



**METHODOLOGICAL QUALITY AND PUBLICATION OF
DOCTORAL DISSERTATIONS IN EDUCATION:
AN EVALUATIVE STUDY OF TEN YEARS OF K-12
DOCTORAL DISSERTATIONS IN THE UNITED STATES**

Amy J Catalano*	Hofstra University, Hempstead, NY, United States	Amy.Catalano@Hofstra.edu
Marilyn M DePietto	Hofstra University, Hempstead, NY, United States	Marilyn.M.DePietto@Hofstra.edu
Alexander J Lord	Queens College, City University of New York, Flushing, NY, United States	Alexander.Lord@qc.cuny.edu
Susan T Radin	Hofstra University, Hempstead, NY, United States	Susan.T.Radin@Hofstra.edu
Lydia Williams	Hofstra University, Hempstead, NY, United States	lydiadwilliams@yahoo.com

* Corresponding author

ABSTRACT

Aim/Purpose	Despite an increase in the numbers of K-12 educators pursuing doctoral degrees, it is unclear if the field of education has been significantly impacted by the research resulting from their doctoral dissertations. Accordingly, the quality of doctoral programs and dissertations and rate of publication after defense, warrants examination.
Background	There have long been discussions regarding concerns about the quality of doctoral dissertations that come in concert with an increase in degrees awarded. This paper presents findings from a study examining dissertation quality for K-12 education doctorates awarded over a ten-year period in the United States.
Methodology	The researchers randomly selected and evaluated the methodological quality of more than 500 dissertations during the last ten years from the Proquest dissertation database using a rubric adapted from Ronau et al. (2014). Using this rubric,

Accepting Editor Nicole A. Buzzetto-Hollywood | Received: January 11, 2024 | Revised: May 19, June 24,
July 11, 2024 | Accepted: July 24, 2024.

Cite as: Catalano, A. J., DePietto, M. M., Lord, A. J., Radin, S. T., & Williams, L. (2024). Methodological quality and publication of doctoral dissertations in education: An evaluative study of ten years of K-12 doctoral dissertations in the United States. *International Journal of Doctoral Studies*, 19, Article 10. <https://doi.org/10.28945/5343>

(CC BY-NC 4.0) This article is licensed to you under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/). When you copy and redistribute this paper in full or in part, you need to provide proper attribution to it to ensure that others can later locate this work (and to ensure that others do not accuse you of plagiarism). You may (and we encourage you to) adapt, remix, transform, and build upon the material for any non-commercial purposes. This license does not permit you to use this material for commercial purposes.

Methodological Quality of Dissertations

	the team described the quality of dissertations with respect to sampling, design, and validity.
Contribution	A new rubric was developed and tested to assess the quality of quantitative, qualitative, and mixed methods studies. These results provided large-scale data on dissertation quality that may be used to guide doctoral programs in implementing program changes.
Findings	Although the majority of the dissertations studied used qualitative methods of inquiry, quantitative research had significantly higher quality scores. Regardless of the methods used, many studies failed to employ rigorous validity controls.
Recommendations for Practitioners	Doctoral programs may wish to make use of these results to improve programming, particularly in methods courses. These results may also be used to guide dissertation advisors and dissertation committee members, along with increasing publication rates among doctoral students.
Recommendations for Researchers	Researchers may wish to extend this study by looking more closely at the methodological quality in specific subject areas. Research is also needed to improve the rubric developed for this study further and to determine other metrics of impact beyond publication.
Impact on Society	Stronger methods sections may lead to more publications of dissertation studies, which can improve impact on educational practice and evidence-based decision-making.
Future Research	Future research should assess the quality of research in the context of usefulness, impact on practice beyond publication, and topic currency. Further, the motives behind publishing dissertation work in fields that do not require publication should be examined.
Keywords	dissertation quality, research methods, qualitative research, quantitative research

INTRODUCTION

In 2022, 57,596 doctorates were awarded in the United States, a growth of over 10% since the pandemic, according to the National Science Foundation (Hoffer et al., 2009; National Science Foundation, 2023). In 2019, doctorates in education accounted for 8.8% of all awarded, a stark decline from 15.9% in 1999. While representation in total doctoral degrees awarded has decreased, the number of education doctorates awarded has increased. Currently, 1% of adults in developed countries possess a doctorate, but current trends show a predicted increase to 2.3%, although tracking of awarded degrees has not been well documented (Sarrico, 2022). In the last five years, access to doctoral education has also increased, particularly in the form of online programs from for-profit institutions. Among several online Ed.D. programs, enrollment increased by 13% overall in 2021, and at one institution increased by 50% (Alfaro, 2023). There have long been discussions regarding concerns about the quality of doctoral dissertations that have come in concert with an increase in the number of doctoral degrees awarded (Zumeta, 1982). The proliferation of programs and increase in enrollment in education doctoral programs prompts questions as to program quality and, particularly, dissertation preparation quality. Accordingly, this study sought to examine a sample of dissertations in the field of K-12 education (higher education and administration/leadership were excluded) and assess the methodological quality of those dissertations. This study posited that research design type along with other variables, including institution type, may influence dissertation quality as measured by methodological rigor.

In addition, a trend in research on dissertations has shown that publication rates after the defense are low (i.e., Evans et al., 2018). Thus, the research team also sought to quantify the rate at which the sample of dissertations is published and how methodological quality may be correlated with publication. The utility of a doctoral degree goes beyond K-12 and higher education, as staff in policy academies, corporations, youth organizations, and museums may also possess a doctoral degree in education. The doctorate is no longer seen only as a pathway to academia, as other professions may value the added credential (Halse & Mowbray, 2011). Beyond the scarcity of publications among doctoral graduates, the question of how and whether publication is an indicator of the impact on practice remains. As other researchers have noted, existing structures such as the What Works Clearinghouse have not always worked as intended to promote research to practice (Schoenfeld, 2006).

Preparing doctoral-level experts in education is essential to improving research in curriculum, school policy, educational psychology, and educational administration. Doctoral training can also have an impact on the field prior to graduation as students work to implement newly learned practices (Kumar & Dawson, 2013). To address the value of the doctoral degree in the context of quality and impact, this study seeks to examine the methodological quality of doctoral dissertations produced over a 10-year period. In addition, this study examined the extent to which doctoral dissertations have been published as either monographs, peer-review articles, or book chapters.

LITERATURE REVIEW

DISSERTATION QUALITY

Research on the quality of education doctoral dissertations is scant, and even in adjacent fields, there is not a preponderance of studies (Kyvik & Thune, 2015). The literature that does exist is somewhat conflicted on the overall quality of doctoral dissertations. Ten Cate et al. (2017) presented a framework for assessing the quality of award-nominated doctoral dissertations in health professions education. Their examination found that the 32 dissertations they assessed averaged only 22.3 out of 35 points (64%) on a quality assessment, indicating that the value of the candidates for an award was based on criteria other than methodological quality.

Some researchers agree that the quality of doctoral dissertations is lacking in rigorous research methodology (Anney, 2014; Thompson, 1994). This lack of rigor may be due to the fact that students completing dissertations are obligated to demonstrate more expertise in writing and research methods than they actually possess at the end of their doctoral programs; expertise in academic scholarship is unlikely to develop through a single research project (Casanave, 2019). The quality of research instruments used in studies has also been called into question by researchers examining methodological rigor in dissertations (Karadağ, 2011).

