

PhD Theses at the Margin: Examiner Comment on Re-examined Theses

Allyson Holbrook, Sid Bourke, Terry Lovat,
Kerry Dally

University of Newcastle

Abstract

It is rare for a PhD candidate who submits a thesis for examination to fail outright. If a thesis exhibits significant flaws the candidate may be required to make major revisions and re-submit the work for re-examination. The written comments of examiners before and after resubmission can provide important insights into the process of examination and the qualities examiners identify in a marginal thesis. Drawing on 101 of the most recent, completed theses across fields in one Australian university, this article investigates the differences in examiner comment on the qualities of theses by the same candidates before and after major revision and re-submission ($N = 6$), and between these theses and those that were 'passed' at the first examination ($N=95$). Critical comments about the literature review and the degree to which the examiner moved into a supervisory role were found to be strong indicators of theses 'at the margin'.

Since the 1980s there has been a growing interest in the 'visibility' of doctoral processes, particularly with respect to supervision, but more recently with respect to examination. Questions are being asked that encompass a range of issues from examiner selection through to the rigour and credibility of assessment procedures (Lawson et al. 2003, Powell and Green 2003). Many commentators have pointed out that doctoral examination, and doctoral study generally, is an exceedingly complex phenomenon that has yet to be subjected to sustained and systematic research. How students achieve success, the role the supervisor plays in getting a candidate's thesis to submission stage, or through an oral defence, and what constitutes quality in postgraduate research are all areas that are receiving attention in the field of research training in higher education.

There are few empirical studies addressing the written examination of doctoral theses or dissertations in the literature. As Morley et al. (2002) have indicated, studies of the assessment process and its consistency tend to be rare because access to examination documentation is difficult. In addition, many universities do not call for extensive documentation of process. Jackson and Tinkler (2001) investigated examination procedures and student and staff responses to examination in the UK. They obtained documentation from 20 universities (based on a stratified sample of old and new institutions) and drew on questionnaire responses from some 106 examiners and candidates from two of the 'old'. With respect to the viva (oral examination) they found there was 'no consensus' about the 'roles' played by the viva and there were inconsistencies and contradictions at the levels of policy and practice (p. 364).

In Australia a compulsory oral examination is not the norm, rather examination hinges on the written examiner reports on the thesis. In an attempt to explore this process Mullins and Kiley (2002) collected interview data from 30 experienced examiners about examination in Australia. Johnston (1997) undertook a content analysis of the text of 51 examiner reports from one Australian university across five faculties over several years. Pitkethly and Prosser (1995) utilised the reports of 74 thesis candidates at one Australian university. The findings include general agreement among examiners and about the core expectations, namely that they expect the thesis will demonstrate originality and make a contribution to the field. Evidence from a comparative cross-national survey by Kouptsov (1994) further bears out general widespread agreement on this point. However, some polarisation occurs around the issue of what is more important - the contribution, or the training (Powell and Green 2003).

Johnston (1997) found examiners tended to follow university guidelines or recommendations about how to report on a thesis, whereas Mullins and Kiley (2002) reported the opposite on the basis of interview data. They found examiners had established their own criteria, and that they noted, but did not use, guidelines provided.

It would seem that those who examine are inherently interested in doing so and approach the task in a positive light (Johnston 1997, Jackson and Tinkler 2001, Mullins and Kiley 2002). However, a poorly written thesis generally had a negative effect on the examiner (Johnston 1997, Mullins and Kiley 2002). A panel of 67 scholars from the USA, UK, Australia and Canada identified writing quality as one of the most problematic issues about PhD study (Noble 1994). Most researchers in the field have pointed out that editorial errors and presentation issues attract a substantial proportion of examiner comment.

Mullins and Kiley (2002) noted that examiners appeared very clear in the distinctions they made between poor, acceptable and outstanding theses, but they also detected that examiners approached the examination process anticipating that students would pass. It has been remarked by examiners that they rarely fail a thesis outright (Mullins and Kiley 2002, Grabbe 2003), however, they may suggest major revisions and re-examination. Becher's (1993 p.134) similar comment re-inforces this view. An exception is Johnston's (1997) study comprising all 16 theses that had been examined in a 'newer university', of which six theses were required to be re-examined. Despite the latter small study, the expectation in Australian universities is that a candidate will not present their thesis for examination unless it is ready (i.e. of pass standard). Reports on those theses that do require re-examination provide a rare and important opportunity to identify the qualities that identify, and are used to arbitrate, 'readiness' for examination.

This article reports the findings from an investigation of doctoral examination at one middle-ranking research university in Australia and focuses on four main aspects related to PhD theses that require major revision and resubmission. First, the characteristics of candidates who were required to resubmit their thesis were compared with the characteristics of candidates whose theses were considered acceptable. Secondly, for the resubmitted theses, examiner comments on the first and subsequent examination are compared. Thirdly, examiner comment on initially 'unacceptable' theses was compared with comment on other theses from the same university which were 'acceptable' on the first examination. And finally, how examiners respond to 'unacceptable' standards and to attempts to revise a thesis to an acceptable level was explored through an analysis of examiner discourse in both initial and re-examination reports. The questions that guide the analysis extend beyond the identification of differences and similarities in examiner comment and emphasis pre and post resubmission, to an exploration of the process of PhD examination, how examiners respond to the task, and what standards they apply.

Approach

The research questions guiding this project are grouped with an emphasis on examination process and outcome but also extend to what we can learn from process and outcome about the skills and knowledge required at PhD level. Data sources include three reports per thesis (the reports average about three and a half pages in length) and a large number of different student variables. *QSR N6* software is utilised for the text analyses of the examiner reports and the generation of text unit counts. In the core phase the examiner reports are coded into five primary

categories and their sub-categories. Analysis takes place in a sequenced fashion to facilitate cross-checking and systematic treatment of different layers of information. Coding is supported by an extended set of coding notes initially developed during the pilot phase, and subsequently refined to capture new elements of comment upon discovery. The process of coding is informed by team review and inter-coder checks. The proportion of total text units in each core category is tabulated and merged with candidature and other data using SPSS.

