Citation analysis of Ph.D. theses with data from Scopus and Google Books

Paul Donner¹

Received: 21 January 2021 / Accepted: 30 September 2021 / Published online: 24 October 2021
© The Author(s) 2021

Abstract

This study investigates the potential of citation analysis of Ph.D. theses to obtain valid and useful early career performance indicators at the level of university departments. For German theses from 1996 to 2018 the suitability of citation data from Scopus and Google Books is studied and found to be sufficient to obtain quantitative estimates of early career researchers' performance at departmental level in terms of scientific recognition and use of their dissertations as reflected in citations. Scopus and Google Books citations complement each other and have little overlap. Individual theses' citation counts are much higher for those awarded a dissertation award than others. Departmental level estimates of citation impact agree reasonably well with panel committee peer review ratings of early career researcher support.

Keywords Early career researchers · Ph.D. thesis · Doctoral students · Doctoral degree holders · Citation analysis · Validation · Research performance · Research assessment

Introduction

In this article we present a study on the feasibility of Ph.D. thesis citation analysis and its potential for studies of early career researchers (ECR) and for the rigorous evaluation of university departments. The context is the German national research system with its characteristics of a very high ratio of graduating Ph.D.'s to available open job positions in academia, a distinct national language publication tradition in the social sciences and humanities and slowly unfolding change from a traditional apprenticeship-type Ph.D. system to a grad school type system. The first nationwide census in Germany reported 152,300 registered active doctoral students in Germany (Vollmar 2019). In the same year, 28,404 doctoral students passed their exams in Germany (Statistisches Bundesamt 2018). Both universities and science and higher education policy attach high value to doctoral training and consider it a core task of the university system. For this reason, doctoral student performance also plays an important role in institutional assessment systems.

✉ Paul Donner
donner@dzhw.eu

¹ Abteilung 2, Forschungssystem und Wissenschaftsdynamik, Deutsches Zentrum für Hochschul- und Wissenschaftsforschung, Schützenstr. 6a, 10117 Berlin, Germany

While there is currently no national scale research assessment implemented in Germany, all German federal states have introduced formula-based partial funding allocation systems for universities. In most of these, the number of Ph.D. candidates is a well-established indicator. Most universities also partially distribute funds internally by similar systems. Such implementations can be seen as incomplete as they do not take into account the actual research output of Ph.D. candidates. In this contribution we investigate if citation analysis of doctoral theses is feasible on a large scale and can conceptually and practically serve as a complement to current operationalizations of ECR performance. For this purpose we study the utility of two citation data sources, Scopus and Google Books. We analyze the obtained citation data at the level of university departments within disciplines.

Doctoral studies

The doctoral studies phase can theoretically be conceived as a status transition period. It comprises a status passage process from apprentice to formally acknowledged researcher and colleague in the social context of a scientific community (Laudel and Gläser 2008).¹ The published doctoral thesis and its public defense are manifest proof of the fulfilment of the degree criterion of independent scientific contribution, marking said transition. The scientific community, rather than the specific organization, collectively sets the goals and standards of work in the profession, and experienced members of a community of peers judge and grade the doctoral work upon completion.² Yet the specific organization also plays a very important role. The Ph.D. project and dissertation are closely associated with the hosting university as it is this organization that provides the environmental means to conduct the Ph.D. research, as a bare minimum the supervision by professors and experienced researchers, but often also formal employment with salary, workspace and facilities. And it is also the department (*Fakultät*) which formally confers the degree after passing the thesis review and defense.

As a rule, it is a formal requirement of doctoral studies that the Ph.D. candidates make substantial independent scientific contributions and publish the results. The Ph.D. thesis is a published scientific work and can be read and cited by other researchers. The extent to which other researchers make use of these results is reflected in citations to the work and is in principle amenable to bibliometric citation analysis (Kousha and Thelwall 2019). Citation impact of theses can be seen as a proxy of the recognition of the utility and relevance of the doctoral research results by other researchers. Theses are often not published in an established venue and are hence absent from the usual channels of communication of the research front, more so in journal-oriented fields, whereas in book-oriented fields, publication of theses through scholarly publishers is common. We address this challenge by investigating the presence of dissertation citations in data sources hitherto not sufficiently considered for this purpose in what follows.

¹ As a case in point for the status change approach we note that, in Germany, holding a Ph.D. degree is a requirement for applying for funding at the national public science funding body Deutsche Forschungsgemeinschaft.

² While in other countries it is the norm that doctoral theses are evaluated by experts external to the university, this has traditionally not been the case in Germany (Wissenschaftsrat 2002). In fact, grading of the thesis by the Ph.D. supervisor is not considered inappropriate by the German Rectors' Conference (Hochschulrektorenkonferenz, 2012, p. 7).

Research contribution of early career researchers and performance evaluation in Germany

Almost all universities in Germany are predominantly tax-funded and the consumption of these public resources necessitates a certain degree of transparency to establish and maintain the perceived legitimacy of the higher education and research system. Consequently, universities and their subdivisions are increasingly subjected to evaluations. The pressure to participate in evaluation exercises, or in some cases the bureaucratic directive to do so by the responsible political actors, in turn, derives from demands of the public, which holds political actors accountable for the responsible spending of resources appropriated from net tax payers. Because the training of Ph.D. candidates is undisputedly a core task of universities, it is commonly implemented as an important component or dimension in university research evaluation.

While there is no official established national-scale research evaluation exercise in Germany (Hinze et al. 2019), the assessment of ECR performance plays a very important role in evaluation and funding of universities and in the systems of performance-based funding within universities. In the following paragraphs we will show this with several examples while critically discussing some inadequacies of the extant operationalizations of the ECR performance dimensions, thereby substantiating the case for more research into the affordance of Ph.D. thesis citation analysis.

The Council of Science and Humanities (*Wissenschaftsrat*) has conducted four pilot studies for national-scale evaluations of disciplines in universities and research institutes (*Forschungsrating*). While the exercises were utilized to test different modalities³, they all followed a basic template of informed peer review by appointed expert committees along a number of prescribed performance dimensions. The evaluation results did not have any serious funding allocation or restructuring consequences for the units. In all exercises, the dimension of support for early career researchers played a prominent role next to such dimensions as research quality, impact/effectivity, efficiency, and transfer of knowledge into society.⁴ In all four exercises, the dimension was operationalized with a combination of quantitative and qualitative criteria.

As the designation ‘support for early career researchers’ suggests, the focus was primarily on the support structures and provisions that the assessed units offered, but the outcomes or successes of these support environments also played a role. Yet, some of the applied indicators are more in line with a construct such as the performance, or success, of the ECRs themselves, namely, first appointments of graduates to professorships, scholarships or fellowship of ECRs (if granted externally of the assessed unit), and awards.⁵ As for the difference between the concept of the efforts expended for ECRs and the concept of the

³ The assessed disciplines were chemistry, sociology, electrical and information engineering, and English and American language and literature. Possibly the most important experimentally varied factor was the criterion by which research outputs were accredited to units: either all outputs in the reference period by researchers employed at the unit at assessment time (“current potential” principle) or all outputs of the units’ research staff in the period, regardless of employment at assessment time (“work-done-at” principle).

⁴ As an exception, the committee on the evaluation of English and American language and literature considered the support of ECRs as a sub-dimension of the ‘enabling of research’ dimension, alongside third-party funding activity and infrastructures, and networks. However, within this category, it was given special importance.

⁵ Other applied indicators are extra-scientific in that they are indicators of compliance to science-external political directions, such as the share of female doctoral candidates.

performance of ECRs, it appears to be implied that the efforts cause the performance, but this is far from self-evident. There may well be extensive support programs without realized benefits or ECRs achieving great success despite a lack of support structures. For this implied causal connection to be accepted, its mechanism should first be worked out and articulated and then be empirically validated, which was not the case in the Forschungsrating evaluation exercises.⁶

No bibliometric data on Ph.D. theses was employed in the Forschungsrating exercises (Wissenschaftsrat 2007, 2008, 2011, 2012). However, it stands to reason that citation analysis of theses might provide a valuable complementary tool if a more sound operationalization of the dimension of the performance of ECRs is to be established in similar future assessments. As for the publications of ECRs besides doctoral theses, these have been included in the other dimensions in which publications were used as criteria without special consideration.⁷

There is a further area of university evaluation in which a performance indicator of ECRs, namely the absolute number of Ph.D. graduates over a specific time period, is an important component. At the time of writing, systems of partial funding allocation from ministries to states' universities across all German federal states are well established. In these systems, universities within a state compete with one another for a modest part of the total budget based on fixed formulas relating performance to money. The performance based funding systems, different for each state, all include 'research' among their dimensions, and within it, the number of graduated Ph.D.'s is the second most important indicator after the acquired third party funding of universities (Wespel and Jaeger 2015). In direct consequence, similar systems have also found widespread application to distribute funds across departments within universities (Jaeger 2006; Niggemann 2020). These systems differ across universities. If only the number of completed Ph.D.'s is used as an indicator, then the quality of the research of the graduates does not matter in such systems. It is conceivable that graduating as many Ph.D.'s as possible becomes prioritized at the expense of the quality of ECR research and training.

