft Academic is one year old: The Phoenix is ready to leave the nest. *Scientometrics*, *112*(3), 1887–1894. https://doi.org/10.1007/s11192-017-2454-3

Haustein, S., & Larivière, V. (2014). Mendeley as a source of readership by students and postdocs? Evaluating article usage by academic status. In *IATUL Conference*, Espoo, Finland, June 2–5, 2014. https://docs.lib.purdue.edu/cgi/viewcontent.cgi?article=2033&context=iatul

Hug, S. E., & Brändle, M. P. (2017). The coverage of Microsoft Academic: Analyzing the publication output of a university. *Scientometrics*, *113*(3), 1551–1571. https://doi.org/10.1007/s11192-017-2535-3

Hug, S. E., Ochsner, M., & Brändle, M. P. (2017). Citation analysis with Microsoft Academic. *Scientometrics*, *111*(1), 371–378. https://doi.org/10.1007/s11192-017-2247-8

Kim, S., Hansen, D., & Helps, R. (2018). Computing research in the academy: Insights from theses and dissertations. *Scientometrics*, *114*(1), 135–158. https://doi.org/10.1007/s11192-017-2572-y

Kousha, K., & Thelwall, M. (2009). Google Book Search: Citation analysis for social science and the humanities. *Journal of the American Society for Information Science and Technology*, *60*(8), 1537–1549. https://doi.org/10.1002/asi.21085

Kousha, K., & Thelwall, M. (2015). An automatic method for extracting citations from Google Books. *Journal of the Association for Information Science and Technology*, *66*(2), 309–320. https://doi.org/10.1002/asi.23170

Kousha, K., & Thelwall, M. (2018). Can Microsoft Academic help to assess the citation impact of academic books? *Journal of Informetrics*, *12*(3), 972–984. https://doi.org/10.1016/j.joi.2018.08.003

Kousha, K., & Thelwall, M. (2019). Can Google Scholar and Mendeley help to assess the scholarly impacts of dissertations? *Journal of Informetrics*, *13*(2), 467–484. https://doi.org/10.1016/j.joi.2019.02.009

Kousha, K., Thelwall, M., & Abdoli, M. (2018). Can Microsoft Academic assess the early citation impact of in-press articles? A multi-discipline exploratory analysis. *Journal of Informetrics*, *12*(1), 287–298. https://doi.org/10.1016/j.joi.2018.01.009

Kousha, K., Thelwall, M., & Rezaie, S. (2011). Assessing the citation impact of books: The role of Google Books, Google Scholar, and Scopus. *Journal of the American Society for Information Science and Technology*, *62*(11), 2147–2164. https://doi.org/10.1002/asi.21608

Larivière, V. (2012). On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics*, *90*(2), 463–481. https://doi.org/10.1007/s11192-011-0495-6

Larivière, V., Sugimoto, C. R., & Bergeron, P. (2013). In their own image? A comparison of doctoral students' and faculty members' referencing behavior. *Journal of the American Society for Information Science and Technology*, *64*(5), 1045–1054. https://doi.org/10.1002/asi.22797

Larivière, V., Zuccala, A., & Archambault, É. (2008). The declining scientific impact of theses: Implications for electronic thesis and dissertation repositories and graduate studies. *Scientometrics*, *74*(1), 109–121. https://doi.org/10.1007/s11192-008-0106-3

Lee, W. M. (2000). Publication trends of doctoral students in three fields from 1965–1995. *Journal of the American Society for Information Science and Technology*, *51*(2), 139–144. https://doi.org/10.1002/(SICI)1097-4571(2000)51:2<139::AID-ASI5>3.0.CO;2-1

Martín-Martín, A., Orduna-Malea, E., Thelwall, M., & López-Cózar, E. D. (2018). Google Scholar, Web of Science, and Scopus: A systematic comparison of citations in 252 subject categories. *Journal of Informetrics*, *12*(4), 1160–1177. https://doi.org/10.1016/j.joi.2018.09.002

Mohammadi, E., Thelwall, M., & Kousha, K. (2015). Can Mendeley bookmarks reflect readership? A survey of user motivations. *Journal of the Association for Information Science and Technology*, *67*(5), 1198–1209. https://doi.org/10.1002/asi.23477

Mohammadi, E., Thelwall, M., Haustein, S., & Larivière, V. (2015). Who reads research articles? An altmetrics analysis of Mendeley user categories. *Journal of the Association for Information Science and Technology*, *66*(9), 1832–1846. https://doi.org/10.1002/asi.23286

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Orduna-Malea, E., Martín-Martín, A., & López-Cózar, E. D. (2017). Google Scholar and the gray literature: A reply to Bonato's review. arXiv preprint arXiv:1702.03991.

