DOCUMENT RESUME

ED 282 499 HE 020 327

AUTHOR	Thompson, Bruce
TITLE	Peer Review of Doctoral Dissertations as a Quality Control Mechanism: Some Methods and Examples.
PUB DATE	31 Jan 87
NOTE	13p.; Paper presented at the Annual Meeting of the Southwest Educational Research Association (Dallas, TX, January 31, 1987).
PUB TYPE	Reports - Descriptive (141) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	*Academic Standards; Departments; *Doctoral Dissertations; Higher Education; *Peer Evaluation; *Quality Control

ABSTRACT

The use of peer review as a way to improve dissertation quality is addressed. A data set involving ratings of 52 dissertations illustrates possible design and analytic choices that may be helpful in using a peer review model. Strategies for evaluating rating consistency and validity are covered, along with analysis of four types of program policy analysis questions. The requirement that current students evaluate dissertations produced by recent cohorts of previous students may result in more attention to quality by doctoral students, since they will anticipate future systematic reviews of their work. This practice may also stimulate doctoral faculty to impose higher standards of quality because of possible identification with studies judged either systematically best or worst. Fourteen dissertations were rated by all raters, and the remaining dissertations were typically rated by two or three raters. Policy questions include: the issue of overall quality of the dissertations at a school and the homogeneity of quality across studies; systematic variations in dissertation quality across program areas; and whether dissertation quality is perceived as improving over time. (SW)

docdiss.rno 1/24/87

PEER REVIEW OF DOCTORAL DISSERTATIONS AS A QUALITY CONTROL MECHANISM: SOME METHODS AND EXAMPLES

Bruce Thompson

University of New Orleans 70148
and
Louisiana State University Medical Center

Paper presented at the annual meeting of the Southwest Educational Research Association, Dallas, January 31, 1987.

ABSTRACT

Dissertation quality may be improved by use of student peer review as a quality control mechanism. An actual data set involving ratings of 52 dissertations is employed to illustrate possible design and analytic choices that may be helpful in using such a model. Strategies for evaluating rating consistency and validity are illustrated. Analyses to address four types of program policy analysis questions are also illustrated.

Perhaps the major challenge confronting doctoral programs is maintaining high standards of quality while also respecting student rights and prerogatives. For example, NCATE accreditation regulations require that students must have a significant voice in the assignment of advisors and committee members. Yet some students may be prone to select committee members for their flexibility rather than for their expertise.

The quality of dissertations must remain a fundamental concern of doctoral faculty, however. Even if a faculty member's own inherent interest in scholarship is not sufficient to warrant interest in the quality of dissertations being produced under the faculty member's own direction or under the direction of colleagues, interests in the survival of the program may itself warrant concern.

External reviewers who make recommendations to state boards regarding program expansion or termination tend to pay particular attention to the quality of dissertations produced by doctoral programs. For example, in a recent review of education doctoral programs in Louisiana, the review team spent the preponderance of its time during the review reading dissertations. The team noted that the quality of dissertations varied markedly and that, in particular, studies often were not adequately grounded in theory. In any case, in its report to the Louisiana Board of Regents the review team seemed to base its major recommendations upon the quality issue:

> The program at UNO [University of New Orleands] deserves more support than it has been receiving. In the judgment of the team, UNO has done the most

> of any institution in the State with what they have. They have certainly been the most efficient. They produce the best dissertations and the single best doctoral program, namely, Special Education... The faculty is underpaid and the team wonders how long they will stay. It is strongly recommended that UNO's doctoral program in education be allowed the same level of resources as obtained at LSU. (p. 87)

Various models for maintaining dissertation quality have been developed. For example, some programs require that advisors must be selected from a pool of persons with widely acknowledged expertise as researchers. Other programs require the Dean of Graduate Studies to supervise every committee. Some programs vest responsibility in the Dean of Graduate Studies for personally certifying that each dissertation meets accepted standards.

An additional model might be used to compliment any of these various other models. Current students themselves might be routinely required to evaluate the dissertations produced by recent cohorts of previous students. The theory of such a model is that even less scholarly doctoral students might be more diligent in their attention to quality if they anticipated future systematic reviews of their work. The review might be an anticipatory check for at least some students who may be sensitive to embarrassment; the process of serving as a reviewer advises current students that their works will be closely scrutinized. Futhermore, doctoral faculty might be stimulated to impose higher standards of quality if they felt that they might be identified with studies that were either systematically best or systematically worst.