The quality of dissertations has been questioned historically. For example, Thompson (1994) found many common flaws in quantitative dissertations in the field of behavioral science, which he argued should not have been present since dissertations are filtered through multiple methodology experts and are not subject to time and length constraints, as is the case with journal articles. Specifically, he noted poor assumptions about sampling, misuse of language around reliability and statistical significance, misinterpretation of results, and underuse of multivariate methods in research projects. Ultimately, he claimed that if institutions want to survive accreditation and scrutiny, scholars ought to manage the statistical accuracy of their students' dissertations more carefully. In a study conducted over 20 years ago, Cleary (2000) sought to examine the changes in the quality of dissertations in the field of public administration and their relationship to graduate program quality. Somewhat in opposition to previously mentioned research, he found that quality trended upward over time and that greater quality dissertations tended to come from "high-quality" programs. Notably, he found that dissertations written on topics considered "central" to the field were not more likely to demonstrate a high-quality, rigorous design. This research raises the question of whether quality or impact on the field can or should be considered equal criteria in evaluating dissertations.

More recently, Ronau et al.'s (2014) study, which served as the model for the current study's methodology, compared the scope and quality of dissertations in mathematics education technology to journal articles in the same field (N=480). This mixed-methods study used a researcher-designed rubric to measure theoretical connections, design and validity, and instrument reliability. MANOVA results revealed a significant main effect of both research type and publication type on quality score. Specifically, by a small margin, qualitative research scored significantly higher than both mixed methods and quantitative research for reporting theoretical connections. By a large margin, quantitative research scored significantly higher on both the validity of the design and reliability than other methods.

Some studies have found that flaws that are apparent in qualitative dissertations tend to appear when analyzing trustworthiness. Both Anney (2014) and Eryilmaz (2022) reported that although the papers being examined used qualitative methods, there were many instances where students used quantitative trustworthiness criteria to evaluate their instruments. The authors asserted that these deficiencies were a result of poor instruction from the respective institutions, where experts should be correcting flawed methodology. Hamilton et al. (2010) contended that, in some cases, the quality of dissertations may be a good indicator of degree program quality.

Factors beyond student and program characteristics can also impact dissertation quality. Kyvik and Thune (2015) investigated whether examiner characteristics affect evaluation scores of the research quality of dissertations granted in Norway. A sample of 1,159 examiners from various countries showed that only regional affiliation and previous experience significantly affected quality assessment. Further, Randolph et al. (2014) investigated the number of pages, research methods, author gender, methods type (qualitative, quantitative, or mixed), and university type (in person or online) in a random sample of 107 dissertations in the field of education. They found that the average dissertation was about 183 pages long, with significant differences among types of methods. They also found that of the 107 dissertations, 53% were quantitative, 33% were qualitative, and 13% were mixed methods.

While the methodological focus on methods type varies by culture and institution, some researchers have noted an emphasis on quantitative methods (Juodaitytė & Kazlauskienė, 2008), and others have been similarly descriptive of the methods and topics covered in education dissertations (Drysdale et al., 2013; Durak et al., 2018) but not evaluative. Walser and Trevisan (2016) explored the field evaluation in education as a dissertation topic and found that most dissertations of this nature used rigorous models of assessment. Based on the limited research available on dissertation quality, there appears to be little consensus as to the factors that are related to dissertation quality other than lack of sufficient instruction in research methods.

PUBLICATION AND IMPACT ON PRACTICE

Researchers have argued that if a dissertation represents an investigation of an issue pertaining to one's field, then that dissertation should be disseminated to the scientific basis of knowledge in the form of publication (Evans et al., 2018). Multiple studies have discussed the impact of dissemination of dissertation research findings and their importance to the field of study (Conn, 2008; Evans et al., 2018; Hochbein & Perry, 2013; Odendaal & Frick, 2018). Publication has been deemed a factor in terms of the overall impact of the work, leading to the call for graduate programs to include explicit instruction on the dissemination of research findings (Hochbein & Perry, 2013). Some universities require students to submit publication-based theses, in which they complete a dissertation that comprises several already-published peer-reviewed journal articles. While requiring a graduate student to publish multiple articles before earning a doctorate might be a high bar to meet, including publication in the dissertation process might be beneficial to the student and also result in a greater impact on the field (Odendaal & Frick, 2018). Their study examined 1,128 theses at a South African university, finding that the majority of their sample (68.08%) was made up of unpublished monographs and publication-based theses, while less than half the total (47.25%) of the authors in the sample published, presented, or submitted some form of their work during candidacy.

Impact and publication have also been assessed through the examination of critical and current issues in education. Weber and Allen (2016) examined the topics covered by published education doctoral dissertations and found that 88% of topics were classified as “inside school” issues. Furthermore, they noted that five topic subgroups the researchers deemed critical to education had no dissertations examining them. While not detracting from the importance of “inside school” research, this study called for a balance of dissertation topics to also address issues of government and politics (such as privatization of schools) and school policies (such as classroom discipline and justice).

The lack of dissertation publication is not only reserved for education research; other fields report a similar dearth of publication, particularly in fields where publication is not required (i.e., clinical practice; Evans et al., 2018). For example, Evans et al. (2018) found that only 25% of psychosocial intervention dissertations were published in peer-reviewed journals within seven years of completion. Furthermore, the published dissertations generally appeared in journals that have moderate to high influence. The researchers speculated that “publishability” should be added as a criteria to programs looking to support students in making an impact on their respective fields. Conversely, Moyer et al. (2010) found no significant differences in quality between 107 published and unpublished dissertations, suggesting that even if dissertations are not publishable due to negative results, they may still be useful in meta-analyses.

NEED FOR CURRENT STUDY

While many of the above studies pointed out the variations in quality among dissertations and lack of publication, none of the literature consulted examined a randomly selected sample of all dissertations in K-12 education. Ronau et al. (2014) evaluated a large sample focused on mathematics technology. Most studies reviewed above included smaller sample sizes, non-random samples, specific fields, or analysis of one institution’s or organization’s body of dissertations. In addition, in this study, the researchers adapted the rubrics developed by Ronau et al. (2014) by expanding how validity is measured in qualitative work using Lincoln and Guba’s (1985) framework for trustworthiness. The researchers also adapted the rubrics to measure different types of mixed methods research and more nuanced levels of statistical rigor in quantitative research. In addition, other research examining topical coverage in dissertations addressed the need for K-12 school improvement studies (Biddle, 2015) and the need to evaluate research use in K-12 settings (Lawlor et al., 2019).

In addition, publication data on dissertations is scant to non-existent in the field of education. In particular, education dissertation quality research has not been updated in nearly ten years, and certainly not since the increase in the availability of online doctoral programs in 2020/2021. In this study, a randomly selected sample of over 900 dissertations from a 10-year period were examined to determine the publication rate. From this sample, over 500 dissertations were selected to assess for quality using a rubric adapted and expanded from the one used by Ronau et al. (2014). This study sought to determine the general quality of dissertations in education published in the last ten years, the differences in quality by methodological paradigm and institution type, and the impact of methodological quality on whether a dissertation was ultimately published.

METHODS

This study examined the quality of methods employed in K-12 education dissertations in the U.S. using a researcher-designed rubric as an assessment tool. Additionally, the publication rates of the sample of dissertations retrieved were determined to examine the impact on the field of K-12 education (as just one metric). To achieve these ends, randomly selected dissertations that were uploaded to the Proquest dissertation database from a ten-year period were downloaded and assessed. With respect to methodological quality, the researchers assessed whether there was variation in quality and use of methodology over time, whether there was a difference in quality across institution types (public, private, for-profit, online-only), and whether there was a difference in quality among methodological paradigms (quantitative, qualitative, mixed methods).

