AARE FOCUS CONFERENCE:
QUALITY IN EDUCATIONAL RESEARCH

Paper Ref. No: BO05011Y *(Revised 26/6/05)*

Paper title:
Using examiner reports to identify quality in PhD theses

Sid Bourke, Allyson Holbrook & Terence Lovat
University of Newcastle

Abstract

Judgements about quality in doctoral research have two, sometimes competing, criteria to consider: the quality of the research training received and the quality of the research output, that is the thesis. Research reported in PhD theses represents a major component of all research undertaken in Australia. In Education, as many candidates are also recently or currently practising teachers and administrators, PhD research often has an immediate impact on schools and schooling. Thus the quality of theses is important as a contribution to our national research profile and professional practice.

A national study in progress has collected candidature information and examiner reports for 804 PhD candidates at eight universities across all discipline areas who submitted a thesis for examination – of these, 84 theses were in the field of Education. Currently the texts of examiner reports have been analysed for 401 candidates at four of the eight universities.

Evaluative comment (summative and other judgement) on each of the assessable areas covered that could be clearly recognised as positive (on average 28% of each examiner report) or negative (4%), have been analysed. The focus in this paper is on theses identified by examiners at both ends of the quality scale – high quality theses and marginal theses, first across all discipline areas and subsequently for theses in Education. The evaluative comments are used to illustrate what examiners write in their reports when assessing high quality and marginal theses. Suggestions for using examiner comment categories to distinguish thesis quality are made and discussed.

Presenter and contact:
Professor Sid Bourke
Faculty of Education & Arts
University of Newcastle
Callaghan NSW 2308
AUSTRALIA

Phone: 02 4921 5901
Email: sid.bourke@newcastle.edu.au

Using examiner reports to identify quality in PhD theses

INTRODUCTION

In the mid 1990s when the word 'quality' was used in relation to research it tended to be tied in a more of less explicit way to a peer-review of grants and articles. Since then the word has become the catch-all for 'benchmarking', 'standards', 'sound practice', 'good research' and a host of other manifestations of 'performance' evaluating research in tertiary settings including the training and education of researchers. The prevalence of 'quality' in government and management discourse in recent times is strongly related to the desire to measure all performance, including research performance. It is an instrumental treatment which is strangely at odds with the fundamental nature of quality. In essence, 'quality' captures what a community values (Harvey & Green, 1993). As Yates (2004) points out, a good doctoral thesis is 'good relative to a particular research community'. Quality in doctoral research is what is valued by thesis examiners and the standards and expectations of the academic community they represent. The value an examiner places on a thesis is evident in the recommendation made, and is explicated by the foci and emphases of the detailed comments provided.

Clearly not all theses exhibit the same level of quality. But a continuum of quality underlying examiner assessment of a thesis is premised on the assumption that the act of submitting a thesis for examination is in itself the first check of thesis quality. Examiners enter the examination process anticipating that the thesis will pass (Mullins & Kiley, 2002), perhaps for two main reasons. Normally a thesis would not be submitted for examination without at least one established academic colleague (the supervisor) believing it was ready for submission.[1] Secondly the amount of work involved in getting the thesis to submission stage is respected by examiners, who would normally also be supervisors of research students and know what was involved.

This paper is premised on the view that examiners can and do make distinctions between theses of varying quality. At one end of the continuum, an examiner judging a thesis to be of very high quality (or excellent) would probably see it as exhibiting all the valued and desirable criteria of good research as they define it, and this would be reflected in their report. At the other extreme, a thesis judged by an examiner to be unacceptable would be a thesis that does not at all meet the valued and desirable or even the essential criteria – and these are very rare indeed (Holbrook, Dally, Bourke, Lovat, Monfries & Lu, 2004). Slightly above the unacceptable, in the eyes of examiners, are theses that we have called 'marginal' (Holbrook, Bourke, Lovat & Dally, 2004a). Typically these theses exhibit some of the valued and desirable criteria, but lack others. These tend to be the theses for which examiners recommend that significant revision and perhaps re-examination of the thesis be required before it could be considered by the examiner to have sufficient quality to be considered worthy of a PhD.

We first say a little about the context of the studies reported here, then move to an overview of how PhD theses are assessed, before briefly describing a study of PhD examination, which includes the detailed analyses of the text of examiner reports. Following the identification of high quality or excellent theses and low quality or marginal theses based on examiner recommendations, the texts of examiner reports are used to identify specific areas of positive and negative comment by examiners that distinguish between the excellent and marginal

---

[1] We recognise that most university regulations allow that a student may submit a thesis for examination without their supervisor's approval. But this circumstance is rare and the possibility would be unlikely to be at the front of the examiner's mind.

theses. Next we focus on theses in the Broad Field of Study of Education, and a range of examples of these distinguishing comments is provided. Finally, suggestions are made of how thesis quality might be further evaluated, questions are raised concerning possible discipline differences in categories found to be important for thesis quality, and more general questions are addressed concerning what we might wish to do about any marked range in thesis quality.

THE CONTEXT

There are several elements of context in this area that are important for any consideration of assessing PhD thesis quality. These relate to the place of the PhD in our formal education system, the research impact of PhD programs, concerns about research training and output, and the forthcoming Research Quality Framework evaluation of Australian research (Roberts, 2005).

