Reviewing the Review: An Assessment of Dissertation Reviewer Feedback Quality

Tara Lehan
Northcentral University, tlehan@ncu.edu

Heather Hussey
Northcentral University

Eva Mika
Northcentral University

Reviewing the Review: An Assessment of Dissertation Reviewer Feedback Quality

Abstract

Throughout the dissertation process, the chair and committee members provide feedback regarding quality to help the doctoral candidate to produce the highest-quality document and become an independent scholar. Nevertheless, results of previous research suggest that overall dissertation quality generally is poor. Because much of the feedback about dissertation quality provided to candidates, especially those in online learning environments, is written, there is an opportunity to assess the quality of that feedback. In this study, a comparative descriptive design was employed using a random sample of 120 dissertation reviews at one online university. Common foundational errors across dissertations and strengths and growth areas in reviewer feedback were noted. Whereas reviewer feedback quality was acceptable overall, there were significant differences across reviewers. Based on the findings, increased discourse, standardization of psychometrically sound measures that assess reviewer feedback quality, and ongoing training for faculty members who review dissertations might be warranted.

Introduction

As an integral part of the peer-review process used by many academic journals, reviewers are charged with identifying foundational flaws and providing useful feedback with the goal of improving quality (Caligiuri & Thomas 2013). Acting as gatekeepers, they play a key role in determining what work is deemed to contribute to the scholarly literature (Caligiuri & Thomas 2013; Min 2014). This process helps authors refine and advance the document and aids in maintaining standards of scientific quality (Onitilo, Engel, Salzman-Scott & Doi 2014).

However, not all reviews are perceived as being equally helpful (Suls & Martin 2009). There appears to be consensus among scholars regarding not only the importance of the peer-review process but the need to improve it (Caligiuri & Thomas 2013; Min 2014; Onitilo et al. 2014; Schoroter, Tite, Hutchings & Black 2006; Suls & Martin 2009; Szekely, Kruger & Krause 2014). According to Caligiuri and Thomas (2013), reviewer comments that are deemed to be the most helpful include those in which reviewers include suggestions for improvement, advice to solve problems, alternate ways to analyze data and feedback regarding the manuscript's contribution to the field. Unfortunately, such comments are uncommon.

In general, there often is inconsistency across reviews in terms of helpfulness, thoroughness and use of evidence versus opinions (Caligiuri & Thomas 2013; Min 2014; Onitilo et al. 2014; Schoroter et al. 2006). Kumar, Johnson and Hardemon (2013) reported that the feedback offered by reviewers frequently is difficult to understand. Szekely and colleagues (2014) suggested that many reviews are biased, inconsistent and sometimes outright wrong.

Because few reviewers are trained to review, or even receive feedback about their reviews, they often do not realise that they are biased (Caligiuri & Thomas 2013; Min 2014). Consequently, there is a need to examine reviewer feedback (Szekely et al. 2014). Just as scholars benefit from feedback on their work, so should reviewers. Snell and Spencer (2005) found that reviewers would appreciate such feedback. Helpful reviewers go beyond identifying problems with the manuscript and offer specific suggestions regarding how to improve the methodology or analyse the data in another way (Caligiuri & Thomas 2013; East, Bitchener & Basturkmen 2012). This process also helps to enhance reviewer accountability and ensure that reviews are constructive and informative on how to move forward.

Whereas much of the research on review quality has involved journal reviewers, feedback from dissertation chairs and committee members about dissertations also warrants scholarly attention. Such feedback is an integral part of doctoral education, as it helps to train doctoral candidates to learn about the writing process, improve their critical-thinking skills and understand the expectations of the academic community (Basturkmen, East & Bitchener 2014; Kumar & Stracke 2007). Many dissertation-committee members state that they can recognise a quality dissertation when they see it, adding that they can describe general characteristics of outstanding, very good, acceptable and unacceptable dissertations (Lovitts 2005). This perspective is consistent with the apprentice model, which is based on the assumption that dissertation advisors can mentor candidates without additional training (Barnes & Austin 2008). Similarly, many faculty members report making holistic decisions about a dissertation versus using some type of rubric or standardised checklist (Lovitts 2005). However, much as with manuscripts submitted to academic journals, a standardised process for document review might improve quality (Lovitts 2005; Onitilo et al. 2014; Ronau 2014).

There is a lack of research on the quality of the feedback given to candidates (Basturkmen et al. 2014; Bitchener & Basturkmen 2010; East et al. 2012), especially online doctoral students, for whom written feedback is especially crucial (Kumar et al. 2013). Inconsistencies in dissertation quality have been noted (Basturkmen et al. 2014; Nelson, Range & Ross 2012). Boote and Beile (2005) found variable quality across dissertations, with overall quality being low. Similarly, many faculty members note that it is uncommon to find an exceptional dissertation (Boote & Beile 2005; Lovitts 2005). Given that dissertation quality commonly is poor and that quality across dissertations is inconsistent, the quality of dissertation-reviewer feedback warrants attention. To address this critical gap in the literature, the current study aimed to examine the quality of reviewer feedback on dissertations at various stages.