The mixed method design has three dimensions (see Figure 1). Dimension I focuses primarily on the quantitative information, initially obtained from university records, and subsequently from the quantifiable elements of the text in the examiner reports (core coding categories). Dimension II is primarily about the features and attributes of the reports. The reports are subjected to text analyses that identify core content and explore patterns, emphases, discursive and other communicative qualities. Dimension III moves further into the symbolic realm and the deconstruction of the reports. The culture and language of the doctorate, what it is to become accepted as 'Doctor', and the disciplinary 'knowing' that this assumes, contribute layers to the examination process and text that extend from clearly articulated expectation to assumption and myth. The dotted lines in the figure indicate the back and forth flow of information and interpretation across the different sources of findings. The solid lines indicate possible avenues of explanation.

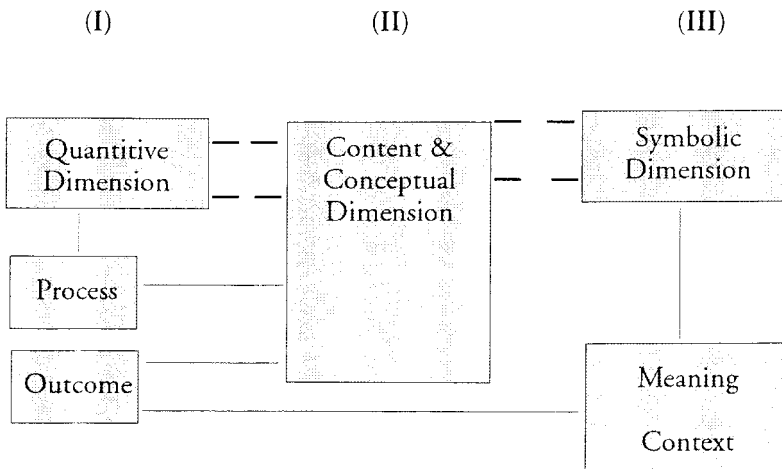


Figure 1. Dimensions of the study

In the discussion above a sequence of analysis was mentioned. Prior to the analysis undertaken in this article the team's focus was the initial examiner reports. All the reports including the resubmissions had been coded together. The next step involved separating out and comparing the core text analysis for the initial reports with the text analysis of the resubmission reports, and then comparing the resubmission group with the total. The results of that analysis are reported here. But in addition there are the results of some extended (i.e. not core) analyses where we went back to the text to see what else would emerge from closer inspection of the way in which the examiners presented their comments and a finer-grained analysis of their main issues with the thesis prior to and after re-submission (a process denoted by the top level of arrows moving from I to III in Figure 1 above). This combined core and extended analysis provided some interesting insights into the emotive aspects of examiner language, the execution of examiner role and also how examiners approached the concept of originality.

Examiner Recommendation

Examination procedures vary across Australia. In the university reported here three external, independent examiners are appointed for each thesis. Examiners are given only a broad guideline, being asked to provide a critical appraisal of the originality and significance of the contribution of the thesis, focusing on strengths and weaknesses as appropriate. Each examiner is required to provide a written report (length unspecified) and an overall recommendation on the thesis. Five possible categories of recommendation are offered to the examiners.

Category 1: Accept the thesis without amendment ('Accept')

Category 2: Accept the thesis but invite the candidate to make minor corrections ('Invite' minor correction)

Category 3: Accept the thesis only after specified corrections have been made to the satisfaction of the supervisor and Head of School ('Require' correction)

Category 4: Require the candidate to revise and resubmit the thesis for further examination ('Revise & resubmit')

Category 5: Fail, without the opportunity for revision ('Fail')

The University's Research Higher Degree Committee then considers the examiner recommendations and reports and makes its decision, which is one of the five possibilities offered to the examiners. Of the 101 candidates who had most recently submitted a thesis at this university in 2001, six were required to 'revise and resubmit' their thesis for re-examination (Cases A to F, see Table 1). In only two cases had an examiner recommended a 'fail' at the initial examination. The spread of

the examiner recommendations before and after resubmission is detailed below in Table 1 and the far right hand column shows the final committee decision on each re-submitted thesis.

Candidate Characteristics

Of the six candidates required by the University's Committee to revise and resubmit their thesis for further examination, no candidate was failed. The other 95 candidates were not required to resubmit the thesis to be re-examined and may or may not have been required to revise their thesis. Some candidate characteristics and the examiner recommendations on these six theses are summarised in Table 1. Age was that recorded at the beginning of candidature, total time was measured as the equivalent of full-time candidature, and full-time percentage indicated the mix of full-time and part-time candidature. None of these six candidates took leave from their studies during candidature.

Table 1. Descriptive information on six candidates required to resubmit their theses for re-examination, individual examiner recommendations and final committee decision

BROAD FIELD OF STUDY	GENDER & AGE	TIME: Total & Full-time ^a	RECOMM. EXAM ± RE-EXAM			FINAL DECISION
			Examr.1	Examr.2	Examr.3	
Case A						
Arts & SS M.	40 yrs	5.3 yrs, 31%	4 ± 1	2 ± 1	4 ± 5	2 (Invite)
Case B						
Arts & SS M.	38 yrs	4.8 yrs, 0%	4 ± 2	3 ± 1	4 ± 1	2 (Invite)
Case C						
Arts & SS F.	45 yrs	5.8 yrs, 87%	5 ± 1 ^a	4 ± 1	3 ± 2	2 (Invite)
Case D						
Education	F, 41 yrs	1.3 yrs ^b , 67%	2 ± 1	4 ± 2 ^a	5 ± 1 ^a	2 (Invite)
Case E						
Health	F, 57 yrs	3.5 yrs, 17%	4 ± 3	4 ± 2	1 ± 3	3(Require)
Case F						
Science	M, 45 yrs	3.5 yrs, 56%	1 ± 1	4 ± 1	4 ± 3	2 (Invite)
MEDIAN 43 yrs 4.1 yrs, 43%						

Notes:

- a. In each of these three cases the re-examination was done by a different examiner, after the initial examiner declined to re-examine the thesis.
- b. Enrolment time at this University – presumably the candidate had a previous enrolment elsewhere.

With a median age of 43 years, the resubmission group was somewhat older on commencing their degree than the other 95 candidates who were not required to resubmit their theses (33 years). Although three candidates had been enrolled for much more than the 'normal' period of candidature, one had been enrolled for less. The median period of candidature for the resubmission group (4.1 years) was a little more than that for the others (3.5 years). In terms of type of enrolment, this group had a lower proportion of full-time candidacy (with a median of 43%) than the others (78%).