The place of the PhD

The PhD is clearly the pinnacle of formal qualifications in our education system – leading it to be termed the 'gold standard' (Scott, Brown, Lunt & Thorne, p.149) for research study, representing excellence and attracting both resources and prestige. Candidates are highly valued and hold a privileged position within universities, and in turn the expectation is that their research will provide an original and significant research outcome, in the form of a thesis.

PhD research impact

The numbers of PhD degrees awarded, across disciplines, has risen dramatically in recent years, as evidenced by Evans, Macauley, Pearson & Tregenza (2003) who reported an increase of almost 140 per cent over the decade between 1990 and 2000. As further evidence of the quantum of research being undertaken within PhD programs, there were in excess of 34,000 PhD students enrolled in 2004 (DEST, 2005). Thus while the potential magnitude of the impact of PhD research undertaken is huge, concern for the quality of the research is at least equally important. Do we want research to have an impact if it is of a quality identified by examiners as marginal?

For the discipline of Education, research reported in PhD theses represents a major component of all educational research undertaken in Australia. Postgraduate research students in faculties of education (and PhD candidates account for more than three-quarters of all these students) are the largest group of educational researchers in Australia, exceeding numbers of university staff and other educational researchers in 1999 by a factor of more than two (DETYA, 2000). Further, the estimated time spent by postgraduate students on research and development in education and training approached three times that of academic staff in universities in 1996 (Holbrook, Ainley, Bourke, Owen, McKenzie, Misson & Johnson, 2000, p. 64). Thus the quality of theses is important as a major contribution to our national research profile in Education. It is likely that this is also the position for other professional disciplines where the impact of research is particularly important.

If research studies in the professional disciplines are to have an impact, allied with the need for high quality is a need for thesis relevance to the profession. Again in Education, the topic areas chosen for research theses indicate at least potential relevance. A national study by Holbrook et al (2000, p. 82) reported that the two dominant research areas of theses in Education were educational processes and structures (36%) including classroom processes

and teaching methods, and subject-specific curriculum areas (33%), both with potential application to schools and schooling. Given that most of the research students were either currently practising teachers or educational administrators, it is not surprising that more than two-thirds of theses were in these potentially relevant and useful areas.

Research training and output

It was essentially only in the mid-1980s that a debate emerged about what a PhD was for. And the debate eventually settled into three strands competing for recognition – the PhD as a generator of research, the PhD as a source of training, and then the purpose of the training *vis-à-vis* the destination of the graduates. As the PhD became an increasingly popular degree, a focus on its provision and resourcing followed. By the late 1990s the PhD was attracting a significant amount of commentary that linked its continued existence to performance, funding and standardisation:

> The old apprenticeship model … is a reflection of a less costly, less complex, less competitive and better resourced time … there are skills that are required of today's graduates that are different from ours and are skills that will be best developed in an environment that is more structured and more coherent than the one we saw (Chubb, 2000, p.18).

There were strong moves to push the emphasis wholly toward training, that is, that a PhD should be less about asking a student to produce a fine piece of research that contributed to the field, and more about satisfying criteria such as ability to write comprehensively and coherently about a field, demonstrate technical and methodological proficiency and show a creative spark. But it is a fine line between showing these abilities and producing good research. Ideally they occur together and clearly both the research training and the thesis itself are important. The research project undertaken is normally the sole, or at least it is the major, application of the research training received. In large part, the quality of the thesis is a measure of both the research training and the research project outcome. The assessment of thesis quality is the focus of this paper.

Research Quality Framework

Most recently, an evaluation of research conducted at Australian universities and other research organizations is currently being planned for implementation over the period 2006 – 2007. The evaluation is known as the Research Quality Framework (RQF), and it seems that higher degree research is to be included in the exercise, or at least associated with the RQF in some form yet to be determined (Roberts, 2005). Consequently there is likely to be an intensification of interest in assessing the quality of higher degree research studies, to a greater extent than is done at present.

HOW THESIS QUALITY IS ASSESSED

In Australia a written thesis (or dissertation) is the only examinable outcome of most PhD candidatures. The thesis is normally sent for examination to either two or three independent, external examiners who each write a report and make a recommendation on the thesis. There is usually no oral examination. Currently most universities simply record a pass or fail result for each PhD candidate, although a few also record whether the thesis examined was considered by examiners to be among the top 5 or the top 10 per cent of all the theses they have examined.

The process of using external examiners should be considered a strength of research training in Australia. The external examiners are selected for their knowledge, expertise and experience in the field, and normally provide an international as well as a national perspective on the thesis evaluation. It is perhaps worth noting that almost half of the examiners of Australian PhD theses come from overseas (Bourke, Hattie & Anderson, 2004; Pitkethly & Prosser, 1995).

The written examiner reports and recommendations are normally considered by a university committee (a Higher Degree Research Committee, or equivalent) which makes the final decision on the fate of the thesis. The decision is normally one of five alternatives that had previously been offered to examiners in tick-box form: (1) Accept the thesis without alteration, (2) Accept the thesis following minor (invited) corrections, (3) Accept the thesis subject to more major (required) corrections, (4) Require the thesis to be revised and resubmitted for further examination, (5) Fail the thesis without the opportunity for revision.