Method

Context of study

Although the focus of this study is on the continuous-improvement process as opposed to the specific review process, it is helpful to understand the latter to understand the former. The review process employed in this study was implemented at a completely online university that primarily grants doctoral degrees. The model included a full-time dissertation chair, subject-matter expert (SME) and reviewer who engaged in a single-blind review process. The reviewer served a similar role to that of a journal reviewer, with limited ongoing interaction with either committee members or students beyond milestone reviews. However, dissertation chairs could correspond with reviewers if there were questions about reviewer feedback. Both dissertation chairs and reviewers had demonstrated expertise in both quantitative and qualitative research methods. In addition, they received ongoing training based on findings of continuous-improvement initiatives.

Candidates completed their dissertation in three phases: concept paper (CP), dissertation proposal (DP) and dissertation manuscript (DM). At each stage, once the chair, SME and candidate believed that the document was of sufficient quality to pass onto the next phase, the chair submitted it for review by an academic reviewer. Upon receiving the document, the reviewer could either choose to give it a full review or defer it because the document was of such poor quality that it was not ready for a full review. Reviewers were expected to use the defer disposition when a CP or a DP either had a foundational error that affected all other components of the document, such as a poorly articulated or unsubstantiated problem statement, or contained numerous foundational errors that seriously affected the quality of the work or violated some rule of research. Reviewers did not have the option to use the defer disposition at the DM stage or after one full review had already occurred at the CP or DP stage.

For CPs and DPs that did not have a foundational error, reviewers had the option of using either a resubmit or a final-feedback disposition. They were told that final feedback was only to be given in a first full review when no foundational errors existed, although final feedback had to be given at the second full review. Regardless of the disposition, reviewers were expected to go through the document, highlight any issues and offer suggestions, reflective questions and resources on how the noted issues might be addressed. Each document was only given two full reviews (not including deferrals).

Under the model employed by this university, reviewers had a limited amount of time (approximately two hours) to devote to each review. The prescribed time limit was based on the intended focus to fine-tune the document. The assumption was that the documents submitted for

review were free of foundational errors, so two hours should have been sufficient to provide substantive feedback in most cases.

Population

The population comprised all 818 dissertation reviews completed in 2014 between January 1 and May 5 to include those with a defer, resubmit or final-feedback disposition. This included theoretical (PhD) and applied doctoral dissertations from the four schools within the university (Education, Marriage and Family Sciences, Psychology and Business). These dissertations were all reviewed by one of six reviewers whose sole responsibility at the university was to provide feedback on the quality of dissertations and provide a disposition.

Sample

Of these 818 reviews, 20 were selected for each of the six reviewers (n = 120). Each reviewer had approximately the same number of CPs, DPs and DMs. In the sample, there were 56 CP reviews, 33 DP reviews and 31 DM reviews. This distribution was consistent with that of the larger population, which included 445 reviews of CPs, 227 reviews of DPs and 146 reviews of DMs completed from January 1 through May 5, 2014. In terms of disposition, 26 milestone documents were deferred, 44 required resubmission and 50 contained final feedback. This distribution was consistent with that of the larger population, which included 155 defer, 263 resubmit and 319 final feedback dispositions given from January 1 through May 5, 2014.

Instrument

The instrument used in this study was developed in alignment with the three dissertation milestone documents (CP, DP and DM) submitted by the chair for academic review. The items on the instrument aligned with the dissertation templates and guidebooks provided to doctoral candidates and their chairs, and encompassed all foundational components (feasibility of problem statement; alignment of problem, purpose, and methods; quality of data collection and analysis; and evaluation and implication of findings). The items also reflected the deferral criteria that reviewers used to assess a dissertation milestone document's foundational components. The three-point Likert-type scale in the instrument consisted of Needs Improvement (reviewer did not detect shortcomings), Acceptable (reviewer detected shortcoming and provided general advice), and Exceptional (reviewer identified shortcoming and provided specific feedback, recommendations, and resources), which reflected both the basic quality-assurance function of the review process and the added function of educating the doctoral candidate. If no foundational error was present, the raters were instructed to select Not Applicable. In addition, they were asked to give an overall rating of Sufficient/Acceptable or Insufficient/Unacceptable. Prior to its use in this study, this instrument was piloted, and revisions were made based on the results.

