



Patterns of dissertation dissemination: publication-based outcomes of doctoral theses in the social sciences

Anastasiya-Mariya Asanov¹ · Igor Asanov¹ · Guido Buenstorf¹ · Valon Kadriu¹ · Pia Schoch¹

Received: 28 August 2023 / Accepted: 18 January 2024 / Published online: 29 February 2024
© The Author(s) 2024

Abstract

Dissemination of knowledge through the publication of findings is a cornerstone of the academic research system. Doctoral dissertations document the findings made by early-stage researchers during their doctoral studies. However, prior research suggests that dissertations may not be effective in disseminating these findings to the broader community of researchers. We study how knowledge documented in doctoral dissertations is disseminated. Specifically, we investigate which dissertation characteristics and institutional factors are related to the number of journal publications based on these dissertations and the number of citations that these publications receive. Our analysis uses a random sample of doctoral dissertations from German universities in economics, political science, and sociology. We find that “cumulative” dissertations—dissertations consisting of a number of separate articles—are turned into three times more publications which receive three times more citations than monographic dissertations. We also find explorative evidence that dissertations written in English and empirical dissertations have higher publication-based outcomes. We conclude that a policy allowing doctoral candidates to write their dissertations in a cumulative format provides them with an opportunity to share the results of their research through publications in peer-reviewed journals.

Keywords Ph.D. thesis · Doctoral students · Citation analysis · Research performance · Knowledge dissemination

✉ Anastasiya-Mariya Asanov
noha@incher.uni-kassel.de

Igor Asanov
igor.asanov@uni-kassel.de

Guido Buenstorf
buenstorf@uni-kassel.de

Valon Kadriu
kadriu@uni-kassel.de

Pia Schoch
p.schoch@uni-kassel.de

¹ INCHER, University of Kassel, Kassel, Germany

Introduction

Doctoral candidates contribute in many ways to the productivity of their universities—for example, as research and teaching assistants (Kifor et al., 2023; Larivière, 2012; Rodríguez-Montoya et al., 2023). However, their primary task and the key prerequisite for being awarded a doctoral degree is to make a research contribution to the respective discipline, which is documented in the doctoral dissertation. Moreover, dissertations turn into published papers serving dissemination of the acquired and documented knowledge. Yet, not many factors are known that determine the success of knowledge dissemination from dissertations (Mayir et al., 2017; Paglis et al., 2006; Rojko et al., 2020).

Depending on the country and discipline, it takes about 3 to 6 years for a doctoral candidate to fulfil all requirements for the doctoral degree, including submitting a comprehensive dissertation of 100 to 400 pages and to graduate (Günauer et al., 2013; Siegfried & Stock, 2001). Prior evidence indicates that dissertations contain high-quality research. For example, one review (McLeod & Weisz, 2004) shows that the methodology in experiments from dissertations was stronger, while the mean effect size reported in dissertations was smaller than in the published articles in the same field, suggesting the dissertations were less susceptible to overestimating the effect size than the published articles. Despite this, compared to journal publications, the citation rate of dissertations in the scientific literature has declined over time (Larivière et al., 2008). Less than 12% of the dissertations produced by UK doctoral candidates have at least one citation on platforms like Scopus, Microsoft Academic, or Google Books (Kousha & Thelwall, 2020).

Even though dissertations themselves are not cited much in the academic literature, one might expect that the research and data that dissertations are based on contribute to publications in peer-reviewed journals. However, research shows that only about 25–29% of dissertations in psychology, counseling, and social work (Evans et al., 2018; Maynard et al., 2014; Osborn et al., 2023) ended up with at least one article derived from them published in a peer-reviewed journal. Similarly, only around 40% of electronic theses and dissertations in engineering produced at a South African university received at least one citation, and only 16.8% of them were converted into research outputs such as books, journal articles, or conference proceedings (Bangani, 2018). Studies of medical theses dissemination show similarly low publication rates of 17% in France (Salmi et al., 2001), 17.6% in Peru (Arriola-Quiroz et al., 2010), and 23.8% in Finland (Nieminen et al., 2007). Finally, 53.2% of dissertations in Turkish language education are turned into journal publications (Karagöz & Şeref, 2021). These prior findings indicate that substantial resources are dedicated to producing high-quality research that is documented in doctoral dissertations but often not disseminated to the broader community of researchers. Yet, little is known about how universities can enhance dissertation-based research dissemination. Several studies have explored individual factors associated with the research productivity of doctoral candidates. Paglis et al. (2006) found no significant association between advisor mentoring and research productivity, which was defined as the total number of conference papers, journal publications, book chapters, and grant proposals accepted. Rojko et al. (2020) did not find a significant difference in the average publication performance of doctoral candidates before and after the implementation of the Bologna reform in Slovenia.

Other research has explored differences in the extent of doctoral dissertation dissemination. For example, making doctoral dissertations available through open-access repositories at universities resulted in higher citation counts (Ferrerías-Fernández et al., 2013). Mayir et al. (2017) did not reveal an association between the publication rate and citation

counts of dissertations in surgery and the type of study on which a dissertation was based (e.g., randomized study, case study, cross-sectional study). Closest to our study, Smaldone et al. (2019) compared the number of peer-reviewed publications based on dissertations from the Columbia University School of Nursing that were written in a monographic or an article-based (i.e., “cumulative”) format. The study found that article-based dissertations were associated with larger numbers of publications in peer-reviewed journals. Similarly, a survey of Australian students and alumni (Thomas et al., 2016) from instructional technology programs found that those who chose an article-based dissertation format reported receiving more citations on their dissertations.

In this study, using a random sample of German dissertations, we investigate dissertation characteristics and institutional factors that may be related to higher research output. We quantify research output as the number of papers based on the dissertation that are published in peer-reviewed journals and the number of citations from these papers.

We aim to answer the following four questions: First, do rates of publications based on dissertations and their citations differ between economics, sociology, and political science; second, do monographic and “cumulative” (article-based) dissertations differ in publication and citation rates; third, do dissertations from universities with and without an established graduate school or graduate academy differ in publication and citation rates; and fourth, do dissertations from universities that were successful in the German excellence initiative differ in publication and citation rates.

