

Consistency and inconsistency in PhD thesis examination

Allyson Holbrook
Sid Bourke
Terry Lovat
Hedy Fairbairn

Centre for the Study of Research Training and Impact, University of Newcastle

This is a mixed methods investigation of consistency in PhD examination. At its core is the quantification of the content and conceptual analysis of examiner reports for 804 Australian theses. First, the level of consistency between what examiners say in their reports and the recommendation they provide for a thesis is explored, followed by an examination of the degree of discrepancy between examiner recommendations and university committee decisions on the theses. Two groups of discrepant recommendations are identified and analysed in depth. Finally the main sources of inconsistency are identified. It was found that the comments of a small minority of examiners were inconsistent with each other or with the committee decision in a significant way. Much more commonly the texts of examiner reports were highly consistent and were closely reflected in the final committee decision.

Introduction

When a PhD thesis is examined, whatever the process, examiners typically require something more from the student, whether it be the correction of typographical errors in the written document or more substantial changes. It is also in the nature of doctoral candidature that there is the expectation that the new researcher can always learn more, and that the thesis is but a step on this path (Mullins & Kiley, 2002). As one commentator notes it is an ‘apprenticeship in the art of discovery’ (Kwiram 2006, p. 141). This explains why most examiners provide some comment that is instructive or formative at the final point of a thesis examination (Holbrook, Bourke, Lovat & Dally, 2004a; Tinkler & Jackson, 2004). Furthermore, there is evidence that the type, extent and tone of this comment reflect the overall evaluation of the qualities of the thesis (Bourke, Hattie & Anderson, 2004). This paper focuses on the standards that examiners apply and how consistently they apply them.

The literature on examination and doctoral thesis quality has mostly emerged from the UK and Australia and on the whole indicates that, except in all but general terms (Morley et al., 2002; Shaw & Green, 2002), doctoral ‘qualities’ and

'standards' have proved very difficult to 'tie down' (Tinkler & Jackson, 2004, p. 8). In the past when there were relatively few candidates destined for scholarly pursuits, thesis quality was not a public or pressing issue, as evidenced by the lack of research and publication on the issue. With the rapid 'massification' of doctoral degrees and the growing need for cross-disciplinary work to tackle emerging social and scientific problems, there has come the realisation that not only is there an absence of benchmarks, but an absence of information about the degree and its evaluation (Morley et al., 2002; Shaw & Green, 2002; Jackson & Tinkler, 2001; Tinkler & Jackson, 2004; Lovitts, 2007). A heavy reliance in Australia and elsewhere is placed on the collective experience of examiners and knowledge of 'peer review' (Hoddell et al., 2002 p. 64; Grabbe, 2003 p. 130; Tinkler & Jackson, 2004 p.119; Holbrook, Bourke, Fairbairn & Lovat, 2007). But what is the correspondence between peer review and thesis examination? Evidence is also emerging from the USA that few students publish in peer-reviewed journals (Nettles & Millett, 2006). For most students, thesis examination constitutes the only instance of independent scholarly review of their complete project.

While supervisors and examiners play a pivotal role in defining and shaping the practices in their disciplines, including how and what candidates need to learn to be successful, there is very little in the literature that explores the connection between expectation, judgement and outcome (Mullins & Kiley 2002; Denicolo, 2003; Powell & McCauley, 2002, 2003). The lack of a formally articulated 'curriculum' in relation to assessment (Gilbert, 2004) prompts concerns about consistency in procedures and judgements (Sloboda & Newstead, 1997; Tinkler & Jackson, 2000; Morley et al., 2002; Shaw & Green, 2002; Lawson et al., 2003; Denicolo, 2003). Although procedures for adjudication where examiners differ have been described (Tinkler & Jackson, 2004; Lawson et al., 2003), until the study on which this paper is based, there has been no sustained analysis of examiner recommendations on the same thesis and the relationships between examiner recommendations and the official committee decision.