Procedure

Given the purpose of this study, a comparative descriptive design was employed. Several steps were taken to enhance validity and reliability. Only one review per doctoral candidate was included in the sample to ensure that observations were independent. To begin, every possible combination of school (Business, Marriage and Family Sciences, Education, Psychology), degree type (applied, PhD), stage (CP1, CP2, DP1, DP2, DM1, DM2) and disposition (deferred, resubmit, final) was generated using Excel. All identifying information, including the names and contact information of the candidate, chair and SME, was removed from the milestone documents. For the

first two rounds of selection, the dissertation coordinator randomly selected one review to represent each possible combination when at least one existed; however, there was not always a review in the population for each combination. In particular, there were very few reviews in 2014 for documents written by candidates in the School of Marriage and Family Sciences. For the first two rounds, after stratifying the sample by school/degree type/stage/disposition, the dissertation coordinator generated random numbers for each review and selected every tenth one to be rated. For the subsequent rounds, after it was ensured that all possible combinations were represented by at least one review, the focus shifted to having an even number of reviews per reviewer. Therefore, the dissertation coordinator stratified the sample by reviewer and randomly selected reviews in a similar manner to the first two rounds.

Three research directors within the Graduate School served as blind raters of the reviewers' feedback. They did not know who the candidate, chair or SME were while completing their ratings. Further, raters received two trainings on how to use the instrument consistently, one before and one after the first round of ratings.

Results

Inter-rater reliability

Three independent raters used the developed scale to assess the quality of reviewer feedback. To determine the level of agreement between raters, reviewer feedback in 25 documents was rated by a fourth independent rater. Of those 25 sets of ratings, 16 had ratings that were reliable in terms of both item ratings (user missing=N/A, Needs Improvement=1, Acceptable=2, Exceptional=3) and overall ratings (Sufficient/Acceptable=1, Insufficient/Unacceptable=0). For the documents to be included in the sample, the overall ratings had to be the same. In addition, at least 50% of the item ratings had to be exactly the same. Given that the scale used was ordinal, but included a nominal rating (N/A), commonly used reliability coefficients would be misleading. Although calculating percentage of exact agreement is an underestimate of inter-rater reliability, this strategy was used. Percentage agreement on all item ratings per review ranged from 50% to 100%, with the average being 66.94%.

Descriptive statistics

Most common foundational errors. Several foundational errors were present in most of the documents examined, which means that the candidate, SME and chair all failed to recognise and address them prior to submitting the documents for review. In some cases, the reviewer also did not highlight one or more foundational errors. The following section includes a description of the most common foundational errors in the reviewed documents, including both those generally highlighted and those generally not highlighted by reviewers.

The most common foundational errors in CPs and DPs were also the ones that were frequently highlighted by reviewers. They included a lack of alignment of core components (present in 54out of 56 CPs and 29 out of 33 DPs) and lack of articulation and substantiation of the problem statement (present in 49 out of 56 CPs and 28 out of 33 DPs). The most common foundational errors in the DMs were frequently highlighted by reviewers. They included insufficient explication of a rationale for the design, including use of seminal authors (present in 21 out of 31 DMs); improper presentation and organisation of results (present in 20 out of 31 DMs); and issues with recommendations (present in 20 out of 31 DMs).

Table 1. Foundational errors generally highlighted in reviewer feedback

Foundational Error	No. of Documents with Error	No. of Documents with Acceptable Comment	No. of Documents with Exceptional Comment	No. of Documents with No Comment	% of Documents with Error Correctly Highlighted
Of 56 CPs					
Lack of articulation & substantiation of problem statement	49	25	10	14	71.4
Lack of feasibility & relevance of topic	22	14	1	7	68.2
Lack of alignment of core components	54	27	8	19	64.8
Of the 33 DPs					
Lack of articulation & substantiation of PS	28	13	5	10	64.3
Inaccurately operationalised variables/ constructs	18	9	2	7	61.1
Lack of alignment of core components	29	14	3	12	58.6
Of the 31 DMs					
Improper presentation & organisation of results	20	12	4	4	80
Issues with recommenda-tions	20	13	0	7	65

[a]An acceptable comment is one in which the specific foundational error was highlighted with general advice about how to move forward.
[b]An exceptional comment is one in which the specific foundational error was highlighted with specific advice about how to move forward and recommendations/resources.

Foundational errors frequently highlighted by reviewers. To determine which foundational errors reviewers frequently highlighted in general, measures of central tendency for each item were examined. Those with a median of 2.0 (sample median) or greater and a mode of 2 or greater were included in the lists, as 2 corresponded with an Acceptable rating. Table 1 shows the foundational errors that were generally highlighted in reviewer feedback, the number of documents that contained that error, the number of documents in which the reviewer highlighted the error with general as well as specific advice, the number of documents in which the reviewer did not highlight the error and the percentage of documents containing that error in which the reviewer at least highlighted it and provided general advice about how to move forward. As the table shows, reviewers generally highlighted the two same foundational errors (lack of alignment of core components and lack of articulation and substantiation of the problem statement) at the CP and the DP stage. The only commonly highlighted foundational error for which no reviewer provided specific advice and recommendations/resources related to issues with the recommendations in the DM.