Materials and methods

Pre-analysis plan

Prior to conducting the empirical analysis underlying the present paper, we specified in a pre-analysis plan the process of sampling, data collection, a set of outcomes and explanatory variables, a set of control variables, a set of hypotheses, and the empirical strategy. We store the pre-analysis plan and a replication package at Open Science Framework: <https://doi.org/10.17605/OSF.IO/U7M2A>.

Sampling

We selected a sample of 1500 doctoral dissertations from 2004 to 2006 and 1500 doctoral dissertations from 2012 to 2014, drawn randomly from the German National Library’s database of all dissertations published in Germany during those years. We determined the desired sample size and sampling strategy based on statistical power calculations before data collection to have a representative sample of dissertations for the sampling frame years. We focused on dissertations classified under the fields of Economy (“Wirtschaft”), Politics (“Politik”), and Social Sciences (“Sozialwissenschaften, Soziologie, Anthropologie”). To ensure accuracy, our team manually classified the dissertations in the sample, particularly distinguishing between those in economics and those in management sciences, which are both classified as “Wirtschaft” in the database. After this classification process, we were left with a total of 1840 dissertations from 73 German universities. Given 1840 dissertations (observations) across 73 universities (clusters) with observed intra-university (-cluster) correlation (ICC) of 0.04 for the main outcomes, assuming a conventional

significance level of 5% and 80% of statistical power, the minimum detectable effect (MDE) size is 0.19 of a standard deviation. Thus, we have sufficient statistical power even to detect a correlation of about 0.1 (r), which can be considered a small effect size given empirically observed effect sizes in observational research in economics (Ioannidis et al., 2017).

Outcome variables

We were interested in two primary outcomes—the number of publications based on the doctoral dissertation and the total number of citations from these publications. Relevant publications were identified as follows. Initially, our team of research assistants, under the supervision of the research team, exhaustively searched for one peer-reviewed publication of each author in various sources (the author’s personal website, the author’s university page, Web of Science, Google Scholar, and WISO) and recorded the author identifiers. To ensure the correct person was found, we verified that the author’s publication lists contained the dissertation as well. Author identifiers were then matched with the Scopus database, and the list of publications was cross-referenced with the dissertation itself to identify which publications were based on the dissertation. The identification of dissertation-based publications was conducted manually by research assistants based on a formalized algorithm created by us, which entailed comparing titles, abstracts, and, if necessary, the introductions of the dissertation and each publication. Research assistants were encouraged to leave comments regarding uncertainties, which were then resolved by a member of the research team. To additionally ensure accuracy, the research assistants’ work was systematically and independently double-checked at random by senior researchers, i.e., a random sample of 30% dissertations was extracted to check for potential mistakes. In both instances—resolving an uncertainty and double-checking a random sample of dissertations—the senior researchers followed the same formalized protocol: check whether the title, abstract, and if necessary, introduction match. To determine whether the paper and a dissertation match, the algorithm required to compare the topic and the object of the study, the study sample, the method and the location of the study. When there was a significant overlap in these categories between the paper and a dissertation/dissertation chapter, the paper was classified as a match. After independent double-checking, the rate of agreement between research assistants’ and senior researchers’ classification was 94%.

Most of the publications from Scopus matched to the dissertations were classified as journal articles. We base our analysis on these observations. However, some of the Scopus items classified as conference proceedings, reviews, book chapters, notes, etc., were (later) also transformed into journal articles. Therefore, two senior researchers independently reclassified these cases manually (agreement between researchers in this classification was more than 95%). We performed a robustness check based on the dataset including these cases (Online Appendix S12).

Explanatory variables

We pre-specified four explanatory variables:

Cumulative dissertation

A “cumulative” doctoral dissertation is a dissertation written in a specific format. In addition to introductory and concluding chapters, it includes three or more chapters written in the format of journal articles. Writing cumulative dissertations is a recent and still relatively uncommon practice at German universities. According to the German Federal Statistical Office, only 13% of doctoral candidates in 2021 pursued a cumulative dissertation, with the remaining doctoral candidates opting for the traditional monographic dissertation format. For doctoral candidates in law, economics, and social science, the share of cumulative dissertations in 2021 was 18% (Bildung & Kultur, 2021). Cumulative dissertations have been suggested to address the low rates of dissertation citations observed in the past (Francis et al., 2009; Larivière et al., 2008). We accordingly hypothesize that cumulative dissertations have a higher number of publications and citations (hypothesis 1.1).

Graduate academy

Graduate academies are specialized institutions within universities that offer comprehensive support and guidance to doctoral candidates from all academic disciplines. The first graduate academies in Germany were established in 2000 and have since become an integral part of most German universities (Bundesbericht Wissenschaftlicher Nachwuchs, 2017). In addition to offering general support and advice, graduate academies typically provide additional quality assurance measures and offer specialized training and mentorship programs for doctoral candidates. These programs are designed to enhance the academic and professional skills of doctoral candidates and to help them succeed in their respective fields.

We hypothesize that dissertations written at universities with established graduate academies have higher numbers of publications and citations (hypothesis 1.2.a).

Graduate school

The traditional format of doctoral education at German universities relied on on-the-job training under the supervision of an individual doctoral advisor. The adoption of the graduate school model in Germany originated from the establishment of the first Research Training Groups (“Graduiertenkollegs” in German), funded by the German Research Foundation (DFG) in the mid-1980s.

In Research Training Groups, a team of professors and post-doctoral researchers jointly provide guidance and supervision to a number of doctoral candidates, all working on dissertations within the group’s common thematic focus. In addition to on-the-job learning opportunities, research training groups offer specialized training programs to enhance doctoral candidates’ academic and professional skills. Research Training Groups emphasize the training of early-career researchers subsequently embarking on academic careers (DFG, 2010). In the 2000s, they provided the template for graduate schools funded by the German Excellence Initiative.

