

Draft copy: Please do not cite without permission.

HOL04678



Centre for the Study of Research Training & Impact

AN INVESTIGATION OF INCONSISTENCIES IN PhD EXAMINATION DECISIONS

Allyson Holbrook, Sid Bourke, Terry Lovat & Kerry Dally
The University of Newcastle

Abstract:

The process of PhD examination in Australia across academic disciplines requires that the PhD thesis be assessed by two or three external examiners who make a recommendation on the thesis and write detailed supporting comments. The recommendation at one end of the spectrum is to pass the thesis outright, the other end is a terminating fail. However most examiner recommendations fall in the middle (71%) and range from inviting minor corrections through to requiring revision and resubmission. These recommendations are then usually adjudicated by a committee, which can in some cases differ substantially from one or more of the examiner recommendations. The disparities in adjudication prompt concerns about the visibility of the examination process. This paper draws on a sub-sample of reports on 400 candidates where such disparities between examiner recommendation and institution decision are evident and then explores the texts of the reports to attempt to identify what may have triggered the discrepancy. Situations where the examiners gave a lower recommendation than the committee were compared to those where the committee recommendation was lower than that of the examiners. From this analysis it could be determined that specific instructional emphases and qualities in the examiner comments appeared to influence the outcome.

Paper presented at the Annual Conference of the AARE,
Melbourne, 28 November – 2 December 2004

Presenter and contact:

Associate Professor Allyson Holbrook
Phone: 02 4921 5945 & 02 4968 6710
Fax: 02 4968 6713
Email: allyson.holbrook@newcastle.edu.au

An investigation of inconsistencies in PhD examination decisions

Background

In Australia when a PhD thesis is submitted for examination supervisors and students breathe a sigh of relief and then settle in to wait some weeks for the examiner reports, which in most cases will require something more from both of them. Examiners usually expect some amendments to the thesis and this expectation is relayed through their written comments. The two main forms of evaluative comment made by examiners in their written reports are instructive and summative in nature. Most of the instructive comment is 'formative' and formative comment usually outweighs the summative comment (Holbrook et al. 2004b). Examiners also present instructive forms of comment during a PhD viva (Tinkler & Jackson 2004, p.16, p.25). In assessment terms formative assessment is typical of the progressive assessment that occurs during the 'course' of a subject, to help build the student's knowledge and skills. The existence of both formative and summative evaluative comment in PhD reports reflects the nature of assessment procedures (examiners are asked for comments to justify their recommendation) and the nature of the PhD which consists of process and outcome, i.e. research training and research thesis. There is a perception among examiners that the candidate will and should continue to develop as a researcher in their field (or at least in their expertise in reading or using research), hence the emphasis on the 'formative' (Mullins & Kiley 2002), while at the same time recognising that the thesis (as the product of their research) has to exhibit specific qualities and meet an accepted set of standards (the summative). However, the qualities and 'standards' have proved difficult to 'tie down'.

...there is no consensus about precisely what a PhD is, and this lack of consensus has implications for the examination process. After all, it is difficult to be sure that a thesis and/or candidate has met the requirements of a PhD when one cannot define exactly what a PhD is in the first place. Examination and definition of PhDs are, therefore, inextricably intertwined...examiners' perspectives about what constitutes a PhD are of central importance... (Tinkler & Jackson 2004 p.8)

Denicolo (2003) found on the basis of a survey of 62 staff and students that 'the degree of consensus about the criteria for assessment is low' (p. 89).

This vagueness and 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, Leonard & David 2002; Shaw & Green 2002; Lawson, Marsh & Tansley 2003; Denicolo 2003). A heavy reliance in Australia and elsewhere is placed on the collective experience of examiners and knowledge of 'peer review' (Wisker, Robinson, Trafford, Warnes & Creighton 2003, p.64; Grabbe, 2003, p.130; Tinkler & Jackson 2004, p.119).

Concerns with respect to the examination of the thesis tend to coalesce around four themes relating to examiners and their judgements: 'examiner selection', 'examiner conduct', 'examiner response to examination criteria', and 'examiner consistency'.

Examiner selection

The issues with respect to examiner selection are how examiners are chosen, their number and their 'qualifications' to examine. Most Australian universities call for three examiners, the rest two (Lawson, Marsh & Tansley 2003). Two Australian studies have shown that

approximately fifty per cent of examiners are selected from institutions outside Australia (Pitkethly & Prosser 1995; Bourke, Hattie & Anderson 2004). This proportion of international examiners is uncommon. In other countries a combination of internal and external (but often same nation) examiners is more common.

As well as ensuring the examiners have the relevant experience and reputation, the selection of external examiners is intended to contribute an 'independent' perspective. That 'independence' might not be achieved through this strategy is also presented in the literature (Kamler & Threadgold 1997).

In Australia most universities require 'the examiner to have a degree equivalent to that which they are examining' (Lawson, Marsh, & Tansley 2003, p. 33). Formal training in examination is not mandated and therefore is not common (Tinkler & Jackson 2002). It is worth noting that 'training' for peer review is not mandatory either. There is a very strong assumption that discipline expertise translates into effective assessment practice. Recent work with supervisors about the viva, however, suggests they are not completely confident about their readiness (Cryer & Mertens 2003; Denicolo 2003; Powell & MacCauley 2003).

In a study of one Australian university the reasons for examiner selection were investigated (Bourke, Scevak & Cantwell 2001). Of the multiple reasons listed for the selection of each of 300 examiners across disciplines it was found that 77% were chosen on the basis of 'expertise' in the topic or methods, 59% by 'reputation', 40% by their publications, and 36% by their experience as examiners and/or supervisors. Certainly examiners believe they are chosen on the basis of expertise and because they can uphold standards (Jackson & Tinkler 2000, Mullins & Kiley 2002).

Examiner conduct

There is a growing literature that provides evidence from various perspectives on examiner conduct, that includes studies reporting on student experience of engagement with examiners during the viva and also examiner perspectives (Hartley & Fox, 2002; Grabbe 2003; Jackson & Tinkler, p.38, pp.70-74). There is an unease expressed in some of the literature that the assessment responses of some examiners are less than objective and are driven by agendas that extend too far outside the thesis under review, for example protecting academic territory or personal reactions to particular approaches or topics.

