

ProQuest Dissertations & Theses Global™: Text and Data Mining

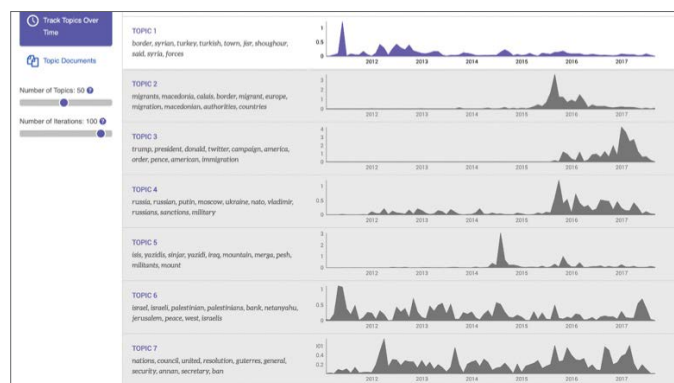
The Evolution of Research. Right at your fingertips.

Many of the most forward-thinking and world-changing discoveries come from the universities that set their sights high with a vision that is dedicated to the transformative power of ideas. ProQuest is dedicated to supporting the success of these institutions. We believe the value of individual research is enhanced exponentially when it is considered as part of a whole.

When historical and current research is analyzed together, it can serve universities and researchers to make future discoveries that impact society and improve human lives. ProQuest is committed to connecting the graduate work from colleges and institutions from one side of the globe to the other. With cutting-edge content and data insight, ProQuest provides students and researchers with access to a full corpus of dissertations and theses content so they can see the evolution of research. Through the mining of the ProQuest Dissertation & Theses Global database, scholars can identify trends and generate hypotheses that they might have not had otherwise.

The ProQuest Dissertations & Theses (PQDT) Global database has nearly 5 million records, from 1637 to present day, each of them with carefully indexed metadata. In addition, over 200,000 new graduate works are added every year from all over the world, making it the perfect resource to use to understand relationships between research studies. PQDT Global is intended for use by scholars, institutions, and the government to help enhance their research.

The humanities and social sciences influence almost every aspect of daily life; from economic policy, to health and social policy, and from the environment and energy to technology and innovation. Understanding these trends and assessing the impacts on daily life takes skilled analysis — and given the diversity and complexities of the data, such analysis requires detailed modeling. By using text and data mining, you can extract large quantities of data and recombine them into patterns and topics to help the researcher further enhance the examination of the scholar's topic of choice.



TDM Studio is a ProQuest tool that can support the analysis efforts researchers are already pursuing within the PQDT Global database.



To talk to the sales department, contact us at
1-800-779-0137 or sales@proquest.com.



Avoiding Bias by Starting at the Source

Research Case Study: *Outcome-Reporting Bias in Education Research*, Loyola University Chicago

Dr. Terri D. Pigott, Ph.D., is the former Dean of the School of Education at Loyola University Chicago, and is currently a professor in the School of Public Health and the College of Education at Georgia State University. Terri's research focuses mostly on methods of research synthesis and meta-data analysis. Because all research needs to be verified, Terri relies on resources like ProQuest Dissertations & Theses to help her further develop her research. Terri also recommends PQDT Global to her students as well.

"All systematic reviews of research in the social sciences and in health need to include a search of the content in PQDT. The published research literature is biased toward statistically significant results, and the content in PQDT represents the most current research in an area that is not biased toward statistical significance. When I review systematic reviews for journals, I insist that PQDT is searched for relevant research."

— Dr. Terri Pigott, Professor in the School of Public Health and the College of Education, Georgia State University

Terri, along with fellow professors and researchers, Jeffrey C. Valentine, Joshua R. Polanin, Ryan T. Williams, and Dericka D. Canada, used the PQDT services to help verify research for their paper, *Outcome-Reporting Bias in Education Research*, where the study examines outcome-reporting bias in educational research by comparing the reports of educational interventions from dissertations to their final published versions. They searched the PQDT database between the years of 2001 and 2005 inclusive to draw conclusions. According to their paper, a large percentage of dissertations are available via electronic databases (e.g., ProQuest Digital Dissertations & Theses) to which many university libraries subscribe, and as such are easily retrievable. The paper also asks the important question of "What steps can we take to increase the quality of reporting of primary studies, and thus the completeness of syntheses using these studies?"