Some of the most powerful mythology in doctoral examination derives from stories of substantial differences in judgement between examiners of the same doctoral thesis (Morley et al., 2002). High levels of inconsistency would cast doubt on the reliability and quality of doctoral assessment and, in this current international climate of quality assurance and research quality assessment, there is intensified interest in the outcomes of doctoral research. So, are examiners frequently inconsistent in their judgements as anecdote would have us believe?

The paper deals with three questions:

- How consistent are PhD examiner comments and recommendations across institutions and disciplines?
- If inconsistency is evident, what forms does it take?
- What aspects of the written examiner reports influence the subsequent decision of the committee when it is faced with conflicting recommendations on a thesis?

Most Australian universities (58 per cent) use three external examiners, others use two; the process does not normally involve a viva voce or oral examination. Almost

half of Australian doctoral examiners are drawn from other countries, and the majority of these are located in the USA and the UK. This information has been published elsewhere together with other details related to supervision, and other individual information such as candidate and examiner gender (Bourke et al., 2007a).

In this paper the level of consistency between what examiners say in their reports and the recommendation they provide for a thesis are explored, followed by an examination of the degree of discrepancy between committee decision and examiner recommendation. Two groups of discrepant recommendations are identified and analysed in depth.

Approach

This study arises out of a larger project investigating doctoral examination through the use of examiner reports and recommendations, and candidate information. This paper is concerned only with questions of consistency, but to reach the stage where we could deal with this issue, many other analyses were conducted. The collaborative mixed methods design we employed has been published in full (Holbrook & Bourke, 2004; Holbrook, Bourke, Lovat & Dally, 2004b) and utilises what has been identified in the methods literature as a fully integrated mixed model design (Teddlie & Tashakkori, 2003).

Data

The researchers received ethics approval to access candidate records (with identification removed) and examiner reports and recommendations for up to nine Australian universities. During the period 2003–04 the researchers obtained the support of eight universities in providing a complete data set comprising examiner reports, examiner recommendations, institutional decisions and doctoral candidate enrolment and supervision history on record for their 100 most recent doctoral completions. The most substantial source of data for each candidate is the examiner reports that, in standardised form, range on average between 100 and 150 lines. The text of these was scanned and prepared for entry into *QSR N6* software.

Sample

The 804 theses and 2121 examiner reports included represent all broad fields of study. In order to select institutions that reflect a range of research intensiveness, Australian universities were divided into three categories on the basis of research quantum: high (consisting of 8 universities), medium (14) and low (13). The final sample consisted of three institutions in each of the high and medium categories, and two in the low. Given that low research-intensive universities had fewer PhD candidates, the sample was reasonably representative on this dimension (Bourke et al., 2007a). Care was also taken to ensure representativeness by both geographic area and size of institution.

Analysis

In the larger study we coded and analysed the data over many iterations. We undertook analysis institution by institution, comparing and combining them along the way. The core analyses are informed by and contribute, in turn, to ‘extended’ analyses. The latter explore examiner discourse and meta-themes of power, gender, discipline and ‘doctorateness’ (see Holbrook, Bourke, Lovat & Dally, 2004c; Lovat, Monfries & Morrison, 2004). In them we seek to understand how examination process is translated into action and examination culture (Lovat, Holbrook & Bourke, 2007). These extended analyses are still in progress.

At the heart of this study is the conversion of the core categories to quantitative data. The quantified coded text is then correlated with examiner and committee recommendation and candidate history such as whether or not the candidate took leave, years of candidature, number of supervisors, and so on. The analyses are based on the proportions of total text units coded at 29 different categories, as well as instances of code occurrence and text intersection. Examiner reports were standardised to a particular format so that comparable measures (based on line counts) could be produced for a range of features of the reports including proportions of text units coded by category, and the number, pattern and sequence of instances of coded text. All of the text units associated with the examination of the thesis are coded at least at one node (i.e., coding category). The core coding categories were established after an extended process of trialling to establish category stability and coder consistency; detailed coding notes were generated (see Holbrook & Bourke (2004) for further detail of the process developed and followed). The core coding categories capture all of the characteristics and content of the reports.