With respect to these six theses, 15 of the 18 re-examinations were undertaken by the same examiner who had initially examined the thesis. Of these 15 examiners, 12 made a more favourable recommendation when re-examining the thesis, one made the same recommendation (to accept the thesis without amendment), and two made less favourable recommendations, in one case moving from initially requiring resubmission to subsequently recommending a failure for the thesis. The initial and final recommendations made by each examiner are shown in Table 1 separated by an arrow. As only one thesis resubmission per candidate is permitted by this university, any examiner who was still dissatisfied with a resubmitted thesis had no option but to recommend that it be failed.

Comparison of the Initial and Re-examination Comments on Resubmitted Theses

As might be expected, re-examination reports were shorter, being about half the length of the initial reports, and this difference was statistically significant (see Table 2). In some cases a re-examination report simply referred back briefly to comments in the initial report.

When the examiner reports were coded (the pre-standardised) lines of text (text units) were coded at categories (often called nodes in the *QSR N6* software). The core coding categories account for 100 per cent of the text in each examiner report, that is, all report text was coded with one or more of the categories. This allowed for a direct comparison of coded comment as a percentage of total text units. So, one can ask, did examiners comment proportionately more of one type of thing in the initial examination reports than in the reports on a resubmitted thesis?

Data entered in *QSR N6* software can be coded hierarchically. There are primary-level coding categories (parent nodes) and their sub-categories, each having a unique numeric code. The three primary coding categories reported in this article are:

Assessable areas covered (node number 3): captures all comment about the possible outcomes, subject matter and presentation of the thesis under examination;

Dialogic elements (node number 4): captures the specific features of examiner discourse that reflect on the nature of academic communication. In particular this category identifies active dialogue – engagement with, and consciousness of communicating personally with the reader/s;

Evaluative elements (node number 5): captures all comment that includes evaluation and judgment.

An example of a sub-category of a parent node can be found in the first row of Table 2. *Publications arising* (313) is a sub-category of the 'Assessable areas covered' by the examiner and refers specifically to comments about publications arising from the thesis. Throughout the article each sub-category reported in a table will be explained individually.

The analysis shown in Table 2 compares the proportions of text allocated to each category in the initial reports ($n = 18$) and the re-examination reports ($n = 18$) for the four categories where there was a statistically significant difference at the .10 level using an independent samples t-test. The .10 significance level, with a two-tailed test, was chosen in place of the more conventional .05 level because it was considered important to identify categories for which there was at least an indication of difference for the small number of theses at this university where re-examination was required. Further investigation of these and other possible category differences for marginal theses will be undertaken in successive individual university case studies.

One category where the difference between the initial and re-examinations was significant is '*publications arising*' (313). Text in this category captures examiner remarks about possible publication from the thesis. Examiners commented more on the publications that might arise from the thesis in their re-examination report. This stands to reason given that presumably the revised thesis had been improved and the initial presence of major flaws in a thesis would distract examiners from the potential for publication. It needs to be borne in mind that the numbers are small. Mention of publications arising occurred in only six of the re-examination reports compared to three in the initial reports. But in the latter all three examiners were not the ones to require major revision either. It appears from these findings that a thesis judged to meet the standards set by examiners was also deemed ready for dissemination to a wider audience.

Table 2. Comparison of percentage differences by text category between initial examination and re-examination

CATEGORY ^a	INITIAL EXAMINATION Mean% (SD)	RE-EXAMINATION Mean% (SD)
Publications arising (3 1 3)	0.62 (1.80)	4.10 (8.29)
Use of first person (4 3)	17.9 (13.8)	32.1 (22.4)
Neutral summative comment (5 1 2)	2.07 (2.27)	5.84 (6.44)
Other judgement (5 4)	5.68 (6.93)	13.4 (13.9)
All evaluative comment (5 1 to 5 4)	82.0 (21.8)	66.4 (31.5)
Length of reports (number of text units)	165 (137)	85.8 (113)

Note

a. In this and the following table, it was noted that most of the distributions for the categories were positively skewed. However, when distributions were normalised, only minor variations in means were observed, and statistical significance was not altered substantively. Means for the raw data are reported in these tables.

Two sub-categories under 'evaluative elements' (5 1 2 and 5 4) were also more prominent in the re-examination reports. In most thesis reports examiners provide at least one tight summative comment that captures their judgement on the thesis in whole or major part. Such comments might be wholly positive or wholly negative. In the re-examination reports there is more of the sub-category 'summative-neutral' (5 1 2). This type of comment is neither wholly positive nor negative but often neatly straddles the fence. It can also be described as a category of qualified judgement, e.g. 'the revised version of the thesis is considerably improved over the original submission...However, there are a number of modifications to the thesis that I consider should be made before it is accepted'. One can hypothesise that in a re-examination situation examiners are more likely to comment in ways that reflect the fact that the thesis was flawed originally. Also unless a student re-writes and re-designs the whole thesis, which is rare, it is highly unlikely that quality will be so improved as to win unreserved positive evaluation. The straddling-the-fence phenomenon reflects the history of the thesis.

The other evaluative category that was more evident in the re-examination is '*other judgement*' (5 4). This category captures evaluative comment that is not summative, nor contains any element of instruction, so these statements tend to cover a lot of ground. '*Other judgement*' is a catch-all category for stand-alone comments with an evaluative element. They can be emotive and may range from emphatic to casual. Examples are: 'congratulations to the candidate on this thesis', 'I am satisfied that the candidate has mastered the essential aspects of research and writing at this level'. The examiner seeks to convey an impression but not in a way that engages with the thesis subject matter very closely or in detail. Comments such as these tend to occur early or at the end of the report, like a 'trimming' or garnish to the main fare. Sometimes the impression is conveyed that they are used to mediate or soften the impact of a critical, or densely technical, body of comment.

Another difference between the initial and the re-examination reports is evident in *dialogic elements* (4). There is significantly more '*use of the first person*' (4 3) in the re-examination reports. Use of first person in the initial reports was typically linked with comment that demonstrated dissatisfaction, frustration or disappointment: 'I do not believe that the defects could be remedied within a one-year period', 'I believe that this argument is unsuccessful', 'it is unclear to me...', 'I am unsure what to make of the argument', 'I have reservations, I would like to have seen more critical discussion'. Use of the first person delivers emphasis. By contrast the comments in the re-examination, while still containing comment that was negative in tone, also contained a substantially increased proportion of first person statements that were positive and supportive: 'I find these sections quite informative', 'I congratulate X on his production of a convincing and important contribution...', 'I found this chapter to be highly original and thought provoking'. They were also linked to compromise or concession, of the 'I still have concerns in one or two areas', variety. The coding category '*summative-neutral*' (5 1 2) also captured the text that reflected this type of 'qualified' judgement.

