Diving Deep into Dissertation Data: Analyzing Graduate Students' Dissertations to Inform Research Data Services (RDS)

Mandy Swygart-Hobaugh, Raeda Anderson, Denise Dimsdale, Joel Glogowski | Georgia State University Library

GEORGIA STATE UNIVERSITY LIBRARY

RESEARCH DATA SERVICES
LIBRARY.GSU.EDU/DATA

Introduction

KNOWN:
Data for one-on-one consults, workshop attendance, and custom course sessions indicate that **graduate students are the largest group using data support services provided by Research Data Services (RDS) Team.**

DEEPER DIVE:
Using a quantitative content analysis of doctoral dissertations to examine graduate students' method and data practices, we sought insights to the following research questions:

- **Research Question 1:** What method types (quantitative, qualitative, or mixed methods), data types (primary or secondary), and analysis software/coding languages do graduate students employ in their dissertation research?
- **Research Question 2:** Are there differences between method type used when comparing across degree type and data type?
- **Research Question 3:** Are there differences between degree type when comparing across method type and data type?
- **Research Question 4:** Are there differences between data type used when comparing across method type and degree type?
- **Research Question 5:** What are the distributions of method type and data type when broken down my academic fields?

Methods

RESEARCH SAMPLE:
152 doctoral dissertations were gathered from the ScholarWorks@GSU institutional repository for the 2017-2018 academic year that met the below inclusion criteria based on methodology used:

- **Qualitative methods –** analysis of non-numeric data
   - open-ended survey questions; open-ended interviews; analysis of text and audiovisual materials using non-numeric/non-statistical content analyses; case studies; ethnographies.
- **Quantitative methods –** numeric data subjected to statistical analysis
   - close-ended survey/measurement scale data collection and analysis; analysis of primary (self-collected) or secondary (previously-collected) numeric data.
- **Mixed methods –** use of both quantitative and qualitative methods

METHOD:
We conducted a quantitative content analysis on the abstracts, methods, and results/findings sections of dissertations to collect the necessary information for coding methodology and data practices. The following variables and their operational definitions guided our coding:

- **DATA TYPE:** Analysis of **primary data**, **secondary data**, or **both primary and secondary data**
- **METHOD TYPE:** Methodology (**qualitative methods**, **quantitative methods**, or **mixed methods**)
- **SOFTWARE TYPE:** Category of software type (**qualitative**, **quantitative**, **other**, or **not identified**)
- **DEGREE TYPE:** Doctor of Philosophy (**Ph.D.**); Doctor of Education (**Ed.D.**); Executive Doctorate in Business (**E.D.B.**)

Figure 1.
Illustration of Data Conversion from Four Nominal Categorical Variables to Thirteen Dichotomous Variables

Example Dissertations with respective coding for Nominal Categorical Variables:
- **Dissertation 1:** Method Type = qualitative methods; Data Type = primary data; Software Type = qualitative software; Degree = Ph.D.
- **Dissertation 2:** Method Type = quantitative methods; Data Type = secondary data; Software Type = not identified; Degree = Ed.D.
- **Dissertation 3:** Method Type = mixed methods; Data Type = both primary and secondary data; Software Type = other (survey, lab) software; Degree = E.D.D.

Data Spreadsheet Coding for Conversion to Thirteen Dichotomous Variables:

	NOMINAL CATEGORICAL VARIABLES												
	Method Type			Data Type			Software Type				Degree		
	DICHOTOMOUS VARIABLES												
	qualitative methods	quantitative methods	mixed methods	primary data	secondary data	both	qualitative software	quantitative software	other (survey, lab)	not identified	Ph.D.	Ed.D.	E.D.B.
Dissertation 1	1	0	0	1	0	0	1	0	0	0	1	0	0
Dissertation 2	0	1	0	0	1	0	0	0	0	1	0	1	0
Dissertation 3	0	0	1	0	0	1	0	0	1	0	0	1	0

Results

Research Question 1: What method types (quantitative, qualitative, or mixed methods), data types (primary or secondary), and analysis software/coding languages do graduate students employ in their dissertation research?