The core categories referred to in this paper are

- *examiner and process*: such comments capture examiners’ approaches to examination, the interpretation of their role and their expectations
- *assessable areas covered*: this category captures all comment about the possible outcomes, subject matter and presentation of the thesis under examination—the substantive elements of the thesis and the project at its core, e.g., scope and significance, literature, approach, analysis and reporting, communicative competence, and publications arising from the study
- *dialogic elements*: there are specific features of examiner discourse that reflect on the nature of academic communication. In particular this category identifies the notion of ‘active’ dialogue—engagement with, and consciousness of, communicating personally with the reader(s).
- *evaluative elements*: this category captures all comment that contains evaluation and judgement, including different types of instructive/formative comment as well as text identifying positive and negative summative judgements about various aspects of the thesis and the candidate’s capacity to conduct research.

In order to answer the questions posed in this paper we drew on the quantified core data and examiner recommendation and committee decision.

The examiner recommendation

There is a strong similarity between Australian universities in the way examiners are asked to provide a recommendation. Essentially there are five categories ranging from (1) 'accept the thesis as submitted', through (2) minor correction (invited or required), (3) requiring more substantial correction, (4) requiring the thesis to be revised and resubmitted for further examination, to (5) 'fail', the latter without the possibility of resubmission for further examination. (In some cases there were additional categories that could be readily collapsed into these five categories; in other universities only four categories are used with categories 2 and 3 being collapsed, but none of the universities in our sample used only four categories.) As the examiners act independently, clearly there will be theses for which examiner recommendations vary widely across the five categories. The university committee is then faced with a decision that is typically not straightforward. The content of the written reports is considered closely in determining their decision.

When there are widely different recommendations made by examiners on a thesis, there are at least two types of discrepancies possible. First, an examiner may recommend acceptance or only minor change to a thesis, while other examiners recommend that the thesis be revised and resubmitted, and the university committee may decide that the thesis is to be revised and resubmitted. It is the committee's view that these theses are marginal, at best. In this case it seems that the examiner making the favourable recommendation has not influenced the decision made by the university (categorised as Group 1 theses, see Table 1). Thus Group 1 theses are those where at least one examiner has recommended that the thesis be accepted (perhaps after minor amendment) while the university committee decided the thesis should be revised and resubmitted. Secondly, for Group 2 theses, there is the opposite case of at least one examiner recommending that a thesis be revised and resubmitted or failed while the university committee decided that the thesis be accepted as submitted, perhaps with minor change. It is the committee's view that these theses are, at least, satisfactory. In this case the recommendation of the critical examiner or examiners seems to have been disregarded (see Table 2). Given our earlier work indicated that examiner comments differed markedly between high-quality and marginal theses (Bourke et al., 2005), the subsequent comparisons between the examiner text categories for these two groups were done separately.

Comparisons of the content of examiner reports

Once the relevant theses in the two groups were identified, the contents of the examiner reports were compared using the percentages of text coded at each of 29 subcategories of text codes. The overall length of each report was also considered, making a total of 30 possible detailed points of comparison for examiner reports within the two groups.

The proportions of comment that the 'discrepant' examiners devoted to different aspects of the thesis were compared with the content of the 'consistent' examiner reports in order to determine whether there were any significant

differences in the proportions of comment provided by these two groups of examiners. Given the disparities in the examiner recommendations, it was anticipated that the proportions of comment in the subcategories in the reports of examiners requiring revision or resubmission, or recommending failure ('dissatisfied' examiners) would be significantly different from those of examiners who judged the thesis to be acceptable or to require only minor corrections ('satisfied' examiners).

Results

As shown in Tables 1 and 2, only in 33 cases out of a total of 2121 examiners (i.e., less than 2 per cent) made recommendations that were inconsistent according to our definitions. It is also worth noting that there were slightly higher proportions of theses with inconsistent recommendations in the broad field of study of engineering (9 per cent), arts, humanities and social sciences (7 per cent) and education (6 per cent) compared with health (3 per cent) and science (zero).