A working group tasked by the Federal Ministry of Education and Research to work out an indicator model for monitoring the situation of early career researchers in Germany proposed to consider the citation impact of publications as an indicator of outcomes (Projektgruppe Indikatorenmodell 2014). Under the heading of "quality of Ph.D.—disciplinary acceptance and possibility of transfer" the authors acknowledge that, in principle, citation analysis of Ph.D. theses is possible, but citation counts do not directly measure scientific quality, but rather the level of response to, and reuse of, publications (impact). Moreover, it is stated that the literature of the social sciences and humanities are not covered well in citation indexes and theses are generally not indexed as primary documents (p. 136). Nevertheless, this approach is not to be rejected out of hand. Rather, it is recommended that the prospects of thesis citation analysis be empirically studied to judge its suitability (p. 137).

Another motivation for the present study was the finding of the National Report on Junior Scholars that even though "[j]unior scholars make a telling contribution to developing scientific and social insights and to innovation" (p. 3) the "contribution made by junior

⁶ Another issue deserving more scrutiny are the incentive structures promoted by the indicators. Indicators such as the number of granted Ph.D.'s and number of current Ph.D. candidates, which were applied in all exercises, could further exacerbate the Ph.D. oversupply in academia (Rogge and Tesch 2016).

⁷ It would be justified to (also) include publications by ECR other than theses in an ECR performance dimension, ideally appropriately weighted for ECR's co-authorship contribution (Donner 2020).

scholars to research and knowledge sharing is difficult to quantify in view of the available data” (Consortium for the National Report on Junior Scholars 2017, p. 19).

To sum up, the foregoing discussion establishes (1) that there is a theoretically under-developed evaluation practice in the area of ECR support and performance, and (2) that a need for better early career researcher performance indicators on the institutional level has been suggested to science policy actors. This gives occasion to explore which, if any, contribution bibliometrics can make to a valid and practically useful assessment.

Prior research

Citation analysis of dissertation theses

There are few publications on citation analysis of Ph.D. theses as the cited documents, as opposed to studies of the documents cited in theses, of which there are plenty. Yoels (1974) studied citations to dissertations in American journals in optics, political science (one journal each), and sociology (two journals) from the 1955 to 1969 volumes. In each case, several hundred citations in total to all Ph.D. theses combined were found, with a notable concentration on origins of Ph.D.’s in departments of high prestige – a possible first hint of differential research performance reflected in Ph.D. thesis citations. Non-US dissertations were cited only in optics. Author self-citations were very common, especially in optics and political science. While citations peaked in the periods of 1–2 or 3–5 years after the Ph.D. was awarded, they continued to be cited to some degree as much as 10 years later. According to Larivière et al. (2008), dissertations only account for a very small fraction of cited references in the Web of Science database. The impact of individual theses was not investigated. This study used a search approach in the cited references, based on keywords for theses and filtering, which may not be able to discover all dissertation citations. Kousha and Thelwall (2019) investigated Google Scholar citation counts and Mendeley reader counts for a large set of American dissertations from 2013 to 2017 sourced from ProQuest. This study did not take into account Google Books. Of these dissertations, 20% had one or more citations (2013: over 30%, 2017: over 5%) while 16% had at least one Mendeley reader. Average citation counts were comparatively high in the arts, social sciences, and humanities, and low in science, technology, and biomedical subjects. The authors evaluated the citation data quality and found that 97% of the citations of a sample of 646 were correct. As for the publication type of the citing documents, the majority were journal articles (56%), remarkably many were other dissertations (29%), and only 6% of citations originated from books. This suggests that Google Books might be a relevant citation data source instead of, or in addition to, Google Scholar.

More research has been conducted into the citation impact of thesis-related journal publications. Hay (1985) found that for the special case of a small sample from UK human geography research, papers based on Ph.D. thesis work accrued more citations than papers by established researchers. In a recent study of refereed journal publications based on US psychology Ph.D. theses, Evans et al. (2018) found that they were cited on average 16 times after 10 years. The citation impact of journal articles to which Ph.D. candidates contributed (but not of dissertations) has only been studied on a large scale for the Canadian province of Québec (Larivière 2012). The impact of journal papers with Ph.D. candidates’ contribution was contrasted to all other papers with Québec authors in the Web of Science database. As the impact of these papers, quantified as average of relative citations, was close to that of the comparison groups in three of four broad subject areas, it can be

tentatively assumed that the impact of doctoral candidates' papers was on par with that of their more experienced colleagues. The area with a notable difference between groups was arts and humanities, in which the coverage of publication output in the database was less comprehensive because a lot of research is published in monographs, and in which presumably many papers were written in French, another reason for lower coverage.

While these papers are not concerned with citations to dissertations, they do suggest that the research of Ph.D.'s is as impactful as that of other colleagues. To the best of our knowledge, no large scale study has been conducted on the citation impact of German theses on the level of individual works or on the level of university departments. We so far have scant information on the citation impact of dissertation theses, therefore the current study aims to fill this gap by a large scale investigation of citations received by German Ph.D. theses in Scopus and Google Books.

Causes for department-level performance differences

As we wish to investigate performance differences between departments of universities by discipline as reflected by thesis citations, we next consider the literature on plausible reasons for such performance differences which can result in differences in thesis citation impact. We do not consider individual level reasons for performance differences such as ability, intrinsic motivation, perseverance, and commitment.

One possible reason for cross-department performance differences is mutual selectivity of Ph.D. project applicants and Ph.D. project supervisors. In a situation in which there is some choice between the departments at which prospective Ph.D. candidates might register and some choice between the applicants a prospective supervisor might accept, out of self-interest both sides will seek to optimize their outcomes given their particular constraints. That is, applicants will opt for the most promising department for their future career while supervisors or selection committees, and thus departments, will attempt to select the most promising candidates, perhaps those who they judge most likely to contribute positively to their research agenda. Both sides can take into account a variety of criteria, such as departmental reputation or candidates' prior performance. This is part of the normal, constant social process of mutual evaluation in science. However, in this case, the mutual evaluation does not take place between peers, that is, individuals of equal scientific social status. Rather, the situation is characterized by status inequality (superior-inferior, i.e. professor-applicant). Consequently, an applicant may well apply to her or his preferred department and supervisor, but the supervisor or the selection committee makes the acceptance decision. In practice however, there are many constraints on such situations. For example, as described above, the current evaluation regime rewards the sheer quantity of completed Ph.D.'s.

Once the choices are made, Ph.D. candidates at different departments can face quite different environments, more or less conducive to research performance (which, as far as they were aware of them and were able to judge them, they would have taken into consideration, as mentioned). For instance, some departments might have access to important equipment and resources, others not. There may prevail different local practices in time available for the Ph.D. project for employed candidates as opposed to expected participation in groups' research, teaching, and other duties (Hesli and Lee 2011).

Ph.D. candidates may benefit from the support, experience and stimulation of the presence of highly accomplished supervisors. Experienced and engaged supervisors teach explicit and tacit knowledge and can serve as role models. Long and McGinnis (1985)

found that the performance of mentors was associated with Ph.D.'s publication and citation counts. In particular, citations were predicted by collaborating with the mentor and the mentor's own prior citation counts. Mentors' eminence only had a weak positive effect on the publication output of Ph.D.'s who actively collaborated with them. Similarly, Hilmer and Hilmer (2007) report that advisors' publication productivity is associated with candidate's publication count. However, there are multiple professors or other supervisors at any department, which causes variation within departments if the department and not the supervisor is used as a predictive variable. Between departments it is then the concentration of highly accomplished supervisors that may cause differences. Beyond immediate supervisors, a more or less supportive research environment can offer opportunities for learning, cooperation or access to personal networks. For example, Kim and Karau (2009) found that support from faculty, through the development of research skills, lead to higher publication productivity of management Ph.D. candidates. Local work culture and local expectations of performance may elicit behavioral adjustment (Allison and Long 1990).

In summary, prior research shows that there are several reasons to expect department-level differences of Ph.D. research quality (and its reproduction and reinforcement) which might be reflected in thesis citation impact. But it needs to be noted that the present study cannot serve towards shedding light on which particular factors are associated with Ph.D. performance in terms of citation impact. It is limited to testing if there are any department-level differences on this measure.

Citation counts and scientific impact of dissertation theses

We have argued above that citation analysis of theses could be a complementary tool for quantitative assessment of university departments in terms of the research performance of early career researchers. Hence it needs to be established that citation counts of dissertations are in fact associated with a conception of the impact of research.