SEARCH STRATEGY AND SAMPLING STRATEGY

Proquest Dissertation Global was used as a source for this study as it is considered the most comprehensive database of full-text dissertations and theses in the United States. This study chose to limit the retrieval of dissertations to those generated in the U.S., first, to establish a reasonable and controlled sample size and second, to level the comparisons of dissertations made between institutions. One dissertation model used both within and outside of the U.S. is the three-article publication model, in which a dissertation is comprised of multiple published articles, some of which may be different studies. In fact, research has shown that there is no one consistently used model (Mason et al., 2024). While some U.S. universities do use this dissertation model, only one example of a multiple-article dissertation was found in this sample and was excluded from the evaluation. The goal of this study was to evaluate the methodological quality of dissertations with an eye toward their impact as assessed through whether they were published. The universe of education dissertations is in the realm of nearly 700,000 English language dissertations and theses given the subject code “education” in the Proquest database. The focus of this study was on dissertations in K-12 teaching and learning in general rather than administration/educational leadership, as the expertise of the team was in the K-12 domain. As noted in the literature review, no comprehensive search of K-12 dissertations has been evaluated for methodological quality. Two separate and broad searches were conducted using the terms “teaching and learning” and “elementary education” in one search and “teaching and learning” and “secondary education” in the other. Categories of any university or study setting not in the U.S., not in English, or not a dissertation were excluded. The search was limited to ten years from 2010 to 2020 to allow for both currency of the doctoral education field and also to allow for a buffer of time between when a student would defend a dissertation and when it would be reasonably published. Most dissertations are published as journal articles several years after the defense. Despite these search parameters, many master’s theses and non-US studies came up in the search results. In the secondary education set of dissertations, there was a retrieval of over 10,000 dissertations, and for elementary there were over 9,000. To derive a representative sample of years, schools, subjects, and a manageable number of dissertations that the team could reasonably review for this study, a sample of 5% of dissertations was randomly retrieved from these two searches. Therefore, for each dataset, about 500 dissertations were randomly selected. These efforts resulted in a data set of 989 randomly selected dissertations over ten years, 2010 to 2020. Each member of the research team assessed between 75-150 dissertations. This number was determined to be reasonable within the constraints of the timeline of the project, resulting in the evaluation of 553 dissertations from this sample.

INSTRUMENTATION

The structure of this study was based on that of Ronau et al. (2014), who assessed mathematics education technology dissertations and published articles for methodological quality. Ronau et al.’s rubric was based on those of Shakeshaft (1982) and Shadish et al. (2002), although those sources focused on quantitative methods with relatively high levels of evaluation of statistical analyses. Ronau et al.’s (2014) quantitative rubric assessed how well the study was grounded in the literature, whether the design had a control, was experimental or not, and whether there were research questions and a purpose statement. Similarly, for qualitative and mixed methods research, Ronau et al.’s rubric indicated that all stated research designs were given the same point value. Instrumentation used in the dissertation was also assessed for both quantitative and qualitative studies with respect to whether reliability or validity was addressed. For both types of methods, the rubric assessed whether construct, external, and internal validity were addressed. Statistical conclusions were assessed by Ronau et al. for quantitative research only. A draft of the rubric was shared with several other experts in dissertation methodology, and additional edits were made to address the evaluation of qualitative research. Ronau et al.’s rubric was then piloted with a small sample of randomly selected dissertations, resulting in further refinement of the measures. These rubrics were amended for the current study as described below.

As Ronau et al. (2014) did, the research team created entirely separate rubrics for quantitative, qualitative, and mixed methods research, resulting in three rubrics. All three methods were rated on research questions, research design, and sampling rigor, albeit on different scales for research type (see the Appendix for all three rubrics).

Qualitative

The validity of qualitative studies was assessed based on Lincoln and Guba's (1985) framework for establishing trustworthiness, which is commonly used in qualitative studies. This framework accounts for credibility, confirmability, dependability, and transferability. The rubric outlined which data collection methods or modes of analyses would signal that any of these types of validity were addressed. Lincoln and Guba describe a series of techniques that can be used to conduct qualitative research that achieves the criteria they outline. These definitions and techniques are listed in the rubric in the Appendix. Primarily, this study looked for discussions of how the researcher addressed each of these validity and trustworthiness assurances in the methods section. However, this study did not assess whether these ends were achieved by fully reviewing the findings. At times, methods sections in qualitative research included a "revisiting validity section," which detailed how the data met trustworthiness metrics. For qualitative studies and mixed methods studies, researchers received 1 point for each type of validity threat addressed (Credibility, Confirmability, Dependability, Transferability), and additional points were given for each method of analysis used to assess for validity (i.e., member checking, triangulation). For example, if a researcher used triangulation, member-checking, and peer review to assure the validity of the analysis, they were awarded 3 additional points. These choices were validated by consulting Creswell (2007) and Creswell and Plano Clark (2007).

Among qualitative dissertations, generally, any "stated" design (indicating 'qualitative' as a design was not sufficient) received 2 out of 2 points. If the study only indicated qualitative methods were used without a stated design, it received only 1 point for design. For qualitative sampling strategies, purposive sampling (or another stated/criterion-based strategy) received a 2 out of 2, and convenience sampling received 1 point. In qualitative research, criterion-based selection allows the researcher to select inclusion and exclusion criteria in line with the research question. Convenience sampling is generally recommended to be avoided as it may not be representative of the population of interest or meet the inclusion and exclusion criteria that would best define a sample for a particular study (Johnson & Christensen, 2017). Because most researchers used some type of criterion-based sampling strategy, there was little variation in scores among qualitative designs. This was a limitation of the adapted rubrics in that it was not possible to value one type of qualitative research over another in the same way the quantitative research design may be evaluated (e.g., experimental design is more rigorous than descriptive). Therefore, most qualitative research, regardless of the scope and quality, were awarded the same scores, and variation in scores rested largely on assessing validity.

Quantitative

Using Ronau et al.'s (2014) categories, quantitative research was assessed for four types of validity with some minor variations in scoring. These scores focused on the quality of the data analysis – construct validity (was the stated construct defined and measured), external validity (are the results generalizable via a large enough sample size), internal validity (controlling for confounding variables), and statistical validity (up to three points for selecting the correct statistical test, testing assumptions, and appropriate sample size). When judging statistical validity, for example, points would not be awarded for using descriptive statistics when inferential statistics could have been conducted, not testing for assumptions appropriate for the statistical test shown (e.g., assessing homogeneity of variances through a Levene's test; checking that data are normally distributed), and for not using a sufficiently sized sample. Since sample size can only be assessed in relation to the project, statistical power, and population, and given the variation in designs in this study, studies with sample sizes of 100 or more were awarded one point.

In addition to assessing validity and research designs, points were awarded for whether instrumentation was assessed for reliability and validity. The mixed methods rubric used the characteristics from both the quantitative and qualitative rubric for a maximum of 21 points.

For quantitative and mixed method designs, an experimental design received a 3 out of 3. Quasi-experimental, regression or causal-comparative studies were deemed on par with any stated method for qualitative, earning a 2. Descriptive or correlational studies received a 1. Thus, there was more variation in scores among quantitative research designs. For quantitative sampling strategies, a 3 was awarded for experimental designs with random assignment. Like with qualitative designs, quantitative and mixed methods studies were given 2 points for purposive sampling and 1 point for convenience sampling.