For Group 1 (i.e., the theses that were considered marginal), of a total of 64 examiner reports for 22 theses, 24 reports were discrepant. Of the 29 subcategories

Table 1 Committee decision 'Revise & resubmit' where one or more examiners recommended 'Accept as submitted' or 'Minor correction' (n = 22 theses with 24 discrepant examiner reports in italics)

<i>Candidate*</i>	<i>Committee decision</i>	<i>Examiner 1 recommend</i>	<i>Examiner 2 recommend</i>	<i>Examiner 3 recommend</i>
1301	Revise & resubmit	Revise & resubmit	Minor correction	Revise & resubmit
1509	Revise & resubmit	Minor correction	Revise & resubmit	Fail
1708	Revise & resubmit	Revise & resubmit	Revise & resubmit	Accept as submitted
1906	Revise & resubmit	Accept as submitted	Revise & resubmit	Revise & resubmit
4005	Revise & resubmit	Fail	Accept as submitted	Revise & resubmit
4006	Revise & resubmit	Accept as submitted	Revise & resubmit	Revise & resubmit
4014	Revise & resubmit	Accept as submitted	Revise & resubmit	Accept as submitted
5023	Revise & resubmit	Fail	Minor correction	No 3rd examiner
5029	Revise & resubmit	Minor correction	Revise & resubmit	No 3rd examiner
5060	Revise & resubmit	Fail	Fail	Minor correction
6086	Revise & resubmit	Minor correction	Revise & resubmit	Revise & resubmit
6100	Revise & resubmit	Accept as submitted	Fail	Revise & resubmit
7007	Revise & resubmit	Accept as submitted	Accept as submitted	Revise & resubmit
7095	Revise & resubmit	Revise & resubmit	Revise & resubmit	Accept as submitted
8004	Revise & resubmit	Accept as submitted	Revise & resubmit	Revise & resubmit
8020	Revise & resubmit	Revise & resubmit	Minor correction	Major correction
8051	Revise & resubmit	Major correction	Revise & resubmit	Accept as submitted
8056	Revise & resubmit	Major correction	Minor correction	Revise & resubmit
8059	Revise & resubmit	Revise & resubmit	Revise & resubmit	Accept as submitted
8067	Revise & resubmit	Revise & resubmit	Minor correction	Major correction
8073	Revise & resubmit	Accept as submitted	Revise & resubmit	Revise & resubmit
8090	Revise & resubmit	Revise & resubmit	Revise & resubmit	Accept as submitted

* The first digit of the candidate number indicates the university.

Table 2 Committee decision ‘Accept as submitted’ or ‘Minor correction’ where one or more examiners recommended ‘Revise & resubmit’ or ‘Fail’ (n = 11 theses with 13 discrepant examiner reports in italics)

<i>Candidate*</i>	<i>Committee decision</i>	<i>Examiner 1 recommend</i>	<i>Examiner 2 recommend</i>	<i>Examiner 3 recommend</i>
2041	Minor correction	Revise & resubmit	Major correction	Revise & resubmit
2049	Minor correction	Minor correction	Revise & resubmit	Accept as submitted
2080	Minor correction	Revise & resubmit	Accept as submitted	Accept as submitted
2090	Minor correction	Minor correction	Minor correction	Fail
3060	Minor correction	Fail	Minor correction	No 3rd examiner
5008	Minor correction	Minor correction	Fail	Accept as submitted
5025	Minor correction	Fail	Minor correction	Accept as submitted
5026	Minor correction	Fail	Minor correction	Minor correction
7006	Minor correction	Revise & resubmit	Revise & resubmit	Minor correction
8064	Accept as submitted	Accept as submitted	Revise & resubmit	Accept as submitted
8087	Accept as submitted	Revise & resubmit	Accept as submitted	Accept as submitted

* The first digit of the candidate number indicates the university.

of report comment, only 8 differed between the satisfied and dissatisfied examiners. Most categories did not differ between these examiners. Four subcategories that differed related to the assessable areas covered. (i.e., significance and contribution; comments about publications that might arise; literature coverage and analysis and reporting). There was more comment from the satisfied examiners concerning significance and contribution of the thesis and potential publications, and less substantive comment on the content of the literature review and on the analysis and reporting of findings. The other four subcategories that differed related to the evaluative elements of the reports. As would be expected, there was more positive comment (summative and other) from the discrepant examiners who were satisfied with the thesis and less formative instruction and prescriptive comment. Overall, the satisfied examiners wrote much shorter reports than the dissatisfied examiners of the theses that were required to be revised and resubmitted (see Table 3).