Some literature reports perceptions about the degree of examiner experience and their relative 'toughness' as assessors (Lawson, Marsh & Tansley 2003). The common assumption in the Australian context is that inexperienced examiners are 'tougher' (Mullins & Kiley 2002). However, in a study of 101 candidates' reports at one institution it was found that it was the examiners who were chosen because they were 'experienced' tended to make less positive comment and were more prescriptive in their instructive comment. Nonetheless overall examination experience did not impact on examiner recommendation. In the same study there was no evidence that the examiners chosen because they were experienced supervisors made significantly different comments or recommendations than those not chosen for those qualities (Holbrook, et al. 2003). In the UK Trafford (2003) found, as a participant-observer for 25 vivas that questions tended toward the 'diagnostic', in general however, relatively inexperienced examiners asked different sorts of questions – less conceptual ones, 'leaving the more experienced examiners to explore contributions to knowledge and the conceptualisation of the research process' (p.122).

Examiner response to criteria

‘Do examiners follow guidelines?’ is a question that raises some concern among administrators. The more powerful question is whether examiners are responding consistently to the same set of criteria in assessing a thesis.

Another critical set of questions are how much are examiners responding to the thesis as they would to completed research in a peer review context, how much to the demonstrated learning, and how much to other ‘indeterminate’ factors? (Tinkler & Jackson 2004, p. 170)

Denicolo (2003) notes that there is aversion among academics in England (where procedures to develop standardised criteria have developed furthest) to fixed criteria (p. 87). But it stands to reason that if a consistent pattern of application of criteria were to be discovered, and if consensus was reached among academics that the pattern provided an appropriate baseline set of standards, this would provide the clearest framework for PhD assessment, examiner training, and the articulation of ‘curriculum’ (National Research Council 2001; Gilbert 2004). Several attempts have been made to articulate such criteria but there has been no replicable, empirical validation within or between disciplines. Hence standards remain undetermined and a matter of serious concern to Quality Agencies (Tinkler & Jackson 2004, p.119).

Examiner consistency

Possibly some of the most powerful mythology in PhD examination derives from stories of substantial differences between examiners, but how frequent are such differences and are they mainly differences in recommendation or in the comments? We also rarely ask whether the committee is consistent in their treatment of such differences or how re-examination or adjudication functions with respect to achieving a ‘consistent’ re-examination outcome. In one rare case of re-examination where one the original examiners chose not to participate in the re-examination and was replaced, an insight was gained into the degree to which examiners become co-researchers. The first examiner gave fulsome advice, and it was in those areas of advice that the new examiner found much to support in the re-written thesis. (Holbrook, Bourke, Lovat & Dally 2004a). In Australia the examination emphasis is on the thesis alone. A viva is neither mandatory nor commonly employed, although there is evidence from a study in England (Tinkler & Jackson 2000; Jackson & Tinkler 2001) to show that a viva (which is held after the examiners have read the thesis) is unlikely to produce information that will change examiner recommendation.

The selection and procedures of the institutional committee that makes the ‘official’ decision on examiner selection and thesis outcome is rarely raised as an issue in the literature, although procedures for adjudication where examiners differ have been described (Lawson, Marsh & Tansley 2003; Tinkler & Jackson 2004). Until the study described below in this paper, there has been no sustained analysis of examiner recommendations on the same thesis and between examiner recommendations and the official committee decision.

This paper focuses on examiner comment in relation to examiner and ‘official’ committee recommendations, and specifically examines ‘inconsistencies’ in recommendations made about theses from Australian universities.

Approach

This paper arises from a national project investigating doctoral examination through the use of examiner reports and candidate information. The research questions guiding the larger 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. The mixed methods design has been described in full elsewhere (Holbrook and Bourke. 2004). The complete data set will comprise the examiner reports, examiner and institutional recommendations and the doctoral candidate enrolment and supervision history on record for 100 of the most recent doctoral completions in eight universities. The candidates represent all broad fields of study.

The most substantial source of data is the examiner reports, which range, on average, between two and three pages in length in standardised format. The content of the reports is coded into categories and entered into *QSR N6* software. In the core phase the analyses are based on the proportions of total text units coded at categories, as well as instances of code occurrence and text intersection. There was an extended process of trialling the categories, the analyses and coder consistency. Detailed coding notes were generated. The core coding categories capture all of the characteristics and content of the reports

The core analyses are informed by and contribute, in turn, to ‘extended’ analyses. The different strands of data and analysis enable the researchers to integrate and validate findings and systematically and creatively explore new directions and emerging theories. The core coding categories are built on a hierarchical structure of five parent coding categories. There are sub-categories for each parent category. The act of coding occurs at the sub-category level. Each coding category has a name as well as a numeric designation and these in turn represent the levels of coding (see appendix).

The primary coding categories are:

1. Report organisation

Examiners can approach report writing in many and varied ways. This category focuses not on text content, but how the examiner organizes and structures their report, including the impact of institutional instructions.

2. Examiner and process

Examiners may make comments about what they know or anticipate about examination and standards, they may comment on what they believe the candidate’s institution expects, and they may talk about themselves – e.g. the extent of their expertise. Such comments capture their approach to examination, the interpretation of their role and their expectations. They may discuss the expectations or ‘state of play’ within the discipline or field, and its methods. They may acknowledge regional and institutional differences and individuals, including the supervisor. All such areas of comment are captured under this broad heading.

3. 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 heart. The topics that are typically addressed include scope and significance, the literature review, the methods used and the substance of the findings (including interpretation and analysis of the data, also the presentation and communicative competence of the candidate. There are no absolute surprises in the areas identified here,

and they might be found in any guide to writing a thesis. They have also been identified under similar headings in other research into report content (Hansford & Maxwell, 1993; Johnston, 1997).

4. 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). This category concentrates on how examiners convey their response, once again in broad ways at this core level, e.g. such as the use of first person, the depth and type of engagement with the thesis, and the directness of their approach to the reader.

