

## **An investigation of PhD examination outcome in Australia using a mixed method approach**

Allyson Holbrook<sup>1</sup> & Sid Bourke  
The University of Newcastle<sup>2</sup>

### **ABSTRACT**

This paper focuses on doctoral assessment as an area that has been relatively neglected in higher education research. It then describes and justifies a mixed-method approach to the study of PhD examination processes and outcomes in Australia. The design is reported of a study including candidate and candidature information for approximately 800 PhD students across all discipline areas at eight Australian universities, some examiner information, and the 2100 examiner reports on their theses. Examination process, outcome and discourse are discussed in relation to the study design and a number of research questions to be investigated. The sampling method and data collection are described. A particular focus of this paper is how the categories were developed, tested and refined for coding the texts of the examiner reports. The overall aims of the study are to contribute new knowledge about doctoral study and provide a firm empirical foundation for enhancing research performance.

**Keywords:** mixed method, doctoral assessment, examiner reports, text analysis, thesis outcome

### **INTRODUCTION**

There is much about the PhD that has set it apart. The degree is the acknowledged ‘gold standard’ (Scott et al., 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 make an original and significant contribution to their field in the form of a research thesis – a research outcome. The award, which confers an internationally recognised public title (‘Doctor’) signifies both this elite status and the substance of the individual’s achievement. The assessment of the degree has been weighted toward achievement in research. Much of the practice and tradition that has shaped the nature of the PhD in Australia had its roots in the experiences of Post World War II academic life, which included the candidate’s articulation from an Honours degree, through a Masters degree and then into a PhD. The PhD was the route into academe and the apprenticeship was a lengthy one. By and large it was expected that doctoral candidates would develop the necessary research skills to undertake the research by working through their doctoral project with their supervisor or academic adviser(s).

In recent years research degrees have become more accessible, and in response to candidate, employer and national demands, more varied. They now serve a broader range of

---

<sup>1</sup> Contact

Associate Professor Allyson P. Holbrook,  
School of Education,  
University of Newcastle, NSW 2308, Australia.  
Email: [Allyson.holbrook@newcastle.edu.au](mailto:Allyson.holbrook@newcastle.edu.au)

<sup>2</sup> The authors would like to acknowledge the assistance of Distinguished Carolina Professor Lorin Anderson, University of South Carolina and Professor Mark Wilson, University of California, Berkeley for their reading of pilot stage material and helpful suggestions during the trial stages of core coding, analysis development.

purposes, and research enrolments have grown significantly. In the light of such developments, the emphasis on thesis contribution to knowledge has been questioned (O'Brien, 1995) and the importance of research training has gained prominence especially as the traditional pathways into the PhD have been eroded. Concern about high attrition and workplace relevance of the PhD have added fuel to the debate about what the degree should signify, and whether the demonstration of research skills provides a more apt assessment focus than the thesis itself (Pearson, 1999; Wright & Cochrane, 2000; Hoddell et al., 2002; Denicolo, 2003; Gilbert et al., 2004; Johnston & Murray, 2004). Gilbert draws attention to the need for doctoral curriculum in such circumstances - the 'systematic articulation of experience' in order to 'produce the intended outcomes' (p.303).

Regardless of whether the focus is on training or research, the question of quality lies at the heart of most concerns about the doctorate, not least because the elucidation of doctoral level outcomes, the pinnacle of academic endeavour, has proved unusually difficult in all but the most general terms (Morley, Leonard & David, 2002; Shaw & Green, 2002). In the past when there were relatively few candidates destined for scholarly pursuits, this was not a public or pressing issue. With the rapid 'massification' of the degree there has come the realisation that not only is there an absence of benchmarks, but an absence of information about the degree and its evaluation (Morley, Leonard & David, 2002; Shaw and Green, 2002; Jackson & Tinkler, 2001; Tinkler & Jackson, 2004, p. 8). While supervisors and examiners play a pivotal role in defining and shaping the practices in their disciplines, including how and what candidates need to learn to be successful, there is very little in the literature that explores the connection between expectation, judgement and outcome (Mullins & Kiley, 2002; Denicolo, 2003; Powell & McCauley, 2002, 2003). In an education context this is most unusual.

In any discipline there is a 'fundamental relationship' between assessment and learning that needs to be expressed:

... every assessment is grounded in a conception or theory about how people learn, what they know and how knowledge and understanding progress over time... each assessment embodies certain assumptions about which kinds of observations, or tasks, are most likely to elicit demonstrations of important knowledge and skills from students [and]...is premised on certain assumptions about how best to interpret the evidence from the observations and draw meaningful inferences about what students know and can do (National Research Council, 2001, p.20).

Examiner judgements capture what can be demonstrated and achieved in a research degree. In the absence of a clear expression of this relationship, the thousands of theses that are examined annually and globally can provide a vital source of information about what is learned and also, possibly, the quality of that learning as well as the quality of the research (Tinkler & Jackson, 2004).

This paper presents a collaborative mixed methods approach used to investigate the content and nature of the comment produced by each examiner to support their recommendation on doctoral theses by candidates of Australian universities. The three year (2003-5) study has received funding through a *Discovery Grant* awarded to three Chief Investigators (Holbrook, Bourke and Lovat) by the *Australian Research Council*. The paper will present the research questions and literature that shaped the study design, and proceed to a discussion of its mixed method elements and the management and integration of those elements.