We classify both the Research Training Groups funded directly by the German Research Foundation and the graduate schools funded by the Excellence Initiative as graduate schools. We hypothesize that dissertations written at universities with a graduate school in the respective discipline have a higher number of publications and citations (hypothesis 1.2.b).

Excellence university

The Excellence Initiative is a large-scale funding program that was jointly established by Germany's federal government and the individual federal Länder in 2006. Its objective was to promote the best German universities to top positions in international university rankings and increase collaboration between German universities and the non-university research sector. The Excellence Initiative encompassed three funding lines: graduate schools, clusters of excellence funding thematically focused research centers connecting universities and research institutes or businesses, as well as university-wide development strategies ("future concepts"). Success in the Excellence Initiative entailed substantial resource and reputation effects on the respective universities. In particular, winning universities in the funding line for development strategies were often considered "excellence universities" (Buenstorf & Koenig, 2020; Möller et al., 2016).

There is mixed evidence of the changes in universities that received Excellence Initiative funding. Some evidence shows a decrease in the number of citations per researcher at universities funded in the first round of the Excellence Initiative compared to universities that did not receive funding (Menter et al., 2018). Other evidence points out that universities funded for their development strategy attracted students with higher GPAs—the effect remained for three years after the funding was awarded—and that students perceived these universities as having higher quality (Fischer & Kampkötter, 2017). Based on these findings, we hypothesize that dissertations written at excellence universities have a higher number of publications and citations (hypothesis 1.3).

Control variables

We collected control variables available from the dissertations, the German National Library portal, and the university websites for each dissertation in the sample. We used the post-double-selection Lasso procedure (Belloni et al., 2014) to select relevant control variables from the set of available control variables. This machine learning procedure relies on a two-step method to identify control variables for inclusion: (1) fitting a lasso regression to predict the outcome variable and (2) fitting a lasso regression to predict the explanatory variables of interest. The union of the variables selected by the procedure is included in the regression. The post-double-selection Lasso procedure reduces the risk of omitted variable bias, while at the same time avoiding overfitting in the presence of many potential control variables (Belloni et al., 2014). It is popular in many social sciences (Kreif & DiazOrdaz, 2019) and in medical research (Dukes & Vansteelandt, 2020), but has not yet been widely adopted in the scientometric literature despite the abovementioned advantages. The full list of control variables is available in Table S1 in the Online Appendix.

We pre-specified that a set of available control variables used in the post-double-selection Lasso procedure will consist of variables with less than 20% missing values. Most variables like language, university and field were retrieved from the German National Library portal and have no missing values. Some explained and controlled variables that could only be obtained from the dissertation text have missing values as 23 dissertations could not be obtained. The analysis included these variables due to a very low missingness rate.

Empirical strategy

We use the following main regression to estimate the relationship between publication-based outcomes and the format of the dissertation (cumulative or monographic), as well as the presence of graduate academies, graduate schools, or excellence funding:

$$Y_{ij} = f(\beta_0 + \beta_{CD} * CD_{ij} + \beta_{GS} * GS_{ij} + \beta_{GA} * GA_{ij} + \beta_{EUI} * EU_{ij} + Controls_{ij})$$

where Y_{ij} is the publication-based measure for dissertation i in university j ; CD_{ij} is a binary variable equal to 1 if dissertation i at university j is in a cumulative format; GS_{ij} is a binary variable equal to 1 if dissertation i comes from a university j with an established graduate school in economics/sociology/political science; GA_{ij} is a binary variable equal to 1 if dissertation i comes from a university j with an established graduate academy; EU_{ij} is a binary variable equal to 1 if dissertation i comes from an excellence university j ; $Controls_{ij}$ is a vector of control variables selected through the post double-selection Lasso procedure (Belloni et al., 2014). We cluster at the university level. f stands for a general functional form in regression analysis. We mainly use negative binomial regression as the number of papers and citations (publication-based measures) is prone to have a skewed distribution and to be overdispersed. We also estimate the Poisson regression model following Azoulay et al. (2019) and a simple linear regression as robustness checks. We performed control variables selection based on the post-double Lasso procedure for each hypothesis tested to see the sensitivity of the results with respect to the second stage of the procedure. In addition, we estimated negative binomial regressions including (i) the full set of institutional variables collected, (ii) the full set of individual author-dissertation variables collected, and (iii) the union of them as additional robustness checks. All estimations were done with R software version 4.0.3. except we had to use STATA for the robustness check regressions with full sets of controls (to ensure model convergence with many controls).

Additional outcome variables

While we hypothesize that cumulative dissertations convert into more journal publications that receive more citations, monographic dissertations may get more citations themselves. We test this conjecture and supplement the main pre-specified analysis using the methodology developed by Donner to estimate the number of citations the dissertations received (Donner, 2021). We followed the algorithm he described and searched for citations to the dissertations in Google Books (using Webometric Analyst) and Scopus and combined the results (Table S4 in the Online Appendix).

Specifically, we used a snapshot of the Scopus data from April 2022. In the first step, we restricted the cited reference data to the publication years of our dissertation sample ± 1 year. In the second step, we looked for exact matches with our dissertation sample regarding the author's surname and initials and the dissertation publication year being around ± 1 year. Lastly, we compared the dissertation title to the Scopus cited item title and cited source title after standardizing them to the length of the shorter title. We calculated the similarity using the Optimal String Alignment (OSA) method and divided the result by the length of the standardized dissertation title, which led to outcomes between 0 and 1 (with 0 being an exact match and 1 being no match). If the outcome was between 0.00 and 0.25, we deemed the citation valid. After manually checking some dissertations at random,

we observed that some authors were occasionally stored in the reference data with their full first name instead of their initials, so we also considered those cases.

Furthermore, we considered names containing German Umlaute (ä, ö, ü) by turning those into a, o, u, and ae, oe, ue. As we focus on the dissertations, we do not include the non-dissertation cited source titles containing words like “Journal”, and “Conference”. Following Donner’s (2021) approach, we applied the same approach to find indexed Scopus source publications matching with our dissertation sample. Lastly, to make sure that we do not have false-positive matches, we sampled 100 citations from the 2376 citations found and manually checked in the actual publications if the references list contains matched dissertations. We did not find any mismatch.