5. Evaluative elements

This category captures all comment that contains evaluation and judgement, including different types of instructive comment as well as text identifying positive and negative judgements about various aspects of the thesis and the candidate's capacity to conduct research.

Examiner consistency

One of the key questions about process in the study is the degree of consistency between examiners and between the examiner recommendations and the institutional recommendation. This paper takes the opportunity to focus on specific differences in recommendations based on a sample of 401 PhD candidates at four Australian universities.

Depending on individual university regulations, PhD theses in Australia are examined by two or three independent examiners, who are normally external to the university. A university committee then makes a decision on the thesis based on the examiner recommendations made and on their written reports. Each of the recommendations recorded here is one of five categories ranging from 'accept the thesis as submitted' to 'fail' without the possibility of revision and resubmission for further examination. The reports typically average 132 lines of text or almost three pages, but range from 1 line to 1272 lines in length. 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 usually not as straightforward as it is when there is clear agreement between examiners on the worth of a thesis. The university committees normally state clearly that examiner recommendations are not simply 'averaged out' to obtain a decision, but that the content of the written reports are considered closely in determining their decision, particularly when there are wide discrepancies in examiner recommendations on a thesis.

When there are widely different recommendations made by examiners on a thesis, there are at least two types of discrepancies possible. First, an examiner making a favourable recommendation on a thesis while the other examiners recommend that the thesis be revised and resubmitted may apparently not influence the decision made by the university committee when that decision is an unfavourable one. Secondly, there is the opposite case of an examiner who is more critical than others of the thesis. In this case, a favourable decision by the committee suggests that the recommendation of the critical examiner may have been disregarded. Employing data collected from four Australian universities, this paper investigates differences in the content of examiner reports for 12 theses with these two types of discrepancy. The current investigation attempts to identify what it is about the

examiner reports that influences the subsequent decision of the committee when it is faced with examiners making conflicting recommendations on a thesis.

Taking the first group where at least one examiner was far less critical – there were six theses where one examiner recommended that the thesis be accepted or required only minor amendment while, with one exception, the other two examiners recommended and the committee required that the thesis be revised and resubmitted for further examination. In the exceptional case one of the other examiners recommended that the thesis be failed outright. There was one other thesis where two examiners recommended the thesis be accepted as submitted but one examiner recommended and the committee required that the thesis be revised and resubmitted. In these cases, the favourable report of one or, in the last-mentioned case, even two examiners was not convincing for the committee. There were 21 examiner reports associated with these seven theses which require detailed content examination when investigating cases with discrepant examiner recommendations and less favourable committee decisions.

For the second group where at least one examiner was far more critical, there were four theses where one examiner recommended revision and resubmission for further examination or that the thesis be failed outright, while the other two examiners recommended the thesis be accepted unchanged or that it required no more than minor correction, and the university committee required only minor correction. The report of one examiner in each of these cases would seem not to have influenced the committee decision. In addition, there was one thesis where two examiners recommended that the thesis be revised and resubmitted, the other examiner recommended it required significant correction but the committee required only minor correction. In this case, the similar recommendations of two examiners failed to influence the committee and, unusually, the committee decision was more favourable than the recommendation of any of the three examiners. Thus there were five theses in the second category (and 14 examiner reports) that require scrutiny when investigating cases of discrepant examiner recommendations and more favourable committee decisions.

In summary, the distribution of examiner recommendations and committee decisions for the 12 theses with discrepancies is shown in Table 1. The 14 discrepant, and presumably largely disregarded, examiner recommendations are shown **in bold**. Group 1 is comprised of those cases where the committee was more demanding than at least one of the examiners and Group 2 is comprised of those cases where the committee was more lenient than at least one of the examiners.

Once the relevant theses were identified, the contents of the examiner reports were compared using the percentages of text coded at each of 29 sub-categories across four major categories – (1) Examiner and process, (2) Assessable areas covered, (3) Dialogic elements and (4) Evaluative elements. 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.

For each group, the proportions of comment that the ‘discrepant’ examiners devoted to different aspects of the thesis were compared with the content of ‘consistent’ examiners 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 disparity in the examiner ratings, it was anticipated that the comments of the examiners requiring revision or resubmission (hereafter referred to as ‘dissatisfied’ examiners) would be significantly different to the comments of the examiners who judged the thesis to be

acceptable or to require only minor corrections (hereafter referred to as ‘satisfied’ examiners). However, contrary to these expectations, the quantitative analysis of the report content for both groups revealed that there were very few differences between the reports of ‘satisfied’ and ‘dissatisfied’ examiners.

Table 1. Summary of committee decision and examiner recommendations for 12 theses

CANDIDATE/ THESIS	COMMITTEE DECISION	EXAMINER 1	EXAMINER 2	EXAMINER 3	
G R O U P 1	1301	Revise & resubmit	Revise & resubmit (1258)	Invite correction (1259)	Revise & resubmit (1260)
	1509	Revise & resubmit	Invite correction (1158)	Revise & resubmit (1228)	Fail (1250)
	1708	Revise & resubmit	Revise & resubmit (1066)	Revise & resubmit (1074)	Accept as submitted (1080)
	1906	Revise & resubmit	Accept as submitted (1159)	Revise & resubmit (1202)	Revise & resubmit (1254)
	4005	Revise & resubmit	Fail (1013)	Accept as submitted (1014)	Revise & resubmit (1015)
	4006	Revise & resubmit	Accept as submitted (1017)	Revise & resubmit (1018)	Revise & resubmit (1019)
	4014	Revise & resubmit	Accept as submitted (1041)	Revise & resubmit (1042)	Accept as submitted (1043)
G R O U P 2	2041	Invite correction	Revise & resubmit (116)	Require correction (117)	Revise & resubmit (118)
	2049	Invite correction	Invite correction (132)	Revise & resubmit (133)	Accept as submitted (134)
	2080	Invite correction	Revise & resubmit (238)	Accept as submitted (239)	Accept as submitted (240)
	2090	Invite correction	Invite correction (264)	Invite correction (265)	Fail (266)
	3060	Invite correction	Fail (1060)	Invite correction (2060)	No 3 rd examiner