As outlined by Hemlin (1996), “[t]he idea [of citation analysis] is that the more cited an author or a paper is by others, the more attention it has received. This attention is interpreted as an indicator of the importance, the visibility, or the impact of the researcher or the paper in the scientific community. Whether citation measures also express research quality is a highly debated issue.” Hemlin reviewed a number of studies of the relationship between citations and research quality but was not able to make a definite conclusion: “it is possible that citation analysis is an indicator of scientific recognition, usefulness and, to some unknown extent, quality.” Researchers cite for a variety of reasons, not only or primarily to indicate the quality of the cited work (Aksnes et al. 2019; Bornmann and Daniel 2008). Nevertheless, work that is cited usually has some importance for the citing work. Even citations classified in citation behavior studies as ‘perfunctory’ or ‘persuasive’ are not made randomly. On the contrary, for a citation to persuade anyone, the content of the cited work needs to be convincing rather than ephemeral, irrelevant, or immaterial. Citation counts are thus a direct measure of the utility, influence, and importance of publications for further research (Martin and Irvine 1983, sec. 6). Therefore, as a measure of scientific impact, citation counts have face validity. They are a measure of the concept itself, though a noisy one. Not so for research quality.

Highly relevant for the topic of the present study are the early citation impact validation studies by Nederhof and van Raan (1987), Nederhof and van Raan (1989). These studied the differences in citation impact of publications produced during doctoral studies of physics and chemistry Ph.D. holders, comparing those awarded the distinction ‘cum laude’ for

their dissertation based on the quality of the research with other graduates without this distinction (cum laude: 12% of $n = 237$ in chemistry, 13% of $n = 138$ in physics). In physics, “[c]ompared to non-cumlaudes, cumlaudes received more than twice as many citations overall for their publications, which were all given by scientists outside their alma mater” (Nederhof and van Raan 1987, p. 346). In fact, differences in citation impact of papers between the groups are already apparent before graduation, that is, before the conferral of the cum laude distinction on the basis of the dissertation. And interestingly, “[a]fter graduation, citation rates of cumlaudes even decline to the level of non-cumlaudes” (p. 347) leading the authors to suggest that “the quality of the research project, and not the quality of the particular graduate is the most important determinant of both productivity and impact figures. A possible scenario would be that some PhD graduates are chosen carefully by their mentors to do research in one of the usually rare very promising, interesting and hot research topics currently available. Most others are engaged in relatively less interesting and promising graduate research projects” (p. 348). The results in chemistry are very similar: “Large difference in impact and productivity favor cumlaudes three to two years before graduation, differences which decrease in the following years, although remaining significant. [...] Various sceptics have claimed that bibliometric measures based on citations are generally invalid. The present data do not offer any support for this stance. Highly significant differences in impact and productivity were obtained between two groups distinguished on a measure of scientific quality based on peer review (the cum laude award)” (Nederhof and van Raan 1989, p. 434).

In Germany, a system of four passing marks and one failing mark is commonly used. The better the referees judge the thesis, the higher the mark. Studies investigating the association of level of mark and citation impact of theses or thesis-associated publications are as of yet lacking. The closest are studies on medical doctoral theses from Charité. Oestmann et al. (2015) provide a correlational study of medical doctoral degree marks (averages of thesis and oral exams marks) and the publications associated with the theses from one institution, Charité University Medicine Berlin. Their data for 1992–2014 shows a longitudinal decrease of the incidence of the third best mark and an increase of the second best mark. For samples from 3 years (1998, 2004, 2008) for which publication data were collected, an association between the level of the mark and the publication productivity was detected. Both the chance to publish any peer-reviewed articles and the number of articles increase with the level of the mark. The study was extended in Chuadja (2021) with publication data for 2015 graduates. It was found that the time to graduation covaries with the level of the mark. For 2015 graduates, the average 5 year Journal Impact Factors for thesis-associated publication increase with the level of the graduation mark in the sense that theses awarded better marks produced publications in journals with higher Impact Factors. As little as these findings say about the real association of thesis research quality and citation impact, they suggest enough to motivate more research into this relationship.

Research questions

The following research questions will be addressed:

1. How often are individual Ph.D. theses cited in the journal and book literature?
2. Does Google Books contain sufficient additional citation data to warrant its inclusion as an additional data source alongside established data sources?

3. Can differences between universities within a discipline explain some of the variability in citation counts?
4. Are there noteworthy differences in Ph.D. thesis citation impact on the institutional level within disciplines?
5. Are the citation counts of Ph.D. theses associated with their scientific quality?

To test whether or not dissertation citation impact is a suitable indicator of departmental Ph.D. performance, citation data for theses needs to be collected, aggregated and studied for associations with other relevant indicators, such as doctorate conferrals, drop-out rates, graduate employability, thesis awards, or subjective program appraisals of graduates. As a first step towards a better understanding of Ph.D. performance, we conducted a study on citation sources for dissertations. The present study is restricted to monograph form dissertations. These also include monographs that are based on material published as articles. However, to be able to assess the complete scientific impact of a Ph.D. project it is necessary to also include the impact of papers which are produced in the context of the Ph.D. project, for both cumulative publication-based theses and for theses only published in monograph form. Because of this, the later results should be interpreted with due caution as we do not claim completeness of data.

Data

Dissertations' bibliographical data

There is presently no central integrated source for data on dissertations from Germany. The best available source is the catalog of the German National Library (Deutsche Nationalbibliothek, DNB). The DNB has a mandate to collect all publications originating from Germany, including Ph.D. theses. This source of dissertation data has been found useful for science studies research previously (Heinisch and Buenstorf 2018; Heinisch et al. 2020). We downloaded records for all Ph.D. dissertations from the German National Library online catalog in April 2019 using a search restriction in the university publications field of "diss*", as recommended by the catalog usage instructions, and publication year range 1996–2018. Records were downloaded by subject fields in the CSV format option.⁸ In this first step 534,925 records were obtained. In a second step, the author name and work title field were cleaned and the university information extracted and normalized and non-German university records excluded. We also excluded records assigned to medicine as a first subject class which were downloaded because they were assigned to other classes as well. As the dataset often contained more than one version of a particular thesis because different formats and editions were cataloged, these were carefully de-duplicated. In this process, as far as possible the records containing the most complete data and describing the temporally earliest version were retained as the primary records. Variant records were also kept in order to later be able to collect citations to all variants. This reduced the

⁸ The field of medicine was not included, because medical theses (for a Dr. med. degree) typically have lower research requirements and are therefore generally not commensurable to theses in other subjects (Wissenschaftsrat 2004, pp. 74–75). It was not possible to distinguish between Dr. med. theses and other degree theses within the medicine category, which means that regular natural science dissertations on medical subjects are not included in this dataset if medicine was the first assigned class.

dataset to a size of 361,971 records. Of these, about 16% did not contain information on the degree-granting university. As the National Library's subject classification system was changed during the covered period (in 2004), the class designations were unified based on the Library's mapping and aggregated into a simplified 40-class system.⁹ If more than one subject class was assigned, only the first was retained.

Citation data

Citation data from periodicals was obtained from a snapshot of Scopus data from May 2018. Scopus was chosen over Web of Science as a source of citation data because full cited titles for items not contained as primary documents in Web of Science have only recently been indexed. Before this change, titles were abbreviated so inconsistently and to such short strings as to be unusable, while this data is always available in unchanged form in Scopus if it is available in the original reference. Cited references data was restricted to non-source citations, that is, references not matched with Scopus-indexed records. Dissertation bibliographical information (author name, publication year and title) for primary and secondary records was compared to reference data. If the author family name and given name first initial matched exactly and the cited publication year was within ± 1 year of the dissertation publication year, then the title information was further compared as follows. The dissertation's title was compared to both Scopus cited reference title and cited source title, as we found these two data fields were both used for the cited thesis title. Before comparison, both titles were truncated to the character length of the shorter title. If the edit distance similarity between titles was greater than 75 out of 100, the citation was accepted as valid and stored. We furthermore considered the case that theses might occasionally be indexed as Scopus source publications. We used the same matching approach as outlined above to obtain matching Scopus records restricted to book publication types. This resulted in 659 matched theses. In addition, matching by ISBN to Scopus source publications resulted in 229 matched theses of which 50 were not matched in the preceding step. The citations to these matched source publications were added to the reference matched citations after removing duplicates. Citations were summed across all variant records while filtering out duplicate citations.