KEY INSIGHTS:
- Most used quantitative methods (61.54%), slightly over a 1/4 used qualitative methods (27.56%), and 10.90% used mixed methods.
- Students largely used solely primary data in their dissertations (60.26%), but a substantive number of students used solely secondary data (28.21%) and a small percent (11.54%) used both primary and secondary data.
- Students do not consistently identify software used for data analysis and/or collection, making it difficult to draw meaningful insights from the "software type" data we collected and thus not worthwhile to perform and report further analyses using software type variables.

TABLE 1: Descriptive Statistics (N = 156).

	Mean %
Method Type	
Qualitative Methods	27.56%
Quantitative Methods	61.54%
Mixed Methods	10.90%
Degree	
Ph.D.	87.82%
Ed.D.	5.13%
E.D.B.	7.05%
Data Type	
Primary Data	60.26%
Secondary Data	28.21%
Primary and Secondary Data	11.54%
Software Type[a]	
Qualitative	14.74%
Quantitative	47.44%
Other (Survey, Lab)	19.23%
Not Identified	30.77%

[a] Students could use qualitative software, quantitative software, and/or other (survey, lab) software for their dissertation analysis, thus these percentages do not equal 100.00%.

Research Question 2: Are there differences between method type used when comparing across degree type and data type?

TABLE 2: Degree Type and Data Type Across Method Type with ANOVA Test Statistics (N = 156).

	Qualitative Methods	Quantitative Methods	Mixed Methods	ANOVA	
	N = 43	N = 96	N = 17	N = 156	
	Mean %	Mean %	Mean %	F statistic	sig[a]
Degree					
Ph.D.	79.07%	92.71%	82.35%	2.90	
Ed.D.	11.63%	2.08%	5.88%	2.84	
E.D.B.	9.30%	5.21%	11.76%	0.70	
Data Type					
Primary Data	69.77%	58.33%	47.06%	1.51	
Secondary Data	16.28%	33.33%	29.41%	2.16	
Primary & Secondary Data	13.95%	8.33%	23.53%	1.81	

KEY INSIGHTS:
- Use of solely primary data continued to dominate, regardless of method type.
- The ANOVA tests did not indicate statistically significant differences in method type for the three degree types or three data types.

Research Question 3: Are there differences between degree type when comparing across method type and data type?

KEY INSIGHTS:
- When parsing by degree type, **the pattern shifted among Ed.D.s to qualitative methods dominating.**
- For quantitative methods, the ANOVA test indicated a statistically significant difference across degree type; an LSD post hoc test indicated that, while Ph.D.s and E.D.B.s were ostensibly equal, as were Ed.D.s and E.D.B.s, **Ph.D.s were statistically more likely to use quantitative methods when compared to Ed.D.s.**
- For dissertations using both primary and secondary data, the ANOVA test indicated a statistically significant difference across degree type; a Games-Howell post hoc test indicated that, while Ph.D.s and Ed.D.s were ostensibly equal, as were Ed.D.s and E.D.B.s, **Ph.D.s were statistically more likely to use both data types when compared to E.D.B.s.**

TABLE 3: Method Type and Data Type Across Degree Type with ANOVA Test Statistics (N = 156)

	Ph.D.	Ed.D.	E.D.B.	ANOVA	
	N = 137	N = 8	N = 11	N = 156	
	Mean %	Mean %	Mean %	F statistic	sig[a]
Method Type					
Qualitative Methods	24.82%	62.50%	36.40%	2.97	
Quantitative Methods	64.96%	25.00%	45.50%	3.27	*
Mixed Methods	10.00%	12.50%	18.20%	0.34	
Data Type					
Primary Data	61.31%	50.00%	54.55%	2.78	
Secondary Data	28.47%	0.00%	45.45%	2.41	
Primary & Secondary Data	10.22%	50.00%	0.00%	7.11	***

[a] Significance level (* p < .05; ** p < .01; *** p < .001).

Research Question 4: Are there differences between data type used when comparing across method type and degree type?