Of the 30 points of comparison there were only three significant differences between the satisfied and dissatisfied examiners in Group 1 and only one significant difference in Group 2. Across both groups there were no differences between ‘satisfied’ and ‘dissatisfied’ examiners in the amount of text they devoted to aspects such as negative judgement about the thesis, formative instruction aimed at assisting the candidate to improve the thesis, or comment about the originality and significance of the study. The lack of significant differences in the current comparison of reports from ‘satisfied’ and ‘dissatisfied’ examiners suggests that although these groups of examiners gave substantially different recommendations, the content of their reports was quite similar. Since a previous study investigating theses that were required to be revised and resubmitted analysed the reports of four of the candidates in Group 1 (Holbrook et.al 2004a), the following discussion briefly examines the nature of the inconsistencies in the Group 1 reports on the theses that the committee decided should be revised and resubmitted, and provides a more detailed analysis of the reports from Group 2 (theses with correction invited).

Reports on theses to be revised and resubmitted

For each of the 7 candidates in Group 1, the committee decision was for the candidate to 'revise and resubmit' the thesis. Of the 21 examiners for these 7 candidates, there were 13 'dissatisfied' examiners whose recommendation was consistent with the committee decision (including two examiners who recommended 'Fail') and 8 'satisfied' examiners whose recommendation was inconsistent with the final rating. The inconsistent ratings included six recommendations to accept the thesis as submitted and two recommendations to accept the thesis with minor corrections. As stated above, the quantitative analysis suggested that despite the discrepancies in the recommendations, the reports of the satisfied and dissatisfied examiners were remarkably similar. The only significant differences between the examiners reports, were that satisfied examiners wrote more about positive aspects of the thesis and its potential for publication, while dissatisfied examiners gave more prescriptive advice or instruction about how to 'fix' mistakes.

A qualitative analysis of the report text for this group revealed that although the 'satisfied' examiners often provided glowing endorsements of the merits of a thesis, they also devoted a considerable proportion of their reports to describing inadequacies in the thesis and providing 'formative' instruction as to how candidates could improve their work 'for publication'. The quantitative analysis indicated that satisfied and dissatisfied examiners devoted the same proportion of text to negative judgements or instructive advice about aspects such as the literature review, the methodology and the conclusions. However, the main difference between the reports of the satisfied and dissatisfied examiners was that the satisfied examiners provided their judgements and instruction in the context of 'advice to be considered for publication of the thesis', while the dissatisfied examiners stipulated that the same recommendations needed to be attended to in order to meet the requirements for awarding the degree.

In the case of candidate 4005, there was one "Accept as Submitted", one "Revise and Resubmit" and one "Fail" recommendation. Despite the differences in the ratings provided, all three examiners commented on the preponderance of secondary literature and a failure to cite primary sources. The two dissatisfied examiners questioned whether the candidate had conducted a sufficiently 'scholarly exploration' of the primary sources and listed other sources that should have been included. The satisfied examiner commended the candidate and her supervisor for the 'very high quality of the intellectual, scholarly and presentational aspects of the thesis' but noted that the thesis was 'strange' because of its 'non-fashionable framework'. Although this examiner was prepared to accept the thesis as submitted, the examiner made a number of suggestions as to how the 'book version' of the thesis could be improved by incorporating many of the same theoretical omissions noted by the other examiners. The committee decision of 'revise and resubmit' appears to suggest that both the 'Accept' and 'Fail' recommendations were disregarded. However, the examiner who made the recommendation of 'Fail' was in agreement with the other examiners that the candidate had put an extraordinary amount of effort into the study. 'If doctorates were awarded for industry alone, this thesis would receive a distinction' (Examiner 1015). While evidence of 'indefatigable work ethic' was apparently insufficient to convince this examiner that the thesis was worthy of the award, the unanimous acknowledgement of the candidate's commitment may have influenced the committee towards leniency. Although the final decision by the committee appeared to be consistent with only one of the examiners recommendations, closer inspection of the content of the three reports reveals that the committee may well have considered the judgements of all three examiners.

The case of candidate 4006 reveals that there was a common concern among the examiners about the adequacy of the statistical methods employed and, as a consequence, about the reliability of the conclusions drawn from this 'dubious' analysis. All three examiners urged the candidate to rewrite some of the 'questionable conclusions' and to demonstrate the significance of these findings to the wider field. As in the previous case, the satisfied examiner prefaced his/her comments and suggestions with the proviso that these points were made 'with a view to possible publication of the thesis' while the two dissatisfied examiners recommended deferral of the award so the thesis could be revised and 'radically improved'.

The final candidate in Group 1 was the most extreme of the cases in which the committee required resubmission as the committee decision was inconsistent with two examiners who both recommended that the thesis should be accepted without change. As was evident in the previous cases, despite the favourable ratings, the comments in the satisfied examiners reports were couched in the same admonishing tone as those in the dissatisfied examiners report. The following are excerpts from the 'satisfied' examiners reports:

I think this student makes gratuitous generalizations that result in "stereotypes".

I found the discussion...a bit incoherent.

*It would be valuable for X to be a bit more thorough in her discussion of...
I wonder if one might not need to study...over a much longer period of time...in
order to get really accurate results.*

*There is a need for sustained contact, otherwise one is left with "impressions" and
"anecdotes" rather than a convincing body of hard, statistical facts.*

The work is very subjective.

*Although the thesis as a whole does not demonstrate brilliance in my opinion, and
would not lead to publication, I recommend that it be awarded a pass.*

The committee decision to require revision and resubmission, apparently in accord with only one examiner's recommendation, is understandable in light of the relatively numerous negative comments provided in the 'satisfied' examiners reports. Thus, the preceding analysis of a sample of the resubmission cases, suggests that an examiner's positive recommendation may be undermined by the extent of instructive comment or negative evaluation contained in the written report. The subsequent analysis of the Group 2 inconsistencies reveals further evidence that although the committee decision, at times, appears to be inconsistent with the recommendations of some examiners, the content of both the satisfied and dissatisfied examiner reports reveals a surprising consistency in tone and comment, if not in final judgement.