EVALUATOR TRAINING AND INTER-RATER RELIABILITY

All of the researchers were raters in this study. Each rater was a graduate of a doctoral program in education with an emphasis on research design and curriculum studies. The first author is a methodologist and teaches research design courses at the doctoral level. Among the raters, three are quantitative methodologists, and two are qualitative methodologists. The first author double-checked random selections of dissertations and served as a mediator when the raters could not agree on a rating.

Training

All raters were involved in two training sessions where the scope of the research, the design of the rubric, and discussions of ratings were involved. Each session was recorded for review. All raters were initially given the same five dissertations to assess using the rubrics developed for this study. From these efforts, initial norming results were computed, and a subsequent meeting was held to address areas of disagreement and to refine the rubrics. A coding spreadsheet was also developed to operationalize each variable being assessed and to describe the dimensions of scoring. After several additional meetings to refine the rubric, the final rubric that appears in the appendices was used as the data collection tool. Raters used the rubric to derive scores and then input those scores onto a shared spreadsheet.

Interrater reliability

After the training period, the raters engaged in parallel rating in which pairs of raters evaluated groups of dissertations together. After these practices, all pairs were given 15-20 dissertations to rate separately in order to assess inter-rater reliability (IRR). Each item was checked for agreement, and subscale scores were computed for reliability for research design, validity, and instrumentation. Pair 1 had an overall IRR of 91% of agreement. The areas with the lowest agreement on individual items were sampling strategy (88%), internal validity (78%), and validity of instruments (78%). Most items for pair 1 had an agreement of 1.0. Pair two largely worked on qualitative studies and had a slightly lower agreement on all items. The score for research design was 87%, for qualitative validity 83%, and for instrumentations 85%, with an overall IRR of 85%. The first author was asked to confirm all quantitative studies for these two raters since they were less familiar with quantitative methods. Most sources note that a minimum agreement of 80% is standard for interrater reliability (McHugh, 2012).

DATA COLLECTION

In addition to the evaluation of the research methods, the research team collected data on the type of institution, the year of publication of the dissertation database, and information about whether the dissertation was published in a book or article, as described below.

Members of the research team searched Google Scholar and library databases to determine whether dissertations had been published as books, book chapters, or peer-reviewed scholarly articles. This information was retrieved by searching for the name of the dissertation author and keywords from the dissertation title and abstract. For all published dissertations, the following data were collected: citation, year of publication, and journal impact factor. While there are vast variations in quality

among peer-reviewed journals that publish scholarly articles, an item to be published if in a journal designated as peer-reviewed in Ulrich's periodical directory. All journals were then searched to determine if they were indexed in Cabell's and by Ebsco. These indices were also used to determine if journals were predatory.

DATA ANALYSIS

After all evaluations were completed, they were entered into an Excel spreadsheet. Frequency counts were used to describe the data, including the number of each design type and method used. The data were then coded for institution type, research design type, and data collection methods used to allow for Analysis of Variance testing. For example, all public institutions were coded with a 1, private with a 2, and for-profit schools with a 3. After data were organized, they were input into SPSS v. 28 for inferential analysis. Because score ranges for each design type varied, ratios were computed to allow for comparison of methodologies by dividing the total quality score by the highest possible score for that rubric. Descriptive statistics were used to describe the dataset, and analysis of variances tests were used to analyze whether there were statistically significant differences between methodological types, institution types, and the year the dissertation was published in the dissertation database.

RESULTS

RESEARCH DESIGN TYPE AND QUALITY

A total of 553 dissertations were evaluated for this study. Because the rubrics used for the quantitative, qualitative, and mixed methods studies each included different domains specific to that paradigm, the range of possible scores varied. The highest possible score for quantitative studies was 16, for qualitative it was 17, and for mixed methods it was 21. Therefore, when comparing scores for all three methods, score ratios were computed to allow for comparing the following variables: year of publication, type of design, and institution type. Table 1 reports the raw mean scores for each type of design, while Table 2 represents the ratio scores for comparisons. Ratios were computed by dividing the raw score for research design by the highest possible score so that all research design types would be comparable. The majority of methods used were qualitative ($N = 292$), while there were 142 quantitative dissertations and 119 mixed methods.

An ANOVA was conducted to determine if there were differences between design types on quality ratio scores. Levene's test of homogeneity was not significant, indicating that the data were suitable for ANOVA testing. There was a significant difference between quantitative and qualitative studies and quantitative and mixed methods studies $F(2, 550) = 26.56, p < .001$. There was no significant difference between qualitative and mixed methods on the total quality score, $p = .08$. Quantitative research rated higher quality scores than both qualitative and mixed methods.

Table 1. Raw mean scores for each research design type

Design type	N	Maximum score possible	M	SD
Quantitative	142	16	9.65	2.76
Qualitative	292	17	8.50	2.70
Mixed Methods	119	21	9.85	3.33

Note. Raw scores are not comparable as each design type had a different range of scores

Table 2. Ratio scores for research design type

Design type	N	M	Sd
Quantitative	142	.60	.17
Qualitative	292	.50	.16
Mixed Methods	119	.47	.16
Total	553	.52	.17

Note. Ratio scores are comparable

DESCRIPTION OF RESEARCH DESIGNS

Over 40 types of research designs were used in these studies, many of which used multiple designs within any one study. For example, a mixed method study might be explanatory sequential, using an experimental design as the quantitative part and phenomenology as the qualitative method, although mixed methods often do not name the designs for both parts of the research. Seventy-two studies did not name a design and instead stated either quantitative, qualitative, or mixed methods as their design with no further explanation.

For quantitative dissertations, the most commonly used designs were surveys (N = 36), although some additional survey research was alternately categorized as causal comparative), quasi-experimental (N = 35), descriptive or correlational (N = 33), and experimental (N = 8). For qualitative dissertations, case study (N = 157) was the most commonly cited method, often in concert with another method, followed by phenomenology (N = 44), ethnography (N = 21), and action research (N = 20). Explanatory sequential was the most frequently used design (N = 40) for mixed methods, although many mixed methods studies did not name a mixed method design and rather named one quantitative design or one qualitative design with a descriptor of the data collection methods for the complementary part (for example, survey, observations, or interviews). Other less commonly used design types included the Q method, Delphi study, secondary data analysis, convergent parallel, and predictive/regression.

INSTITUTION TYPE AND QUALITY

Dissertations were also categorized by institution type to determine whether there was an impact on dissertation quality, particularly in recognition of the number of doctoral programs available from for-profit, online-only schools. There were 311 public/state schools, 131 private not-for-profit schools, 73 for-profit schools, most of which are also online-only schools, and 38 religious schools, which were also private schools. An ANOVA showed no significant differences between institutions overall, and no significant differences were found in the post hoc analysis. A preponderance of the schools sampled were Walden (N = 32) and Capella (N = 20), representing a total of 9.4% of the sample, while all other schools had no more representation than 1-5 dissertations in the sample.

YEAR OF PUBLICATION TO DISSERTATION DATABASE AND QUALITY

To determine whether quality and type of dissertations have changed over time, comparisons were made using an ANOVA. Overall, there was no significant difference between the year of publication and quality score, although post hoc analysis showed that 2011 and 2015 had significantly lower quality scores when compared to other years (see Table 3).

Crosstabs were used to investigate the relationship between the year of publication and design type. Chi-square showed no significant relationship (Pearson's Chi-square = .627), indicating no shift in the type of design used over time (Table 4).