When all the '*evaluative elements*' (sub-categories 5 1 through to 5 4) were summed it was found that the initial examination reports included a significantly greater proportion of evaluative comment than the re-examination reports. This indicates that the proportions of evaluation categories other than neutral comment and other judgements, although not significantly different when taken singly, were significantly different collectively. The most common evaluative category, '*formative evaluation*' (5 2), particularly exhibited this difference constituting 39 per cent of initial report content compared with 22.5 per cent of re-

examination reports. Because there was wide variation in these proportions for each examination, they were not statistically significantly different individually ($p = 0.105$).

A Comparison of the Initial Reports on Resubmitted Theses and Reports on Theses not requiring Re-examination

For the initial examination, the proportions of text coded at the core coding categories were compared between the 18 examiner reports on the 6 theses that were required to be re-examined and the 285 examiner reports on the other 95 theses that were not re-examined. The mean proportions of text coded to categories in reports that were found to differ at the .05 significance level (two-tailed) using independent sample t-tests are shown in Table 3.

In reports on theses that did not require re-examination the '*scope*' (311) of the thesis featured more prominently than in those that required resubmission. Scope statements are the type of statements that tend to elaborate on the ground covered by the thesis and are usually adjacent to statements on '*significance*'.

In theses not destined to be resubmitted, examiners wrote proportionally more about the use of the literature (that is, how the candidate extracts and applies the knowledge gained from the literature), and less about the extent to which the literature in the field was covered, than in cases where the candidate was required to revise and resubmit. The most common criticisms in regard to the latter were that important literature had been omitted, the literature employed was not current, or the research questions were not grounded in the literature.

Table 3. Proportions of initial examination report text categories that differed between theses requiring re-examination (N=18) and theses not requiring re-examination (N=285)

CATEGORY	RE-EXAM REQUIRED Mean% (SD)	RE-EXAM NOT REQUIRED Mean% (SD)
Thesis scope (3 1 1)	0.75 (1.26)	3.59 (8.16)
Literature coverage (3 2 1)	7.77 (10.9)	3.78 (8.02)
Literature use/theory (3 2 3)	0.49 (1.04)	1.81 (4.15)
Topic related issues (3 4 2)	21.6 (20.4)	9.66 (17.3)
Positive summative comment (5 1 1)	3.16 (3.94)	9.70 (13.5)
Instructive comment (5 3 1)	20.4 (22.7)	8.51 (12.7)
Other judgement (5 4)	5.68 (6.93)	14.8 (17.3)
Length of report (lines)	165 (137)	111 (114)

Discussion of 'topic related issues' (3 4 2) comprised 22 per cent of the total report for the theses which required re-examination, more than double the proportion that was devoted to these issues in the theses which did not require re-examination. Typically, the examiners spent a considerable amount of time either questioning or contradicting the candidates' assertions, or providing detailed information about the topics being discussed. Often the topic related sections were several paragraphs and sometimes several pages long. The examiners appeared to adopt a 'teaching' and 'expert' role in these lengthy monologues in which they endeavoured to 'enlighten' the candidate or expand candidate knowledge and understanding of particular issues. Candidates were often directed to read more widely or to read specific literature, even basic beginner literature. For example:

A serious reading of this basic text (often used in undergraduate courses) would have forestalled quite a few misunderstandings.

With respect to evaluative comment, proportionally more 'instructive comment' (5 3 1) was written on theses that were subsequently

re-examined, and there was proportionally less '*positive summative*' (5.1) comment and '*other judgement*' (5.4) on these same theses. The theses which required re-examination contained more than double the proportion of '*instructive comment*' (20%), compared to the theses for which re-examination was not required (9%). Instructive comment refers to statements by the examiners which suggest that the thesis could or should be better, and were generally directed at perceived problems in the methodology, results or conclusions sections of the theses. Common methodological problems identified by the examiners included methods which were poorly explained, not justified, inadequately thought through, inappropriately applied or fundamentally flawed. In regard to results, examiners expressed dissatisfaction when results were overstated, misreported or misinterpreted or when there was insufficient information provided:

There are a number of areas where the work could be significantly improved, and there are some areas where the author seems to have over-interpreted the data to an extent that requires revision.

Conclusions were criticised if they were not explicitly related to the research questions or to the 'real world', were not adequately discussed, or if the strengths and limitations of the study were not acknowledged or explained: 'Many emotive statements are made with little or no academic support'.

Lack of originality was also cited as a major obstacle. In one instance 'originality' was cited as the required standard:

The thesis is a well-constructed body of substantial work but it has to meet the criterion in the guidelines of 'original and significant contribution' for a degree at this level.

There was significantly less '*positive summative comment*' and '*other judgement*' in the theses requiring re-examination. This is not completely surprising as these comments typically reflect an overall satisfaction with the worth of a thesis.

As also shown in Table 3, reports on theses requiring re-examination were significantly longer than other reports. Two possible reasons are suggested. First, it may be that examiners felt a need to write more in justification of a harsh report than a favourable report. If the possible examiner recommendations were considered as a five-point rating scale, the correlation between the rating given to a thesis and the length of the written report was 0.480 ($p = 0.044$) for the 18 initial reports on re-examined theses, and was 0.412 ($p < 0.001$) for the other 285 reports,

in both cases supporting this suggestion. Alternatively, as discussed above, more lengthy reports for the theses requiring re-examination may be caused by the additional instruction that examiners included for candidates who they felt needed more guidance in the conduct of their study.

Differences in Reports of Examiners Recommending Change and those not requiring Change

For the six theses that were re-examined, there were a total of 36 examiner reports (made up of 18 initial and 18 re-examination reports). Of these, 32 reports required revisions to the thesis and four did not. When the text coded at the core categories for these two groups of reports were compared, there were two categories included in the reports of examiners requiring revisions that did not appear at all in the reports of examiners not requiring revisions. No statistical tests were applied to these differences owing to the very small number of cases and complete lack of any variation in the second group. The two categories that differed in this way were those that captured evaluative comment of a 'prescriptive' instructional type, which on average comprised 10 lines of each report requiring revision, and, to a lesser extent, use of the literature.