Foundational errors frequently not highlighted by reviewers. To determine which foundational errors reviewers frequently did not highlight, the measures of central tendency for each item were examined. Those with a median lower than 2.0 (sample median) and mode lower than 2 were included in the lists, as 2 corresponded with an Acceptable rating. Table 2 shows the foundational errors that were generally not highlighted in reviewer feedback, the number of documents that contained that error, the number of documents in which the reviewer highlighted the error with general as well as specific advice, the number of documents in which the reviewer did not highlight the error and the percentage of documents containing that error in which the reviewer at least highlighted it and provided general advice about how to move forward. None foundational errors that reviewers generally did not highlight at the CP stage were generally highlighted at the DP stage (lack of an explication of the rationale for the design, potential ethical issues/breaches and lack of synthesis and critical analysis in the brief literature review). In addition, two of those foundational errors (lack of an explanation of the rationale for the design and potential ethical issues/breaches) were generally not highlighted at all three stages. Two of the foundational errors that were generally not highlighted by reviewers in DPs were also generally not highlighted in DMs (lack of alignment across chapters/core components and issues with the sampling protocol). Further, reviewers infrequently provided exceptional feedback for the foundational errors that were generally not highlighted. Notably, whereas a majority (21 of 31) of the DMs lacked a sufficient explanation of the rationale for the selected design, in only one document did the reviewer highlight it.

Overall ratings. Of the 120 reviews in the sample, 31 (25.8%) received an overall rating of Sufficient/Acceptable. That is, at a minimum, the reviewer highlighted every foundational error in the document and provided general advice (Acceptable). In some cases, the reviewer also provided specific advice within the context of the study as well as recommendations and resources when appropriate (Exceptional). As previously stated, even if a reviewer did not highlight just one foundational error, the review had to be rated Insufficient/Unacceptable overall. Further, even if the review was exceptional, if the disposition was not appropriate, the review had to be rated Insufficient/Unacceptable overall. In two reviews, the item ratings all met or exceeded 2 (Acceptable), but the reviews were deemed to be Insufficient/Unacceptable overall because the reviewer failed to highlight just one foundational error. In four reviews, the item ratings met or exceeded 2, but the reviews were deemed to be Insufficient/Unacceptable overall because they should have been deferred due to the number and severity of the foundational errors.

Table 2. Foundational errors generally not highlighted in reviewer feedback

Foundational Error	No. of Documents with Error	No. of Documents with Acceptable Comment	No. of Documents with Exceptional Comment	No. of Documents with No Comment	% of Documents with Error Correctly Highlighted
Of 56 CPs					
Lack of explanation of rationale for design	45	16	6	23	48.8
Potential ethical issues/breaches	10	4	0	6	40
Lack of synthesis and critical analysis in literature review	32	7	2	23	28.1
Of the 33 DPs					
Lack of feasibility and relevance of topic	12	5	1	6	50
Inappropriate level of detail provided in the methods section	20	4	6	10	50
Issues with sampling protocol	25	10	2	13	48
Lack of explanation of rationale for design	22	6	4	12	45.5
Potential ethical issues/breaches	13	4	1	8	38.5
Inappropriate theoretical/ conceptual framework	15	4	1	9	33.3
Lack of synthesis and critical analysis in literature review	18	4	1	13	27.8
Of the 31 DMs					
Insufficient comparison of study findings to existing literature	25	10	2	13	48

Lack of alignment across chapters and core components	18	7	1	10	44.4
Lack of clarity and integration of conclusions	17	8	0	9	47.1
Statistical analysis and/or analytical strategy that is not aligned with hypotheses and/or research questions	18	7	0	11	38.9
Insufficient discussion of limitations	18	7	0	11	38.9
Presentation of findings that is unrelated to the conceptual/ theoretical framework	18	7	0	11	38.9
Potential ethical issues/breaches	3	1	0	2	33.3
Issues with sampling protocol	15	3	0	12	20
No pilot studies/ field tests for instruments/ protocols	14	2	0	12	14.3
Lack of explanation of rationale for design	21	1	0	20	4.8

[a]An acceptable comment is one in which the specific foundational error was highlighted with general advice about how to move forward.
[b]An exceptional comment is one in which the specific foundational error was highlighted with specific advice about how to move forward and recommendations/resources.

Differences in the number of Sufficient/Acceptable overall ratings were noted for each milestone stage. Of the 56 CPs in the sample, 19 (33.9%) were rated as Sufficient/Acceptable overall. Of the 33 DPs in the sample, 11 (33.3%) were rated as Sufficient/Acceptable overall. However, of the 31 DMs in the sample, only 1 (3.2%) was rated as Sufficient/Acceptable overall.