Pigott, et. al suggest that all educational research would start with a highly operational, publicly available protocol that guides the research. While this is a lofty goal, Pigott and the rest of the researchers suggest that we start the development of professional norms that hold researchers responsible for fully documenting the research methods and analytic choices, including reporting all outcomes they measure and expand from there. By using resources like PQDT Global, Pigott and other researchers have maximized the depth of their research.



Analyzing the Past to Understand the Future

Research Case Study: *Sociology of Science, Diversity Effects, and Research Innovation, Stanford University*

In this day and age, researchers need rich longitudinal datasets that accurately reflect a population of subjects. With the digitalization, many research professors can now tap into datasets like these that weren't available 20-30 years ago. Professors, like Daniel McFarland, a professor of Education at Stanford University, see resources like ProQuest Dissertations & Theses Global as a nice example of such high quality data and use it in their research.

McFarland studies the social and organizational dynamics of educational systems like universities and disciplines. McFarland is currently engaged on several different projects including writing a textbook on Social Network Analysis in R with Craig Rawlings, Jeff Smith, and James Moody. He also has generated a line of articles on the sociology of science, diversity effects, and research innovation. That's where using a product like PQDT Global has helped McFarland. Using PQDT Global and other resources, he's able to make his research as well-rounded as possible for publication.

"What I like about PQDT is that affords us a sizeable sample of beginning researcher's careers. Using National Center of Education Statistics census of all US PhDs, we can weight the PQDT sample and make inferences about all US PhDs. That's pretty great as most datasets on research reflect found data, and we can only guess how representative those samples are. With PQDT, we can argue they reflect a known population."

– Daniel McFarland, Professor of Education and Sociology, Stanford University

According to McFarland, the standards for being published have risen so researchers have to check their models and inferences many different ways. McFarland also stated that "compared to 20 years ago, research standards for publication are higher. That's a good thing, but it also means we have to find rich longitudinal information that can be anchored in known populations – then our inferences make sense."

McFarland and his research team turned to PQDT Global because they wanted to study scholar careers and they wanted data that accurately reflected the pool of potential scholars. "What I liked about PQDT Global is that it wasn't biased like some journals are and it's evenly representative across disciplines. We don't get only journal science, or conference proceedings, or books, but rather see people of all fields who write theses. PQDT also doesn't select on the outcome. By this I mean most studies look at faculty or persons who got PhDs and succeeded in becoming faculty. With PQDT we have recourse to following all PhDs into various jobs and out of academia – so the entire pool of potential faculty." McFarland and his team needed a data set of at least 30 years and all potential future faculty to develop models of "successful" academic careers. This is something that PQDT can offer with content that dates back to the 1700s.

"ProQuest Dissertations & Theses Global is a wonderful corpus for scholars to work with, and especially if they are interested in following the pool of potential researchers and their careers," McFarland stated.



More TDM Research Examples Using Dissertations

David Zeitlyn, Daniel W. Hook used the PQDT Global database to expand on the network theoretical social model to explore and understand the dynamics of prestige in the academic hierarchy. Their paper, on Perception, prestige, and PageRank, was published in *PLOS ONE*, May 2019.

John Meier, Angela Raw Davis, and Vanessa Lyn Eyer used data from full-text engineering dissertations to determine the change in engineering literature referenced in Google Scholar over the past 10 years and published an article in PennState ScholarSphere on the same topic.

Julia Lane, et. al explored the UMETRICS data initiative and related data sets, including the PQDT Global database to explore linked data on research investments, published in *Research Policy* in November 2015.