The fact that only the examiners requiring revision included these categories in their reports suggests that prescriptive comments and comments on the use of the literature were generally important for examiners when they were deciding whether to require revision of theses. Prescriptive comments direct candidates to fix specific aspects of the thesis. Generally, there is no discussion or negotiation about these directives which are issued as a command rather than a suggestion, for example, 'The Latin translations... need to be the candidate's own and they must be correct'. Clearly, the four examiners who judged the theses already to be at an acceptable standard did not indicate that there was any need for such action to be taken. It should be pointed out, however, that these examiners may have identified such weaknesses but did not consider they warranted mention. With respect to the use of literature, a typical comment was 'The discussion is very unsatisfactory and devoid of reference to the literature'.

When the lengths of examiner reports were compared for the two groups, the examiners requiring revision of the thesis wrote reports (averaging 134 lines) that were more than twice as long as the reports by examiners who did not require revision. As discussed above, this may have been due to examiners justifying their request for re-submission or to the additional instruction the examiners provided for candidates who did not demonstrate a required level of mastery.

The degree to which the literature and its scope and use cropped up in the resubmissions is extremely interesting. Are examiners weighting this aspect more in their consideration of the worth of the thesis, or is it an indicator of deeper flaws, much like a litmus test? At this point therefore we extend the analysis to attempt to elicit as fully as possible what examiners mean, what expectations they bring to the process and what they understand by thesis examination. Concentration on these few scripts, in the full knowledge of how they differ from a large number of others, provides the opportunity to test and hone emerging theories about research culture and assessment parameters.

Examiner language changes between an initial and a re-examination report. The use of the first person and more prescriptive discourse flags this. Are there other indicators? In theses with significant flaws examiners provided more detail and engaged in more instructional commentary. One feature of instructional commentary is the overarching nature of such comment, sometimes tending to the general, but often general in the sense '... if it were me I would do this a different way'. Not all examiners have the same expertise or set of understandings, the same expectations, or even the same tolerance for ambiguity, error or methodological scope. Moreover different elements of a thesis may trigger these in different measure within the one examiner. This also turns our attention to the 'role' of the examiner. How examiners play out their role appears to be triggered by the quality of the thesis as they perceive it. So role may be another litmus test of quality.

Examiner 'knowing'

If perception of quality is the end point of this puzzle then a starting point is how examiners 'know'. The 'Ways of Knowing' thesis of Jürgen Habermas (1972) suggests that there is a consistent pattern across discipline areas by which knowledge is revealed and further negotiated. Human perception arises from a series of 'cognitive interests', interests which are part and parcel of the way the human mind works. These interests are three-fold. First, there is an interest in technical control which impels an 'empirical analytic' type of knowing. Second, the interest in understanding meanings gives rise to an 'historical hermeneutic' way of knowing. Third, there is an interest in being emancipated, a free agent as it were, which issues in a 'critical', or 'self-reflective' form of knowing. As far as Habermas is concerned, all three interests are operative regardless of the discipline area.

Where empirical/analytic, or technical, knowing is operative, the supervisor or examiner is most likely to be the 'expert'. The expert represents and stands as the custodian of the body of technical and conventional knowing to which the learner must conform. Hence when

examiner comment falls within the realm of expert it has an hierarchical feel to it. When an academic is called upon to examine a thesis it stands to reason that their expertise is being called upon. They expect this, as they see themselves in the role of upholding standards (Mullins and Kiley 2002). Expertise manifested in an assessing or accrediting role such as that of 'examination' finds quintessential form in the decision given.

For many examiners specific elements of the thesis are open for negotiation. Where historical/hermeneutic knowing is operative, the supervisor and examiner are more like partners with the candidate, communicating about meanings and negotiating about understandings. Herein, the concern is not with 'right' or 'wrong' knowing but with knowing that results from interpretive understanding. Examination will not only exude negotiability but will have a co-learner 'feel' to it. The question is how likely is it that examiners will demonstrate this quality of 'knowing' in a situation where the thesis falls well short of expectations?

When dealing with knowing of the critical/self-reflective type, the traditional roles of supervisor/examiner and learner are potentially reversed, with the learner being acknowledged as the one who is in control of their own knowing, and the role of the former being as listener. The challenge here for any traditional modes of teaching/learning relates to the fairly obvious truth that learners may often 'know' in ways that are beyond the knowing of the teacher. This is the way of knowing which, it is said, is a necessary precursor to the stretching of the boundaries of knowledge, to genuinely new knowing taking place. Granted the elevated status which the PhD enjoys in the learning system, and especially granted the mandatory tenet related to originality, one might hypothesize that this type of knowing would be fairly prominent in the average dissertation, but very likely to be absent in a thesis deemed unsatisfactory.

Examiner Response to Resubmission

The report on a thesis is a vehicle for justifying the examiner recommendation. Theses that have flaws attract more 'instructive' comment from examiners and more specifically attract prescriptive instruction. 'Should' is a word that very frequently crops up in relation to instructions about how to improve a thesis, along with '[you] need to', 'have to', etc.

Moreover, where resubmission is called for by an examiner we also find the input of the expert and 'colleague' marked by the language of deficit and of frustration, disappointment and even anger. The selected excerpts in Figure 2 show the range of examiner responses to what they see as flaws in the thesis they are examining.

There is also another frame – that of reconciled expectations. This ranges from happiness and relief that the thesis meets all of most of the examiner's concerns, through to qualified acknowledgment that the thesis has stumbled over the line into the acceptable category. But where is that line drawn?

The following cases provide illustration of the way examiners apply their knowledge, execute their role and some glimpses of how 'acceptable' and 'original' are defined.