There is a clear progression between these five categories which provide an indication of thesis quality, particularly at the extremes. Recommending that a thesis be passed without amendment or that a thesis required major revision and further examination are very clear indicators of perceived quality. However, the recommendations made by examiners and the decisions of committees have a strong formative aspect and are not uniformly intended as a summary indicator of thesis quality. The examiner, in many cases, may see the thesis as a work-in-progress and, for example, may recommend category 2 (invite correction) or even category 3 (require correction) for what the examiner believes to be a high quality thesis because it is seen as worth the effort to make the thesis even better. Conversely an examiner may recommend category 1 (accept without amendment) or category 2 for a marginal, barely passable thesis, on the basis that it reaches an acceptable level but that it is not worth the additional effort necessary to raise it to the level of mediocrity.[2] However, these are the exceptions – more normally the high quality theses (in the opinion of the examiner) receive a recommendation in category 1 or 2, and the marginal theses receive a recommendation of either category 4 (revise and resubmit) or 5 (fail). But the written examiner reports tell us much more than this.

OUTLINE OF THE STUDY

This paper arises from a national project investigating doctoral examination through the use of examiner reports and candidate information that is due to be completed early in 2006. 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. Detail about the aims and mixed methods design has been published previously (Holbrook & 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 each of eight universities. The candidates represent all broad fields of study. The study reported here draws on the completed core analysis for one half of those institutions.

---

[2] Although a single examiner may be uncritical, it is extremely unlikely that more than one examiner of a thesis that was not of high very quality would respond by indicating that it should be passed without significant correction or amendment.

The most substantial source of data is the examiner reports, which range in length from one line to about 27 pages averaging between two and three pages 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 with detailed coding notes being 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 (see Lovat, Monfries & Morrison, 2004).

The initial unit of data collection and coding is the institution. We undertake core coding of examiner reports for each university separately. Institutional data are then merged and the unit of analysis becomes the candidate within a Broad Field of Study. In this paper we initially present merged institutional data for all candidates and then switch focus to examiner comment on Education theses.

Analysis of the examiner reports

The core coding categories are built on a hierarchical structure of five parent coding categories, with 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. The primary coding categories or parent 'nodes' are presented and illustrated in Figure 1.

The substantive text of the assessment is contained within nodes 3 and 5, and these nodes constitute the focus of the analyses that follow. Node 3 (assessable areas covered) was intersected with two of the sub-categories within evaluative elements (node 5). They are summative comment (node 5 1) and other judgement (node 5 4). The comment in both is about overall judgement, that is about drawing the key threads of judgement together. In each case the positive and negative 'tone' of the comments can be clearly identified. The selection of the positive and negative 'summative' categories for intensive analysis about 'quality' was informed by the findings from a number of preceding analyses that showed that summative comment predicted examiner recommendation at both ends of the continuum of quality (Holbrook, Bourke & Dally, 2003).

The areas assessed by examiners

In the parent node 'Assessable areas covered' there are five sub-categories as well as some lower levels of categorisation. The categories reported in this paper are identified in bold text below.

1. Scope, significance and contribution
There are four sub-categories in this node. **Scope** captures examiner reference to the candidate's research questions and aims. The examiner may paraphrase or quote what the candidate claims the scope of the study to be, or they may elaborate on the questions and summarise what the study is about.

Comments about **Contribution and Significance** are typically brief and sometimes restricted to simple statements and common word use such as 'useful', 'interesting' 'important', 'timely' and 'worthwhile.' It is what examiners go on to say about the components of the thesis and their relative strengths that will give a sense of the contribution they identify.

Publications are a specific type of contribution. Two types of reference to publication are categorised in this study. They are where **existing publications** by the candidate are noted by the examiner, or where the examiner suggests or anticipates **publications that may arise.** It needs to be noted, in reference to Table 1, that in Education specifically, examiners do not often refer to existing publications. This is not surprising in that 'impact' in Education, unlike the hard science disciplines is not tied to rapid publication of findings.

Node 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.

Node 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.

Node 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).

Node 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.

Node 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.

Figure 1: Parent Nodes used in the PhD Examination Study

2. Review of the literature

'Systematic acquisition and understanding of a substantial body of knowledge' is a generally accepted learning outcome of doctoral study (Shaw & Green 2002, p.117). Examiners' discussion of the literature comprises three broad categories. One is **coverage**, that is the depth, breadth, recency, and adequacy of the literature used in the thesis. Another is **use and application** where the examiner explores the candidate's application of the literature and its role in the thesis, reference to theoretical contribution and the candidate's immersion in, and

their depth of understanding of the literature. A further category relates to **errors or inaccuracies** in referencing or citation.

3. Approach
Approach encompasses examiner comment on how the candidate approached the research, their methods and design, and variations in design or reflections on method. This category covers the nature of the experiments, instruments and mechanics of approaches to analysis (for example, the computer software employed) through to reasons given for their use. Further, this category also captures any comments on the methodological and epistemological position taken in the thesis. It does not capture findings *per se*, so if some comments on findings are indistinguishable from comments on method of analysis or interpretation, the latter are captured in the next category.