For Group 2 (i.e., theses that were considered at least satisfactory), of a total of 32 examiners for 11 theses, 13 reports were discrepant. For these theses, only three of the 29 subcategories in total differed between the satisfied and dissatisfied examiners. The only assessable areas subcategory that differed was ‘publications arising’ where the dissatisfied examiners wrote less. There were two evaluative elements that differed: both summative and other positive judgements where the examiners who were dissatisfied wrote less (see Table 3).

It seems that, contrary to expectations, the quantitative analysis of the report content for both groups revealed that there were few differences between the reports of satisfied and dissatisfied examiners for theses where there was at least one inconsistent examiner report. The overall lack of significant differences in the comparison of reports from satisfied and dissatisfied examiners of the same thesis suggests that, although these groups of examiners made

Table 3 Comparisons of comment by examiners with whom the committee agreed and disagreed (discrepant examiners): percentages of text where there was a significant difference ($p < .05$) between the two groups of examiners

<i>Text sub-category code & description</i>	<i>Satisfactory theses (Committee decision)</i>		<i>Marginal theses (Committee decision)</i>	
	<i>Examiners agree (%)</i>	<i>Discrepant examiners (%)</i>	<i>Examiners agree (%)</i>	<i>Discrepant examiners (%)</i>
Significance and contribution			5.7	11.1
Publications arising	4.0	0.1	0.4	4.9
Literature coverage			11.6	5.8
Analysis & reporting			43.7	28.1
Summative positive comment	16.0	3.0	3.3	24.0
Formative instruction			33.3	16.4
Prescriptive comment			18.1	6.4
Positive other judgement	13.1	1.3	2.8	9.1
Text units: report length (no. of lines)			211	109

substantially different recommendations, the content of their reports was generally synchronous.

Cases with a major inconsistency between the committee and the majority of examiners

There were four thesis examinations where the committee decision was inconsistent with the recommendations of a majority of examiners. These instances exhibit the extremes of inconsistency between the committee and examiners. Although clearly a rare occurrence, these cases have been selected for individual description of individual examiner report content because, in looking at the extreme cases, we hope to understand more about the examination process.

Marginal theses We first turn to reports on two of these where two examiners ticked the box ‘accept as submitted’ but the committee agreed with a third examiner who ticked the ‘revise and resubmit’ box (see Table 1). For candidate 4014, when the text of examiner reports is considered, we find agreement in the content of the ‘revise and resubmit’ report (Examiner 2) and one of the ‘accept’ reports (Examiner 3). Both referred very centrally to a lack of depth and analytical strength. The examiner who recommended the thesis be revised and resubmitted produced a detailed and authoritative report that demonstrated how the level of originality and contribution of the thesis was questionable. Moreover, the other examiner (Examiner 1), who also recommended ‘accept’ provided some very tentative statements about the substantiveness of the work, for example noting some incoherence, over-generalisation and lack of understanding of important points.

For candidate 7007, the two examiners who recommended ‘accept’ (Examiners 1 and 2) presented the type of positive summative comment that is

typical of reports on high-quality theses. The reports are of average length and are complimentary about all the key aspects of the thesis. They note originality, effectiveness of the positioning of the study within the field, effectiveness of the methods and the care with which they are detailed, and also the up-to-date and competently executed treatment of the literature. The other examiner who recommended 'revise and resubmit' (Examiner 3) was strongly critical of every aspect of the study, including the fact that the candidate has misread Examiner 3's own work. It is a very aggressive, detailed and authoritative report. The other two reports by comparison provide fewer specifics and virtually no clear 'warrants' of examiner expertise.

Satisfactory theses There were two cases where the committee decided the theses required only 'minor correction' although examiner recommendations were much harsher, being either 'revise and resubmit' or 'fail' (see Table 2).