In addition, we investigated the utility of Google Books as a citation data source. This is motivated by the fact that many Ph.D. theses are published in the German language and in disciplines favoring publication in books over journal articles. Citation search in Google Books has been made accessible to researchers by the Webometric Analyst tool, which allows for largely automated querying with given input data (Kousha and Thelwall 2015). We used Webometric Analyst version 2.0 in April and May 2019 to obtain citation data for all Ph.D. thesis records. We only used primary records, not variants, as the collection process takes quite a long time. Search was done with the software's standard settings using author family name, publication year and six title words and subsequent result filtering was employed with matching individual title words rather than exact full match. We additionally removed citations from a number of national and disciplinary bibliographies and annual reports because these are not research publications but lists of all publications in a

⁹ Classification mapping according to <https://wiki.dnb.de/download/attachments/141265749/ddcSachgruppenDNBKonkordanzNeuAlt.pdf> accessed 07/18/2019. It should be noted that some classes were not used for some time, for example electrical engineering was not used between 2004 and 2010 but grouped under engineering/mechanical engineering alongside mining/construction technology/environmental technology.

discipline or produced by a certain research unit. We also removed Google Books citations from sources that were indexed in Scopus (10,958 citations) as these would otherwise be duplicates.

Google Scholar was not used as a citation data source, because it includes a lot of grey literature and there is no possibility to restrict citation counts to citations from journal and book literature. It has alarming rates of incorrect citations (García-Pérez 2010, p. 2075), however Kousha & Thelwall (2015, p. 479) found the citation error rate for Ph.D. theses in Google Scholar to be quite low.

Dissertation award data

For later validation purposes we collected data on German dissertation awards from the web. We considered awards for specific disciplines granted in a competitive manner based on research quality by scholarly societies, foundations and companies. A web search was conducted for awards, either specifically for dissertations or for contributions by early career researchers which mention dissertations besides other works. Only awards granted by committees of discipline experts and awarded by Germany-based societies etc. were considered, we did not include internationally oriented awards. In general, these awards are given on the basis of scientific quality and only works published in the preceding one or 2 years are accepted for submission. We were able to identify 946 Ph.D. theses that received one or more dissertation awards from a total of 122 different awards. More details can be found in the “Appendix”, Table 5.

A typical example is the Wilhelm Pfeffer Prize of the German Botanical Society, which is described as follows: “The Wilhelm Pfeffer Prize is awarded by the DBG’s Wilhelm Pfeffer Foundation for an outstanding PhD thesis (dissertation) in the field of plant sciences and scientific botany.”¹⁰ The winning thesis is selected by the board of the Foundation. If no work achieves the scientific excellence expected by the board, no prize is awarded.¹¹

Results

Citation data for Ph.D. theses

In total we were able to obtain 329,236 Scopus citations and 476,495 Google Books citations for the 361,971 Ph.D. thesis records. There was an overlap of about 11,000 citations from journals indexed in both sources, which was removed from the Google Books data. The large majority of Scopus citations was found with the primary thesis records only (95%). Secondary (variant) thesis records and thesis records matched as Scopus source document records contributed 5% of the Scopus citations. Scopus and Google Books citation counts are modestly correlated with Pearson’s $r = 0.20$ ($p < 0.01$, 95% CI 0.197–0.203). Table 1 gives an overview of the numbers of citations for theses published in

¹⁰ <https://www.deutsche-botanische-gesellschaft.de/en/about-us/promoting-young-scientists/wilhelm-pfeffer-prize> accessed 26/04/2021.

¹¹ <https://www.deutsche-botanische-gesellschaft.de/ueber-die-dbg/nachwuchsfoerderung/wilhelm-pfeffer-preis/satzung-pfeffer-stiftung> §4, accessed 26/04/2021.

Table 1 Sums of citations by Ph.D. thesis publication year and data source

Year	Ph.D. theses	Mean Scopus citations	Mean Google Books citations	Mean (SD) combined citations per thesis
1996	14,712	1.39	2.25	3.64 (15.08)
1997	14,994	1.35	2.32	3.67 (7.54)
1998	14,967	1.36	2.28	3.64 (10.08)
1999	15,548	1.44	2.19	3.62 (8.01)
2000	15,317	1.47	2.20	3.68 (7.92)
2001	14,584	1.46	2.09	3.55 (7.08)
2002	14,144	1.50	2.11	3.62 (9.44)
2003	14,181	1.48	2.03	3.51 (6.36)
2004	14,865	1.38	1.92	3.30 (6.22)
2005	14,918	1.28	1.75	3.02 (5.52)
2006	15,118	1.26	1.61	2.87 (5.44)
2007	15,548	1.21	1.47	2.67 (6.15)
2008	16,236	1.01	1.20	2.21 (4.02)
2009	16,062	0.89	1.04	1.94 (3.96)
2010	16,266	0.80	0.96	1.76 (3.54)
2011	16,000	0.68	0.84	1.52 (2.92)
2012	16,526	0.54	0.73	1.27 (2.37)
2013	16,856	0.44	0.66	1.09 (1.96)
2014	16,602	0.34	0.59	0.92 (1.66)
2015	18,434	0.21	0.42	0.63 (1.30)
2016	15,659	0.09	0.33	0.42 (0.95)
2017	18,071	0.03	0.18	0.21 (0.63)
2018	16,363	0.00	0.09	0.09 (0.38)
Total	361,971	0.91	1.32	2.23 (6.22)

different years.¹² One can observe a minor overall growth in annual dissertations and modest peak in citations in the early years of the observation period. Overall, and consistently for all thesis publication years, theses are cited more often in Google Books than in Scopus by a ratio of about 3 to 2. Hence, in general, German Ph.D. theses were cited more often in the book literature covered by Google Books than in the periodical literature covered by Scopus. The average number of citations per thesis seems to stabilize at around 3.5, as the values for 1996–2003 are all of that magnitude. As citations were collected in 2019, these works needed about 15 years to reach a level at which they did not increase further. Whereas Kousha and Thelwall (2019) found that 20% of the American dissertations from the period 2013–2017 were cited at least once in Google Scholar, the corresponding figure from our data set is 30% for the combined figure of both citation data sources.

We studied author self-citations of dissertations, both by comparing the thesis author to all publication authors of a citing paper (all-authors self-citations) and only to the first author. We used exact match or Jaro-Winkler similarity greater than 90 out of 100 for author name comparison (Donner 2016). Considering only the first authors of publications

¹² These figures are lower than the official numbers for completed doctoral exams as published by the Federal Statistical Office in Series 11/4/2 because medicine is not included.

Fig. 1 Citation count distribution for German Ph.D. theses from 1997–2003 (values greater than 10 citations not shown), $n = 118,447$

is more lenient in that it does not punish an author for self-citation that could possibly be suggested by co-authors and is at least endorsed by them (Glänzel and Thijs 2004). For the Google Books citation corpus we find only 8366 all-authors self-citations (1.7%) and 5711 first author self-citations (1.2%).¹³ In the Scopus citations there are 52,032 all-author self-citations (15.6%) and 31,260 first-author self-citations (9.4%). Overall this amounts to an all-author self-citation rate of 7.5% and a first-author self-citation rate of 4.6%, quite lower rates than Yoels (1974). We do not exclude any self-citations in the following analyses.

Figure 1 displays the citation count distribution for theses published before 2004, in which citation counts are likely to not increase any further. In this subset, 58% of dissertations have one or more citations. Ph.D. theses exhibit the typical highly skewed and long-tailed citation count distribution.

The distributions of theses and citations by data source over subject classes are displayed in Table 2. There are striking differences in the origin of the major part of citations across disciplines. Social sciences and humanities such as education, economics, German language and literature, history, and especially law are cited much more often in Google Books. The opposite holds in natural science subjects like biology or physics, and in computer science and mathematics, where most citations come from Scopus. This table also indicates that the overall most highly cited dissertations can be found in the humanities (history, archeology and prehistory, religion) and that natural science dissertations are poorly cited (biology, physics and astronomy, chemistry). Much of this difference is probably because in the latter subjects, Ph.D. project results are almost always communicated in journal articles first and citing a dissertation is rarely necessary.

Validation of citation count data with award data

In order to judge whether thesis citation counts can be considered to include a valid signal of scientific quality (research question 5) we studied the citation impact of theses that received dissertation awards compared to those which did not. High citation counts, however, can not simply be equated with high scientific quality. As a rule, the awards for which we collected

¹³ As this is a very low figure, we manually checked the results and did not notice any issues. We note that about 58,000 (about 17%) of the Google Books citation records did not have any names of authors.