4. Analysis and Reporting
The category **analysis and reporting** captures text covering the candidate's findings and the results of analysis and interpretation as well as the examiner's 'take' on, or general response to, the candidate's endeavours. Here the examiner conveys, and comments on, the way in which findings have been arrived at, treated and presented.

5. Communicative competence
There is examiner comment that captures **substantial issues** about 'communication and presentation' (that is holistic appraisal and/or summary of communication skills). These comments may not necessarily be critical. The examiner may state there is a major flaw, or many flaws, in communication or presentation, or conversely that there are none. They may convey frustration with weaknesses in communication or style throughout. Conversely they may praise the candidate for a well-developed conceptual map, well-honed argument, the absolute absence of errors and clear linkages throughout.

The evaluative elements in the examiner reports and their 'tone'

It emerged from successive readings of the examiner reports that there are three broad categories of evaluative comment. The largest category is 'instructive' (on average 44% of text is devoted to this). That is, there is something in what the examiner is saying that is instructional. In the most basic sense the examiner is telling the candidate, and/or possibly the supervisor, to do something to the thesis or its products. The other two forms of evaluation do not contain instruction, one is 'summative', and another category we refer to as 'other judgement'. Later we refer to these as stand-alone judgement. The difference is highlighted below.

Summative comments are in fact statements that sum up the examiner's judgements about the whole thesis or a substantial sub-section of it. Such statements can be positive, negative or neutral. *Positive summative* comment identifies strengths. *Negative summative comment* identifies deficiencies and failings. However, there are often situations where the examiner endeavours to balance their comments. They may say something like '…on the one hand, [something is] good and on the other [that same something, or something else] is flawed. An examiner attempting to communicate a *neutral* position in a summary way may provide guarded praise or skirt around direct criticism by using an oblique turn of phrase. This type of comment, on average, comprises only 3 per cent of the total report. In 82 per cent of reports examiners summed up something positive about the thesis, only 25 per cent of reports contained negative summation, and 43 per cent of the examiners contrived comments that balanced the positive with the negative.

The comments captured in the final category of judgement, **other judgement,** typically convey a sense of responsiveness to the thesis by the examiner that range from the formal to the personal. They occur in 86 per cent of reports. Other judgement comments are not summative, or instructive. Unlike summative evaluation which positions the thesis (in whole or part) along a continuum of 'very good' to 'very poor', the 'other judgement' category captures such dimensions as 'satisfaction' (or not), 'worth' (or not), 'effort' (or not) through to 'interest', 'displeasure', 'frustration', 'amazement', etc. This category of comments may very clearly exhibit the examiner's involvement or engagement with the thesis in their adjudication of its value or worth. The comment may be celebratory, passionate or declarative in nature, suggesting strong feelings about some feature or aspect of the thesis. They may also capture the reiteration of something positive or negative that the examiner has written at more length about elsewhere in the report. At 16 per cent they constitute a relatively large proportion of the average report.

As indicated above the statements that capture overall judgement about the thesis are not only significantly related to recommendation but they have the most direct potential for identifying different levels and areas of 'quality'.

The two groups were merged providing the sum of positive and negative comments for each of the 10 assessable area sub-categories. The proportion of text in each examiner report of each of these was then calculated to provide estimates of the balance of examiner comment on each of the assessable area sub-categories.

IDENTIFYING THE HIGH QUALITY AND MARGINAL THESES

Given the argument presented above that the examiner recommendations provided a good guide to thesis quality at the extremes, the high quality and marginal theses in the total group of 401 theses[3] were identified. The high quality theses were those for which all examiners and the committee uniformly ticked the top box, indicating the thesis should be accepted as submitted. A total of 24 theses with 59 examiner reports were in this category. These high quality theses represent 6 per cent of the 401 theses for which the examiner reports have been coded and analysed to date. The marginal theses were those for which at least half the examiners recommended that the thesis be revised and resubmitted for further examination or be failed, and the committee decision was that the thesis at least required major revision. Only 12 theses (3%) with 35 examiner reports were in the marginal category. The difference in stringency of selection of the high and marginal quality theses arose because few theses received a revise and resubmit recommendation and very few were recommended to fail. In fact no thesis in the study was failed by the committee, at least at the first examination.

The mean proportions of the categories in each report for the assessable areas (node 3) intersected with evaluative comment that had either a positive or negative tone were compared for the high quality and the marginal theses, using independent samples t-tests and the 0.05 statistical significance level. The categories exhibiting statistically significant differences were taken as distinguishing between theses of at the two quality extremes. The proportions in each category are shown in Table 1.

It will be noted that there were more significant differences between the high quality and marginal theses for the positive comments than for the negative comments in the examiner reports. In part this was because there were more positive than negative summative comments

<hr>

[3] These 401 theses generated 1103 examiner reports.

and other judgements in the reports, as would be expected given the fact that most theses are passed by the examiners.[4] However, negative comments about the scope and the significance and contribution of the thesis and the level of communicative competence do signpost differences between theses at opposite ends of the quality continuum. All positive categories except thesis scope and literature coverage differentiated between the two groups of theses. These results are similar, but not identical, to those we previously obtained more simply by identifying all comments that distinguished significantly between high and low quality theses (Holbrook, Bourke, Lovat & Dally, 2004b). The previous analysis had identified the following as important – aspects of thesis topic and approach, literature use and reporting, communicative competence and comment about publications arising from the thesis.