There was also a clear trend in terms of the number of Sufficient/Acceptable overall ratings across reviewers (Table 3), with most (77.4%) of the Sufficient/Acceptable reviews being associated with three reviewers. Of the 31 documents that received an overall rating of Sufficient/Acceptable, one reviewer did not have any. On the other hand, one reviewer had 11 reviews that were deemed to be Sufficient/Acceptable overall. This same reviewer had ratings of 3 (Exceptional) on all applicable items on another document, but the review received an overall rating of Insufficient/Unacceptable because the document should have been deferred due to the number and severity of the foundational errors. Similarly, in some cases reviewers had ratings of 3 (Exceptional) on all applicable items, but the review received an overall rating of Insufficient/Unacceptable because the document should have been deferred due to the presence of one or more foundational errors.

Table 3. Number of Insufficient/Unacceptable and Sufficient/Acceptable reviews by reviewer

Reviewer	No. of Insufficient/ Unacceptable Reviews	No. of Sufficient/ Acceptable Reviews
1	17 (85%)	3 (15%)
2	16 (80%)	4 (20%)
3	9 (45%)	11 (55%)
4	13 (65%)	7 (35%)
5	14 (70%)	6 (30%)
6	20 (100%)	0 (0%)

To determine if there were significant differences across reviewers in terms of the number of documents that received an overall rating of Sufficient/Acceptable, a chi-square test was conducted using the contingency table above. Results showed that the number of Sufficient/Acceptable documents across reviewers was significantly different, χ^2 (5, n = 120) = 18.49, p = .002. Upon review of the standardised residuals and using a critical value of 1.96 (α = .05), it was found that Reviewer 3 had significantly more and Reviewer 6 had significantly fewer reviews deemed to be Sufficient/Acceptable than the other reviewers (as shown by a standardised residual of 2.6 and 2.3, respectively).

Item ratings. Given the ordinal scale of measurement and the positive skewness of the data, the median was the most meaningful measure of central tendency. Across all reviews examined for this project, the median item rating was 2.0 (IQR: 1.0), which corresponded with an Acceptable rating. Because a review could have an overall rating of Insufficient/Unacceptable, despite Acceptable and/or Exceptional item ratings, it was important to examine both item ratings and overall ratings. In addition, the scale was treated as ordinal, as items rated as N/A were coded as user missing data. A Kruskill-Wallis test was employed to determine if there were significant differences across reviewers in terms of median item ratings. Results showed that item ratings differed significantly across reviewers, χ^2 (5, n = 120) = 35.72, p < .001. Given that the overall test yielded significant results, post-hoc tests were conducted using the Mann-Whitney U Test. Because multiple comparisons were made, the a priori alpha level was set at .003. Results showed

that significant differences existed between:
- Reviewer 2 and both Reviewer 3 (U = 80.0, p < .001, r = .38) and Reviewer 4 (U = 94.0, p = .002, r = .28).
- Reviewer 6 and Reviewer 3 (U= 39.0, p <.001, r = .44), Reviewer 4 (U = 56.0, p < .001, r = .40), and Reviewer 5 (U = 75.0, p < .001, r = .34).

For the most part, these results seem to be consistent with those of the chi-square test of overall ratings. Specifically, Reviewer 3 had significantly more reviews with Sufficient/Acceptable overall ratings than the other reviewers and significantly higher item ratings than both Reviewer 2 and Reviewer 6. Reviewer 6 had significantly fewer reviews with Sufficient/Acceptable overall ratings than the other reviewers and had significantly lower item ratings than Reviewers 4 and 5 (in addition to Reviewer 3). Further, it was found that Reviewer 2 had significantly lower item ratings than Reviewer 4 (in addition to Reviewer 3 as stated above).

Discussion

Although the structure and roles associated with dissertation committees can vary across universities, all committee members serve as guides and advisors through offering feedback to doctoral candidates on their dissertations (Bloomberg & Volpe 2012). Nevertheless, little is known about the quality of this feedback, especially that given to online doctoral candidates, for whom written feedback is especially important. If quality feedback is not provided to candidates, they might not produce high-quality dissertations or develop into independent scholars. To address this gap in the literature, the current study involved independent raters' inspecting each dissertation review for any foundational errors that might have been missed and any inappropriate dispositions made by reviewers. Despite the use of a deficit approach, it was found that many aspects of reviewer feedback were acceptable or exceptional. At the same time, several areas for improvement became evident.

Strengths of reviewer feedback

In approximately one-fourth of the reviews, the reviewer highlighted and provided feedback on all foundational errors. Given that the median item rating was 2.0, it seems that the quality of the reviewer feedback was generally acceptable, which is consistent with the findings of previous studies on quality of journal-article reviews (Black, van Rooven, Godlee, Smith & Evans 1998; Shroter et al. 2004). However, existing evaluations of dissertation reviews indicate that it is not uncommon for reviewers to miss key errors (Evans et al. 1993; Shroter et al. 2004). Similarly, in this study, whereas two reviewers did not consistently highlight foundational errors, the other four reviewers did.