Reports on theses with correction invited

For each of the five candidates in Group 2, the committee decision was to 'invite correction' of the thesis. Of the 14 examiners for these five candidates, there were eight examiners whose recommendation was consistent with the committee decision and six examiners whose recommendation was inconsistent with the final rating. The inconsistent ratings included four recommendations to revise and resubmit a thesis and two recommendations that the candidate should fail. The quantitative analysis revealed that the

only significant difference between satisfied and dissatisfied examiners in this group was that 'satisfied' examiners made more comments referring to the supervisor, typically congratulating the candidate and their supervisor on the successful completion of the degree.

A qualitative analysis of the report text revealed that, despite the discrepant examiners giving a 'harsher' recommendation, their reports often contained effusive remarks about the thesis, including its contribution, the way the experiments and analyses were conducted and the 'critical thought' contained in the literature review. Conversely, despite rating the thesis as 'acceptable', the satisfied examiners also identified inadequacies in the theses and it was often difficult to differentiate the tone and content of the 'satisfied' examiner's reports from the reports of examiners who were 'dissatisfied' with the quality of the thesis. The following examples illustrate that despite their recommendation, examiners were remarkably consistent in identifying areas of weakness in a particular thesis, with similar concerns raised by both the satisfied and dissatisfied examiners. The common concerns typically included the methodological approach, the writing style, the depth or breadth of the literature review and the reliability and applicability of the results. The main differentiation in the assessment outcomes appeared to reside in the degree of importance that each examiner ascribed to the identified weakness. Some examiners treated the deficiency as a crucial and sometimes irremediable flaw, while others noted that there were significant problems but acknowledged that there were other redeeming features of the work that outweighed the poorer components. At times the same feature was judged to be both a strength and a weakness (even by the same examiner).

The three examiners for Candidate 2049 all mentioned the 'broad sweep' of the literature review. While the 'dissatisfied' examiner gave the harshest pronouncement, even the satisfied examiners suggested that the literature review needed to be more focused and the irrelevant details omitted. Perhaps the committee is more likely to condone what appear to be deficiencies in a thesis if some examiners also noted these flaws but have nonetheless judged the thesis to be acceptable. In the cases under consideration, the concerns raised by the dissatisfied examiners were also canvassed by the satisfied examiners, but were judged to be either acceptable, or at least inconsequential, in determining the overall merit of the thesis. In the case of Candidate 2080, the discrepant examiner criticized the way the thesis was written and objected to the 'fragmentary way in which themes are treated', the "superficial" analysis of the literature and the "glaring omissions from the primary material cited". For this examiner, these deficiencies were of sufficient magnitude to recommend that the thesis be revised and resubmitted so that the candidate could develop the themes "more fully" and locate the study in a "wider context". The satisfied examiners, who both recommended that the thesis should be accepted as submitted, also made numerous comments about the lack of depth in the analysis of the literature and its "narrow focus". For example:

These remarks about...are too sweeping.

This is an important point and should be developed.

This discussion should be enlarged.

The conclusion to the opening paragraph could well be taken further.

This opportunity to speculate on... is not taken up. It looks serious and deserves a further discussion.

He raises some interesting issues but sometimes does not carry the discussion of them far enough.

However, despite these often repeated criticisms, both examiners were prepared to accept the thesis without change. The examiners were not oblivious to the flaws in the study but perceived their role as 'weighing up' the strengths and weaknesses of the thesis. For these two examiners at least, the balance came out in the candidate's favour. There was sufficient evidence of competency in other areas to convince these examiners that, despite some evident deficiencies in the thesis, the candidate was worthy of the award.

What might have been an excessive narrowness of focus is redeemed by the fact that he introduces comparative discussion of....

These 'lapses' do not take away from the fact that this is an original contribution to knowledge.... Its strengths as a good piece of scholarship ... easily outweigh the criticisms made above.

In light of the evidence provided to the university committee it is understandable that the committee acceded to the majority opinion of the 'satisfied' examiners. These examiners demonstrated that they were aware of the deficiencies identified by the discrepant examiner, but judged that these 'lapses' did not detract from the overall worth of the thesis.

The situation of examiners identifying similar features of concern but coming to different judgements about the impact the faults have on the overall quality of the thesis was also evident in the case of Candidate 2090. The critical issues for the dissatisfied examiner were inadequacies in the literature review, a lack of the requisite details about the methodology and a 'selective' interpretation of the data. This examiner judged these flaws, particularly the extensive but unfocused breadth of the literature review to be 'fatal' and gave a recommendation of 'Fail'. Although the two satisfied examiners identified these same aspects and commented on them at length (with one examiner providing seven pages of suggestions and corrections), these examiners concluded that the "quantity of errors" did not outweigh the quality of the thesis. The breadth of the literature review, while attracting considerable criticism from all examiners, was commended by one of the satisfied examiners as being "refreshingly broad", with this same examiner indicating that this fault was not beyond repair. "It would take very little to focus the literature review on the topic". For the satisfied examiners the crucial indicator of quality appeared to be the significance of the thesis and its potential for publication. The satisfied examiners were consistent in their judgement that the thesis made a "substantive original contribution to knowledge" and that the candidate demonstrated the "capacity to conduct research independently and at a high level of originality". As was evident in the previous case, the satisfied examiners clearly identified the same weaknesses as the dissatisfied examiner, but instead of judging the thesis as irremediable, provided extensive information about how to 'fix' the problems. Thus, for these examiners, the thesis was not only 'repairable' but the perceived flaws could be forgiven because of the compensating strengths of the thesis, in particular, its significance and contribution. As in the previous case, the committee decision to invite correction is understandable, given the majority view of the examiners was that the flaws were remediable and the thesis had sufficient merit to warrant the award.