Table 1: Percentages of comments with positive and negative tone for each assessable area covered in examiner reports compared for high quality and marginal theses (all discipline areas are included)

ASSESSABLE AREAS	POSITIVE TONE			NEGATIVE TONE		
	HIGH Q %	MARG Q %	SIGNIF	HIGH Q %	MARG Q %	SIGNIF
Scope and significance of thesis						
Scope	4.4	1.0	NS	0	0.8	.032
Significance & contribution	10.4	2.6	.002	0	1.8	.025
Publications arising	2.5	0.2	.001	0.1	0	NS
Existing publications	0.7	0	.007	0	0	NS
Review of literature						
Literature coverage	1.5	1.4	NS	0	0.5	NS
Literature inaccuracy[1]	0	0	-	0	1.4	.007
Literature use & application	1.4	0.1	.006	0	0.4	NS
Approach	4.3	1.2	.005	0.2	0.9	NS
Subject matter and findings						
Analysis & reporting	9.3	1.9	.000	0.8	2.4	NS
Communicative competence						
Substantial	4.4	1.2	.001	0.3	2.4	.020

Note:
1. Clearly there are no positive comments at all in examiner reports about inaccuracies in the literature review.

The significant differences noted above were obtained from examiner reports across all discipline areas. Our work to date has indicated that, in the main, the types of comments examiners make do not differ by discipline area (Bourke, Hattie & Anderson, 2004). However, given that this conference is principally concerned with educational research, quality indicators for both the high quality and marginal theses given as illustrations in the

[4] Previously we have published findings from this study indicating that there was a higher proportion of negative comments in examiner reports than positive comments (Holbrook et al, 2003). These results are not inconsistent with that. Most of the negative comments were formative rather than summative in nature, that is they were intended to encourage or require the candidate to amend the thesis. Formative comments are not included here.

next section of this paper have been taken only from examiner reports on theses in the Broad Field of Study of Education.

<center>EXAMINER INDICATORS OF QUALITY IN EDUCATION THESES</center>

In this paper, we take the opportunity to move the analysis forward for the project, particularly for Education theses. We now have a large enough group of candidates to explore some combinations of nodes (that is, where text codes intersect). In this case, we have intersected the assessable areas covered in the examiner reports with the combined judgement categories (hereafter referred to as stand-alone judgement or judgement), as either positive or negative comments.

The following brief extracts and commentary from examiner reports that differentiated between high quality and marginal theses in Education, based on this intersection, are grouped according to the assessable areas shown in Table 1. The comments identified here are on the whole less rich and textured than many other examples of evaluative comment. This is not surprising as we are drawing on the most succinct of the comments, identified as those where examiners pinpoint central strengths and weaknesses of theses in Education. Limiting this analysis to Education examiner reports, we excluded many comments from reports in other disciplines that were more evocative or expressive of what examiners saw as high quality and marginal theses.

Thesis scope

With respect to scope of thesis, there were very few distinguishing comments of the stand-alone judgement type, and all were negative. In the example below, the scope of the study may not have been inadequate, but it was not clear to the examiner.

> *It was not always clear ... what the actual aim of the study was. For example, the aim is written in different styles on [several listed pages].*
> (Examiner 408412530)

In another case, the examiner felt that the thesis title was not reflected in the thesis which was 'parochial', that is, the title was appropriate for a doctoral thesis but the scope of the thesis itself was inadequate.

Thesis significance and contribution

There were many more distinguishing judgements that were positive about thesis significance and contribution, than there were negative comments for either the high or marginal theses. With respect to positive comments, in most cases the examiner was referring to an original contribution to knowledge or to theory in the field.

> *Impressive scholarly contribution ... rich in insight and provides a helpful heuristic for the advancement of knowledge in the field... highly original and thought-provoking.*
> (401510440)

Other stand-alone judgements about contribution also crossed over into the analysis and reporting and the approach sub-categories when the examiner commented on a contribution to methodology.

> *... has made a valuable and interesting contribution to knowledge because of the comprehensive nature of the data-gathering and also the unique analysis.* (150911580)

Some examiners in Education also commented on the contribution of the thesis to practice, or as 'adding to practical debates'. Positive comments about the contribution of the thesis to professional practice were also made in conjunction with another sub-category, related to the importance of publications arising from the thesis.

Publications arising from the thesis and existing publications

The focus of the positive stand-alone judgements by Education examiners related to potential publications arising from the thesis tended to be about making the 'information accessible to others', both researchers and practitioners, indicating that the results 'deserve to be widely promulgated'. There were no examiner comments at all on Education theses related to existing publications, although these were present in examiner comments on theses in other disciplines.

Review of literature: coverage, use and application

Apart from stand-alone negative judgement on inaccuracies in the literature review, there was little negative and positive judgement on the literature review for either the high quality or marginal theses. Although literature coverage was not found to differentiate between high quality and marginal theses across all disciplines, positive comment about coverage was more common in examiner reports on theses in Education. Both comprehensiveness and location of literature in the field were important for examiners.