After obtaining the dissertation citations from Google Books and Scopus, we searched for the overlap between both sources. We found 18 citations present in both citations retrieved from the Google Books database and the Scopus database and removed them from the Google Books citations. We then combined the citations from both sources as described in (Donner, 2021).

We provided the analysis on combined Scopus and Google Books citations in line with the pre-specified empirical strategy above, being interested in whether monographic dissertations receive more citations. Finally, we manually collected Google Scholar citations, which have been used before to estimate the scholarly impact of dissertations (Kousha & Thelwall, 2020), and applied the above empirical strategy to assess if the results also hold for Google Scholar citations.

Table 1 Number of papers based on a dissertation and total number of citations of these papers

	<i>N</i>	Number of papers		Number of citations	
		Mean	Variance	Mean	Variance
Field					
Economics	1058	0.52	1.11	14.63	4397.53
Sociology	511	0.18	0.42	8.74	6201.15
Political science	271	0.1	0.18	1.28	49.72
Type of dissertation					
Cumulative	358	1.34	2.06	39.49	13,289.79
Monographic	1459	0.13	0.23	4.22	1893.93
Language					
English	614	0.89	1.64	26.27	8317.34
German or other	1226	0.1	0.19	3.39	2084.61
Dissertation topic					
Empirical	628	0.79	1.6	25.91	11,396.15
Other	1189	0.15	0.27	3.38	428.98
Total	1840	0.37	0.81	11.03	4277.52

The number of papers based on the dissertation and their citation counts are split by field, type of dissertation, language, and topic of the dissertation. The total number of dissertations is 1840; however, for the type of dissertation and dissertation topic, the number of dissertations adds up to 1817 due to 23 dissertations that could not be obtained

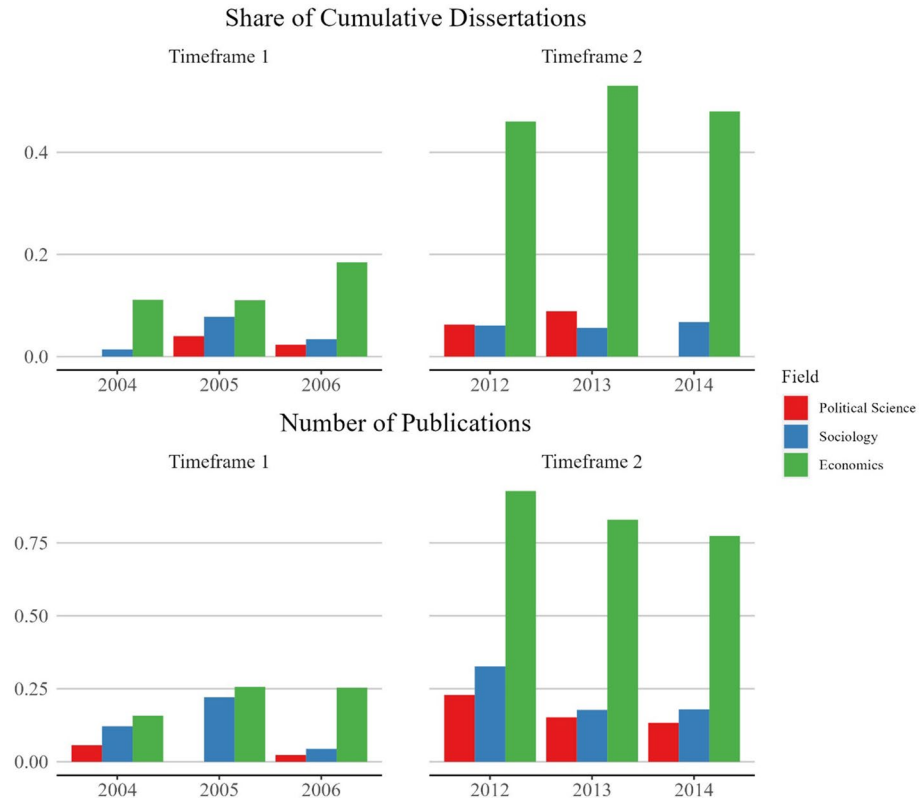


Fig. 1 Share of cumulative dissertations and average number of publications based on the dissertations per field over time

Results

Our data shows that 26% of the dissertations in economics, 11% in sociology, and 7% in political sciences end up with at least one publication in a peer-reviewed journal. Additionally, the average number of papers based on the dissertations is 0.52 for economics, 0.18 for sociology, and 0.1 for political science. The corresponding citation counts of papers resulting from these dissertations are 14.63, 8.74, and 1.28, respectively (Table 1). We also observe that variance exceeds the mean on both primary outcomes—overdispersion for the number of papers and citations from these papers—suggesting that the negative binomial is the preferred specification.

Interestingly, we observe a considerable increase in the number of publications for dissertations in economics during the years 2012–2014, as shown in Fig. 1.

Results based on tests of the pre-specified hypotheses

Results based on tests of the pre-specified hypotheses with and without control variables selected by the double-lasso selection algorithm (Belloni et al., 2014) can be found in Table 2. We find that cumulative dissertations are associated with a significantly higher

Table 2 Negative binomial regression for the number of papers based on the dissertation and their citation counts