In the current study, reviewers in general frequently highlighted the lack of articulation and substantiation of the problem statement and lack of alignment across core components/chapters in CPs and DPs. However, it is not clear why after the candidates had received feedback on them at the CP stage, these foundational errors continued to be common at the DP stage. Similarly, reviewers generally highlighted issues relating to the feasibility and relevance of the dissertation topic to the candidate's degree and discipline at the CP stage. Yet, it is unclear why a document that was not clearly feasible and relevant would even be submitted for review by the chair. It begs the question of how SMEs, chairs, and reviewers can help candidates to ensure that these issues are addressed earlier in the dissertation sequence. It also seemingly highlights the importance of identifying strategies to increase committee-member collaboration for the benefit of the doctoral

candidate's academic development (Lee & Mitchell 2011). Further, it seems that committee members should work both collaboratively and as checks and balances so that dissertation candidates can have the best experience and produce the highest-quality document possible (Cassuto 2012).

In this study, reviewers also generally highlighted inaccurately operationalised variables and/or constructs at the DP stage. Such feedback is critical, as doctoral candidates must be able to explicate exactly what it is they are measuring to conduct sound dissertation research. Further, at the DM stage, reviewers generally commented on the improper presentation and organisation of results. In addition, they frequently noted issues relating to the recommendations for research and practice. This feedback is important, as it is difficult, if not impossible, for candidates to discuss implications for practice and recommendations for future research (as well as present findings in the context of existing research and the selected framework, for that matter) when the results are not properly presented and organised (Bloomberg & Volpe 2012).

Growth areas of reviewer feedback

In the present study, it appears that many milestone documents were submitted for review before they were ready, as evidenced by the presence of foundational errors in many of the documents. Nevertheless, the reviewer also bears responsibility for not highlighting foundational issues. As previously discussed, reviewers were not always successful in accomplishing this goal, which can affect both the quality of the document and candidates' personal and professional development. For example, without a well-articulated and substantiated problem statement and clearly explicated guiding framework, which are the bases of the entire study, it is difficult, if not impossible, for candidates to develop the subsequent components of the dissertation (Ellis & Levy 2008).

Overall, if reviewers did not highlight a foundational error at an earlier stage, they did not highlight it at a later stage, meaning that these errors persisted. This finding shows why it is important to address foundational errors at the earliest stage possible. For example, reviewers generally did not highlight a lack of synthesis and critical analysis in the review of the literature in the CP and the DP. This is a common writing issue among doctoral candidates that needs to be addressed early, as synthesis and analysis help them to identify the gaps in the literature and guide research decisions (Bair & Mader 2013). Similarly, reviewers generally did not highlight potential research ethics issues or breaches at the CP, DP or DM stage. However, this finding may be explained by differing perceptions of the complex and often ambiguous nature of research ethics (Eysenbach & Till 2001) or wariness of only using hunches (Rosenfeld 2010). It could also be that reviewers sometimes highlighted a foundational error at an earlier stage, but not at a later stage. For example, although reviewers often highlighted the issue in CPs, they generally did not point out a lack of feasibility and relevance of the dissertation topic at the DP or DM stages. It is possible that they did not believe that they should provide feedback on these issues at a more advanced stage in the dissertation review process (Cassuto 2012). Similarly, reviewers frequently did not comment on the first three chapters of the DM in this study. Cassuto noted that committee members may hesitate to offer feedback on areas they deem approved by the chair.

Reviewers in this study also missed some critical methodological flaws. For example, they often did not highlight insufficient detail about the proposed methods at the DP stage. Relatedly, reviewers generally did not point out issues with the sampling protocol at either the DP or the DM stages. Further, they frequently did not highlight when there was no pilot study or field test for an instrument or protocol, which is problematic given that many doctoral candidates are novice

researchers (Kwan 2013). Many times, reviewers did not highlight an insufficient explication of a rationale for the selected design, including seminal authors, at the DP or DM stages, despite the finding in this study that this foundational error was common in DMs. It is also not clear why this foundational error was not more common at the DP stage. One possibility is that candidates did not follow through with the research plan presented in the DP (Aceme 2014). Reviewer feedback in some reviews suggested that this was sometimes the case.

In the present study, as evidenced by the findings that reviewers generally highlighted foundational issues in CPs and DPs and that only one DM review was deemed to be Sufficient/Acceptable overall, there seems to have been a breakdown at the DM stage. For example, whereas reviewers generally highlighted the improper presentation and organisation of results and issues relating to recommendations for research and practice in DMs, they frequently did not point out when candidates failed to present findings in relation to the existing literature and the selected framework. According to Gall, Borg and Gall (1996), it is common for researchers to fail to clearly relate the findings of the literature review to their own study. It seems that reviewers of scholarly manuscripts, including dissertations, should be on the lookout for this common foundational error. In this study, reviewers generally did not highlight an insufficient discussion of the study limitations. Moreover, they frequently did not point out a lack of clarity and integration of the conclusions. This is problematic, given that many candidates struggle with writing these sections; therefore, academic programs might include more writing support around these skills (Bair & Mader 2013).