Table 3. Quality ratio score by year

Year	M	N	SD
2010	.54	49	.17
2011	.47	38	.18
2012	.54	45	.29
2013	.51	52	.17
2014	.53	62	.16
2015	.50	70	.17
2016	.50	59	.15
2017	.52	49	.18
2018	.57	46	.15
2019	.55	44	.17
2020	.52	39	.16
Total average	.52	553	.17

Table 4. Research design type over time

Year	Quantitative		Qualitative		Mixed methods		Total
	n	%	n	%	N	%	n
2010	17	35	25	51	7	14	49
2011	8	21	23	61	7	18	38
2012	12	27	23	51	10	22	45
2013	9	17	32	62	11	21	52
2014	12	19	35	56	15	24	62
2015	22	31	32	46	16	23	70
2016	12	20	27	46	20	34	59
2017	13	27	26	53	10	20	49
2018	16	35	22	48	8	17	46
2019	11	25	24	55	9	20	44
2020	10	26	23	59	6	15	39
Total	142		292		119		553

RESEARCH DESIGN SUBSCALE SCORES

The research design (R.D.) subscale, which assessed whether the R.D.s fit the research question, the rigorousness of the design, and the rigorousness of the sampling strategy, was separately analyzed for each study. Using ratio scores, the research design subscale scores were significantly different among design types; scores were higher for qualitative studies ($m = .72$) than quantitative ($m = .62$) and mixed methods ($m = .56$); $F(2, 552) = 58.96, p < .001$.

SUBSCALE SCORES FOR RELIABILITY AND VALIDITY

Reliability

Where applicable, the members of the research team looked for discussions of reliability scores or discussions of assurance of reliability, including inter-rater reliability or internal consistency reliability.

If scores were reported, a value of .70 or greater was considered acceptable. Fourteen percent ($N = 41$) of qualitative dissertations reported reliability scores or discussions of assurance of reliability, 61% ($N = 86$) of quantitative dissertations reported a reliability score, and 38% ($N = 45$) of the mixed methods studies reported a reliability score. It is generally expected that quantitative studies would be more likely to use reliability measures, although a qualitative dissertation using any type of protocol should discuss how they are assessed for validity or reliability. If variables used in studies do not typically include reliability scores, then some discussion about how or whether those variables are reliable is warranted.

Quantitative validity

For quantitative studies, two parts of the rubric were used to derive validity scores. First, this study looked for scores related to validity or a discussion of how variables were validated. There was some expectation that even if a scale or measurement tool was not used as a variable, there would be discussion and assurance of the validity of the variables. For mixed methods studies using quantitative methods, 98 (82%) studies reported validity measures, while 114 (80%) of quantitative studies reported validity measures. In addition, all quantitative studies were assessed for statistical validity for a maximum of seven points. Quantitative studies had a mean validity score of 4.05 points, while mixed methods studies had a mean score of 1.94. There was a statistically significant difference in validity scores between quantitative and mixed methods $F(1, 259) = 84.565, p < .001$.

Qualitative validity

For qualitative studies, validity scores were based on the framework presented by Lincoln and Guba (1985), where confirmability, credibility, dependability, and transferability are assessed. In essence, a high score reflected whether the researcher stated they were assuring for these types of validity using specific methods (i.e., member checking and peer review to assure credibility; thick description to assure transferability). These scores ranged from 0 to 4. Points were added for assurance of reliability (i.e., inter-rater reliability or reliability of protocol) and validity of analysis measures (thick description, triangulation, peer review, etc.). Seventy-eight percent ($N = 228$) of qualitative studies reported validity assurances of trustworthiness, and 86 (72%) of mixed methods studies reported these assurances.

In addition, qualitative studies were given points for methods used to assess the validity of the data analysis. Studies were given 1 point for each data analysis method used to assess or assure validity, including triangulation, peer review, member checking, and thick description as the most commonly used data analysis validation tools. The average validity score for qualitative studies was 4.02, and for the qualitative portion of the mixed methods was 2.93. There was a statistically significant difference between validity scores on mixed methods and purely qualitative dissertations $F(1, 408) = 19.69, p < .001$.

Ratios were computed for validity scores for all dissertations to allow for comparison between all three methods. The mixed methods ratio was a computation of the quantitative and qualitative validity scores combined and then converted into a ratio. The mean validity ratio for quantitative research was .58, for qualitative it was .40, and for mixed methods, it was .29. An ANOVA showed that there was a statistically significant difference between the three validity scores $F(2, 549) = 58.6, p < .001$.

To explore the issues of validity for mixed methods, these studies tended to do one method type better than the other. Using the total ratio score for validity, mixed methods studies scored significantly lower when compared with quantitative and qualitative research studies.

PUBLICATION OF DISSERTATIONS

From the final data set of 553 dissertations, 22 (4%) were found to be published. Six were quantitative studies, 12 were qualitative, and four were mixed methods. Two dissertations were published as a series of articles (between 2 and 3 articles) and one as a book. Of the full set of 980 dissertations randomly retrieved from the dissertation database, three additional dissertations were found to have

been published. Using the final data set of 553 dissertations, an ANOVA did not show significant differences between published ($m = 50.8$) and unpublished dissertations ($m = 52.12$) on quality, $p = 0.711$. The lack of significant differences may be explained by the uneven sizes of the groups. With respect to quality indexing, the analysis showed that only two journals were indexed in Cabell's, and 12 out of 20 were indexed by EBSCO. Four of the published articles appeared in journals that either no longer exist or are no longer accepting manuscripts, one of which was suspected as a predatory journal. One publication was in a teacher trade journal that is peer-reviewed but not blind reviewed or scholarly. Only one publication of the 22 was in a top-tier journal.

DISCUSSION

This study assessed and compared the methodological quality of dissertations defended from 2010 to 2020 in the subject area of education. Analyses of scores on a new researcher-designed rubric revealed that quantitative designs had higher ratio scores, in part due to the differences in the way validity is measured for both types of design. Validity controls are more often present in quantitative research designs than in qualitative research designs. Variances in the assessment of validity for qualitative studies relied on assurances of trustworthiness using Lincoln and Guba's (1985) framework. Mixed methods studies had significantly lower quality scores, likely because researchers tended to do one type of method better than the other. Further, dissertations using mixed methods designs that focused on the quantitative part of the study rarely assessed qualitative validity and trustworthiness. Methods courses should train students completing mixed dissertations to master both methods at once.

The study also sought to determine whether institution type affected dissertation quality given the expanded availability of online doctoral degrees. There were no significant differences in quality scores between types of institutions.

In the sample, 22 doctoral graduates published research from their dissertations. It also became clear, through the course of searching each dissertation for peer-review publication status, that those who did publish their dissertations also had other publications and, in most cases, were university faculty members. Given the low rate of publication, one might assume that since most professionals with doctorates stay within the K-12 field or other non-academia roles, publication is not an integral career goal. Previous research also shows that across disciplines, even beyond education, few doctoral graduates publish their dissertations (Evans et al., 2018; Moyer et al., 2010). Moyer et al. (2010) found no difference in quality over those dissertations published or unpublished supporting the current findings.

Variance in quantitative scores came from threats to validity; often validity of an instrument was claimed with no data to support the statement. For studies with low statistical validity scores, tests were often appropriate but rarely did an author state whether assumptions were met or if data normality was assessed. Construct validity was universally addressed. Construct validity was supported if the construct was properly explained, defined, and measurable based on the literature. Internal validity was often scored lower as homogeneity or heterogeneity was difficult to establish or control in classroom research. If a researcher did not attempt to control (through group manipulation or matching) or code for a covariate, then they did not earn this point. Similar to the current study, Ronau et al. (2014) found that quantitative research scored higher on validity and reliability than qualitative research.