	Dependent variable			
	Number of papers		Number of citations	
	(1)	(2)	(3)	(4)
Cumulative dissertation (H.1.1)	2.24*** (0.14)	1.13*** (0.20)	2.28*** (0.32)	1.11*** (0.40)
Graduate academy (H.1.2.a)	0.25** (0.13)	0.03 (0.13)	− 0.16 (0.34)	0.28 (0.34)
Graduate school (H.1.2.b)	0.22 (0.17)	0.08 (0.16)	0.17 (0.45)	− 0.10 (0.35)
Excellence university (H.1.3)	0.20 (0.12)	− 0.06 (0.15)	− 0.004 (0.29)	0.35 (0.30)
English		0.94*** (0.20)		1.07*** (0.28)
Empirical dissertation		0.45*** (0.14)		1.21*** (0.46)
Online dissertation		0.25 (0.16)		0.45 (0.46)
Number of pages		− 0.002*** (0.001)		− 0.01*** (0.001)
Field: political science		− 0.33 (0.29)		− 1.37*** (0.52)
Field: sociology		0.03 (0.20)		0.06 (0.36)
Constant	− 2.17*** (0.13)	− 2.17*** (0.31)	1.44*** (0.37)	2.58*** (0.62)
Year FE	No	Yes	No	Yes
Observations	1817	1814	1817	1814
θ	0.98*** (0.16)	1.43*** (0.27)	0.04*** (0.003)	0.05*** (0.003)

Estimations were made without and with control variables selected by the double-lasso algorithm, including year-fixed effects (Belloni et al., 2014). The control variables reported were selected (for cumulative dissertation as explanatory variables at the second stage) by the double-lasso procedure. Standard errors are clustered at the university level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

number of journal articles than monographic dissertations (p value < 0.00001). Furthermore, the total citation count of papers based on cumulative dissertations is also significantly higher than for monographic dissertations (p value $= 0.06$).

On average, cumulative dissertations turn into three times as many publications as monographic dissertations (Table 2 Model 2 : $\beta_{CD} = 1.13; e^{\beta_{CD}} = 3.1$), even if we account for a large set of controls algorithmically selected by double lasso. Moreover, the average total citation count of the papers from the cumulative dissertations is more than three times as high as for monographic dissertations (Table 2 Model 4 : $\beta_{CD} = 1.11; e^{\beta_{CD}} = 3.03$). In

addition, we observed a notable increase in the share of cumulative dissertations in economics during the second period, followed by a higher number of publications, as shown in Fig. 1.

Our analysis indicates that dissertations from universities with established graduate academies are initially associated with a higher number of publications in peer-reviewed journals. This association becomes insignificant with the inclusion of algorithmically selected control variables and we do not observe any significant difference in citation counts with or without control variables. We also investigated whether the presence of graduate schools in the respective discipline or recognition as an excellence university was related to the publication-based outcomes. However, we did not find statistically significant difference at any conventional level of significance in the number of publications or citation counts between dissertations from universities with or without graduate schools and universities with or without excellence status.

Out of all journal publications for which the year of publication is known, 62.2% were published in the years after the dissertation defense, and 37.8% were published in or before the year of the defense. We re-ran Model 2 from Table 1 separately for publications from the years before and after the defense (Table S2 in the Online Appendix). The significantly positive relationship between cumulative dissertations and the number of publications holds for publications both before and after the defense.

One might expect that, while cumulative dissertations are turned into more journal publications that receive more citations, monographic dissertations receive more citations themselves. We scrutinize this conjecture using the same empirical strategy as before on the following two outcome variables: (a) dissertation citations in Google Books and Scopus constructed following (Donner, 2021) and (b) Google Scholar citations.

We find a significantly negative relationship between cumulative dissertations and the number of dissertation citations. On average, a cumulative dissertation receives 36% fewer citations in Google Books and Scopus (Table 3 Model 2 : $\beta_{CD} = -0.45; e^{\beta_{CD}} = 0.64$). In other words, monographic dissertations receive only 1.5 times more citations in Google Books and Scopus than cumulative ones. We also find a significantly negative relationship between cumulative dissertations and Google Scholar citations. The average number of Google Scholar citations is 63% lower for cumulative dissertations than for monographic dissertations (Table 3 Model 4 : $\beta_{CD} = -0.99; e^{\beta_{CD}} = 0.37$), which implies less than a threefold increase in Google Scholar citations for monographic dissertations compared to cumulative ones. Finally, we do not see a stable association between dissertations from excellence universities and the number of dissertation citations.

We assess our results using Poisson and linear regressions as robustness checks. All results hold with these alternative specifications, both for primary outcomes (Tables S4 and S5 in the Online Appendix) and dissertation citations (Tables S6 and S7 in the Online Appendix).

In addition to estimating the regressions with covariates selected by the post-double Lasso selection procedure, we also estimate regressions with all institutional (Kifor et al., 2023; Rojko et al., 2020) and/or individual author-dissertation factors (Larivière, 2012; Mayir et al., 2017; Maynard et al., 2014; Paglis et al., 2006) as robustness checks (Tables S8, S9, S10 and S11 in the Online Appendix, columns 1–3). The results remain robust to the inclusion of these control variables. They are also robust to the inclusion of covariates selected at the second stage of the post-double Lasso algorithm for each pre-specified explanatory variable, except for the association between graduate academy and the number of papers (Tables S8, S9, S10 and S11 in the Online Appendix, columns 4–7).

Table 3 Negative binomial regression model for the number of dissertation citations

	Dependent variable			
	Dissertation citations Google Books and Scopus		Dissertation citations Google Scholar	
	(1)	(2)	(3)	(4)
Cumulative dissertation (H.1.1)	– 0.47*** (0.13)	– 0.45*** (0.13)	– 1.20*** (0.19)	– 0.99*** (0.18)
Graduate academy (H.1.2.a)	– 0.74*** (0.10)	– 0.15 (0.11)	– 0.48*** (0.11)	– 0.17 (0.13)
Graduate school (H.1.2.b)	– 0.13 (0.15)	– 0.11 (0.12)	0.06 (0.14)	0.07 (0.14)
Excellence university (H.1.3)	– 0.11 (0.16)	0.31** (0.13)	– 0.18 (0.21)	0.21 (0.18)
English		0.42*** (0.13)		0.67*** (0.15)
Empirical dissertation		– 0.09 (0.11)		0.05 (0.10)
Online dissertation		– 0.24* (0.14)		– 0.41*** (0.10)
Number of pages		0.0000 (0.001)		0.002*** (0.0004)
Field: political science		– 0.18 (0.14)		0.16 (0.14)
Field: sociology		0.12 (0.12)		0.75*** (0.12)
Constant	1.52*** (0.08)	1.74*** (0.19)	2.41*** (0.06)	1.67*** (0.15)
Year FE	No	Yes	No	Yes
Observations	1811	1808	1802	1799
θ	0.33*** (0.01)	0.36*** (0.02)	0.39*** (0.01)	0.44*** (0.02)

Negative binomial regression model for the number of dissertation citations calculated based on Donner (2021) algorithm and extracted from Google Scholar without and with control variables selected by the double-lasso algorithm, including year fixed effects (Belloni et al., 2014). The control variables reported were selected (for cumulative dissertation as an explanatory variable at the second stage) by the double-lasso procedure. Standard errors are clustered at the university level

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Thus, we consider the association between the number of papers and the presence of a graduate academy non-robust.