Limitations

The findings of this study should be considered in light of several limitations. The examination of reviewer feedback was limited to relatively few dissertations from one online university with a somewhat unique committee structure and dissertation-review process. The extent to which these findings would generalise to brick-and-mortar universities is unclear. Although a randomisation procedure was used, the degree of representativeness of the reviewer feedback in this sample to all current dissertations at this particular online university and beyond is unknown. Because the study was foundational in nature, it seemed important to limit the focus to the assessment of the quality of reviewer feedback in general before introducing additional variables or constructs. Nevertheless, it is possible that factors that were not included in this study affected the quality of reviewer feedback, including reviewer fatigue and reviewer demographics (Donaldson et al. 2010). Also, only a snapshot of reviewer feedback was examined in this study. It is possible that reviewers did highlight foundational errors during an earlier or later review of a dissertation milestone document that was not included in this study; however, the overall trend across all documents reviewed might suggest otherwise.

Implications for practice in distance higher education

In the present study, the dissertation milestone documents for which reviewers provided feedback generally contained numerous foundational errors that had not been addressed by the doctoral candidate, chair or SME prior to submission for review. It seems that all members of the committee failed to uphold the basic gatekeeping function, which supports the need to develop explicit quality-assurance standards. These findings suggest that the presumption that committee members in online higher-educational environments will recognise quality dissertations without explicit standards may be debatable.

Based on the findings of this study, there seem to be some significant differences in terms of

quality of feedback across dissertation reviewers. As a result, it seems that there might be a need for increased discourse and standardisation of psychometrically sound measures that assess reviewers' feedback quality in online education. Further, faculty members who are tasked with reviewing dissertations might be offered training on common foundational errors in dissertations, the characteristics of high-quality feedback and research methodology/design and statistics. During this training, the importance of addressing foundational errors at the earliest stage possible might be highlighted, as the errors seem to persist in future stages of the dissertation-milestone sequence. In general, these recommendations are consistent with a larger trend away from the apprentice model, which is based upon the assumption that chairs can mentor candidates without additional training, to one that focuses on explicit standards and ongoing training.

Recommendations for future research

Based on the findings and stated limitations of this study, several recommendations for research are offered. Future researchers might investigate the psychometric soundness of instruments used to assess the quality of reviewers' feedback. Further, training for faculty members who are charged with reviewing dissertations might be developed and tested to determine whether it enhances the quality of reviewer feedback. In addition, reviewer feedback from other online universities as well as brick-and-mortar institutions might be examined to determine whether the quality differs significantly. Future researchers might conduct longitudinal research to gain a better understanding of reviewer feedback quality over time and across dissertation milestone stages, as well as the factors that influence reviews. Moreover, in future research, raters might be selected from outside of the institution(s) at which reviewer feedback is being investigated to limit the possibility that their ratings will be influenced by their concern that reviewers might face consequences. Subsequent research might also address the experiences of faculty members who review dissertation milestone documents, as they may hold implicit assumptions that guide their approach to written feedback that should be explicated and discussed.

Conclusions

If foundational errors, such as the ones described in this study, are not highlighted by faculty members who review dissertations, there could be consequences for candidates, committee members and the university. In addition to missing a valuable learning opportunity, candidates might be unable to conduct meaningful research and face delays in time to completion at each stage as errors pile up. For committee members, it is much more challenging to help candidates to develop an acceptable dissertation if any foundational errors escape notice and if documents are approved before they are ready. In addition, a greater number of reviews might be required that contain more-extensive feedback, and thus require greater time and effort from the reviewer. Moreover, if these documents move forward without the chairs, SMEs or reviewers highlighting foundational errors, it can reflect negatively upon the university and potentially affect accreditation. Clearly, it is more efficient and beneficial to all stakeholders for committee members to highlight all foundational errors as they occur, provide specific research advice, offer ideas for next steps and provide links/references to scholarly resources as appropriate.

References

Aceme, N 2014. Postgraduate research methodological flaws detected at the final examination stage: Who is to blame?. *South African Journal of Science,* vol. 110, no. 3/4, pp. 14-20.

Bair, M A & Mader, C M 2013. Academic writing at the graduate level: Improving the curriculum through collaboration. *Journal of University Teaching & Learning Practice,* vol. 10, no. 1, pp. 1-12.

Basturkmen, H, East, M & Bitchener, J 2014. Supervisors' on-script feedback comments on drafts of dissertations: Socialising students into the academic discourse community. *Teaching in Higher Education,* vol. 19, no. 4, pp. 432-445.

Barnes, B J & Austin, A E 2008. The role of doctoral advisors: A look at advising from the advisor's perspective. *Innovations in Higher Education,* vol. 33, no. 5, pp. 297-315.