The committee decision for Candidate 7006 was to invite corrections despite the recommendations of two examiners that the thesis should be revised and resubmitted (Examiners 1 and 2). The committee followed the 'minor correction' recommendation of Examiner 3 but all three examiners found the thesis lacked interpretative and analytical treatment of the material. One examiner commented that a 'compendium does not equate to a doctoral thesis' (Examiner 2). One possible explanation for the recommendation of Examiner 3 and its acceptance by the committee was a greater willingness to accept an unconventional thesis—one examiner acknowledging that it had been a difficult thesis to examine because of its non-traditional form—but even Examiner 3 expressed a desire to see the theoretical section expanded and more rigorously argued and a greater degree of coherence to be given to the work.

Candidate 2041 presents a truly anomalous case where the committee decision to invite corrections is inconsistent with all three examiners' reports. Two of the reports recommended revise and resubmit (Examiners 1 and 3) and the third report (Examiner 2) recommended major corrections. All three were scathing about the literary quality of the thesis and what they saw as serious problems with editing and presentation. All examiners provided pages of corrections extending well beyond a list of typographical errors. Two examiners were highly critical of shortcomings and limitations in the experimental methodology: one commented that this constituted a major limitation of the study. While there were a few positive comments, such as the contribution the thesis makes to the field, such comments were tentative and seemed to be insufficient to outweigh the advice of all three examiners that the thesis was fraught with content, methodological and presentation errors. Why the committee overruled the expert opinion of all three examiners and allowed the work to stand without any required corrections is inexplicable.

Discussion of results

The paper has focused on 'consistency' in examination process and is one of the first studies to undertake a sustained analysis of examiner recommendations on

the same thesis, and of the relationships between examiner recommendations and the official committee decision. For the same thesis it can be anticipated that inconsistency may occur between the recommendations made by examiners and also between the examiners and the committee decision. We found an extremely high level of consistency in examiner recommendations. Only 33 of the 804 theses (4 per cent) had one or more discrepant examiner reports according to our definition. Of the 2121 examiners' reports, only 37 examiners (less than 2 per cent) showed a marked discrepancy from the other examiners and the committee. It is also worth noting that there were slightly higher proportions of theses with inconsistent recommendations in the fields of arts, humanities and social sciences, education and engineering compared with health and science, and that one institution showed more discrepancy than the others. A total of four cases showed a discrepancy between examiner recommendations and the university committee decision.

Where inconsistencies were detected between examiners, there was at least one discrepant examiner for 11 satisfactory theses, and 22 for very problematic, that is, marginal theses. Even in this very small number of cases, the language of examiners and the content of their written reports were generally synchronous even if their recommendations were discrepant. In providing guidance for PhD supervisors and candidates, we have noted this previously, and have suggested that the written report is one of the strengths of the examination process, together with the independence of examiners (Bourke et al., 2007b, pp. 239–240).

With the exception of the four cases showing major inconsistency, the committee that adjudicated on the differences between examiners drew on the examiner comments rather than the inconsistent recommendations, and the official outcome indicated this. Of the four cases, only one is inexplicable. In the other three cases at least two of the examiners commented similarly in their reports, so there was majority agreement in the text even if the recommendations were anomalous. The one notable exception is by its very nature one of those that goes to reinforce the 'myth' that the process is faulty, whereas the many more that are not discrepant suggest that the multiple reporting mechanism proves to be extremely robust in facilitating the achievement of equitable outcomes. The one case in 804 theses that does seem inexplicable perhaps should not be asked to carry too much weight in the debate about examiners' consistency. This notwithstanding, a note of concern needs to be raised in relation to all four cases where the examiners and committee did differ considerably. From the point of view of assessment ethics, equity and fairness, this situation is a worst-case scenario and highlights the importance of there being in place a clear code of practice and transparent procedures.