In summary, our findings suggest that cumulative dissertations are turned into more publications in peer-reviewed journals and receive more citations to these peer-reviewed publications. In contrast, monographic dissertations receive more citations as separate works.

Explorative results

We also exploratively examine the variables selected as control variables by the double-lasso procedure. Of the dissertations in our sample, 33.4% were written in English, with the rest being written in German (except for four dissertations in French, two in Italian, and two in Spanish). With 45.1%, the share of dissertations in English is highest in economics, while in sociology and political science, the shares were 18% and 16.6%, respectively.

Our findings indicate that dissertations written in English have significantly higher publication-based outcomes. On average, English dissertations turn into almost twice as many published papers as other dissertations (Table 2 Model 2 : $\beta_{English} = 0.94; e^{\beta_{English}} = 2.56$). In addition, the average citation count for papers based on English dissertations is more than three times as high as that for dissertations in German and other languages (Table 2 Model 4 : $\beta_{English} = 1.07; e^{\beta_{English}} = 2.92$). These results are consistent with other research comparing publication and citation levels of dissertations written in English versus the local language (Nieminen et al., 2007; Donner, 2021).

Empirical dissertations comprise 34.6% of our sample, with economics having the highest share at 48%, followed by sociology at 21.2%, and political science at 6.4%. We defined a dissertation as empirical if it contained hypothesis-testing statistical procedures, including moments of statistical distribution (mean, median,

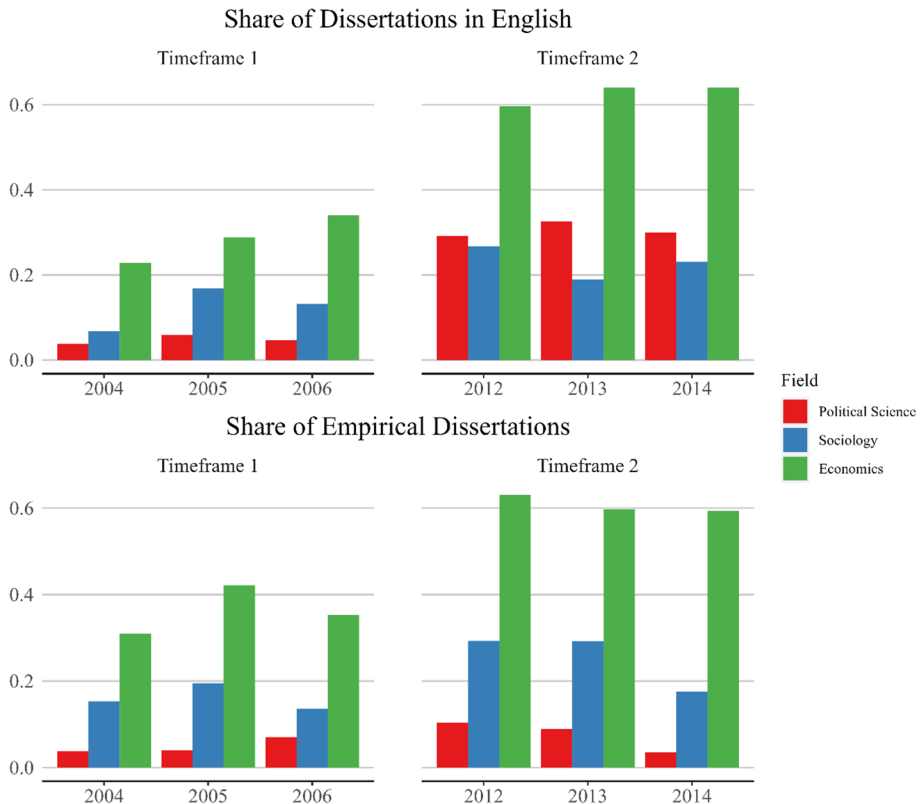


Fig. 2 Share of dissertations in English and share of empirical dissertations per field over time

variance, etc.), regression coefficients, standard errors, p values, t values, or z values. Empirical dissertations had 50% more publications than other dissertations (Table 2 Model 2 : $\beta_{Empirical} = 0.45; e^{\beta_{Empirical}} = 1.57$) and more than three times as many citations to these papers (Table 2 Model 4 : $\beta_{Empirical} = 1.21; e^{\beta_{Empirical}} = 3.35$). Furthermore, our analysis shows an upward trend in the share of dissertations written in English and in the share of empirical dissertations (see Fig. 2).

The algorithm also selected control variables denoting online dissertations, the number of pages, and the field of the dissertation as control variables, but these variables are not consistently associated with significantly different publication-based outcomes in our analysis. Finally, we inspected control variables that were not selected by the algorithm but that we included in additional robustness estimations (with the full set of institutional variables collected, the full set of individual author-dissertation variables collected, and the union of them). We do not observe any associations between these additional control variables and publication-based outcomes that are robust across model specifications. Notably, we do not observe gender differences in publication-based outcomes of dissertations across model specifications.