Bitchener, J & Basturkmen, H 2010. The focus of supervisor written feedback to thesis/ dissertation students. *International Journal of English Studies,* vol. 10, no. 2, pp. 79-97.

Black, N, van Rooven, S, Godlee, F, Smith, R & Evans, S 1998. What makes a good reviewer and a good review for a general medical journal. *JAMA,* vol. 280, no. 3, pp. 231-233.

Bloomberg, L D & Volpe, M 2012. *Completing your qualitative dissertation: A road map from beginning to end* (2nd ed.), Sage, Los Angeles.

Boote, D N & Beile, P 2005. Scholars before researchers: On the centrality of the dissertation literature review in research preparation. *Educational Researcher,* vol. 34, no. 6, pp. 3-15.

Caligiuri, P & Thomas, D C 2013. From the editors: How to write a high-quality review. *Journal of International Business Studies,* vol. 44, pp. 547-553.

Cassuto, L 2012. The adviser and the committee. Viewed at http://chronicle.com/ article/The-Adviserthe-Committee/132841/.

Donaldson, M R, Hasler, C T, Hanson, K C, Clark, T D, Hinch, S G & Cooke, SJ 2010. Injecting youth into peer-review to increase its sustainability: A case study of ecology journals. *Ideas in Ecology and Evolution,* vol. 3, pp. 1-7.

East, M, Bitchener, J & Basturkmen, H 2012. What constitutes effective feedback to postgraduate research students? The students' perspective. *Journal of University Teaching & Learning Practice,* vol. 9, no. 2, pp. 1-16.

Ellis, T J, & Levy, Y 2008. Framework for problem-based research: A guide for novice researchers on the development of a research-worthy problem. *The International Journal of an Emerging Transdiscipline,* vol. 11, pp. 17-33.

Eysenbach, G, & Till, J E 2001. Ethical issues in qualitative research on internet communities. *BMJ,* vol. 323, no. 7321, pp. 1103-1105.

Gall, M D, Borg, W R & Gall, J P 1996. *Education research: An introduction* (6th ed.), Longman, White Plains, NY.

Kumar, S, Johnson, M & Hardemon, T 2013. Dissertations at a distance: Students' perceptions of online mentoring in a doctoral program. *International Journal of E-Learning & Distance Education,* vol. 27, no. 1.

Kumar, V, & Stracke, E 2007. An analysis of written feedback on a PhD thesis. *Teaching in Higher Education,* vol. 12, no. 4, pp. 461-470.

Kwan, B S C 2013. Facilitating novice researchers in project publishing during the doctoral years and beyond: A Hong Kong-based study. *Studies in Higher Education,* vol. 38, no. 2, pp. 207-225.

Lee, T W, & Mitchell, T R 2011. Working in research teams: Lessons from personal experiences. *Management and Organization Review,* vol. 7, no. 3, pp. 461-469.

Lepak, D 2009. Editor's comments: What is good reviewing? *Academy of Management Review,* vol. 34, no. 3, pp. 375-381.

Lovitts, B E 2005. How to grade a dissertation. *Academe,* vol. 91, no. 6, pp. 18-23.

Min, K S 2014. Reviewers are not perfect but could they try harder? *Journal of Business Research,* vol. 67, pp. 1967-1970.

Nelson, J S, Range, L M, & Ross, M B 2012. A checklist to guide graduate students' writing. *International Journal of Teaching and Learning in Higher Education,* vol. 24, no. 3, pp. 376-382.

Onitilo, A A, Engel, J M, Salzman-Scott, S A, & Doi, S A R 2014. A core-item reviewer evaluation (CoRE) system for manuscript peer review. *Accountability in Research,* vol. 21, pp. 109-121.

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 Education Research Journal,* vol. 51, no. 5, pp. 974-1006.

Rosenfeld, R M 2010. How to review journal manuscripts. *Otolaryngology – Head and Neck Surgery*, vol. 142, no. 4, pp. 472-488.

Schroter, S, Black, N, Evans, S, Carpenter, J, Godlee, F, & Smith, R 2004. Effects on quality of peer review: Randomised controlled trial. *BMJ*, vol. 328, no. 7441, pp. 1-5.

Schoroter, S, Tite, L, Hutchings, A, & Black, N 2006. Differences in review quality and recommendations for publication between peer reviewers suggested by authors or editors. *JAMA*, vol. 295, no. 3, pp. 314-317.

Snell, L, & Spencer, J 2005. Reviewers' perceptions of the peer review process for a medical education journal. *Medical Education*, vol. 39, pp. 90-97.

Suls, J, & Martin, R 2009. The air we breathe: A critical look at practices and alternatives in the peer-review process. *Perspectives on Psychological Science*, vol. 4, no. 1, pp. 40-51.

Szekely, T, Kruger, O, & Krause, E T 2014. Errors in science: The role of reviewers. *Trends in Ecology & Evolution,* vol. 29, no. 7, pp. 371-373.