The resemblances between the reports of satisfied and dissatisfied examiners was also evident in the case of Candidate 3060. There were only two examiner reports for this candidate because that was the requirement of the institution. Although the examiner recommendations were disparate (one examiner inviting corrections and the other

recommending Fail) again the report contents were remarkably similar. The critical issue for the dissatisfied examiner was that the study had failed to produce reliable data and therefore this examiner did not believe that the study had made a significant contribution, since the results could not be applied to “practical clinical problems”. Despite providing a recommendation of ‘Fail’, this examiner commended the candidate for his “well conceived and original experimental approach to an important clinical problem” and for the “substantial critical thought” that was evident in the literature review, the presentation of the results and the discussion. This examiner also congratulated the candidate for addressing and acknowledging the limitations and shortcomings of the methodology. “The thesis is well written and the problems encountered are described in a forthright manner. This thesis provides a wealth of information on the problems that can be encountered with the methods and techniques employed”. However, it appears that this examiner values the “product” of a research degree over the “process” of research training that doctoral study entails.

The satisfied examiner also had reservations about the methods employed and questioned why other experimental approaches were not attempted but did not regard the lack of conclusive findings as a major problem. For this examiner, the candidate’s demonstrated capacity to conduct independent research, to devise an original experimental approach and to interpret the findings with a clear understanding of the limitations of the approach were sufficient to merit the award of a doctoral degree. Presumably the university committee concurred with the view that the process of research training is more important than the research results. Although, the committee decision to ‘invite correction’ appears to disregard the “Fail” recommendation of one examiner, the favourable remarks in the dissatisfied examiner’s report about the design and conduct of the study actually corroborated the assessment of the satisfied examiner and are supportive of the committee’s final decision.

The preceding analysis of examiner reports on theses for which there were major inconsistencies between the examiner recommendations, tends to refute the notion that in making their final determination, the university committee ignored the discrepant examiner’s assessment. Although this appeared to be the case when considering the “ticked box” only, a close analysis of the examiners’ written comments revealed that the majority of positive judgements as well as the concerns about perceived weaknesses were common in the reports of both the satisfied and dissatisfied examiners. Even in the most extreme cases, where one examiner gave a recommendation of Fail and another examiner was content to accept the thesis as submitted, there were remarkable similarities in the types of comments provided.

In both types of cases, including both the more harsh and the more lenient committee decisions, the written reports of the inconsistent examiners contained remarks that substantiated the majority judgement and the final decision taken by the committee. For the cases in which the committee was more harsh than at least one of the examiners, the reports of the satisfied examiners typically identified similar shortcomings as the reports of dissatisfied examiners. Although these comments were provided in the context of improving the thesis for publication, the committee nonetheless may interpret these comments as implied criticisms, especially when the same issues are nominated by other examiners as **requiring** amendment.

For the cases in which the committee was more lenient than at least one of the examiners, once again, both satisfied and dissatisfied examiners identified similar weaknesses. In these cases, however, the satisfied examiners typically dismissed the importance of the

perceived inadequacies when weighed against the strengths of the thesis in other areas and typically the reports of the dissatisfied examiners contained as much positive comment as the reports of the satisfied examiners. Thus, the committee's decisions appear to be congruent with what examiners write in their reports and to be influenced more by the examiners' written remarks than by the box they ticked. As discussed in the previous analyses, in 11 of the 12 cases under investigation, the committee appeared to take heed of the written reports of **all** examiners. However, there was one exception.

An anomalous case

The last section of this paper analyses one anomalous case in which the committee decision was not consistent with any of the examiner recommendations. This case provides an appropriate context in which to apply conceptual lenses around the notions of 'expert knowledge' and 'power'. These themes have been explored in previous analyses of the PhD examination reports (Holbrook et al. 2004a; Lovat 2004; Monfries & Lovat 2004) after Habermas (1972; 1974); Corson (1995). It is a fascinating exercise to make such application to the issue of the University Committee's deliberations on the basis of an apparent blanket dismissiveness of the examiners' reports.

The thesis in question (Candidate 2041) received three fairly scathing reports. Two reports recommended revision and re-submission, while the third report, even more damning in part, recommended a huge list of required corrections before acceptance. Yet, the University Committee's decision was the far lesser one of inviting corrections, essentially accepting the situation that the apparently flawed thesis might go into the Library uncorrected. This decision essentially waived as requirements all the essential corrections identified by the three examiners.

It is difficult to detect from the examiner reports any information that might have supported such a decision. There was the occasional positive text, identifying some of the contributions that the thesis would make to the field, but nothing so substantial that it could have outweighed the clear advice from all three examiners that the thesis, as it stood, was flawed in major ways and, for all three of the examiners, in ways that were beyond repair in its present form.

All three reports are intensive, together constituting 29 pages of single spaced 11 point font. Negative comment and corrections are embedded throughout, ranging from significant content error to major presentation flaws. Such comments include the following:

This paragraph is incorrect

It is not one of the factors at all

This paragraph should be deleted

The following references were not found in the bibliography

... incorrect numbering of tables and confusion about the content

This thesis is clearly unsuitable for awarding a PhD

The poor quality of presentation of the thesis means that X should be required to revise the thesis for re-examination.

Following directly on from this last comment is a list of specific errors that runs for 4 pages, including almost 2 pages of references not found in the thesis.

It could be that the committee was aware of an exigency unknown to the examiners and, naturally, to the investigators of the PhD project. Even if so, the fact remains that a body of work purporting to be an original contribution to knowledge in the area was, according to the expert opinion of the examiners, fraught with content error and presentation flaws. The decision of the committee over-ruled this expert opinion and allowed the work to stand as an original contribution to knowledge in the area, without any requirement to correct or amend in any way the content errors and presentation flaws.

As suggested, there are numbers of comments from the examiners around redeeming qualities of the thesis, such that might have moved the committee to soften the majority recommendation around revision towards the minority recommendation that the candidate be required to make corrections. There is, however, no apparent justification for the degree of softening that occurred.

A case like this causes us to consider again notions of ‘expert knowing’ and ‘power’ in relation to the business of the doctorate and its examination. Assuming that there were not extenuating circumstances unknown to us and that this result occurred as a regular part of the process of final deliberation, could it be indicative of a disposition on the part of some University Committees that they possess an expertise about doctoral standards and quality that, if push comes to shove, has the capacity to superordinate the expertise even of the examiners? In this case, what role do the examiner reports truly play in the process of final adjudication?