Discussion

In this study, we investigated how publication-based outcomes of social science dissertations in Germany are associated with dissertation characteristics and institutional factors. Consistent with our hypothesis specified in a pre-analysis plan, we observe that cumulative dissertations lead to a higher number of publications in peer-reviewed journals as well as a higher number of citations from these publications. We also find that the share of cumulative dissertations increased over time in economics. Our analysis does not suggest that the citation advantage enjoyed by publications based on cumulative dissertations is offset by a lower number of citations to the dissertations themselves. While we found that monographic dissertations receive more citations than cumulative ones, their implied advantage in direct citations is smaller than their disadvantage in publication-based citations. We thus conclude from our analysis that results of social science dissertation research documented in cumulative dissertations tend to be disseminated more extensively than results documented in monographic dissertations.

As dissertations are not randomly allocated into cumulative and monographic formats, the patterns we observe in our data cannot be interpreted as causal effects. Indeed, our analysis suggests that publication and citation outcomes for dissertations are affected by dissertation characteristics and institutional factors and that controlling for these variables helps explain some of the differences in outcomes. Regarding institutional factors, we do not see any robust significant difference in publication-based outcomes of dissertations from universities with or without graduate academies, graduate schools in the respective discipline and recognition as “excellence universities”.

Going beyond the hypothesized associations that were specified in our pre-analysis plan, we explored how differences in control variables selected by the double-lasso procedure (Belloni et al., 2014) are related to publication and citation outcomes. These exploratory analyses indicate, first, that dissertations written in English are associated with significantly more publications in peer-reviewed journals and higher citation counts compared to those written in German or other languages. Second, empirical dissertations in our sample also had higher publication-based outcomes compared to other types of dissertations. Overall, shares of dissertations in English and empirical dissertations seem to be increasing over time. In a nutshell, it appears that dissertations written in English or empirical dissertations

increase in number and tend to particularly contribute to the dissemination of knowledge produced by doctoral students in German universities.

Various factors, however, like author characteristics and institutional conditions can affect the choice of dissertation language and topic. Moreover, we did not hypothesize in the pre-analysis plan if dissertations written in English or empirical dissertations are associated with higher publication-based outcomes, barely allowing us to post-hoc speculate about the cause of the higher publication-based outcomes of the dissertations with these characteristics. Thus, we interpret these findings as indicative and encourage further empirical work to probe into their robustness in other settings.

It is hard to distinguish the causal effect of the dissertation features and institutional factors on the publication-based outcomes as this study is based on observational data. More research can be done in the future to identify causal effects and extrapolate the results for other countries. Additionally, automatic, or alternative formal matching algorithms between publications and dissertations could be used to cover more research fields, languages or countries (e.g. Donner, 2022; Echeverria et al., 2015; Heinisch & Buenstorf, 2018). However, based on the results of the study, we can conclude that a policy that allows doctoral students to write cumulative dissertations permits them to strengthen their research output counted as papers published or cited.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s11192-024-04952-1>.

Acknowledgements We would like to thank for excellent research assistance Denise Bornewasser, Maria Paz Catalan Palma, Besmir Daci, Lea Hartwig, Alana Hindiyeh, Iriela Kola, Christina Lerke, James Macumber, Dariia Melnyk, Veronica Vela, Jana Zarembo. Special thanks to Johannes Koenig for finding efficient ways to work with Scopus data and to Maria Mavlikeeva.

Author contributions Conceptualization: GB. Funding acquisition: GB. Investigation: A-MA, VK. Methodology: A-MA, IA, GB, VK, PS. Formal Analysis: Anastasiya-Mariya Asanov. Visualization: Anastasiya-Mariya Asanov. Supervision: IA, GB. Writing—Original Draft Preparation: A-MA. Writing—Review & Editing: A-MA, IA, GB, VK, PS.

Funding Open Access funding enabled and organized by Projekt DEAL. All authors gratefully acknowledge funding by the German Federal Ministry of Education and Research (BMBF) under Grant number 16PH20011 (PI: GB) and via the German Competence Network for Bibliometrics under Grant number 16WIK2101A. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Declarations

Competing interests The authors declare no competing interests.

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

- Arriola-Quiroz, I., Curioso, W. H., Cruz-Encarnacion, M., & Gayoso, O. (2010). Characteristics and publication patterns of theses from a Peruvian medical school. *Health Information & Libraries Journal*, 27(2), 148–154.
- Azoulay, P., Fons-Rosen, C., & Zivin, J. S. G. (2019). Does science advance one funeral at a time? *American Economic Review*, 109(8), 2889–2920. <https://doi.org/10.1257/aer.20161574>
- 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>
- Belloni, A., Chernozhukov, V., & Hansen, C. (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies*, 81(2), 608–650.
- Bildung und Kultur/Statistik der Promovierenden. Wiesbaden. Statistisches Bundesamt. (2021). https://www.statistischebibliothek.de/mir/receive/DESerie_mods_00007247
- Buenstorf, G., & Koenig, J. (2020). Interrelated funding streams in a multi-funder university system: Evidence from the German Exzellenzinitiative. *Research Policy*, 49(3), 103924. <https://doi.org/10.1016/j.respol.2020.103924>
- Bundesbericht Wissenschaftlicher Nachwuchs. (2017). *Statistische Daten und Forschungsbefunde zu Promovierenden und Promovierten in Deutschland*. WBV.
- DFG. (2010). 20 years of research training groups—Matrix for new doctoral cultures: Innovative, interactive, international. https://www.dfg.de/download/pdf/dfg_im_profil/geschaeftsstelle/publikationen/20_jahre_graduiertenkolleg_en.pdf
- Donner, P. (2021). Citation analysis of Ph.D. theses with data from Scopus and Google Books. *Scientometrics*, 126(12), 9431–9456. <https://doi.org/10.1007/s11192-021-04173-w>
- Donner, P. (2022). Algorithmic identification of PhD thesis-related publications: A proof-of-concept study. *Scientometrics*, 127(10), 5863–5877.
- Dukes, O., & Vansteelandt, S. (2020). How to obtain valid tests and confidence intervals after propensity score variable selection? *Statistical Methods in Medical Research*, 29(3), 677–694.
- Echeverria, M., Stuart, D., & Blanke, T. (2015). Medical theses and derivative articles: Dissemination of contents and publication patterns. *Scientometrics*, 102, 559–586.
- 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., Merlo-Vega, J. A., & García-Peñalvo, F. J. (2013, November). Impact of scientific content in open access institutional repositories: a case study of the repository Gredos. In *Proceedings of the first international conference on technological ecosystem for enhancing multi-culturality* (pp. 357–363).
- Fischer, M., & Kampkötter, P. (2017). Effects of German universities’ Excellence Initiative on ability sorting of students and perceptions of educational quality. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*. <https://doi.org/10.1628/093245617X14816371560173>
- Francis, K., Mills, J., Chapman, Y., & Birks, M. (2009). Doctoral dissertations by publication: Building scholarly capacity whilst advancing new knowledge in the discipline of nursing. *International Journal of Doctoral Studies*, 4, 97–106.
- Günauer, F., Krüger, A. K., Moes, J., & Steidten, T. (2013). *Promovieren mit Perspektive—Ein Handbuch von DoktorandInnen für DoktorandInnen. Zum Anliegen des Handbuchs* (pp. 17–55).
- Heinisch, D. P., & Buenstorf, G. (2018). The next generation (plus one): An analysis of doctoral students’ academic fecundity based on a novel approach to advisor identification. *Scientometrics*, 117, 351–380.
- Ioannidis, J. P., Stanley, T. D., & Doucouliagos, H. (2017). The power of bias in economics research. *The Economic Journal*, 605(127), F236–F265. <https://doi.org/10.1111/eoj.12461>
- Karagöz, B., & Seref, İ. (2021). A holistic approach to PhD theses in Turkish Language education (1995–2020). *Selcuk Universitesi Edebiyat Fakultesi Dergisi-Selcuk University Journal of Faculty of Letters*, 46.
- Kifor, C. V., Benedek, A. M., Sîrbu, I., & Săvescu, R. F. (2023). Institutional drivers of research productivity: A canonical multivariate analysis of Romanian public universities. *Scientometrics*, 128(4), 2233–2258. <https://doi.org/10.1007/s11192-023-04655-z>
- Kousha, K., & Thelwall, M. (2020). Google Books, Scopus, Microsoft Academic, and Mendeley for impact assessment of doctoral dissertations: A multidisciplinary analysis of the UK. *Quantitative Science Studies*, 1(2), 479–504. <https://doi.org/10.1002/asi.23170>