Qualitative research had significantly higher research design scores, unsurprisingly, since nearly all qualitative studies are not leveled by rigor but by purpose. As Ronau et al. (2014) point out, qualitative dissertations tended to do a better job of establishing theoretical foundations and thus improving the quality of work done in the field of education. Quantitative studies had more variation in scores as there are hierarchies of rigor in this paradigm type. The case was similar for the sampling strategy as a stated purposive method; in qualitative, it achieved a score of 2/2, while in quantitative, the only

strategy to receive a 3/3 was random assignment. Most studies received a 2 for research questions. An assumption can be made, too, that experimental design (the most rigorous research design) is more difficult for doctoral students to conduct in a K-12 setting.

As noted by other researchers, the quality of doctoral dissertations has often been questioned. As was found in this study, higher scores were given to dissertations that had more validity measures. Similar to the current study, Anney (2014) found that qualitative theses used fewer reliability and validity measures, and studies that were higher quality often used quantitative measures of validity to describe their instruments. Other researchers posed that this may be a result of poor instruction in research methods (Hamilton et al., 2010; Thompson, 1994). Accordingly, Kyvik and Thune (2015) and Hilmer and Hilmer (2011) found that evaluator quality, as measured by affiliation and experience, positively influenced dissertation quality. In addition, some doctoral degrees focus less on research design and more on curriculum and leadership. Excellence in research develops over time over multiple projects (Casanave, 2019; Casanave & Li, 2015). However, as Ronau et al. (2014, p. 33) note, “research community must in turn begin to demand greater quality in its published studies, both through how researchers write about their own studies and how they review the works of others.”

While the scores on dissertations demonstrate variability among and between research designs, the more important question is the impact that dissertation work has on practice. This study only articulates one manner by which dissertations may influence practice; that is by publication. The question remains whether research design flaws are reflective of program quality or if the dissertation is viewed by doctoral programs as a capstone and an opportunity to put research design instruction into practice.

LIMITATIONS

The most significant limitation of this study was the lack of variance among scores for qualitative research as measured by the rubric. There was no hierarchy related to the design or number of validation tools used for data analysis. The variance was largely related to addressing validity threats. In some cases, there was no mention of validity threats. In others, a description of all possible threats was made, but no plan, or a limited plan, to address them was demonstrated. The highest quality qualitative research extensively addressed validity threats, trustworthiness, and researcher bias within the methods and results sections. In addition, this study relied on the methods section to establish whether validity was addressed; however, it did not assess whether these ends were achieved by fully reviewing the findings.

It is important to reiterate that quantitative and qualitative research are not comparable, although attempts were made to measure both design types equitably through rubrics for each design type and scoring that reflected the different values placed on research design decisions in each design type. They have different purposes and values. Mixed methods tended to have lower quality for both paradigms. In quantitative studies, since sample size can only be assessed in relation to the project, statistical power, and the population, and given the variation in designs in this study, awarding a point for sample sizes of 100 or above was a somewhat subjective threshold.

With respect to the analysis of publication rates, the quality of the journals in which the articles were published was not assessed. Publication quality of journal articles, particularly for early career researchers, may vary. It is possible that the dissertations that were published into journal articles were in predatory or low-quality journals. Future research should examine and correlate methodological quality to journal metrics that describe quality and rigor.

IMPLICATIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

As a result of this study, many other questions about doctoral dissertation quality come to light. Studies comparing subjects and levels (elementary vs. secondary) in dissertations may yield different results. In addition, more qualitative approaches to examine dissertation quality are warranted. For example, future research should assess the quality of research in the context of usefulness, impact on

practice beyond publication, and topic currency. Further, other aspects of dissertations should be examined. In this study, through cursory reviews, a lack of quality in literature reviews and writing was noted overall. Boote and Beile (2005) assessed literature reviews via rubrics but not recently nor through random samples of a nationwide population.

Given the dearth of dissertations published after the defense, the question remains why so few graduates publish their work beyond our own explanations here of the nature of K-12 career trajectory. More studies examining how doctoral programs support and encourage publication (or not) would better explain this phenomenon. An investigation into the personal motivations of doctoral graduates who do publish would also extend this line of inquiry.

This study has implications for doctoral programs and institutions to consider. The methodological (and other) aspects of quality in dissertations have been called into question here. What can doctoral programs do to improve quality? Or is it student quality that most affect dissertation quality? Is a publication of value to doctoral programs? If so, which types of programs? Moving to a model of publication-based theses and dissertations would most certainly improve publication rates and accordingly improve dissertation impact on the field (Odendaal & Frick, 2018), but may inhibit completion of doctoral degrees.

CONCLUSION

This study examined the extent to which there are methodological quality differences between types of research designs in dissertations. The study also explored whether dissertations are published after they are defended and whether there is a relationship between methodological quality and publication. In this study, the researchers expanded and adapted a rubric developed by Ronau et al. (2014) by extending the measure of qualitative research validity using Lincoln and Guba's (1985) framework for trustworthiness. The rubric developed for this study treats mixed methods as a separate paradigm rather than a combination of quantitative and qualitative studies. Ronau et al.'s (2014) rubric differs from the one used for this study by also measuring the quality of the theoretical framework and literature review. As noted in the discussion, the researchers acknowledge that the assessment of literature reviews has been extensively assessed by Boote and Beile (2005). The researchers also included more nuanced measures of statistical rigor and accounted for the variation in mixed methodologies. While there was variation in quality among research designs, the strongest indicators of quality were those studies that addressed threats to validity. Quantitative designs more often included measures to account for validity. Because quantitative designs include more types of statistical and design controls, the quantitative design ratio scores were often higher than qualitative and mixed methods scores. Mixed methods studies had higher ratio scores than qualitative research; often, researchers conducted either the qualitative or quantitative part rigorously, but not both.

This study also demonstrated that research is still needed to demonstrate how dissertations and doctoral work impact practice. In this study, only a small percentage of dissertations were published into monographs or peer-reviewed articles, but more research is needed to assess the relationship between the quality of dissertations and the quality of the journals in which they have been published. This study also showed that doctoral programs should work to improve writing and publication instruction in addition to promoting methodological rigor. Rigor and attention to detail varied overall among dissertations, and no significant pattern emerged in trends over time or by institution type.

REFERENCES

-
- Alfaro, I. (2023, March 15). *The future of the Ed.D. degree – and how it will change over time*. Fortune Education. <https://fortune.com/education/articles/the-future-of-the-ed-d-degree-and-how-it-will-change-over-time/>
- Anney, V. N. (2014). Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. *Journal of Emerging Trends in Educational Research and Policy Studies*, 5(2), 272-281.