In similar fashion, could there be a disposition in the University committee towards seizing and maintaining the power over the process, perhaps brought to the fore when this power is challenged? In the case at hand, the examiners not only criticized and challenged the quality of the thesis, but also made a number of references to the quality of the processes of the host institution. Comments of this type included from one examiner: “This reflects poorly on the University ... ;” and from another examiner: “I strongly recommend that the University of X more closely scrutinize PhD theses sent for external examination to ensure that they at least contain minimal spelling and simple grammatical error.” In the context of a competitive regime around expertise, power and power maintenance, is it possible that such direct criticism could have the effect of affronting a committee, in turn leading to a disposition towards defence of the thesis on the basis that, at the end of the day, the committee ‘knows best’?

Discussion and Conclusions

It is not only PhD assessment practices that have been called into question in recent years, university assessment generally appears to be subject to many ‘imponderables’ (Elton 2004, p. 53). The qualities of ‘graduateness’ are proving as difficult to pin down as ‘postgraduateness’ and ‘doctorateness’. This is not entirely surprising given the focus for assessment studies have tended to be school level achievement where the importance of good practice is perceived as very pressing (National research Council 2001). Those who have reached tertiary education are seen to be in a different league and so funding for large

scale empirical studies is not easily achieved. In the current climate of Quality Assurance, however, the state of affairs in higher education has been thrown into sharper relief.

In the doctoral degree at least, to affix thesis assessment too hard to specific criteria, may eventuate in the loss of some very important qualities, possibly those linked to higher order thinking. Supervisors and examiners utilise language that indicates that they can both ‘feel’ and articulate when a thesis demonstrates the appropriate combination of properties (Mullins & Kiley 2002; Cantwell & Scevak 2004).

The assessment discourse of examiners suggests they share the same expectations about thesis quality and there are processes in place (e.g., the reading of reports by committee) that ensure these understandings are equated in terms of judgements. The lack of clear standards under these circumstances may be more frustrating to those outside the community than actually indicative of poor processes or variable standards. The cases presented in this study suggest a high degree of consensus about substantive issues of quality.

This paper focussed on ‘consistency’ in examination process. In summary, the distribution of examiner recommendations and committee decisions for the 12 theses with discrepancies were explored by means of an approach that combined text analysis, text counts, and correlation analyses. It is notable that there were only 12 theses with significant discrepancies in recommendation in 401 candidate cases. With one exception the language of examiners was synchronous with the outcome, whereas their recommendations were not. The committee that adjudicated on the differences, drew on the comments over the discrepant recommendations and the official outcome indicated this. The one exception is inexplicable and is by its very nature one of those that goes to reinforce the ‘myth’ that the process is wholly at fault, whereas the 11 other cases, and the many more that are not ‘discrepant’, suggest the multiple reporting mechanism proves to be extremely robust in facilitating the achievement of equitable outcomes. The one case in 401 that does seem inexplicable perhaps should not be asked to carry too much weight in the debate about examiner consistency.

It remains, however, that similar issues, concerns and evaluations noted by examiners in their comments on theses sometimes lead to discrepant recommendations – in this case in just 3 per cent of cases. It may be that the recommendation framework commonly used in Australia is at the basis of this ‘problem’. What we are coming to know about the criteria PhD examiners consistently apply could be used to modify or add to the framework. At least one of the outcomes of our work applying examiner commentary to notions of thesis quality would desirably be a set of stable indicators that allows the academic community and others to distinguish clearly between theses of threshold quality and those of higher and the highest quality.

References

- Bourke, S., Hattie, J. & Anderson, L. (2004). Predicting examiner recommendation on PhD theses. Paper presented at the EARLI, Higher Education Conference, Stockholm, July.
- Bourke, S., Scevak, J. & Cantwell, R. (2001). Ph. D. examination and examiner characteristics. Paper in a symposium, *Frontiers in Research Training: Evaluating Ph. D. examination and supervision*, at the Annual Conference of the Australian Association for Research in Education, Fremantle. Published on the internet at <http://www.aare.edu.au/01pap/BOU01588.pdf>
- Cantwell, R. & Scevak J.(2004). Discrepancies between the ‘ideal’ and ‘passable’ doctorate: Supervisor thinking on doctoral standards, Full refereed paper presented at the Annual conference of the AARE, 28 November-3 December.