There is no mandated curriculum for the PhD in Australia or elsewhere; thesis examination is as subjective as each thesis is unique and yet the findings of this study demonstrate the innate robustness of the 'invisible' doctoral curriculum and evidence of consistently applied standards. Kwiram (2006) in an essay produced for the Carnegie volume on the future of the doctorate in the USA, notes that while there are differences in expectations, quality and performance across candidates,

faculties, disciplines, departments and nations 'there seems to be a tacit understanding of what constitutes a well-prepared Ph.D. student', and that 'in the complete absence any central repository of rules or a cosmic accrediting agency' there is 'extraordinary stability' (p. 142). Drawing on the empirical evidence reported here for Australia, we concur. The consistency between examiners and the stability in what they are looking for are remarkable and present some exciting and heartening possibilities. In the first instance this finding has implications for cross-disciplinary thesis projects. There has been concern expressed in many forums including the Carnegie Initiative on the Doctorate (Golde et al., 2006) that there are significant disciplinary impediments to cross-disciplinary work being undertaken because of the 'silo-like' nature of the disciplines but, in fundamental ways, thesis examiners are privileging the same generic qualities in their reports. This suggests there is a solid foundation of consistency, agreement and translatability across disciplines to build on. In the second instance, the assessment discourse of examiners shows they share the same expectations about thesis quality and that, in Australia, there are processes in place (e.g., independent examiner reports that are read by committee) that ensure these understandings are considered when making judgements. The examination process enables examiners to set the standards for PhD theses and the examiners seem to be successful at doing this. What this suggests is that there is no reason to try to write standards from first principles or impose them from without, because they can be reliably identified from what examiners say in their extensive reports. At least one of the outcomes of our work that applies examiner commentary to notions of thesis quality would be the development of a set of stable indicators that allows the academic community and others to distinguish between theses of threshold quality and those of higher and highest quality (see also Lovitts, 2007). As a consequence, a framework for thesis quality could be articulated and used to improve learning processes and dispel the mystery that has prevailed for so long. This is currently being explored by the authors.

Keywords

doctoral examination
higher education

examiner consistency
graduate education

thesis assessment

References

- Bourke, S., Hattie, J., & Anderson, L. (2004) Predicting examiner recommendations on PhD theses. *International Journal of Educational Research*, 41(2), 178–194.
- Bourke, S., Holbrook, A. & Lovat, T. (2005, June). *Using examiner reports to identify quality in PhD theses*. Paper presented at the AARE Focus Conference: Quality in educational research, Cairns. Retrieved 17 January 2008 from <http://www.aare.edu.au/05papc/bo05011y.pdf>
- Bourke, S., Holbrook, A., & Lovat, T. (2007a). *Relationships of PhD candidate, candidature and examination characteristics with thesis outcomes*. Paper presented at the Annual Conference of the AARE, Adelaide, November 2006. Retrieved 17 January 2008 from <http://www.aare.edu.au/06pap/bou06655.pdf>