Table 2 Sums of citations by Ph.D. thesis publication year and data source

Subject class	Ph.D. theses	Mean Scopus citations	Mean Google Books cita- tions	Mean (SD) combined citations
Agriculture	3094	1.80	1.20	3.00 (5.33)
Archeology and prehistory	1403	3.46	4.10	7.56 (10.26)
Architecture	2632	0.75	1.94	2.70 (4.59)
Biology	40,232	0.50	0.30	0.80 (5.31)
Chemical engineering	4490	1.39	0.39	1.78 (3.93)
Chemistry	35,684	0.31	0.14	0.46 (2.04)
Computer science	11,873	1.97	0.78	2.75 (18.11)
Economics	36,249	0.84	1.63	2.47 (4.85)
Educational science	7923	0.65	1.95	2.60 (4.33)
Electrical engineering	8570	0.93	0.34	1.27 (4.47)
English language and literature	2109	1.19	2.01	3.20 (6.70)
Environmental protection and engineering	1253	1.18	1.90	3.08 (5.14)
Ethnology	1555	0.69	1.80	2.49 (4.24)
General science and culture	497	1.01	1.83	2.84 (5.43)
Geosciences	10,195	1.93	0.80	2.73 (5.85)
German language and literature	6478	0.99	4.42	5.41 (6.42)
History	5372	2.85	6.33	9.18 (11.27)
Home economics, hospitality management studies	257	1.00	1.89	2.89 (4.85)
Industrial engineering	9537	1.63	0.58	2.20 (4.26)
Journalism	992	1.03	4.00	5.03 (6.23)
Languages and literatures other than English and German	4947	1.49	3.13	4.62 (7.31)
Law	37,219	0.27	2.50	2.77 (3.72)
Library and information science, archiving	522	1.26	2.79	4.05 (7.16)
Linguistics	938	1.32	1.46	2.78 (7.05)
Mathematics, statistics	9354	1.65	0.61	2.26 (6.11)
Mechanical engineering	20,059	1.46	0.48	1.94 (5.12)
Music and performance arts	3584	0.66	2.86	3.52 (5.38)
Natural resources, energy, environment	3057	0.73	0.41	1.14 (2.65)
Natural sciences general	735	0.62	0.65	1.27 (6.51)
Philosophy	3881	1.17	3.46	4.64 (6.67)
Physics, astronomy	31,551	0.52	0.24	0.76 (3.00)
Political science, military	10,266	0.95	2.24	3.19 (5.65)
Psychology	5368	0.92	1.28	2.20 (6.93)
Religion	7640	1.52	4.32	5.84 (7.80)
Sociology	7424	1.20	2.35	3.56 (6.06)
Technology general	4381	1.79	0.65	2.45 (5.05)
Trade, communication, traffic	2793	1.02	1.39	2.41 (4.77)
unknown	470	0.70	5.81	6.51 (37.39)
Veterinary medicine	13,322	0.65	0.23	0.88 (2.14)
Visual arts	4065	0.82	3.16	3.98 (6.19)

data are conferred by committees of subject experts explicitly on the basis of scientific quality and importance. But if the content of theses has been published in journal articles before it was published in a thesis it is possible that awards juries might have been influenced by citation counts of these articles.

Comparing the 946 dissertations that were identified as having received a scientific award directly with the non-awarded dissertations we find that the former received on average 3.9 citations while the latter received on average 2.2 citations. To factor out possible differences across subjects and time we match each award-winning thesis with all non-awarded theses of the same subject and year, and calculate the average citation count of this comparable publication set. Award theses obtain 3.9 citations and matched non-award theses 1.7 citations. This shows that Ph.D. theses that receive awards on average are cited more often than comparable theses and indicates that citation counts partially indicate scientific quality as sought out by award committees. Nevertheless, not every award-winning thesis need be highly cited nor every highly cited thesis be of high research quality and awarded a prize. The differences reported here hold on average for large numbers of observations, but do not imply a strong association between scientific quality and citation count on the level of individual dissertations. This is important to note lest the false conclusion be drawn that merely because a thesis is not cited or rarely cited, it is of low quality. Such a view must be emphatically rejected as it is not supported by the data.

A possible objection to the use of award data for validating the relationship of citation counts and the importance and quality of research is that it might be the signal of the award itself which makes the publications more visible to potential citers. In other words, a thesis is highlighted and brought to the attention of researchers by getting an award and high citation counts are a result not of the intrinsic merit of the reported research but merely of the raised visibility. The study by Diekmann et al. (2012) has scrutinized this hypothesis. They studied citation counts (Social Sciences Citation Index) of 102 papers awarded the Thyssen award for German social science journal papers and a random control sample of other publications from the same journals. The award winners are determined by a jury of experts and there are first, second, and third rank awards each year. It was found that awarded papers were cited on average six times after 10 years, control articles two times. Moreover, first rank award articles were cited 9 times, second rank articles 6 times, and third rank articles 4 times on average. The jury decides in the year after the publication of the articles. The authors argue that publications citing awarded articles in the first year after its publication can not possibly have been influenced by the award. For citation counts of a 1-year citation window, awarded articles are cited 0.8 times, control group articles 0.2 times on average. And again, the ranks of the awards correspond to different citation levels. Thus it is evident that the citation reception of the articles are different even before the awards have been decided. Citing researchers and expert committee independently agree on the importance of these articles.

We can replicate this test with our data by restricting the citations to those received in the year of publication of the thesis and the next year. This results in average citation counts of 0.040 for theses which received awards and 0.012 for theses which did not receive any award. Award-winning Ph.D. theses are cited more than three times as often as other theses even before the awards have been granted and before any possible publicity had enough time to manifest itself in increased citation counts.

Application: a preliminary study of citation impact differences between departments

In this section we consider research questions 3 and 4, which are concerned with the existence of differences in Ph.D. thesis citation impact between universities in the same discipline and their magnitude. This application is preliminary because only the citation impact of the thesis but not of the thesis-related publications is considered here and because we use a subject classification based on that of the National Library.¹⁴ In order to mitigate against these limitations as much as possible, we study here only two subjects from the humanities (English and American language and literature, henceforth EALL) and the social sciences (sociology) which are characterized by Ph.D. theses mainly published as monographs rather than as articles and thus show relatively high dissertation citation counts. These are also disciplines specifically identifiable in the classification system and as distinct university departments. We furthermore restrict the thesis publication years to the years covered by the national scale pilot research assessment exercises discussed in the introduction section which were carried out by the German Council of Science and Humanities in these two disciplines (Wissenschaftsrat 2008, 2012) in order to be able to compare the results and to test if the number of observations in typical evaluation periods yield sufficient data for useful results.

We use multilevel Bayesian negative binomial regressions in which the observations (theses) are nested in the grouping factor universities, estimated with the R package *brms*, version 2.8.0 (Bürkner 2017). By using a multilevel model we can explicitly model the correlation of observations within a group, that is to say, take into account the characteristics of the departments which affect all Ph.D. candidates of a department and their research quality. The negative binomial response distribution is appropriate for the characteristic highly skewed, non-negative citation count distribution. The default prior distributions of *brms* were used. Model estimations are run with two MCMC chains with 2500 iterations each, of which the first 500 are for sampler warmup. As population level variables we include a thesis language dummy (German [reference category], English, Unknown/Other), the publication year, and dummies for whether the dissertation received an award (reference category: no award received). There were no awards identified for EALL in the observation period, so this variable only applies to the Sociology models. For the language variable we used the data from the National Library where available and supplemented it with automatically identified language of the dissertation title using R package *clld3*, version 1.1. Languages other than English and German and theses with unidentifiable titles were coded as Unknown/Other.

Models are run for the two disciplines separately, once without including the university variable (null models) and once including it as a random intercept, making these multilevel models. If the multilevel model shows a better fit, i.e. can explain the data better, this would indicate significant variation between the different university departments in a discipline and higher similarity of citation impact of theses within a university than expected

¹⁴ This is not ideal because its classes are not aligned with the delimitations of university departments. The National Library data do not contain information about the university department which accepted a thesis. As this exercise is not intended to be understood as an evaluation of university departments but only as a feasibility study, we still use the classification in lack of a better alternative. It goes without saying that for any actual evaluation use this course of action is not appropriate. Instead, a verified list of the dissertations of each unit would be required and all publications related to the theses would have to be included.

Table 3 Regression results

	Sociology 2001–05		EALL 2004–10	
	A	B	C	D
Intercept	– 62.62 [– 167.20; 41.35]	– 68.60 [– 177.23; 37.61]	269.08 ^a [146.01; 385.31]	258.86 ^a [143.62; 374.47]
Year	0.03 [– 0.02; 0.08]	0.04 [– 0.02; 0.09]	– 0.13 ^a [– 0.19; – 0.07]	– 0.13 ^a [– 0.19; – 0.07]
Language (ref: German)				
English	– 0.19 [– 0.45; 0.08]	– 0.30 ^a [– 0.58; – 0.02]	0.60 ^a [0.36; 0.84]	0.51 ^a [0.28; 0.75]
Unknown/Other	0.06 [– 0.34; 0.50]	0.06 [– 0.37; 0.50]	– 0.55 [– 1.40; 0.41]	– 0.59 [– 1.41; 0.28]
Award	1.62 ^a [0.63; 2.73]	1.54 ^a [0.52; 2.58]		
R ²	0.07	0.11	0.05	0.08
Num. obs.	1028	1028	569	569
SD: university group effect		0.26		0.34

Negative binomial models for citation counts. Interval in brackets is 95% highest posterior density interval

^aInterval of estimate does not include 0

at random. In other words, the null model assumes independence of observations while the multilevel model controls for correlation of units within a group, here the citation counts of theses from a university. The results are presented in Table 3. The coefficient of determination (row ‘R²’) is calculated according to Nakagawa et al. (2017).