Methodological Quality of Dissertations

- Biddle, J. C. (2015). Dissertation requirements in professional practice education doctoral programs to facilitate K-12 school improvement in the United States. *Work Based Learning e-Journal International*, 5(1), 55-85. <https://eric.ed.gov/?id=EJ1305241>
- Boote, D. N., & Beile, P. (2005). Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher*, 34(6), 3-15. <https://doi.org/10.3102/0013189X034006003>
- Casanave, C. P. (2019). Performing expertise in doctoral dissertations: Thoughts on a fundamental dilemma facing doctoral students and their supervisors. *Journal of Second Language Writing*, 43, 57-62. <https://doi.org/10.1016/j.jslw.2018.02.005>
- Casanave, C. P., & Li, Y. (2015). Novices' struggles with conceptual and theoretical framing in writing dissertations and papers for publication. *Publications*, 3(2), 104-119. <https://doi.org/10.3390/publications3020104>
- Cleary, R. E. (2000). The public administration doctoral dissertation reexamined: An evaluation of the dissertations of 1998. *Public Administration Review*, 60(5), 446-455. <https://doi.org/10.1111/0033-3352.00107>
- Conn, V. S. (2008). The light under the bushel basket: Unpublished dissertations. *Western Journal of Nursing Research*, 30(5), 537-539. <https://doi.org/10.1177/0193945908317602>
- Creswell, J. W. (2007). *Qualitative inquiry and research design* (2nd ed.). Sage.
- Creswell, J. W., & Plano Clark, V. L. (2007). *Designing and conducting mixed methods research*. Sage.
- Drysdale, J. S., Graham, C. R., Spring, K. J., & Halverson, L. R. (2013). An analysis of research trends in dissertations and theses studying blended learning. *The Internet and Higher Education*, 17, 90-100. <https://doi.org/10.1016/j.iheduc.2012.11.003>
- Durak, G., Cankaya, S., Yunkul, E., & Mısırlı, Z. A. (2018). A content analysis of dissertations in the field of educational technology: The case of Turkey. *Turkish Online Journal of Distance Education*, 19(2), 128-148. <https://doi.org/10.17718/tojde.415827>
- Eryılmaz, Ö. (2022). Are dissertations trustworthy enough? The case of Turkish Ph.D. dissertations on social studies education. *Participatory Educational Research*, 9, 344-361. <https://doi.org/10.17275/per.22.70.9.3>
- 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>
- Halse, C., & Mowbray, S. (2011). The impact of the doctorate. *Studies in Higher Education*, 36(5), 513-525. <https://doi.org/10.1080/03075079.2011.594590>
- Hamilton, P., Johnson, R., & Poudrier, C. (2010). Measuring educational quality by appraising theses and dissertations: Pitfalls and remedies. *Teaching in Higher Education*, 15(5), 567-577. <https://doi.org/10.1080/13562517.2010.491905>
- Hilmer, M. J., & Hilmer, C. E. (2011). Is it where you go or who you know? On the relationship between students, Ph.D. program quality, dissertation advisor prominence, and early career publishing success. *Economics of Education Review*, 30, 991-996. <https://doi.org/10.1016/j.econedurev.2011.04.013>
- Hochbein, C., & Perry, J. A. (2013). The role of research in the professional doctorate. *Planning and Changing*, 44(3/4), 181-195.
- Hoffer, T. B., Welch, V., Webber, K., Williams, K., Lisek, B., Hess, M., Loew, D., & Guzman-Barron, I. (2009). *Doctorate recipients from United States universities: Summary report 2005*. National Science Foundation. <https://www.nsf.gov/statistics/doctorates/>
- Johnson, R. B., & Christensen, L. (2017). *Educational research: Quantitative, qualitative, and mixed approaches*. Sage.
- Juodaitytė, A., & Kazlauskienė, A. (2008). Research methods applied in doctoral dissertations in education science (1995-2005): Theoretical and empirical analysis. *Vocational Education: Research & Reality*, 15, 36-45.
- Karadağ, E. (2011). Instruments used in doctoral dissertations in educational sciences in Turkey: Quality of research and analytical errors. *Educational Sciences: Theory & Practice*, 11, 330-334. <https://files.eric.ed.gov/fulltext/EJ919904.pdf>

- Kumar, S., & Dawson, K. (2013). Exploring the impact of a professional practice education doctorate in educational environments. *Studies in Continuing Education, 35*(2), 165-178.
<https://doi.org/10.1080/0158037X.2012.736380>
- Kyvik, S., & Thune, T. (2015). Assessing the quality of PhD dissertations. A survey of external committee members. *Assessment & Evaluation in Higher Education, 40*(5), 768-782.
<https://doi.org/10.1080/02602938.2014.956283>
- Lawlor, J., Mills, K., Neal, Z., Neal, J. W., Wilson, C., & McAlindon, K. (2019). Approaches to measuring use of research evidence in K-12 settings: A systematic review. *Educational Research Review, 27*, 218-228.
<https://doi.org/10.1016/j.edurev.2019.04.002>
- Lincoln, Y. S., & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Mason, S., Frick, L., Castelló, M., Cheng, W., Chong, S. W., Díaz Villalba, L., García-Morante, M., Kong, M. S., Sakurai, Y., Shojaeian, N., Spronken-Smith, R., & Weise, C. (2024). Prominence, promotion and positioning of the 'Thesis by Publication' in six countries. *Higher Education Policy*. <https://doi.org/10.1057/s41307-024-00350-7>
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica, 22*(3), 276–282.
<https://doi.org/10.11613/BM.2012.031>
- Moyer, A., Schneider, S., Knapp-Oliver, S. K., & Sohl, S. J. (2010). Published versus unpublished dissertations in psycho-oncology intervention research. *Psycho-Oncology: Journal of the Psychological, Social and Behavioral Dimensions of Cancer, 19*(3), 313-317. <https://doi.org/10.1002/pon.1561>
- National Science Foundation. (2023). *Doctorate recipients from U.S. universities: 2022*. NSF 24-300.
<https://nces.nsf.gov/pubs/nsf24300>
- Odendaal, A., & Frick, L. (2018). Research dissemination and PhD thesis format at a South African university: The impact of policy on practice. *Innovations in Education and Teaching International, 55*(5), 594-601.
<https://doi.org/10.1080/14703297.2017.1284604>
- Randolph, J., DeWeese, S., Manuel, A. K., Baugher, G., Tessmer, K., Westbrook, A., Shoats, S., Balloun, J., & Crawford, L. (2014). Page length and methodological characteristics of recently published doctoral dissertations in education. *Georgia Educational Researcher, 11*(2), Article 2.
<https://doi.org/10.20429/ger.2014.110202>
- Ronau, R. N., Rakes, C. R., Bush, S. B., Driskell, S. O., Niess, M. L., & Pugalee, D. K. (2014). A survey of mathematics education technology dissertation scope and quality 1968–2009. *American Educational Research Journal, 51*(5), 974-1006. <https://doi.org/10.3102/0002831214531813>
- Sarrico, C. S. (2022). The expansion of doctoral education and the changing nature and purpose of the doctorate. *Higher Education, 84*, 1299–1315. <https://doi.org/10.1007/s10734-022-00946-1>
- Schoenfeld, A. H. (2006). What doesn't work: The challenge and failure of the What Works Clearinghouse to conduct meaningful reviews of studies of mathematics curricula. *Educational Researcher, 35*(2), 13-21.
<https://www.jstor.org/stable/3699951>
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton Mifflin.
- Shakeshaft, C. (1982). *A framework for studying schools as work settings for women leaders* (ED216441). ERIC.
http://www.eric.ed.gov/ERICWebPortal/search/detailmini.jsp?_nfpb=true&_ERICExtSearch_SearchValue_0=ED216441&ERICExtSearch_SearchType_0=no&accno=ED216441
- Ten Cate, O., Derese, A., Durning, S. J., & O'Sullivan, P. (2017). Excellence in PhD dissertations in health professions education: Toward standards and expectations. *Medical Teacher, 39*(9), 926-930.
<https://doi.org/10.1080/0142159x.2017.1302573>
- Thompson, B. (1994). *Common methodology mistakes in dissertations, revisited* (ED368771). ERIC.
<https://eric.ed.gov/?id=ED368771>