Rasuli, B., Schöpfel, J., & Prost, H. (2018). EDT programs after two decades: Exploring impact. In *ETD 2018. Beyond Boundaries of Rims and Oceans: Globalizing Knowledge with ETDs*. https://hal.archives-ouvertes.fr/hal-01885481/document

Sánchez, D., Martínez-Sanahuja, L., & Batet, M. (2018). Survey and evaluation of web search engine hit counts as research tools in computational linguistics. *Information Systems*, *73*, 50–60. https://doi.org/10.1016/j.is.2017.12.007

Slagle, D., & Williams, A. M. (2019). Changes in public affairs and administration doctoral research, 2000 and 2015. *Journal of Public Affairs Education*, *25*(4), 441–456. https://doi.org/10.1080/15236803.2018.1477370

Stewart, P. K., Roberts, M. C., & Roy, K. M. (2007). Scholarly productivity in clinical psychology PhD programs: A normative assessment of publication rates. *Clinical Psychology: Science and Practice*, *14*(2), 157–171. https://doi.org/10.1111/j.1468-2850.2007.00075.x

Thelwall, M. (2016). The precision of the arithmetic mean, geometric mean and percentiles for citation data: An experimental simulation modelling approach. *Journal of Informetrics*, *10*(1), 110–123. https://doi.org/10.1016/j.joi.2015.12.001

Thelwall, M. (2017). Microsoft Academic: A multidisciplinary comparison of citation counts with Scopus and Mendeley for 29 journals. *Journal of Informetrics*, *11*(4), 1201–1212. https://doi.org/10.1016/j.joi.2017.10.006

Thelwall, M. (2018). Early Mendeley readers correlate with later citation counts. *Scientometrics*, *115*(3), 1231–124. https://doi.org/10.1007/s11192-018-2715-9

Thelwall, M. (2020). Mendeley reader counts for US computer science conference papers and journal articles. *Quantitative Science Studies*, *1*(1), 347–359. https://doi.org/10.1162/qss_a_00010

Thelwall, M., Haustein, S., Larivière, V., & Sugimoto, C. R. (2013). Do altmetrics work? Twitter and ten other social web services. *PloS One*, *8*(5), e64841. https://doi.org/10.1371/journal.pone.0064841

Truong, V. D., Garry, T., & Hall, C. M. (2014). Social marketing as the subject of doctoral dissertations. *Social Marketing Quarterly*, *20*(4), 199–218. https://doi.org/10.1177/1524500414546230

Wolhuter, C. C. (2015). The scholarly impact of doctoral research conducted in the field of education in South Africa. *South African Journal of Education*, *35*(3). https://www.ajol.info/index.php/saje/article/download/121858/111321. https://doi.org/10.15700/saje.v35n3a1090

Zahedi, Z., & Haustein, S. (2018). On the relationships between bibliographic characteristics of scientific documents and citation and Mendeley readership counts: A large-scale analysis of Web of Science publications. *Journal of Informetrics*, *12*(1), 191–202. https://doi.org/10.1016/j.joi.2017.12.005

Zahedi, Z., Costas, R., & Wouters, P. (2017). Mendeley readership as a filtering tool to identify highly cited publications. *Journal of the Association for Information Science and Technology*, *68*(10), 2511–2521. https://doi.org/10.1002/asi.23883

Zhang, Y., Lee, K., & You, B. J. (2001). Usage patterns of an electronic theses and dissertations system. *Online Information Review*, *25*(6), 370–378. https://doi.org/10.1108/EUM0000000006536