Regarding the population level variables, it can be seen that the publication year has a negative effect in EALL (younger theses received less citations, as expected) but no significant effect in Sociology. As the Sociology models data is from an earlier time period, this supports the notion that the citation counts used in the Sociology models have stabilized but those from EALL have not. According to the credible intervals, while there is no significant language effect on citations in the Sociology null model, controlling for the grouping structure reveals the language to be significant predictor. In EALL, English language theses received substantially more citations than German language theses. There is a strong positive effect in Sociology from having received an award for the Ph.D. thesis. If we compare the null models with their respective multilevel models, that is A to B and C to D, we can see that introducing the grouping structure does not affect the population level variable effects other than language for Sociology. In both disciplines, the group effect (standard deviation of the random intercept) is significantly above zero and the model fit in terms of R² improved, indicating that the hierarchical model is more appropriate and that the university department is a significant predictor. However, the values of the coefficients of determination are small, which suggests that it is not so much the included population level predictors and the group membership, but additional unobserved thesis-level characteristics that affect citation count. In addition, this means it is not possible to estimate with any accuracy a particular thesis’ citation impact only from knowing the department at which it originated. The estimated group effects describe the citation impact of particular university departments. The distribution of the estimates from models B and D with associated 95% credible intervals are

Fig. 2 Ph.D. thesis citation impact regression group effect estimates for **a** 49 sociology departments (2001–2005) and **b** 52 English and American language and literature departments (2004–2010). Means and 95% posterior probability ranges

Table 4 Comparison of rated levels of early career researcher support and theses citation impact for 47 German sociology departments (2001–2005)

ECR support expert committee rating	Mean (SD) of impact estimates	Observations
not satisfactory	− 0.01 (0.22)	3
satisfactory	− 0.06 (0.11)	14
good	0.00 (0.18)	19
very good	0.10 (0.11)	8
excellent	0.22 (0.04)	3

displayed in Fig. 2. It is evident that while there are substantial differences in the estimates as a whole, there is also large uncertainty about the exact magnitude about all effects, indicated by the wide credible intervals. This a consequence of the facts that, first, most departments produce theses across the range of citation impact and small differences in the ratios of high, middle and low impact theses determine the estimated group effects and, second, given the high within-group variability, there are likely too few observations in the applied time period to arrive at more precise estimates.

The results of the above mentioned research evaluation of sociology departments in the *Forschungsrating* series of national scale pilot assessments allow for a comparison between the university group effects of Ph.D. thesis citation impact obtained in the present study and qualitative ordinal ratings given by the expert committee in the category ‘support of early career researchers.’¹⁵ In the sociology assessment exercise, the ECR support dimension was explicitly intended to reflect both the supportive actions of the departments and their successes. The reviewer panel put special weight on the presence of structured degree programs and scholarships and obtained professorship positions of graduates. Further indicators that were taken into account were the number of conferred doctorates, the list of Ph.D. theses with publisher information, and a self-report of actions and successes.

¹⁵ This is not possible for EALL because early career researcher support was not assessed separately but compounded with acquisition of third party funding and ‘infrastructures and networks’ into a dimension called ‘enabling of research’ and the ratings in this category were further differentiated by four sub-fields (Wissenschaftsrat 2012, p. 19).

Fig. 3 ECR support expert committee rating and mean estimates of Ph.D. thesis citation impact of 47 German sociology departments (2001–2005)

This dimension was rated by the committee on a five point scale ranging from 1, ‘not satisfactory’, to 5, ‘excellent’ (Wissenschaftsrat 2008, p. 22).

For 47 universities both an estimated citation impact score (group effect coefficient from the above regression) and an ECR support rating were available for sociology in the same observation period. A tabulated comparison is presented in Table 4. The Kendall τ rank correlation between these two variables at the department level is 0.36 ($p < 0.01$), indicating a moderate association, but the mean citation scores in the table do not exhibit a clear pattern of increase with increasing expert-rated ECR support. The bulk of departments were rated in the lower middle and middle categories, that is to say, the ratings are highly concentrated in this range, making distinctions quite difficult.

This relationship is displayed in Fig. 3. It can be seen that while the five rating groups do have some association with the citation analysis estimates, there is large variability within the groups, especially for categories 3 and 4. In fact, there is much overlap across all the rating groups in terms of citation impact scores. The university department with the highest citation impact effect estimate was rated as belonging to the middle groups of ECR support. In summary, the association between rated ECR support of a department and the impact of the department’s citation is demonstrable but moderate in size.

Discussion

In this study we have demonstrated the utility of the combination of citation data from two distinct sources, Scopus and Google Books, for citation analysis of Ph.D. theses in the context of research evaluation of university departments. Our study has a number of limitations hampering its immediate implementation into practice. We did not have verified data about the theses produced by departments and used publication sets approximated by using a subject classification. We did not take into account the publications of Ph.D.’s other than the theses, such as journal articles, proceedings papers, and book chapters, which resulted in low citation counts in the sciences. These limitations must be resolved before any application for research evaluation and the present study is to be understood as a feasibility study only.

We now turn to the discussion of the results as they relate to the guiding research questions. The first research question concerned the typical citation counts of Ph.D. theses. We found that German Ph.D. theses were cited more often in the book literature covered by Google Books than in the periodical literature covered by Scopus and that it takes 10–15 years for citation counts to reach a stable level where they do not increase any further. At this stage, about 40% of theses remained uncited. We further found large differences in typical citation rates across fields. Theses are cited more often in the social sciences and humanities, especially in history and in archeology and prehistory. But citation rates were very low in physics and astronomy, chemistry and veterinary medicine. Furthermore, there were distinctive patterns of the origin of the bulk of citations between the two data sources, in line with the typical publication conventions of the sciences and the social sciences and humanities. Science theses received more citations from Scopus, that is, primarily the journal literature, than from the book literature covered by Google Books. The social sciences and humanities, on the other hand, obtained far more citations from the book literature covered in Google Books. Nevertheless, these fields' theses do also receive substantial numbers of citations from the journal literature which must not be neglected. Thus with regard to our second research question we can state that Google Books is clearly a very useful citation data source for social sciences and humanities dissertations and most likely also for other publications beyond dissertations from these areas of research. Citations from Google Books are complementary to those from Scopus as they clearly cover a different literature; the two data sources have little overlap.

Our results of multilevel regressions allow us to affirm that there are clearly observable thesis citation impact differences between departments in a given discipline (research question 3). However, they are of small to moderate magnitude and the major part of citation count variation is found at the individual level (research question 4). Our results do not allow any statements about what factors caused these differences. They could be the result of mutual selection processes or of different levels of support or research capacity across departments. Our results of citation impact of departments, interpretable as collective performance of early career researchers at the university department level, are roughly in line with qualitative ratings of an expert committee in our case study of sociology. This association also does not rule out or confirm any possible explanations, as both department supportive actions and individual ECR successes were compounded in the rated dimension which furthermore showed only limited variation.

Research question 5 concerned the association of citation counts of Ph.D. theses and their research quality. Using information about dissertation awards we showed that theses which received such awards also received higher citation rates vis-à-vis comparable theses which did not win such awards. This result strongly suggests that if researchers cite Ph.D. theses, they do tend to prefer the higher quality ones, as recognized by award granting bodies and committees.