> *...literature review is comprehensive and probably definitive ...it locates and supports the study* (408412520)

> *...very precisely locates her research project in a wider body of scholarship that she carefully and accurately summarises.* (400510140)

Negative comment about literature use and application was generally not a distinguishing characteristic between theses in Education or the other disciplines.

Approach of thesis

Positive judgements about the approach taken in the thesis were reasonably common for high quality theses, although negative judgements were rare for both high quality and marginal theses. As would be expected the positive comments on approach related, in the main, to methodological issues, including data collection and analysis, and in some cases focussed on 'care and attention to detail' being evident in a complex research project.

> *...study [was] methodologically soundly based and comprehensive ... In terms of the data collection and analysis, the thesis is impressive, evidencing insightfulness, originality and sensitivity of ...* (202000570)

Positive comment on the approach may also have focussed on reporting which links with the communication sub-category as shown in these two examples.

good knowledge of the methods that were used and gave good descriptions of … method and procedures… (408412540)

The methodology is well justified, sound and clearly described … (404611400)

Analysis and reporting

Positive judgement about analysis and reporting was common for high quality theses, effectively distinguishing these from marginal theses. One of the key virtues identified by examiners of Education theses was commitment, this was evidenced by 'care and attention to detail', and by reporting that was considered 'challenging' or 'honest, even courageous'.

Other positive comments related to good connections between strands of the thesis or between methods, analysis and theory.

*D*evelops clear connections ... indications of the rich seam of research ... (150111220)

The candidate has [constructed] a strong and effective argument. She has recognised the value of having and developing a thesis. ... has shown a remarkably high level of perceptiveness, insight and intellectual acumen in the ways she used her materials to draw out appropriate and interesting insights and conclusions. (400510140)

There was perhaps slightly more negative examiner judgement in Education on analysis and reporting than was the case in other disciplines. The concern seemed to be focussed on inadequate connections or lack of critical analysis of what was done.

There is a sense throughout that the writer has had to struggle hard to synthesise and integrate his material, and that he is not fully confident with it. (202000580)

It is of concern that a person writing a thesis at this level offers no critical analysis of the information presented. (308201650)

Communicative competence: substantive issues

Both positive and negative judgement on substantial issues of communicative competence of the thesis distinguished between high quality and marginal theses, and positive comments were reasonably common for high quality theses. One of the key issues of communication for examiners of Education theses was enjoyment derived from reading a thesis because it was well written, sometimes expressed as 'easy to read'. Writing that was 'cogent and straightforward' was valued, as was a thesis with 'data and analyses [that] flow sequentially to logical conclusions'. Occasional 'eloquence' was seen as a bonus.

Negative comments on substantial communicative competence were not common in Education examiner reports, but there was a focus on polish that was broader than editorial issues. Theses that were considered too long were seen to require further work to improve communication. Perhaps more importantly, an unpolished, hurried thesis might contain 'broad generalisations [that] are not supported', even though the examiner could see how the assertions could have been supported with more time and care taken.

SUMMARY AND CONCLUSIONS

This paper has outlined ways in which examiners identify quality in PhD theses, first in theses across all discipline areas, and then with a particular focus on Education. From the wide range of comments made, sub-sets of positive and negative summative comments and other judgements (described in the paper as 'stand-alone judgement') on a range of assessable areas were identified that distinguished high quality from marginal theses. High quality theses were those uniformly recommended for acceptance as submitted. Theses were determined to be marginal if at least half the examiners had recommended that they be revised and resubmitted or failed and the committee decided that they required major revision. Most of these marginal theses actually were eventually passed and the PhD awarded, following substantial correction and perhaps re-examination.

It was found that positive and/or negative examiner judgement about several sub-categories of assessable areas – scope and significance of the thesis, the literature review, approach of the thesis, subject matter and findings and communicative competence – were indicative of high quality and marginal theses. Examples of both positive and negative examiner judgements were given in this paper. In particular, the comments of examiners making positive judgements provided a variety of clear indicators of high quality theses. There was a lesser range of negative examiner judgements that indicated marginal theses. The analysis of comment defined as stand-alone judgement reported here has been more successful in identifying clear, succinct indicators of high quality theses than it has been in identifying such indicators of marginal theses. Examiners do not provide much of this type of judgement for theses at the margins because they are more inclined to provide extended evaluative comment of an instructive kind (that is, where the examiner uses their report in an attempt to teach the candidate as well as to evaluate the thesis, as illustrated in Appendix 1). In such categories of evaluative comment, negative judgement is closely integrated with instructive comment. The next step that we need to take is to draw out the judgemental components of instructive comment to assist us to identify a larger variety of evaluative comments that identify marginal theses.

We have selected and used categories of judgement with both positive and negative tones that would specifically assist us to identify the features that significantly distinguished between theses at the two ends of the quality continuum. Our findings may not cause much surprise but they do provide serious food for thought if we are considering the impact of doctoral research on the discipline of Education. When examiners capture what is really good about a PhD, the model PhD if you will, then what we have is the combination of fine research, research talent and well-assimilated research training and education. The very best theses bear all the markers not of documents produced by 'learner researchers', but by persons already acclaimed as 'researchers', perhaps new researchers but those definitely with the potential to become acceptable as peers.