Digging in - becoming Tougher

In the case of Candidate A (see Table 1 above) the third examiner's (A3) initial recommendation was for the candidate to re-submit. On re-examination, however, this examiner 'dug-in' and recommended a fail. This was atypical. Most who re-examined a thesis that still had flaws either accepted the flaws that remained or they asked that they be given further attention without the need for re-examination. In A3's initial report the hierarchical text of the expert was in evidence but was intensified in the second report. In the earlier report, one senses an openness to engage at a more collegial level. This is seen explicitly where the examiner commends the candidate for wide reading and persistence, and more subtly when the examiner momentarily takes the blame, rather than passing it on to the candidate, for not understanding aspects of the thesis. In one section, for instance, sentence after sentence begins with a phrase of the following kind: 'I do not know ...', 'I also do not understand ...', 'I'm not sure ...', 'I also don't follow ...'. Such phraseology provides the candidate with some space for retrieval - a gesture which signals to the candidate a willingness to be convinced about these things. The proviso is that the re-submitted thesis takes full account of the expert critique being offered. In the second report, there is no such concessionary text. Here, the examiner is far more impersonal, detached and closed to further consideration. More commonly, we find phrases such as '... the thesis is quite unconvincing' and 'It is extremely hard to follow'. When the examiner uses the first person in this case this signals the 'expert': 'I find this argument unconvincing', 'I could not follow...', 'I found it impossible ...'.

There was also a case where one examiner (E3) recommended an unqualified pass in the first report because the thesis 'more than adequately meets the requirements of the degree...the thesis is a valuable and original contribution to knowledge' with few limitations that do not prohibit it from being 'commendable', 'original, independent and credible', then provides a more qualified recommendation of acceptance in the second report 'only after corrections have been made'. The second report is longer as well as more reserved in its support, appearing to

suggest that limitations that were tolerable on the first reading are in need of correction on the second reading. There is also the strong element

The language of deficit ...

Systematic failure to consider or discuss alternate theories or ideas
Surveys the field in an idiosyncratic and eclectic way
Much more reading and conceptualization needs to be done
Much of the argument takes place in isolation from current scholarship
Devoid of reference to ...
Many emotive statements are made with little or no academic support
Seriously outdated resources
Shallow understanding
Lacking in depth
Needs more rigour
Serious misunderstanding of theory
It seriously misrepresents the evidence
The claims made are too far in advance of the data
Much of the chapter seems to be purely speculative
One of the biggest problems was over generalization and lack of contextualisation
The whole needs to be designed and rewritten
It is entirely unclear what hypothesis is being tested
Has not demonstrated to me that they have sufficient understanding of experimental design
Unconvinced about the validity of the analysis
...here rests the main methodological problem

Expressions of frustration, disappointment, anger ...

To the extent that I understand this at all, I am left wondering how widely the candidate read...we are all bound to read as widely as possible whether we like it or not
Choice of the literature was ad hoc (I've read it so I'll use it)
This is confusing
Baffled
Puzzled
Irritating
Pretentious
Sloppy
Absurd line of argument
It is an insult to the reader
More credence is given to political developments and opinions than to evidence
Bluntly the work is repetitive, diffuse and over-long
Feeling of being stranded without a line to grasp on to
Hopelessly jumbled
Odd use of evidence
This simply begs for discussion
The discussion has the appearance of being dumped in to make up words. I have no idea what the author means by it.
The arguments are a nonsense
What is your claim?

Figure 2. Excerpts from examination reports on theses requiring resubmission

of 'the student has not followed through thoroughly enough on my comments'. For example, the text reads:

In my previous assessment of this thesis, I referred to some limitations with the qualitative analysis. These criticisms still apply... There has been an attempt to discuss ...[the choice of method] and define ...[the sampling procedures] but the candidate does not adequately explain...

It would seem that what the examiner conceded was a limitation in the candidate's skills initially, was vindicated and more clearly in evidence the second time around. The language suggests their faith is shaken.

One can only speculate as to what is going on in the mind of this examiner. Of course, the answer may be as simple as the fact that a second reading always renders greater insight, or that this particular examiner put more time into the exercise the second time around. However, could it be that he/she felt a little intimidated by the fact that at least one academic peer had set a higher standard in refusing to accept these limitations, and so the second report provided an opportunity for a little face-saving? Whatever the explanation, the semiotics tell us that the regime of expertise is ramped up significantly in the second report.

There would, on the surface, appear to be two threads of expertise evident in thesis assessment. Their existence warns against simple classifications. One level is the expert in an area, field or approach, the other is the expert on what a PhD is or should be within the discipline and even more generally (e.g. as an artifact). Where there are flaws in the latter we see what a fine line there can be between expert opinion and gatekeeping.

Necessary Concessions

Examiners often felt they had to make a concession on the basis of the effort they recognised had gone into the production of a thesis. The extent of the effort combined with redeeming features, such as attention to the examiner's points in the resubmission (often noted specifically by the examiner as having been done) would lead to the examiner taking the position that 'now' after following their advice, the revised version was acceptable. There were four cases where one examiner recommended rating 4 the first time and 1 the second. This is the most extreme change of position in all the cases of resubmission.

In two of the three cases the examiner was still seeking some improvement despite recommending category 1, saying such things as: 'I still have concerns', 'remain unconvinced' (A1); 'still some issues

unresolved', 'argument loses intellectual tension', and 'I am not convinced still that [the argument] holds equally well for all the cases X makes for it' (C2).

The exception was B3. Initially this examiner delivered their report 'with regret' about having to recommend the candidate revise and re-submit. In the first instance the examiner thought it 'not well set in the literature' and methodologically 'flawed'. They also thought the key theoretical model in the particular thesis was inadequately clarified, suggesting:

...[the candidate] needs to consider the question of what all this means...

On re-examination the examiner 'unhesitatingly' recommended acceptance without need for revision, urging the candidate to publish (with some information about where and what the candidate might expect of the market for their work). They believed the candidate 'showed every sign of being an informed, thoughtful and above all committed, scholar'.

Another examiner of the same thesis was a great deal more negative (B1). In the initial examination they were not convinced that the work had the potential to pass given their estimate of the amount of revision necessary. Indeed, the examiner suggested:

it seems to me that much more reading and conceptualisation needs to be done in addition to a substantial rewrite...perhaps the candidate, after discussion with the supervisor, and if a suitable rewrite seems impossible, should consider submitting the case studies in an appropriate contextual setting for a Masters degree. This would I believe, still involve considerable rewriting, but rather less in terms of reviews of the appropriate literature. The model could be presented without being defended in the way it would have to be for a PhD.

The insight this comment provides about the 'level' of PhD study is taken up in the next major section of this article.

The first report by Examiner B1 is harsh and objective, the latter depicted by the persistent use of the third person, e.g. '... these chapters made no sense at all ...', '... the whole needs to be re-designed and re-written'. The only use of the first person is reserved for the most damning comment of all: 'I am not particularly sanguine about the chances of a rewrite being possible within one year', and then going on to recommend the downgrade option.