Appendix

Table 5 Ph.D. awards used in the study

Award granting body (specific award)	Awards in the data set
Adolf-Martens-Fonds e. V.	11
Anatomische Gesellschaft	1
Arbeitsgemeinschaft Magnetismus der DPG	3
Arbeitskreis Prozessanalytik der Fachgruppe Analytische Chemie der Gesellschaft Deutscher Chemiker	2
Arbeitskreis der Hochschullehrer für Messtechnik e.V.	15
Behnken-Berger-Stiftung	2
Berliner Wissenschaftliche Gesellschaft	1
Berufsverband Oecotrophologie e.V.	6
Bundesinstitut für Kultur und Geschichte der Deutschen im östlichen Europa (Immanuel-Kant-Forschungspreis)	1
Con Sozial	5
DECHEMA (Nachwuchswissenschaftler-Preis für Naturstoffforschung)	6
Daimler Benz Stiftung	8
Demokratie-Stiftung der Universität zu Köln	3
Deutsch-Schweizerischer Fachverband für Strahlenschutz (FS)	1
Deutsche Arbeitsgemeinschaft für Mustererkennung e.V.	2
Deutsche Botanische Gesellschaft (Strasburger Preis)	4
Deutsche Botanische Gesellschaft (Wilhelm-Pfeffer-Preis)	8
Deutsche Dendrologische Gesellschaft	5
Deutsche Diabetes Gesellschaft	1
Deutsche Gesellschaft für Biomedizinische Technik/Stiftung Familie Klee	5
Deutsche Gesellschaft für Demographie e.V.	16
Deutsche Gesellschaft für Elektronenmikroskopie	3
Deutsche Gesellschaft für Entwicklungsbiologie	5
Deutsche Gesellschaft für Evaluation e.V.	7
Deutsche Gesellschaft für Geographie	10
Deutsche Gesellschaft für Hygiene und Mikrobiologie (DZIF Promotionspreis der DGHM)	3
Deutsche Gesellschaft für Hygiene und Mikrobiologie (Promotionspreis der DGHM)	6
Deutsche Gesellschaft für Immunologie e.V. (Otto-Westphal-Promotionspreis)	5
Deutsche Gesellschaft für Luft- und Raumfahrtmedizin	4
Deutsche Gesellschaft für Massenspektrometrie	29
Deutsche Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie e.V.	2
Deutsche Gesellschaft für Mykologie e.V.	3
Deutsche Gesellschaft für Nephrologie e.V.	1
Deutsche Gesellschaft für Osteuropakunde e.V.	18
Deutsche Gesellschaft für Parasitologie	2
Deutsche Gesellschaft für Phoniatrie und Pädaudiologie e.V.	2
Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde e.V.	1
Deutsche Gesellschaft für Psychologie	8
Deutsche Gesellschaft für Psychologie (DGPS) (C.F. Graumann Preis)	4
Deutsche Gesellschaft für Psychologie (DGPS) (Dissertationspreis der Fachgruppe Arbeits-, Organisations- und Wirtschaftspsychologie)	10

Table 5 (continued)

Award granting body (specific award)	Awards in the data set
Deutsche Gesellschaft für Psychologie (DGPS) (Gustav A. Lienert-Preis)	11
Deutsche Gesellschaft für Psychologie (DGPS) (Margret-und-Paul-Baltes-Preis)	9
Deutsche Gesellschaft für Psychologie (DGPS) (Nachwuchsförderpreis der Fachgruppe Rechtspychologie)	3
Deutsche Gesellschaft für Psychologische Schmerztherapie und -forschung	1
Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK)(Dissertationsförderpreis)	10
Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK)(Dissertationspreis “Medien-Kultur-Kommunikation”)	6
Deutsche Gesellschaft für Publizistik- und Kommunikationswissenschaft (DGPuK)(Herbert von Halem Nachwuchspreis)	2
Deutsche Gesellschaft für Radioonkologie e.V.	1
Deutsche Gesellschaft für Rechtsmedizin	3
Deutsche Gesellschaft für Schlafforschung und Schlafmedizin (Promotions-Nachwuchsförderpreis “Becker-Carus”)	3
Deutsche Gesellschaft für Soziologie	17
Deutsche Gesellschaft für Sprachwissenschaft (DGfS)	4
Deutsche Gesellschaft für experimentelle und klinische Pharmakologie und Toxikologie e.V.	1
Deutsche Mathematiker Vereinigung	5
Deutsche Physikalische Gesellschaft (Dissertationspreis der Sektion Atome, Moleküle, Quantenoptik und Plasmen der DPG)	9
Deutsche Physikalische Gesellschaft (Dissertationspreis der Sektion Materie und Kosmos)	10
Deutsche Physikalische Gesellschaft (Dissertationspreis der Sektion kondensierte Materie)	11
Deutsche Quartärvereinigung e.V.	10
Deutsche Stiftung für Recht und Informatik	12
Deutsche Vereinigung für Politikwissenschaft (DVPW)	15
Deutsche Vereinigung für Wasserwirtschaft, Abwasser und Abfall DWA e.V.	6
Deutsche Zoologische Gesellschaft	10
Deutscher Arbeitskreis für Analytische Spektroskopie der Fachgruppe Analytische Chemie der Gesellschaft Deutscher Chemiker	6
Deutscher Verein für öffentliche und private Fürsorge	3
Deutsches Aktieninstitut	6
Deutsches Kollegium für Psychosomatische Medizin	1
Deutsches wissenschaftliches Institut der Steuerberater e.V.	9
EHI Stiftung	10
European Marketing Academy (EMAC McKinsey Marketing Dissertation Award)	5
Fachgruppe Bauchemie der Gesellschaft Deutscher Chemiker	5
Fachgruppe Biochemie der Gesellschaft Deutscher Chemiker	2
Fachgruppe Chemie des Waschens der Gesellschaft Deutscher Chemiker	12
Fachgruppe Computer in der Chemie der Gesellschaft Deutscher Chemiker	7
Fachgruppe Elektrochemie der Gesellschaft Deutscher Chemiker	11
Fachgruppe Festkörperchemie und Materialforschung der Deutschen Gesellschaft für Chemiker	4
Fachgruppe Medizinische Chemie der Deutschen Gesellschaft für Chemiker	3
Fachgruppe Nachhaltige Chemie der Deutschen Gesellschaft für Chemiker	2

Table 5 (continued)

Award granting body (specific award)	Awards in the data set
Fachgruppe Nuklearchemie der Deutschen Gesellschaft für Chemiker	12
Fachgruppe Photochemie der Deutschen Gesellschaft für Chemiker	13
Fachgruppe Wasserchemische Gesellschaft der Gesellschaft Deutscher Chemiker	22
Forum InformatikerInnen für Frieden und gesellschaftliche Verantwortung e.V. (FIfF)	1
Gesellschaft - Altern - Medien e.V.	3
Gesellschaft Deutscher Chemiker (Gerhard-Billek-Preis)	8
Gesellschaft Deutscher Chemiker (Klaus-Grohe-Stiftung für medizinische Chemie)	13
Gesellschaft Deutscher Chemiker (Starck Promotionspreis für Festkörperchemie und Materialforschung)	14
Gesellschaft für Angewandte Mathematik und Mechanik (GAMM)	15
Gesellschaft für Biochemie und Molekularbiologie e.V. (Bayer Healthcare Promotionspreis)	5
Gesellschaft für Biochemie und Molekularbiologie e.V. (GBM-Promotionspreis)	4
Gesellschaft für Biochemie und Molekularbiologie e.V. (Karl-Lohmann-Preis)	5
Gesellschaft für Biologische Systematik	6
Gesellschaft für Epilepsieforschung e.V.	2
Gesellschaft für Genetik	11
Gesellschaft für Informatik, Fachbereich Datenbanken und Informationssysteme	15
Gesellschaft für ökologie (Dissertations-Förderpreis)	2
Gesellschaft für Ökologie (Horst-Wiehe-Förderpreis)	3
Goldmedia	2
Gregor Louisoder Umweltstiftung	11
Hans-Lilje-Stiftungspreis	2
Harassowitz Verlag	3
Heinrich Stockmeyer Stiftung	10
Hochschulverband für Geographiedidaktik	8
Horizont Stiftung	5
Klaus-Tschira Stiftung	46
Leibniz-Institut für Primatenforschung	17
Marie-Luise und Ernst Becker Stiftung	3
NEUROWIND e.V.	1
Roman Herzog Institut	10
Schmalenbach-Gesellschaft für Betriebswirtschaft e.V.	19
Schroubek Fonds Östliches Europa	4
Statistisches Bundesamt (Destatis)	24
Stiftung Industrieforschung	1
Stiftung Kapitalmarktforschung für den Finanzstandort Deutschland	1
Stiftung Leben Pur	1
Stiftung Steine-Erden-Bergbau und Umwelt	3
Tyczka Energiestiftung	7
Verein Deutscher Ingenieure	4
Verein von Altertumsfreunden im Regierungsbezirk Darmstadt e. V.	5
Vereinigung für Allgemeine und Angewandte Mikrobiologie	52
Verlag Karl Alber	8

Table 5 (continued)

Award granting body (specific award)	Awards in the data set
Vodafone Stiftung für Forschung	16
Willy-Hager-Stiftung	16
Wolfgang Ritter Stiftung	39