- Kreif, N., & DiazOrdaz, K. (2019). *Machine learning in policy evaluation: New tools for causal inference*. Oxford University Press.
- 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*, 109–121. <https://doi.org/10.1007/s11192-008-0106-3>
- Mayir, B., Bilecik, T., Çakır, T., Doğan, U., Gündüz, U. R., Aslaner, A., & Oruç, M. T. (2017). Analysis of the publishing rate and the number of citations of general surgery dissertations. *Turkish Journal of Surgery*, *33*(1), 33.
- Maynard, B. R., Vaughn, M. G., Sarteschi, C. M., & Berglund, A. H. (2014). Social work dissertation research: Contributing to scholarly discourse or the file drawer? *British Journal of Social Work*, *44*(4), 1045–1062. <https://doi.org/10.1093/bjsw/bcs172>
- McLeod, B. D., & Weisz, J. R. (2004). Using dissertations to examine potential bias in child and adolescent clinical trials. *Journal of Consulting and Clinical Psychology*, *72*(2), 235. <https://doi.org/10.1037/0022-006x.72.2.235>
- Menter, M., Lehmann, E. E., & Klarl, T. (2018). In search of excellence: A case study of the first excellence initiative of Germany. *Journal of Business Economics*, *88*, 1105–1132. <https://doi.org/10.1007/s11573-018-0909-5>
- Möller, T., Schmidt, M., & Hornbostel, S. (2016). Assessing the effects of the German Excellence Initiative with bibliometric methods. *Scientometrics*, *109*(3), 2217–2239. <https://doi.org/10.1007/s11192-016-2090-3>
- Nieminen, P., Sipilä, K., Takkinen, H.M., Renko, M. and Risteli, L., 2007. Medical theses as part of the scientificttraining in basic medical and dental education: experiences from Finland. *BMC medical education*, *7*, 1–7.
- Osborn, C. J., Storlie, C. A., & Ricciutti, N. (2023). Enhancing professional identity through scholarship: Examining the publication of dissertations in counselor education and supervision in peer-reviewed journals. *Journal of Counselor Leadership and Advocacy*. <https://doi.org/10.1080/2326716X.2023.2191254>
- Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*, 451–476. <https://doi.org/10.1007/s11162-005-9003-2>
- Rodríguez-Montoya, C., Zepa-García, C., & Cherubin, M. (2023). Aspiring PhDs: The (un) surprising relation between doctoral students and research productivity. *SN Social Sciences*, *3*(2), 32. <https://doi.org/10.1007/s43545-023-00616-8>
- Rojko, K., Bratić, B., & Lužar, B. (2020). The Bologna reform's impacts on the scientific publication performance of Ph. D. graduates—the case of Slovenia. *Scientometrics*, *124*, 329–356. <https://doi.org/10.1007/s11192-020-03482-w>
- Salmi, L., Gana, S., & Mouillet, E. (2001). Publication pattern of medical theses, France, 1993–98. *Medical Education*, *35*(1), 18–21. <https://doi.org/10.1046/j.1365-2923.2001.00768.x>
- Siegfried, J. J., & Stock, W. A. (2001). So you want to earn a Ph.D. in economics?: How long do you think it will take? *Journal of Human Resources*. <https://doi.org/10.2307/3069663>
- Smaldone, A., Heitkemper, E., Jackman, K., Joanne Woo, K., & Kelson, J. (2019). Dissemination of PhD dissertation research by dissertation format: A retrospective cohort study. *Journal of Nursing Scholarship*, *51*(5), 599–607. <https://doi.org/10.1111/jnu.12504>
- Thomas, R. A., West, R. E., & Rich, P. (2016). Benefits, challenges, and perceptions of the multiple article dissertation format in instructional technology. *Australasian Journal of Educational Technology*. <https://doi.org/10.14742/ajet.2573>