But in the second report Examiner B1 stated:

This is now an excellent thesis that contributes significantly to the area of [X]. The candidate should investigate publishing the work either as a monograph or as a series of papers...

This examiner goes on from these very positive words, which clearly acknowledge sufficient original contribution to warrant publication in relevant professional journals, to identify about 20 quite small errata, half of them to do with commas and apostrophes (in contrast to their previous report where a comment on errata was dealt with in one sentence of 2 lines). It may well be there were fewer typographical and grammatical errors in the copy submitted first, or that errors of this type become more obvious when larger flaws are not clamouring for attention. Nonetheless, as this example illustrates, the major attention was given to the artefact, when all that was substantial about the thesis had been said in four lines. This examiner holds to the position of expert, while conceding major revision of their initial assessment. Most examiners nominate at least some typographical errors in a thesis, regardless of the tone of their discourse, but the examiner who holds to the position of expert regardless of the quality of the thesis, is the unassailable gatekeeper.

Examiner 2 for that same thesis (B2) moved their recommendation from a category 3 to 1. All examiners of this thesis had identified methodological weaknesses and problems with the scope of the literature as the main reasons for their recommendation and were consistent in agreeing there had been, as Examiner 2 put it, 'significant improvement' at resubmission. This examiner also provided comments to assist the candidate to publish, going as far as to nominate specific journals, but was not as unstinting as the third examiner, conceding that the thesis presentation was 'quite good'.

The examination process tends to lock examiners into the language and position of the expert almost without exception but, as already indicated, the subtle differences reveal much about the type and level of expectation. One such exception is when an examiner feels vulnerable in a situation where they had indicated no major problems, only to discover that their peers think differently. F1 believed the thesis met all the criteria for the award the first time. In an unusual twist for the university concerned, this examiner saw the other examiners' initial reports before writing their report for the resubmitted thesis:

I did read the comments of the other referees and the responses and modifications made by ...[the candidate] and found these illuminating. The expertise of the referees in...[the particular

area] is vastly superior to mine and I could appreciate the changes they wanted made to the thesis, I am sure it has benefited from the exercise. I can only repeat my initial assessment that within the areas of my expertise the thesis was more than satisfactory and I am comfortable with my recommendation that a PhD is warranted.

It is useful to point out that in his initial report the examiner was impressed by the care with which the thesis had been produced, its clarity and its contribution in a notoriously difficult area wherein 'all the scientist can do...is employ a statistical approach, interpret the data with care and not overstate the findings', just as the candidate had done. The examiner had also discussed 'a possible different set of interpretations' with the candidate – a collegial approach – no instruction, no deficit, no disappointment and some humility with respect to their own 'expertise'.

Arbiter, Supervisor and Colleague

In previous analyses of the initial examiner reports for all 303 examiners we identified that the examiner appeared to take on three distinct roles, sometimes all within the same report. The first is the arbiter/assessor role, so clearly illustrative of expert knowing defined above. The second is the collegial role where an examiner is exhibiting close engagement with aspects of the thesis, and virtually carrying on a 'dialogue' with the candidate.

This collegial role is less distinct and less in evidence in any form in the resubmission cases, except where the first examiner gave the thesis a category 1 or 2 recommendation. Examiner F1 for instance, commiserates with the candidate about the difficulties of the field and tells them they are 'doing a good job'. Examiner D1 only sought minor revisions and then did not re-examine the thesis. This examiner also commiserated with the candidate on the complexity of the area even while indicating their opinion was 'partly different' on several key points to that of the candidate.

In the report by examiner E3 the language use is very indicative. Instead of using 'should', they consistently use 'could' – 'greater use could have been made', 'could have extended the analysis', 'could have been strengthened'. Apart from these slim examples the only other sign of the collegial role was in advice about publication.

The third and final role is that of 'supervisor'. It is notable that when major flaws are evident, particularly in the reading and identification of the literature, examiners start providing lists of references the candidate needs to read. They also begin to supply common

interpretations of the literature. Assisting the candidate to position their study is the quintessential role of the supervisor. On more than one occasion examiners mention the supervisor in this regard. There can be little doubt the intention is to admonish that individual, however gently, that they have failed in that task. In every case of identifying major flaws the examiner assumed the supervisory role and slipped into instructive mode, including broad commentary about the scope of the flaws and more prescriptive instruction about what to read, how to fix and what to address. It is particularly interesting that these individuals engaged in less 'formative' instruction overall than in cases where the candidate had not been required to revise and resubmit. The committee that appraised the comments must have been struck by the forceful and unadorned nature of comments about paucity of literature, lack of originality, evidence of bias, inadequate interpretation, methodological weakness and inaccuracies (See Figure 2 above) and this may have been instrumental in their decision to require resubmission.

Quality and the Contribution of the Thesis

The guidelines provided to examiners for this university explicitly refer to an original and significant contribution of the thesis, asking them to indicate strengths and weaknesses. But is original and significant work sufficient for a thesis to be judged as of adequate quality? Examiners indicated that the existence of ideas or approaches that were original, novel, important, valuable or interesting were embraced in their understanding of contribution in meeting the guidelines. Nonetheless such features were given little weight if candidates did not apply their skills to make the best of the original or significant elements and prove their grasp of what was going on in the discipline or field. The before-and-after comments on one thesis by two examiners highlight this.

Case C Examiner 1 (initial)

I do not base this judgement on the credibility or lack of credibility of the candidate's thesis (his ideas) so much as on what I see as a number of flaws in the presentation, in the marshalling of evidence, and in the attempted demonstration of the case, offset by an interesting and often attractive approach to the topic... A PhD thesis ...must provide evidence that the candidate has absorbed the major previous discussions of the topic in question.

Case C, replacement for Examiner 1 (re-examination)

X recommends some interpretations that are, to this untutored eye, at least novel. On these grounds, I will concede that the

thesis satisfies the minimal requirements for the successful completion of the doctorate. Central to these I take to be command of a few primary texts; the ability to make wise use of the secondary literature and specific research tools; and the capacity to assemble ones findings – including a novel insight or two – into a readable and pleasing whole.