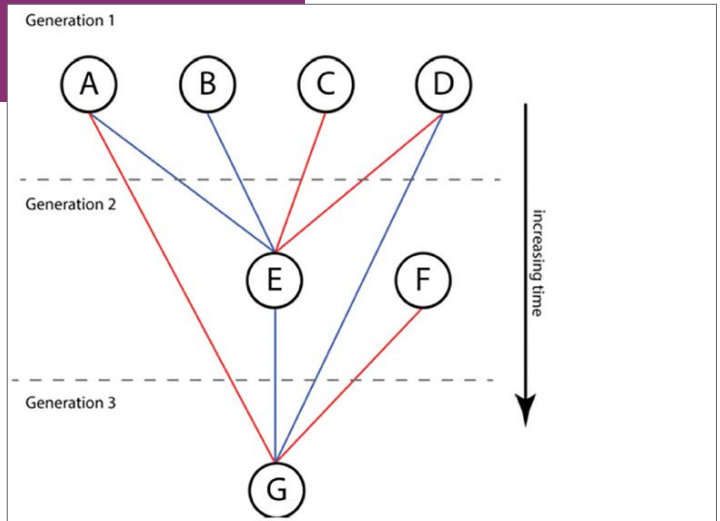


Fig 1. Simple model for academic hierarchy involving both supervisor-student links and examiner-examinee links. Red lines represent examination relationships and blue lines represent supervisor/advisor relationships. In this diagram, A and B supervise E, who is examined by C and D; E then supervises G with D, who is in turn examined by A and F.

John Eric Humphries, Assistant Professor of Economics at Yale University is currently analyzing dissertations within PQDT Global in relation to human capital. He began his research in 2019 and continues to pursue it in 2020 with the use of ProQuest's TDM Studio.

| Dependent variables ↓ | Females | Males | Diff | | | | | | |
|---|------------------|-----------------|--------------------|--------------------|--------------------|------------------|------------------|-----------------|-----------------|
| | (a) | (b) | (c) | (1) | (2) | (3) | (4) | (5) | (6) |
| Employed in industry | 0.40 (0.022) | 0.47 (0.02) | -0.13*** (0.03) | -0.11*** (0.03) | -0.11*** (0.03) | -0.05 (0.03) | -0.05 (0.03) | -0.03 (0.05) | -0.03 (0.05) |
| In wage | 10.50 (0.063) | 10.93 (0.03) | -0.37*** (0.07) | -0.35*** (0.07) | -0.35*** (0.07) | -0.11* (0.07) | -0.11* (0.07) | 0.01 (0.10) | 0.01 (0.10) |
| In wage (with industry controls) | 10.40 (0.057) | 10.71 (0.04) | -0.31*** (0.07) | -0.29*** (0.07) | -0.30*** (0.06) | -0.9 (0.06) | -0.9 (0.07) | 0.02 (0.10) | 0.02 (0.10) |
| University, first year trend, left-censored | | | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Degree year | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Race, Hispanic origin, age, age-squared | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Dissertation topic | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Funding agency | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Married or partnered, presence of children | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Female × (married or partnered + children) | | | | | ✓ | ✓ | ✓ | ✓ | ✓ |
| Observations | 370 | 867 | 1,237 | 1,237 | 1,237 | 1,237 | 1,237 | 1,237 | 1,237 |

Notes: Labor outcomes are taken from one year following graduation or separation from the university payroll, whichever is greater. Wages are in 2012 dollars. Sample includes observations with dissertation topics in a STEM field. Each cell in columns 2–8 displays the estimated coefficient on the FEMALE indicator from a separate regression. Robust standard errors.

***Significant at the 1 percent level.
**Significant at the 5 percent level.
*Significant at the 10 percent level.

Sources: Author calculations. UMETRICS linked to 2010 census, ProQuest, LEHD, W2, LBD, BR, and iLBD.

Catherine Buffington, et.al used the PQDT Global data to verify dissertation publication and graduation against UMETRICS data. The data was then used to analyze how STEM degrees correspond to career outcomes. Their paper, *STEM Training and Early Career Outcomes of Female and Male Graduate Students: Evidence from UMETRICS Data Linked to the 2010 Census*, was published in *The American Economic Review*, May 2016.

The evolution of research is at your fingertips. Enhance your studies with data sets you can't find anywhere else by using the ProQuest Dissertations & Theses Global database. Reach out to one of our ProQuest specialists for more information today!

**To find out more about the services offered by ProQuest
Dissertations & Theses contact your Account Manager or visit:
[dissertations.com](https://www.proquest.com)**



proquest.com

To talk to the sales department, contact us at **1-800-779-0137** or **sales@proquest.com**.