- Bourke, S., Holbrook, A. & Lovat, T. (2007b). Examiners and examination outcomes. In C. Denholm & T. Evans (Eds), *Supervising doctorates downunder*. Melbourne: ACER Press.
- Denicolo, P. (2003). Assessing the PhD: A constructive view of criteria. *Quality Assurance in Education*, 11(2), 84–91.
- Gilbert, R. (2004). A framework for evaluating the doctoral curriculum. *Assessment and Evaluation in Higher Education*, 29(3), 299–309.
- Golde, C., Walker, G., & Associates (2006). *Envisioning the future of doctoral education: Preparing stewards of the discipline*. Carnegie Essays on the Doctorate. San Francisco: Jossey-Bass.
- Grabbe, L. (2003). The trials of being a PhD external examiner. *Quality Assurance in Education*, 11(2), 128–133.
- Hoddell, S., Street, D., & Wildblood, H. (2002). Doctorates: converging or diverging patterns of provision. *Quality assurance in education* 10(2), 16–70.
- Holbrook, A., Bourke, S., Fairbairn, H., & Lovat, T. (2007). Examiner comment on the literature review in PhD theses. *Studies in Higher Education*, 32(3) ((in press: accepted June 14, 2006).
- Holbrook, A., & Bourke, S. (2004). An investigation of PhD examination outcome in Australia using a mixed method approach. *Australian Journal of Educational and Developmental Psychology*, 4, 153–169.
- Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004a) Investigating PhD thesis examination reports. *International Journal of Educational Research*, 41(2), 178–194.
- Holbrook, A. Bourke, S., Lovat, T., & Dally, K. (2004b). Qualities and characteristics in the written reports of doctoral thesis examiners. *Australian Journal of Educational and Developmental Psychology*, 4, 126–145. Retrieved 17 January 2008 from http://www.newcastle.edu.au/group/ajedp/Archive/Volume_4/v4-holbrook-et-al.pdf
- Holbrook, A., Bourke, S., Lovat, T., & Dally, K. (2004c). PhD theses at the margin: Examiner comment of re-examined theses. *Melbourne Studies in Education*, 45(1), 89–115.
- Jackson, C., & Tinkler, P. (2001). Back to basics: a consideration of the purposes of the Ph.D. *Assessment and Evaluation in Higher Education*, 26(4), 355–366.
- Lawson, A., Marsh, H., & Tansley, T. (2003). Examining the examiners. *Australian Universities Review*, 46(1), 32–36.
- Kwiram, A. (2006). Time for Reform? In C. Golde, G. Walker and Associates (Eds), *Envisioning the future of doctoral education: Preparing stewards of the discipline*. Carnegie Essays on the Doctorate. San Francisco: Jossey-Bass.
- Lovat, T., Monfries, M., & Morrison, K. (2004). Ways of knowing and power discourse in doctoral examination. *International Journal of Educational Research*, 41(2), 163–177.
- Lovat, T., Holbrook, A., & Bourke, S., (2007). Ways of knowing in doctoral examination: How well is the doctoral regime? *Educational Research Review*. <http://dx.doi.org/10.2016/j.edurev.2007.06.002>
- Lovitts, B. (2007). *Making the implicit explicit: Creating performance expectations for the dissertation*. Sterling, VA: Stylus Publishing.
- Morley, L., Leonard, D., & David, M. (2002). Variations in Vivas: quality and equality in British PhD assessments. *Studies in Higher Education*, 27(3), 263–272.
- Mullins, G., & Kiley, M. (2002). ‘It’s a Ph.D., not a Nobel Prize’: how experienced examiners assess research theses. *Studies in Higher Education*, 27(4), 369–386.
- Nettles, M., & Millett, C. (2006). *Three magic letters: Getting to Ph.D.* Baltimore, MD: Johns Hopkins University Press.
- Powell, S., & McCauley, C. (2002). Research degree examining—common principles and divergent practices. *Quality Assurance in Education*, 10(2), 104–115.

- Powell, S., & McCauley, C. (2003). The process of examining research degrees: some issues of quality. *Quality Assurance in Education*, 11(2), 73–83.
- Shaw, M., & Green, D. (2002). Benchmarking the Ph.D.—a tentative beginning. *Quality Assurance in Education*, 10(2), 116–124.
- Sloboda, J. A., & Newstead, S. E. (1997, September). Guidelines for examiners: An evaluation of impact. *Psychologist*, 407–410.
- Teddlie, C., & Tashakkori, A. (2003). Major issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (Eds), *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage Publications.
- Tinkler, P., & Jackson, C. (2000). Examining the doctorate: Institutional policy and the Ph.D. examination process in Britain. *Studies in Higher Education*, 25(2), 67–180.
- Tinkler, P., & Jackson, C. (2004). *The doctoral examination process: A handbook for students, examiners and supervisors*. Berkshire, England: Society for Research into Higher Education and Open University Press.

Acknowledgements

The research reported here was supported by an Australian Research Council Discovery Grant.

Authors

Allyson Holbrook is an Associate Professor of Education, Assistant Dean (Research Training), Faculty of Education and Arts, and the Director of the Centre for the Study of Research Training and Impact (SORTI), University of Newcastle.

Sid Bourke is Professor of Education and Assistant Dean (Research) for the Faculty of Education and Arts, University of Newcastle, and a member of the Centre for the Study of Research Training and Impact.

Terry Lovat is Professor in Education in the Faculty of Education and Arts and a member of the Centre for the Study of Research Training and Impact, University of Newcastle.

Hedy Fairbairn is Project Manager for the Centre for the Study of Research Training and Impact, University of Newcastle.

Email: hedy.fairbairn@newcastle.edu.au