- Cryer, P. & Mertens, P. (2003). The PhD examination: support and training for supervisors and examiners. *Quality Assurance in Education*, 11 (2), pp.92-99.
- Denicolo, P. (2003). Assessing the PhD: A constructive view of criteria. *Quality Assurance in Education*, 11 (2), 84-91.
- Elton, L. (2004). A Challenge to established assessment practice. *Higher Education Quarterly*, 58(1), 43-62
- Gilbert, R. (2004). A framework for evaluating the doctoral curriculum. *Assessment & Evaluation in Higher Education*, 29 (3), 299-309.
- Grabbe, L.L. (2003). The trials of being a PhD external examiner. *Quality Assurance in Education*, 11 (2), 128-133.
- Habermas, J. (1972). *Knowledge and human interests*. (Tr. J. Shapiro). London: Heinemann.
- Habermas, J. (1974). *Theory and practice*. (Tr. J. Viertel). London: Heinemann.
- Hansford, B.C. & Maxwell, T.W. (1993). A masters degree program: structural components and examiners' comments. *Higher Education Research and Development*, 12 (2), 171-187.
- Hartley, J. & Fox, C. (2002). The viva experience: examining the examiners. *Higher Education Review*, 35 (1), 24-30.
- Holbrook, A. & Bourke, S. (2004). An investigation of PhD examination outcome in Australia using a mixed method approach. *Australian Journal of Educational & Developmental Psychology*, 4, -109.
- Holbrook, A., Bourke, S., Lovat, T. & Dally, K. (2004a). PhD Theses at the Margin: Examiner Comment on re-examined theses, *Melbourne Studies in Education*, 45 (1) May, pp. 89-115.
- Holbrook, A., Bourke, S., Lovat, T. & Dally, K. (2004b). Qualities and characteristics in the written reports of doctoral thesis examiners. *Australian Journal of Educational & Developmental Psychology*, 4, 110-129.
- Holbrook, A. Dally, K., Cantwell, R., Scevak, J., Bourke S. & Lovat T. (2003) Doctoral examiners as supervisors. Presentation and Proceedings of the 2003 HERDSA conference, Christchurch NZ, 6-9 July, <http://surveys.canterbury.ac.nz/herdsa03/pdfsnon/N1083.pdf>.
- Jackson C. & Tinkler, P. (2000). The PhD examination: An exercise in community-building and gatekeeping, in Ian McNay (ed.) *Higher Education and its Communities*. (pp. 38-50). Buckingham: Society for Research in Higher Education and Open University Press.
- Jackson, C. & Tinkler, P. (2001). Back to basics: a consideration of the purposes of the Ph. D. *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.
- Kamler, B. & Threadgold, T. (1997). Which thesis did you read?, in Z. Golebiowski (ed) *Policy and Practice of Tertiary Literacy*. Selected Proceedings of the First National Conference of Tertiary literacy: Research and Practice. (pp. 42-58). Melbourne: Victoria University of Technology.
- Lawson, A., Marsh, H., & Tansley, T. (2003). Examining the examiners. *Australian Universities Review*, 46 (1), 32-36.
- Lovat, T. (2004). 'Ways of Knowing' in doctoral examination: how examiners position themselves in relation to the doctoral candidate. *Australian Journal of Educational & Developmental Psychology*, 4, 130-136.
- Monfries, M. & Lovat, T. (2004). Power discourse in PhD examination reports: A cross-disciplinary analysis. Paper LOV041000 presented at the Annual Conference of the AARE, Melbourne, 28 November – 2 December.
- 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.
- 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.
- National Research Council (2001). *Knowing what Students Know: The Science and Design of Educational Assessment*, Washington: National Academy Press.
- Pearson, M. (1999). The changing environment for doctoral education in Australia: implications for quality management, improvement and innovation. *Higher Education Research & Development*, 18 (3), 269-287.

- Pitkethly, A. & Prosser, M. (1995). Examiners' comments on the international context of Ph. D. theses. In C. McNaught & K. Beattie (eds) *Research into Higher Education: Dilemmas, Directions and Diversion* (pp.129-36). Melbourne: HERDSA.
- Powell, S. & McCauley, C. (2002). Research degree examining – common principles and divergent practices. *Quality Assurance in Education*, 10 (2), 104-15
- Powell, S. & McCauley, C. (2003). The process of examining research degrees: some issues of quality. *Quality Assurance in Education*, 11 (2), 73-83.
- Quality Assurance Agency for Higher Education (2001). *The National Qualifications Framework*. The National Committee of Enquiry into Higher Education.
- Remenyi, D. & Money, A. (2002) When is a doctorate dissertation ready for submission? *Irish Journal of Management* 23 (2) p.111.
- 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). Guidelines for examiners: An evaluation of impact. *The Psychologist*, September, 407-10.
- 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.
- Trafford, V. (2003), Questions in doctoral vivas: views from the Inside. *Quality Assurance in Education*, 11 (2), 113-21.
- Wisker, G., Robinson, G., Trafford, V., Warnes, M. & Creighton, E. (2003). From Supervisory Dialogues to successful PhDs: strategies supporting and enabling the learning conversations of staff and students at postgraduate level. *Teaching in Higher Education*, 8 (3), 383-97.

APPENDIX: Coding Categories in the PhD Examination Study

1 REPORT ORGANISATION

- 1 1 GENERAL SECTION**
 - 1 1 1 first
 - 1 1 2 last
 - 1 1 3 middle
- 1 2 CHAPTER SECTION**
- 1 3 ERROR SECTION**
- 1 4 REQUIRED AMENDMENTS**
- 1 5 CONCEPT, TOPIC**
- 1 6 METHOD, FINDINGS**
- 1 7 DIRECTIONS FOR READER**
- 1 10 OTHER 1**

2 EXAMINER AND PROCESS

- 2 1 PERSONAL &
PROFESSIONAL CONTEXT**
- 2 2 SPECIFIC & ANTICIPATED
CRITERIA**
- 2 3 THE MODEL PhD**
- 2 4 SUPERVISOR**
- 2 5 RECOMMENDATION**
- 2 10 OTHER 2**

3 ASSESSABLE AREAS COVERED

- 3 1 SCOPE, SIGNIFICANCE &
CONTRIBUTION**
 - 3 1 1 scope
 - 3 1 2 significance &
contribution
 - 3 1 3 publications arising
 - 3 1 4 existing publications
- 3 2 REVIEW OF THE
LITERATURE**
 - 3 2 1 coverage
 - 3 2 2 inaccuracy
 - 3 2 3 utilisation/ theoretical
application
- 3 3 APPROACH**

(cont.)

- 3 4 SUBJECT MATTER,
FINDINGS**
 - 3 4 1 analysis and reporting
 - 3 4 2 topic related issues
- 3 5 COMMUNICATIVE
COMPETENCE**
 - 3 5 1 substantial issues
 - 3 5 2 editorial issues
- 3 10 OTHER 3**
 - 3 10 1 ethics
 - 3 10 2 other 3
 - 3 10 3 ingredients

4 DIALOGIC ELEMENTS

- 4 1 INTELLECTUAL
ENGAGEMENT**
- 4 2 CONVERSATION**
- 4 3 FIRST PERSON**
- 4 10 OTHER 4**

5 EVALUATIVE ELEMENTS

- 5 1 SUMMATIVE**
 - 5 1 1 positive
 - 5 1 2 neutral
 - 5 1 3 negative
- 5 2 FORMATIVE
INSTRUCTION**
- 5 3 OTHER INSTRUCTION**
 - 5 3 1 commentary
 - 5 3 2 prescriptive
- 5 4 OTHER JUDGEMENT**
 - 5 4 1 positive
 - 5 4 2 neither
 - 5 4 3 negative
- 5 10 OTHER 5**
 - 5 10 1 other
 - 5 10 2 resubmission and/or
plagiarism