The problem with such an endeavor is that numerous studies must be evaluated to provide reasonable comparisons, and numerous raters must be employed so that confidence can be vested in rendered judgments. New cohorts of doctoral students provide a natural pool of reviewers that can be used for these purposes.

The purpose of the present paper is to illustrate some design and analytic choices that might be employed in such a model. Actual data from a small sample (n=9) of ratings are employed to illustrate possibilities. All education dissertations (52) produced during a five year period at one university were rated. A list of dissertation authors by program area is presented in Appendix A.

Design Issues

Typically it will not be feasible to require all raters to read all dissertations. Matrix sampling techniques can be employed to make the rating task manageable. However, a core of dissertations that will be rated by all subjects can be identified so that interrater reliability can be investigated. Generally, the core should be selected to represent the full range of dissertations with respect to quality; this allows the ratings to be more variable and thus estimates of interrater reliability become more accurate. In the present example, 14 dissertations were rated by all raters and the remaining dissertations were typically rated by two or three raters.

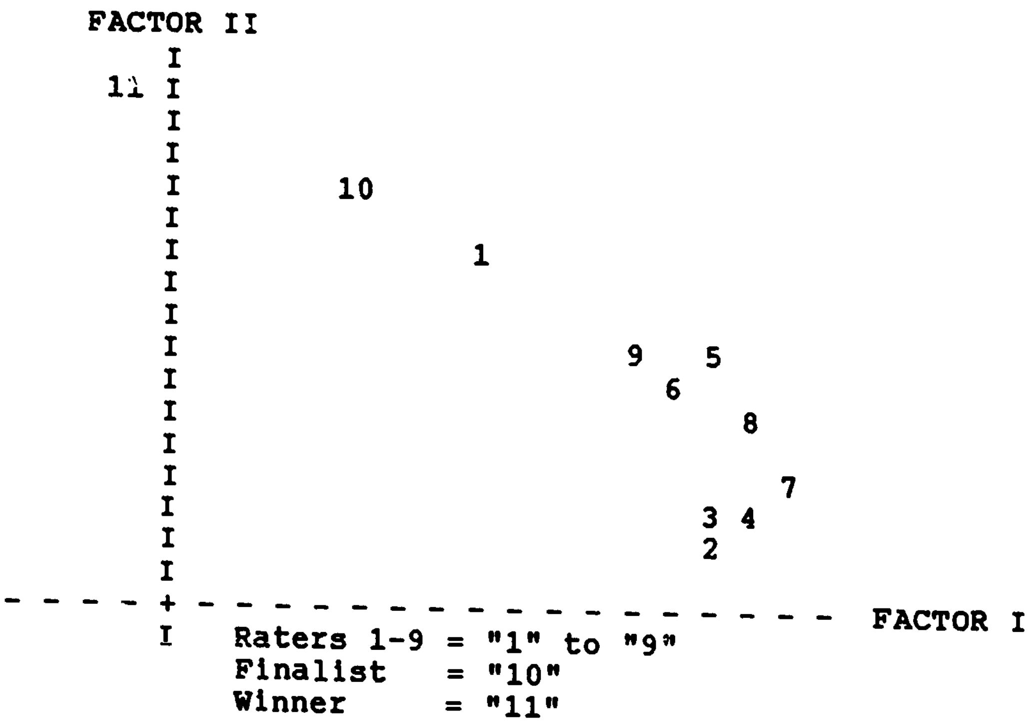
It is important to evaluate the consistency and the validity of the ratings. External ratings data can be employed to investigate concurrent validity. For example, in the present example all dissertations had been reviewed by a doctoral faculty committee for the purpose of designating annual winners and finalists of an outstanding dissertation award. The two variables, finalists and winners, were dummy coded and correlated with each reviewers ratings. These correlation coefficients are reported in Table 1; the coefficients were computed using pairwise deletion of missing cases in the calculation of each coefficient so that each coefficient was based on all available ratings data.