By contrast many examiners struggle with recommendations on theses of marginal quality. But should they? It may be time to distinguish more forcefully between thesis excellence, pass and fail. Examiners do sometimes pinpoint the depth of the flaws in a marginal thesis, for example, when the thesis is not making anything by way of a contribution or appears to have been flawed in some cases from the outset. More than this, perhaps the candidate cannot 'communicate' effectively. The latter is more than the thesis exhibiting problems with conventions, is it a problem of real substance. We have shown elsewhere that ultimately it is often the examiner who provides the key ideas and even the references that allow such candidates to raise their thesis to a marginal pass and scrape through (Holbrook, Bourke, Lovat & Dally, 2004a). Fortunately this group of candidates is small, but their doctoral completion could hardly be claimed as an indicator of research 'quality'.

Across all institutions, the best theses are yielding outstanding research quality, and we should be careful not to jeopardise this status by accepting all theses submitted for examination unless we establish that they reach at least a minimally acceptable standard. Currently individual examiners and university committees define this standard – can we help them further to do this? The PhD process can and normally does yield sound research training and significant research outcomes. Dissemination of only the best research will assist in ensuring the best research outcomes have greater impact.

Based on the results of the work described here, it would be possible to develop a checklist of criteria for assessing theses, once the data coding and analysis for the large study are complete. Examiners could be asked to complete the checklist and forward it with their written report on the thesis. Clearly the inclusion of such a checklist with the thesis would be likely to influence the examiner report, and we might like to ponder the desirability of this occurring. The nature of thesis examination, and thus the assessment of research completed by higher degree research candidates could be changed profoundly.

There currently exists, however, a range of approaches to higher degree research examination that are used by universities. Although checklists are not generally used (except for the overall recommendation), some universities do provide detailed guidelines on what examiners should comment on in their reports, while others do not. Although experienced examiners tend to say that they read, but do not use the guidelines provided (Mullins & Kiley, 2002), there are some indications of subtle differences in the nature of examiner comments between universities that differ on the provision of detailed guidelines (Holbrook, Dally, Bourke, Lovat, Monfries & Lu, 2004).

Although, as mentioned above, the proportions of examiner comment across coded categories does not generally differ by discipline area, further investigation is needed and will be undertaken when all the data are available. When tone is taken into account, however, there are more positive examiner comments on existing publications for theses in science and technology disciplines than in the humanities and social sciences discipline areas. In part this difference reflects a divergence in practice and expectation of candidates between these broad areas, and it is more likely that candidates in science areas will have published something before submitting their thesis.

Another minor discipline difference related to tone was that negative comments on literature coverage by thesis examiners were indicative of marginal theses in Education, but less so in other disciplines, where this relationship was not significant.

There is also the possibility that examiner comments on theses in professional disciplines (notably Education, Engineering and Health) might differ from those from other discipline areas (the humanities, arts and some of the sciences), particularly with respect to comments on significance and contribution of the thesis results. There are some possible indications in the texts of examiner reports that a thesis which is considered useful to the profession, in addition to the research community of scholars, was seen in a more positive light. If such a difference were found to exist consistently, it would need to be taken into consideration in any checklist provided for use by examiners.

Guidelines towards developing a method of determining the quality of theses accepted for the award of a PhD will clearly be of interest to those concerned with reporting to a Research Quality Framework, but we would hope for a more proactive use of guidelines than that

alone. By providing information to research supervisors and students about what aspects examiners privilege in their assessments of thesis quality, the supervisors and students will be better placed to improve what they do when designing research projects and writing theses. It would also be possible to assist inexperienced examiners by providing them with such information gleaned from the reports of more experienced examiners. Thus, the quality of research training itself and of its reporting in theses and the research literature should be enhanced to the overall benefit of Australia's research effort and profile across all discipline areas.

REFERENCES

Bourke, S., Hattie, J. & Anderson, L. (2004). Predicting examiner recommendations on PhD theses. *International Journal of Educational Research, 41,* 2, 178-194.

Chubb, J. W. (2000). The impact of the white paper on universities: some possibilities. In M. Kiley & G. Mullins (eds), *Proceedings of the Quality in Postgraduate Research Conference: Making ends meet*, April 13-24, Adelaide, pp. 15-23.

Department of Education, Training and Youth Affairs (DETYA) (2000). *Publications - Higher Education Statistics Collections*, DETYA (now DEST), Canberra, published at http://www.detya.gov.au/highered/statpubs.htm#studpubs, accessed 11 Mar 2005.

Department of Education, Science and Training (DEST) (2005). *Publications – Higher Education Statistics Collections*, DEST, Canberra, published at http://www.detya.gov.au/highered/statpubs.htm#studpubs, accessed 11 Mar 2005.

Evans, T., Macauley, P., Pearson, M. & Tregenza, K. (2003). A decadic review of PhDs in Australia. Paper presented at the Joint AARE/NZARE Annual Conference, Auckland, Nov/Dec. Accessible at http://www.aare.edu.au/03pap/eva03090.pdf

Harvey, L. & Green, D. (1993). Defining quality. *Assessment & Evaluation in Higher Education, 18*, 1, 9-34.