TABLE 4: Method Type and Degree Type Across Data Type with ANOVA Test Statistics (N = 156)

	Primary	Secondary	Primary & Secondary	ANOVA	
	N=94	N=44	N=18	N = 156	
	Mean %	Mean %	Mean %	F statistic	sig[a]
Method Type					
Qualitative Methods	31.91%	15.91%	33.33%	2.11	
Quantitative Methods	59.57%	72.73%	44.44%	2.38	
Mixed Methods	8.51%	11.36%	22.22%	1.47	
Degree					
Ph.D.	89.36%	88.64%	77.78%	0.96	
Ed.D.	4.26%	0.00%	22.22%	7.15	**
E.D.B.	6.38%	11.36%	0.00%	1.34	

[a] Significance level (* p < .05; ** p < .01; *** p < .001).

KEY INSIGHTS:
- The pattern of quantitative methods dominating continued but became slightly less pronounced among dissertations using both primary and secondary data.
- For Ed.D.'s, the ANOVA test indicated that there was a statistically significant difference across data type; however, a Games-Howell post hoc test to parse the specific between-groups differences then indicated that there were no statistically significant differences.

Research Question 5: What are the distributions of method type and data type when broken down by academic fields?

TABLE 5: Method Type and Data Type by Academic Field (N = 156).

Academic Field[a]	Overall	Method Type			Data Type		
		Qualitative Methods	Quantitative Methods	Mixed Methods	Primary Data	Secondary Data	Primary & Secondary
	N = 156	N = 43	N = 96	N = 17	N = 94	N = 44	N=18
Business	14.1%	22.7%	63.6%	13.6%	54.5%	40.9%	4.5%
Physical Sciences & Math	26.9%	9.5%	88.1%	2.4%	85.7%	4.8%	9.5%
Education	21.8%	61.8%	29.4%	8.8%	76.5%	2.9%	20.6%
Health Sciences	5.1%	0.0%	100.0%	0.0%	50.0%	50.0%	0.0%
Social Sciences	29.5%	23.9%	58.7%	17.4%	28.3%	58.7%	13.0%
Humanities	2.6%	50.0%	0.0%	50.0%	75.0%	25.0%	0.0%

[a] Academic department was recorded/coded for each dissertation but is not reported in the above table.

KEY INSIGHTS:
- Qualitative methods were comparatively more predominant in the Education and Humanities fields.
- There was a comparatively wider variety between the Social Sciences departments in terms of method type: Communication, Political Science, and Sociology had comparatively larger proportions of qualitative methods; Criminal Justice, Economics, Psychology, and Public Management & Policy were dominated by quantitative methods.
- The Social Sciences fields were dominated by secondary data use, which bucked the overall trend of primary data use predominance.

Implications for RDS Support

Quantitative methods dominated overall, but there was a substantive qualitative methods contingent, particularly among certain academic fields/departments.

- Continue offering proportionally more workshops/guides/tutorials on quantitative software/methods/data resources than qualitative.
- Focus collection development efforts on quantitative data, software, and methods resources (e.g., software manuals, methods books, dataset purchases).
- Target market qualitative methods support to specific academic departments.
- Invest in building quantitative skills among current RDS staff (e.g., basic and advanced statistical analysis, data visualization, etc.) and/or hire additional staff with these skills.
- Advocate for wider and free off-campus access to proprietary software (e.g., SPSS, Stata).

Primary data use dominated overall, but there was a substantive secondary data use contingent, particularly among certain fields/departments.

- Explore increasing workshops/guides/tutorials on primary data collection (e.g., Qualtrics; survey design; web scraping; qualitative data collection methods) **AND** secondary data resources (because primary data collection is time consuming and costly, and graduate students should explore more expeditious/efficient and affordable modes of secondary data collection).
- Focus collection development efforts on primary data collection resources (e.g., books on survey design, qualitative interview techniques, etc.) **AND** secondary data resources (e.g., increasing secondary dataset resources, textual and archival resources for qualitative analysis, etc.).
- Increase marketing/outreach around secondary data resources and use.
- Invest in building primary data collection skills among current RDS staff (e.g., survey design skills) and/or hire additional staff with these skills.