Walser, T. M., & Trevisan, M. S. (2016). Evaluability assessment thesis and dissertation studies in graduate professional degree programs: review and recommendations. *American Journal of Evaluation*, 37(1), 118-138. <https://doi.org/10.1177/1098214015583693>

Weber, R. K., & Allen, E. J. (2016). Doctoral dissertation topics in education: Do they align with critical issues? *International Journal of Doctoral Studies*, 11, 403. <https://doi.org/10.28945/3609>

Zumeta, W. (1982). Doctoral programs and the labor market, or how should we respond to the “PhD glut”? *Higher Education*, 11, 321–343. <https://doi.org/10.1007/BF00155622>

APPENDIX

QUANTITATIVE RUBRIC

Domain	0	1	2	3	Score
Research Question	Methods do not align with R.Q.	Some methods align with R.Q.	All R.Q.s are addressed by methods	N/A	/2
Design	None stated or evident	Descriptive statistics or correlation only	Quasi-experiment (no randomization but use of comparison groups); causal-comparative or predictive (use of regression)	Randomized Experiment (has a control group and random assignment)	/3
Sampling Strategy	None indicated	Convenience/snowball	Clearly defined strategy with criteria	Random assignment or random selection	/3
Threats to Validity		Internal External Construct			/3 1 point for <u>each</u> threat addressed
		Correct statistical approach Assumptions tested and met Sample size at least 100			/3 1 point for <u>each</u> statistical conclusion met
Reliability	Not addressed	Internal consistency (Cronbach’s alpha) or alternate forms with value .7 or higher			/1
Validity		Content Concurrent Criterion Predictive Criterion Construct Discriminant Convergent			/1 1 point if <u>any</u> validity type is addressed

QUALITATIVE RUBRIC

Domain	0	1	2	Score
Research Questions	Methods do not align with R.Q.	Some methods align with R.Q. All	R.Q.s are addressed by methods	/2
Design	None named or described	Any named and described phenomenology, narrative; grounded theory; case study; ethnography, etc.		/1
Sampling Strategy	None indicated	Convenience/snowball	Other justified strategy: purposive, quota, etc.	/2
Validity Threats	None named or addressed	Credibility Transferability Confirmability Dependability		/4 1 point for <u>each</u> threat mentioned with a plan for addressing it
Reliability	Not addressed	Internal consistency or interrater		/1 1 point if <u>any</u> type of reliability is addressed and at an acceptable level
Validity	None named or addressed	Persistent observation Member checking/respondent validation Thick description Dependability audit Confirmability audit Reflective journal External moderators Critical friendship Negative case analysis Peer debriefing Triangulation Discrepant evidence Bracketing		/1 1 point for <u>each</u> type of validity addressed

MIXED METHODS RUBRIC

Domain	0	1	2	3	Score
Research Question	Methods do not align with R.Q.	Some methods align with R.Q.	All R.Q.s are addressed by methods		/2
Design	None stated or evident	Explanatory sequential, exploratory sequential, or convergent parallel	Quasi-experiment (no randomization but comparison groups); causal comparative or predictive (use of regression)	Randomized Experiment (has a control group and random assignment)	/3
Sampling Strategy	None indicated	Convenience/snowball	Clearly defined strategy with criteria	Random assignment or random selection	/3
Quantitative Section					

Methodological Quality of Dissertations

Domain	0	1	2	3	Score
Threats to Validity		Internal External Construct			/3 1 point for <u>each</u> threat addressed
		Correct statistical approach Assumptions tested and met Sample size at least 100			/3 1 point for <u>each</u> statistical conclusion met
Reliability	Not addressed	Internal consistency (Cronbach's alpha) or alternate forms with value .7 or higher			/1
Validity		Content Concurrent Criterion Predictive Criterion Construct Discriminant Convergent			/1 1 point if <u>any</u> validity type is addressed
Qualitative Section					
Sampling Strategy	None indicated	Convenience/snowball	Other justified strategy: purposive, quota, etc.		/2
Validity Threats	None named or addressed	Credibility Transferability Confirmability Dependability			/4 1 point for <u>each</u> threat mentioned with a plan for addressing it
Reliability	Not addressed	Internal consistency or interrater			/1 1 point if <u>any</u> type of reliability is addressed and at an acceptable level
Validity	None named or addressed	Persistent observation Member checking or respondent validation Thick description Dependability audit Confirmability audit Reflective journal External moderators Critical friendship Negative case analysis Peer debriefing Triangulation Discrepant evidence bracketing			/1 1 point for <u>each</u> type of validity addressed

AUTHORS



Dr. Amy Catalano, EdD, MLS, MALS, MS, is a Professor of Teaching, Learning, and Technology at Hofstra University. She is also the director of science and STEM education and runs the Girls Summer STEM program at Hofstra. She has been teaching research design and literature review writing to doctoral students for 10 years. In 2019, she was named Distinguished Teacher of the Year in the School of Education. Her research interests include psychometric evaluation of instruments, engineering design in science education, earth science education, and girls and STEM. She is the author of 5 books and over 20 peer-reviewed articles on topics including STEM, engineering, water quality, and information

literacy.



Dr. Marilyn DePietto, EdD., MEd., is an adjunct Professor of Teaching, Learning, and Technology at Hofstra University, where she is also the program director for Secondary Mathematics Education. She has been teaching in New York City public schools for 20 years and has taught common branch subjects in elementary grades as well as secondary mathematics. Over the course of her career in New York City, she has been appointed Math Instructional Lead, New Teacher Mentor, and Citywide Model Teacher. Her research interests include quantitative analysis of constructivist mathematical instructional strategies in both elementary and secondary grades.



Dr. Alexander J. Lord, EdD, MEd is an Adjunct Assistant Professor of Mathematics and Secondary Mathematics Education at Queens College, City University of New York, as well as a secondary mathematics educator. He is both a Math for America and New York State Master Teacher Fellow and has also been awarded educator fellowships through Columbia University, New York University, and CUNY. He has designed curricula for the Algebra for All initiative at CUNY and serves on the Wildlife Conservation Society's Professional Development Teacher Advisory Council. His research interests include secondary mathematics educators' implementation of emerging technologies and methods to blend mathematics, history, and culture within the classroom.



Dr. Susan Turner Radin, EdD, teaches dance and theatre in the performing arts department at Walt Whitman High School in South Huntington, NY, and education at Hofstra University in Hempstead, NY. She serves as the dance representative for the Long Island Scholar Artist Award, is a member of the N.Y. State Education Department's committee for the Individual Arts Assessment Pathway, and is an experienced arts curriculum and grant writer. In 2018, she was honored by N.Y. State's Dance Education Association with the N.Y. State Dance Educator of the Year award. Her research is education-based, with an emphasis on

dance, identity construction, and graduate education.



Dr. Lydia Williams, EdD. is a passionate educational leader with almost 20 years of experience serving diverse communities within the New York City and Long Island areas. Dr. Williams currently serves as the K-12 Director of Curriculum and Instruction for the Wyandanch school district, where her commitment to advancing student outcomes drives her interests to continuously discover best practices in education and guides her efforts to transform theory into practice in ways that best meet the needs of all stakeholders. Her various works have been published in *Educator's Voice*, the *Journal of African American Studies*, and presented at educational conferences internationally. Dr. Williams earned a doctorate in Learning and Teaching and an Advanced Certificate in Educational Leadership from Hofstra University.