Table 1
Rater Consistency and Validity

	1	2	3	4	5	6	7	8	9	Fin
Rater 1										
2	.46									
3	.40	.66								
4	.56	.74	.87							
5	.71	.79	.84	.89						
6	.57	.63	.69	.84	.86					
7	.71	.82	.85	.85	.92	.76				
8	.74	.85	.65	.85	.91	.81	.92			
9	.64	.68	.68	.72	.75	.81	.77	.84		
Finalist	.45	.37	.41	.51	.57	.63	.44	.52	.57	
Winner	.55	.14	.21	.12	.45	.32	.14	.29	.39	.60

Q-technique factor analysis provides a vehicle for evaluating the interrater consistency and validity of the judgments. The method groups "raters" into clusters based upon the similarities of their judgments. All factors with eigenvalues greater than one were extracted. The eigenvalue of the first factor (6.34) indicated that it accounted for 57.6% of the variance in the correlation matric. Factor II accounted for an additional 13.7% percent of the variance in the matrix. Figure 1 presents the location of the raters in the factor space defined by the varimax rotation of the principal components analysis of the Table 1 matrix.

Figure 1
Raters Arrayed in Factor Space



The second factor appears to be a measurement artifact arising from the analysis of two dummy coded variables consisting of ones and zeros together with analysis of intervally-scaled ratings provided by the raters. Overall, inspection of the Table 1 correlation coefficients and interpretation of the Q-technique eigenvalue for Factor I indicates that the raters judgments were reasonably consistent and valid.

Substantive Interpretations of Ratings

In addition to serving as an educative process, the proposed model can also be used for formative program evaluation purposes, or to address policy questions in service of decision-making. Some selected examples of some analyses may illustrate the possibilities. The analyses use forms of the ratings data, such as the data presented in Table 2. The table presents the mean rating of each dissertation, along with the standard deviation of the ratings of each dissertation.

A rather elementary but important question to ask considers the issue of overall quality of the dissertations at a school and the homogeniety of quality across studies. In the present example the mean of the mean ratings of the 52 studies was 11.1 (SD=4.2). The studies were rated on 19-interval scales to maximize reliability of the ratings, and 19 represented the most favorable rating. The dissertation ratings ranged from 2.89 for one study to 17.11 for the most favorably rated dissertation. The coefficient of skewness (-.25) indicated that ratings were symmetrically distributed about the mean rating. These results suggest that dissertation quality was perceived as being highly variable.

Another policy question of potential interest considers possible systematic variations in dissertation quality across program areas. As noted in Appendix A, four program areas were represented in the example data set. Table 3 presents mean ratings for each program area. The results indicate some tendency toward systematic differences in quality across programs. However, the standard deviation for Program 1 suggests that variability in dissertation quality may be particularly troublesome with respect to this degree program.

Table 2
Descriptive Statistics for Rated Dissertations

Mean	SD
2.889	2.759
3.000	2.784
3.500	2.121
4.500	2.121
5.000	.000
5.000	1.414
5.500	.707
6.000	4.243
6.000	2.828
6.667	5.686
7.500	6.364
8.000	8.485
8.000	.000
8.333	4.509
8.333	4.933
8.500	6.364
9.000	6.245
9.000	.000
9.222	5.696
9.333	5.508
9.500	6.364
9.667	4.726
10.000	5.657
10.000	5.657
10.667	5.033
11.000	1.732
11.500	2.121
11.667	5.686
11.667	4.933
12.500	4.950
12.500	4.950
13.000	1.414
13.500	.707
13.500	2.121
13.667	3.873
14.000	2.828
14.000	3.000
14.333	2.872
14.778	2.539
14.889	3.551
15.333	3.279
15.500	3.536
16.000	3.240
16.222	3.270
16.222	2.539
16.333	4.743
16.333	2.915
16.500	3.536
16.500	2.121
17.000	1.414
17.000	.000
17.111	1.167

Table 3
Mean Ratings Across Programs

		Mean	SD
Program	1	9.82	6.20
	2	9.89	3.79
	3	12.33	3.85
	4	13.39	3.36

A third issue of possible interest considers whether dissertation quality is perceived as improving over time. Fourteen semesters were represented within the pool of rated dissertations in the present example. Table 4 reports descriptive statistics for ratings of dissertations conducted during each of these semesters. Polynomial trend ANOVA identified no noteworthy trends in the Table 4 means.

Table 4
Descriptive Statistics Over Time

Time	Mean	SD
1	10.8	5.7
2	8.4	2.4
3	12.4	3.5
4	10.4	6.3
5	12.7	4.5
6	10.9	3.1
7	11.7	0.0
8	7.7	6.0
9	9.6	7.3
10	10.5	7.1
11	12.8	1.8
12	11.7	0.0
13	12.5	5.6
14	10.7	3.2

A fourth analysis considers whether particular major professors are more likely to be associated with dissertations that are perceived as being of higher quality. Twenty faculty had served as major professors for at least one of the dissertations in the rated pool. The mean ratings for studies associated with given faculty ranged from a low of 4.75 (SD=1.77) for one faculty member to a high of 17.00 (SD=.00) for a faculty member who only served as major professor one time during the five year period under consideration. Results from a one-way analysis of variance of these differences is presented in Table 5. These results suggest that the programs may not have been wisely utilizing their faculty during the time period considered in the study.