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.

Holbrook, A., Ainley, J., Bourke, S., Owen, J., McKenzie, P., Misson, S. & Johnson, T. (2000). Mapping educational research and its impact on Australian schools. In DETYA, *The Impact of Educational Research*, Research Evaluation Program, Higher Education Division, Canberra.

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,* 153-169.

Holbrook, A., Bourke, S. & Dally, K. (2003). How examiners define quality in the doctoral thesis. Paper presented at the Biennial Conference of the European Association for Research in Learning and Instruction, Padova, Italy.

Holbrook, A., Bourke, S., Lovat, T. & Dally, K. (2004a). PhD theses at the margin: Examiner comment on resubmitted theses. *Melbourne Studies in Education, 45,* 1, 89-116.

Holbrook, A., Bourke, S., Lovat, T. & Dally, K. (2004b). Investigating PhD thesis examination reports. *International Journal of Educational Research, 41.* 2, 98-120.

Holbrook, A. & Dally, K (2003), Emphases in 603 doctoral examination reports, Investigating Doctoral Assessment Symposium, paper 1, presented at the EARLI Conference, Padova, Italy August 26-30.

Holbrook, A., Dally, K., Bourke, S., Lovat, T., Monfries, M. & Lu, Y. (2004). Evaluating some fundamental features of doctoral assessment. Paper presented in a Symposium, 'Assessment and its implications for research education', at the Quality in Postgraduate Research Conference, Glenelg.

Johnston, S. (1997). Examining the examiners: an analysis of examiners' report on doctoral theses. *Studies in Higher Education, 22*, 3, 333-347.

Lovat, T., Monfries, M. & Morrison, K. (2004). Ways of knowing and power discourse in doctoral examination. *International Journal of Educational Research, 41.* 2, 163-177.

Mullins, G. & 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.

Pitkethly, A. & Prosser, M. (1995). Examiners' comments on the international context of PhD theses. In C. McNaught & K. Beattie (Eds), *Research into Higher Education: Dilemmas, Directions and Diversion.* (pp. 129-136). Melbourne: HERDSA.

Roberts, G. (Chair) (2005). *Research Quality Framework: Assessing the quality and impact of research in Australia. Issues Paper.* DEST, Canberra. Published at http://www.dest.gov.au/resqual/issues_paper.htm

Scott, D., Brown, A., Lunt, I. & Thorne, L. (2004). *Professional Doctorates: Integrating Professional and Academic Knowledge,* Berkshire England: Society for Research into Higher Education and 0pen University Press.

Shaw, M. & Greene, D. (2002). Benchmarking the Ph.D. – a tentative beginning. *Quality Assurance in Education, 10,* 2, 116-124.

Yates, L. (2004). *What does Good Education Research look like?* Open university Press, McGaw-Hill Education, Berkshire, England.

Appendix 1: Text examples of the sub-categories of the category 'evaluative elements'

Category	Example of text
Summative positive	The use of this excellent interview material and other correspondence adds significantly to our knowledge of this subject. Overall this thesis is based on an impressive range of primary sources, and demonstrates wide reading and a good grasp of the secondary literature.
Summative neutral	(1) There are a few omissions in the literature, but they are minor and do not detract greatly from the final product. (2) On the one hand she is commendably succinct in her summaries of enormously complex material. For the most part her arguments work well. On the other hand, in some cases, there is a bit more to be said about the topic.
Summative negative	The underpinning theoretical and methodological precepts are outlined scantily and the whole is not coherent. It is marred by an abundance of examples of poor grammar and inattention to editorial detail
Formative instruction	There are many references to X…some bear suggestive relationships to the iconography found in some of Hopkins's poetry, but given the large proportion of the thesis devoted to these references there needs to be a very clear and firm case made for Hopkins's acquaintance with each of them…there is merely a footnote…which asserts that Hopkins is aware of this …tradition… in the absence of careful arguments establishing Hopkins's knowledge of particular…books, or a clearly stated rationale for their inclusion for purposes of comparison, such material often appears arbitrary and puzzling …. Furthermore the arguments for the importance to Hopkins of [X] , even if we suspend the issue of his acquaintance with them, tend to be tenuous in nature. [examples provided]… On the other hand an indication of the candidate's capacity for independent research is evidenced in the discussion of X…a most interesting and original observation and commentary… [now turning to another matter] is X included because it somehow influenced or inflects Hopkins's poetry and ideas about *[most is negative in tone, a small proportion is not clearly negative or positive and there is one section that is clearly positive]*
Instructional commentary	Why was it necessary to store seeds in a dark cupboard – an opaque container would have been sufficient (indeed probably better than a cupboard which might have been opened for reasons other than access to the seeds). *[negative in tone]*
Prescription	(1) I prefer not to read that a null hypothesis had been accepted, more correctly one should say that it has not been rejected. (2) The abstract is too brief and needs to include reference to those elements of managerialism discussed in the chapter. *[negative in tone]*
Other judgement	I found the opening chapter somewhat stolid and not adding a great deal to the thesis that follows *[negative in tone]*

Source: Holbrook & Dally (2003, p. 8)