It would seem that the second examiner's expectations about such skills and understandings were reasonably modest, but they also provided a strong indication of baseline acceptability. The further quotations mirror similar base-line expectations:

Case B, Examiner 3 (initial)

This thesis addresses an interesting and important topic using materials which have not been carefully scrutinised in the past. The account given is a valuable one. However I have a number of problems with the thesis: it is not well set in the literature; it has methodological flaws; the model needs some modification; the presentation needs minor attention.

Case E, Examiner 1 (initial)

I have major difficulties with this thesis. The major weaknesses are that the contents are not well linked to the research literature and the research component is relatively small and appears to have been used to justify more than it deserves. Indeed, most of the conclusions are based more on findings from previous studies than on the authors own research...I wonder if this thesis is really at the standard required of a PhD? While the central idea is sound, the thesis reads like several poorly-connected small studies that have been put together post-hoc, with poor connection to the academic literature, inadequately thought-through and applied methodology, and poor linkage back to the real world. Despite the obvious work done by the candidate, I think that major revision is necessary.

Regardless of discipline, recognition by the examiner of a command of the literature is a consistent base-line indicator of 'originality and significance'. Another test may be the examiner's admission they have learned something new – the celebration of new knowing. Can we test this hypothesis? If we look to those examiners who did not examine the initial thesis, without exception these individuals indicated they had learned something from the thesis. In effect examiners seek to learn from a thesis and, in that, they also expect the candidate to rival their own knowledge base, their own expert status. They do not want to step in as the supervisor, *post hoc*, to facilitate that process but, if they have

done this and the candidate follows their advice, they will subsequently concede that the outcome meets or at least approximates what they had sought.

Rarely is a candidate's 'original' work original in an individual sense. It is the product of a team effort, the primary team being the candidate and his/her supervisor(s), especially in the stage of identifying and positioning a project's originality. In particular instances it is the examiners, acting as supervisors who alerted them to the original potential of their project. Which does raise another interesting if tangential question: If the candidate responded to the initial examiners' guidance, then is their work as original as when they succeed in doing this without examiner input?

*

*

*

The current analysis of a small number of theses that were required to be resubmitted for re-examination has identified a number of features that appear to be important to examiners when they are making a judgement about the worth of a candidate or the value of a thesis at the margin of acceptability. The issues that differentiated acceptable from non-acceptable standards of research were mainly concerned with the content of the thesis, the candidate's approach to conducting a study or analysing its findings and the contribution of the research to its particular field. The examiners' criticisms focused on the substance of a thesis and whether or not a candidate had the ability to present an hypothesis that was grounded in current literature and was thoroughly and systematically investigated. Of less concern to examiners were features of presentation such as mechanical or editorial errors. Although numerous grammatical and typographical errors were often listed as essential corrections in reports requesting re-submission of a thesis, these reports did not contain a significantly greater proportion of text devoted to editorial comments than reports on theses which did not require revision. This finding suggests that while examiners are concerned with defects in presentation, such flaws are not crucial to an examiner's final evaluation.

To try to get at the notion of what examiners decide on and how, and how others (such as a university committee or the candidate and supervisor) might make sense of their evaluation, it is necessary to undertake very close analyses of examiner reports. Examiners are aware that, in some senses, their report becomes a 'public' document. But from such a public document what can we learn of the process? Although a different study would be required to explore what the examiner had in mind when they wrote and/or cut-and-pasted together their reports (that is, their intent), analyses of initial and re-submission reports provide

some insights into the process. These analyses provide valuable indications of what examiners consider important for acceptability or a base-line 'quality' of a thesis.

If a thesis is flawed the examiner exercises their knowing in specific turns of phrase that also suggest specific roles. All examiners are arbiters but if they detect significant flaws this position solidifies into gatekeeping by the virtually unassailable expert. The language is of deficit and usually marked at several points by frustration bordering on anger or disappointment. Following re-submission, theses do not usually disappoint. For all their expert status, examiners are usually happy to grease the hinges. Their gatekeeping in cases of re-submission, as Jackson and Tinkler (2001) found in relation to the viva, performs a largely ceremonial function. They are prepared to concede if their own knowing is acknowledged in the revision.

If an examiner moves into the supervisory role this is one indicator that the thesis is close to base-line (that is, acceptable) quality, and also may suggest it is not original or significant. It would seem that problems with coverage of the literature may be the strongest indicator that the thesis is not only of poor quality but not original or significant. Clearly these findings are based on few cases from a limited range of disciplines and only one university. Nevertheless, they provide a useful springboard into further cases, and suggest that it may be possible to identify elements of thesis quality that can be illustrated, verified and generalised.

References

- Becher, T., 1993, Graduate Education in Britain: The View from the Ground. In B.R. Clark, ed., *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan*, Berkeley, University of California Press, 115-153.
- Grabbe, L., 2003, The trials of being a PhD external examiner, *Quality Assurance in Education*, 11(2), 128-33.
- Habermas, J., 1972, *Knowledge and Human Interests*, (Tr. J. Shapiro) London, Heinemann.
- Jackson, C and Tinkler, P., 2001, Back to basics: a consideration of the purposes of the PhD viva, *Assessment & Evaluation in Higher Education*, 26(4), 355-366.
- Johnston, S., 1997, Examining the examiners: an analysis of examiners' report on doctoral thesis, *Studies in Higher Education*, 22 (3), 333-347.
- Kouptovg, O., 1994, *The Doctorate in the Europe Regio*. CEPES: UNESCO.
- Lawson, A., Marsh, H. and Tansley, T., 2003, Examining the

examiners, *Australian Universities Review*, 46(1), 32-6.

Morley, L. Leonard, D. and David, M., 2002, Variations in Vivas: quality and equality in British PhD assessments, *Studies in Higher Education*, 27 (3), 263-273.

Mullins, G., and Kiley, M., 2002, 'It's a PhD, not a Nobel Prize': how experienced examiners assess research theses, *Studies in Higher Education*, 27 (4), 369-386.

Noble, K.A., 1994, *Changing Doctoral Degrees: An International Perspective*. Buckingham, The Society for Research into Higher Education & Open University Press.

Pikethy, A. and Prosser, M., 1995, Examiners' comments on the international context of PhD theses. In: C. Mcnaught and Beattie, eds. *Research into Higher Education: Dilemmas, Directions and Diversion*, Melbourne, HERDSA, 129-36.

Powell, S. and Green, H., 2003, Research degree examining: quality issues of principle and practice, *Quality Assurance in Education*, 11(2), 55-63.