Table 5
One-Way ANOVA

Source	SOS	df	MS	Fcalc	Effect Size
Major Professor	584.73	19	30.78	3.10**	64.8%
Error	318.17	32	9.94		
Total	902.91	51			

**$p < .01$

NOTE: Effect size is an r-squared analog indicating the proportion of the SOS total explained by a given effect.

Summary

Dissertation quality may be improved by use of peer review as a quality control mechanism. A small data set involving ratings of 52 dissertations was employed to illustrate possible design and analytic choices that may be helpful in using such a model. Strategies for evaluating rating consistency and validity were illustrated. Analyses to address four types of program policy analysis questions were also illustrated.

Appendix A: Rated Dissertations

	Name	Deg	Dis Abst	U Microfilms
C	FAUST	EdD	41/4056A	81-12,898
A	LANE	PhD	41/4913A	81-12,899
O	LEVITOV	PhD	42/0143A	81-13,580
A	STEEDLEY	EdD	42/0046A	81-13,581
C	ANDERSSON	PhD	42/2993A	81-29,909
S	ARAMBURO	EdD	42/2604A	81-25,880
S	CARLTON	EdD	42/2606A	81-25,881
A	DUMAS	EdD	42/1872A	81-23,012
C	GREENE,J	EdD	42/4816A	82-08,060
O	JEANSONNE	PhD	42/2502A	81-25,882
A	KAPPELMAN	EdD	42/2405A	81-25,884
O	LUMPKIN	PhD	42/2883A	81-26,222
C	MCCULLOCH	PhD	42/2528A	81-25,885
A	MELANCON	EdD	42/1882A	81-23,013
A	MEZA	EdD	42/1882A	81-23,017
A	PHILLIPS	EdD	42/4676A	82-08,010
A	RICHARDSON	PhD	42/2421A	81-25,886
S	SLATE	EdD	42/2628A	81-25,887
A	TAYLOR	PhD	42/2061A	81-23,015
A	BECNEL	EdD	43/2168A	82-27,568
A	DUCOTE	PhD	43/1888A	82-23,262
A	FEOLA	PhD	43/2842A	82-23,263
C	GAUDIN	EdD	43/1914A	82-25,264
A	MILLER,WM	EdD	43/2189A	82-27,570
C	PITTS	PhD	43/1910A	82-23,261
A	QUAIN	PhD		
S	SABRIO	PhD	43/1932A	82-23,267
A	SCAFIDEL	EdD	43/2850A	82-23,268
A	THERIOT	EdD	43/3828A	83-10,891
A	ZEIGLER	PhD	43/2197A	82-27,571
C	BORRELLO	EdD	45/0065A	84-08,136
A	BROOKS	PhD	44/2301A	83-28,108
A	HARRIS	EdD	45/0094A	84-08,137
S	JACKSON	EdD	45/0147A	84-09,292
S	KELLEY	PhD	44/2717B	83-28,209
C	SULLIVAN	EdD	44/2425A	83-28,214
O	TEXIDOR	PhD	44/2369A	83-28,215
O	THOMAS	PhD	44/2682A	83-28,216
S	BANBURY	PhD	45/2062A	84-17,612
A	BARTON	PhD	45/2316A	84-25,266
C	BEALL	PhD	46/2186A	85-07,561
C	CASBERGUE	PhD	46/0110A	85-07,562
O	GARLINGTON	PhD	45/1342A	84-17,613
S	MELVILLE	EdD	46/0123A	85-07,563
S	MILLER,JAY	PhD	46/0128A	85-07,564
A	WILLIAMS	PhD	46/0044A	85-07,566
A	ALEXIS	PhD	46/2200A	85-16,301
A	BUSENBARK	PhD	46/1451A	85-16,302
A	CARTER	PhD	46/1527A	85-16,303
A	GREEN,ROS	PhD	46/1458A	85-16,304
S	KREIG	PhD	46/1902A	85-16,305
S	SISTERHEN	PhD	46/1597A	85-16,307

Program Area Codes
"A"=Administration
"C"=Curriculum
"O"=Counseling
"S"=Special Education