Acknowledgements Funding was provided by the German Federal Ministry of Education and Research [Grant Numbers 01PQ16004 and 01PQ17001]. We thank Beatrice Schulz for help with data collection.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *Sage Open*, 9(1). <https://doi.org/10.1177/2158244019829575>.
- Allison, P. D., & Long, J. S. (1990). Departmental effects on scientific productivity. *American Sociological Review*, 55(4), 469–478. <https://doi.org/10.2307/2095801>.
- Bornmann, L., & Daniel, H.-D. (2008). What do citation counts measure? A review of studies on citing behavior. *Journal of Documentation*, 64(1), 45–80. <https://doi.org/10.1108/00220410810844150>.
- Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1), 1–28.
- Chudaja, M. (2021). *Promotionen an der Charité Berlin von 1998 bis 2015. Qualität, Dauer, Promotionstyp* (Ph.D. thesis). Charité - Universitätsmedizin Berlin.
- Consortium for the National Report on Junior Scholars. (2017). *2017 National Report on Junior Scholars. Statistical Data and Research Findings on Doctoral Students and Doctorate Holders in Germany. Overview of Key Results*. Retrieved from <https://www.buwin.de/dateien/buwin-2017-keyresults.pdf>.
- Diekmann, A., Näf, M., & Schubiger, M. (2012). Die Rezeption (Thyssen-) preisgekrönter Artikel in der “Scientific Community.” *Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 64(3), 563–581. <https://doi.org/10.1007/s11577-012-0175-4>.
- Donner, P. (2016). Enhanced self-citation detection by fuzzy author name matching and complementary error estimates. *Journal of the Association for Information Science and Technology*, 67(3), 662–670. <https://doi.org/10.1002/asi.23399>.
- Donner, P. (2020). A validation of coauthorship credit models with empirical data from the contributions of PhD candidates. *Quantitative Science Studies*, 1(2), 551–564. https://doi.org/10.1162/qss_a_00048.
- 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>
- García-Pérez, M. A. (2010). Accuracy and completeness of publication and citation records in the Web of Science, PsycINFO, and Google Scholar: A case study for the computation of h indices in Psychology. *Journal of the American Society for Information Science and Technology*, 61(10), 2070–2085. <https://doi.org/10.1002/asi.21372>.

- Glänzel, W., & Thijs, B. (2004). Does co-authorship inflate the share of self-citations? *Scientometrics*, 61(3), 395–404. <https://doi.org/10.1023/b:scie.0000045117.13348.b1>.
- Hay, A. (1985). Some differences in citation between articles based on thesis work and those written by established researchers: Human geography in the UK 1974–1984. *Social Science Information Studies*, 5(2), 81–85. [https://doi.org/10.1016/0143-6236\(85\)90017-1](https://doi.org/10.1016/0143-6236(85)90017-1).
- Heinisch, D. P., & Buenstorf, G. (2018). The next generation (plus one): An analysis of doctoral student's academic fecundity based on a novel approach to advisor identification. *Scientometrics*, 117(1), 351–380. <https://doi.org/10.1007/s11192-018-2840-5>
- Heinisch, D. P., Koenig, J., & Otto, A. (2020). A supervised machine learning approach to trace doctorate recipient's employment trajectories. *Quantitative Science Studies*, 1(1), 94–116. https://doi.org/10.1162/qss_a_00001
- Hemlin, S. (1996). Research on research evaluation. *Social Epistemology*, 10(2), 209–250. <https://doi.org/10.1080/02691729608578815>.
- Hesli, V. L., & Lee, J. M. (2011). Faculty research productivity: Why do some of our colleagues publish more than others? *PS: Political Science & Politics*, 44(2), 393–408. <https://doi.org/10.1017/S1049096511000242>.
- Hilmer, C. E., & Hilmer, M. J. (2007). On the relationship between the student-advisor match and early career research productivity for agricultural and resource economics Ph.Ds. *American Journal of Agricultural Economics*, 89(1), 162–175. <https://doi.org/10.1111/j.1467-8276.2007.00970.x>.
- Hinze, S., Butler, L., Donner, P., & McAllister, I. (2019). Different processes, similar results? A comparison of performance assessment in three countries. In W. Glänzel, H. F. Moed, U. Schmoch, & M. Thelwall (Eds.), *Springer handbook of science and technology indicators* (pp. 465–484). Berlin: Springer. https://doi.org/10.1007/978-3-030-02511-3_18.
- Hochschulrektorenkonferenz. (2012). *Zur Qualitätssicherung in Promotionsverfahren*. Retrieved from <https://www.hrk.de/positionen/beschluss/detail/zur-qualitaetssicherung-in-promotionsverfahren/>.
- Jaeger, M. (2006). Leistungsbezogene Budgetierung an deutschen Universitäten - Umsetzung und Perspektiven. *Wissenschaftsmanagement*, 12(3), 32–38.
- Kim, K., & Karau, S. J. (2009). Working environment and the research productivity of doctoral students in management. *Journal of Education for Business*, 85(2), 101–106. <https://doi.org/10.1080/08832320903258535>.
- 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. (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>.
- 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., 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>.
- Laudel, G., & Gläser, J. (2008). From apprentice to colleague: The metamorphosis of early career researchers. *Higher Education*, 55(3), 387–406. <https://doi.org/10.1007/s10734-007-9063-7>.
- Long, J. S., & McGinnis, R. (1985). The effects of the mentor on the academic career. *Scientometrics*, 7(3–6), 255–280. <https://doi.org/10.1007/BF02017149>.
- Martin, B. R., & Irvine, J. (1983). Assessing basic research: Some partial indicators of scientific progress in radio astronomy. *Research Policy*, 12(2), 61–90. [https://doi.org/10.1016/0048-7333\(83\)90005-7](https://doi.org/10.1016/0048-7333(83)90005-7).
- Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R^2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface*, 14(134), 20170213. <https://doi.org/10.1098/rsif.2017.0213>.
- Nederhof, A., & van Raan, A. (1987). Peer review and bibliometric indicators of scientific performance: A comparison of cum laude doctorates with ordinary doctorates in physics. *Scientometrics*, 11(5–6), 333–350. <https://doi.org/10.1007/bf02279353>.
- Nederhof, A., & van Raan, A. (1989). A validation study of bibliometric indicators: The comparative performance of cum laude doctorates in chemistry. *Scientometrics*, 17(5–6), 427–435. <https://doi.org/10.1007/BF02017463>.
- Niggemann, F. (2020). Interne LOM und ZLV als Instrumente der Universitätsleitung. *Qualität in der Wissenschaft*, 14(4), 94–98.
- Oestmann, J. W., Meyer, M., & Ziemann, E. (2015). Medizinische Promotionen: Höhere wissenschaftliche Effizienz. *Deutsches Ärzteblatt*, 112(42), A–1706/B–1416/C–1388.

- Projektgruppe Indikatorenmodell. (2014). *Indikatorenmodell für die Berichterstattung zum wissenschaftlichen Nachwuchs*. Retrieved from <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/indikatorenmodell-endbericht.pdf>.
- Rogge, J.-C., & Tesch, J. (2016). Wissenschaftspolitik und wissenschaftliche Karriere. In D. Simon, A. Knie, S. Hornbostel, & K. Zimmermann (Eds.), *Handbuch Wissenschaftspolitik* (2nd ed., pp. 355–374). Berlin: Springer. https://doi.org/10.1007/978-3-658-05455-7_25
- Statistisches Bundesamt. (2018). *Prüfungen an Hochschulen 2017*. Retrieved from <https://www.destatis.de/DE/Themen/Gesellschaft-Umwelt/Bildung-Forschung-Kultur/Hochschulen/Publikationen/Downloads-Hochschulen/pruefungen-hochschulen-2110420177004.pdf>.
- Vollmar, M. (2019). Neue Promovierendenstatistik: Analyse der ersten Erhebung 2017. *WISTA - Wirtschaft und Statistik*, 1, 68–80.
- Wespe, J., & Jaeger, M. (2015). Leistungsorientierte Zuweisungsverfahren der Länder: Praktische Umsetzung und Entwicklungen. *Hochschulmanagement*, 10(3+4), 97–105.
- Wissenschaftsrat. (2002). *Empfehlungen zur Doktorandenausbildung*. Saarbrücken.
- Wissenschaftsrat. (2004). *Empfehlungen zu forschungs- und lehrförderlichen Strukturen in der Universitätsmedizin*. Berlin.
- Wissenschaftsrat. (2007). *Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Chemie*. Köln.
- Wissenschaftsrat. (2008). *Forschungsleistungen deutscher Universitäten und außeruniversitärer Einrichtungen in der Soziologie*. Köln.
- Wissenschaftsrat. (2011). *Ergebnisse des Forschungsratings Elektrotechnik und Informationstechnik*. Köln.
- Wissenschaftsrat. (2012). *Ergebnisse des Forschungsratings Anglistik und Amerikanistik*. Köln.
- Yoels, W. C. (1974). On the fate of the Ph.D. dissertation: A comparative examination of the physical and social sciences. *Sociological Focus*, 7(1), 1–13. <https://doi.org/10.1080/00380237.1974.10570872>.