Researchers in the Social Sciences have been combining methods for some time, but the literature on mixed methods has only recently attained a critical mass. Teddlie & Tashakkori (2003) suggest that mixed methods [plural] embrace 'mixed method' and 'mixed model' designs. *Mixed method* designs use 'qualitative and quantitative data collection and analysis techniques in either parallel or sequential phases' (p. 11; see also Creswell et al., 2003). The mixing occurs in the method. *Mixed model* research is mixed in many or all stages of the study (from developing questions to the drawing of inferences). In the latter it is possible to

have more than one paradigm and worldview mixed throughout a single study or a series. Teddlie & Tashakkori refer to 'fully integrated' mixed model designs as 'the most advanced and dynamic'. They incorporate 'multiple approach' oriented questions, the collection of qualitative and quantitative data, which may be converted, an example is qualitative data 'quantified' and 'analysed accordingly'. Inferences are made on the basis of the different analyses and the results are 'combined together at the end to form 'meta-inference'. This model 'combines concurrent and sequential possibilities' and is 'interactive' allowing change and modification to occur throughout the project (pp. 689-90) providing the flexibility to achieving a full and effective integration. We were not aware of this 'type' of mixed methods model when we initially devised our study on thesis examination, but it mostly closely approximates the outcome.

## THESIS EXAMINATION IN AUSTRALIA

Australian universities seek multiple written reports for research theses. Very rarely are Australian candidates required to undergo an additional oral examination at the final stage, whereas outside of Australia this tends to be the norm. However, oral examination may not play a determining role in assessment. Jackson & Tinkler (2001) identified frequent instances in English universities where the examination result was decided in advance of the viva, that is on the basis of examining the thesis. Trafford (2003), gathering data on examiner comment in the role of participant observer during a number of vivas, arrived at a similar conclusion.

Many questions are raised about the choice of examiners, their experience, independence and number (Hansford & Maxwell, 1993; Johnston, 1997; Kamler & Threadgold, 1997; Jackson & Tinkler, 2000; Tinkler & Jackson, 2000; Morley Leonard & David, 2002; Lawson, Marsh & Tansley, 2003). Australia is unusual in the degree to which it draws on international examiners, with approximately one half in this category (Pitkethly & Prosser, 1995; Bourke et al., 2004). Some Australian institutions invite examiners to consult with each other, but it is more common for an examiner not to know the identities of the other examiners until the process has been concluded. A very small proportion of examiners are 'qualified' staff from institutions outside of universities (e.g. a research institute, an industrial firm or a government department or, for some fields, an art gallery, conservatorium or museum).

Hansford and Maxwell (1993) and Johnston (1997) have drawn attention to a possible lack of consistency in examination standards, i.e. between different examiner ratings and comments on the same thesis, and between an individual examiner's rating and their specific comments. The same researchers also identified the prevalence of certain types of comment and emphasis in examiner reports, including a disproportionate amount of comment on 'presentation'. Pitkethly and Prosser (1995) in a single institution, phenomenological study noted little difference between the frequency of various types of comment by Australian and international examiners. Evidence from international comparative work on doctoral requirements reveals considerable variety in examination process in established research disciplines, yet quite subtle differences in general expectation of outcome (Clark, 1993; Kouptsov, 1994; Noble, 1994).

A further element in the process is the determination of the final decision by the institution. The examiner recommendation is normally in the form of selection from a series of options provided by the institution with an outright pass at one end and fail at the other. In between there are alternatives specifying the amount and type of changes to the thesis that are expected of the PhD candidate. At one end of this 'in-between' spectrum the comments are about improving an already pass-level thesis, at the other they can be about improving the thesis to reach a pass level. The institution, normally through a committee or panel, draws on the examiner recommendations and reports to determine the decision and advice that goes to the candidate. There may be differences between examiners and also between examiner recommendations and the committee decision based on what the examiners say in their reports as opposed to what they recommend. Differences such as these promote interest in the introduction of a uniform code of thesis examination practice for universities (Lawson et al., 2003). However, given the lack of an explicit, summary measure of thesis quality available,

there is no evidence that differences in examiner judgement between theses are related to differences in university procedures.

In a recent interview-based study (weighted toward the sciences), 30 experienced Australian examiners were asked how they approached thesis examination (Mullins & Kiley, 2002). Examiners saw their role as an important one, particularly with respect to upholding standards - a position echoing the findings of research undertaken by Tinkler & Jackson (2001, see also Jackson & Tinkler, 2000). Given the singular emphasis on the written thesis in Australia, there is an expectation that thesis examination will be thorough, complete, and consistently applied. But to what extent do examiners apply common criteria including those advised by institutions? Mullins & Kiley (2002) noted that examiners appeared to use their own criteria, and were confident in the distinctions they made between poor, acceptable and outstanding theses (see also Winter, Griffiths & Green, 2000). However, there is also evidence that some examiners might not be so sure. In a recent study by Denicolo (2003), a group of 62 UK examiners in the field of Education agreed that while the thesis had the highest priority as a source of evidence for 'quality', there was 'low' consensus about the criteria for assessment (p.89), and a 'high' level of 'insecurity about their knowledge of general standards' (p.90). In their book on doctoral examination in the UK, Tinkler and Jackson (2004) draw attention to the 'broad range of standards embraced by the award of Ph.D.' (p. 119).

The clarity of the role, resource intensiveness and usefulness of thesis examination processes have been raised in relation to both the written report and the viva (Johnston 1997, Jackson & Tinkler, 2001). Mullins and Kiley (2002) found that examiners enter into the process anticipating that students would pass and were quite reluctant to fail a student. Indeed, students who submit a thesis will rarely fail at examination stage. The predominance of a low failure is borne out in earlier literature in the UK (Becher 1993, p. 135) as well as our own study (Holbrook et al., 2004a).

Those familiar with Australian examiner reports will be aware that they provide rich and perhaps unexpectedly diverse layers of information. Drawing on the reports of 1103 examiners across all Broad Fields of Study we found that the average report was between two and three pages in length, ranging from one line to 1272 lines (more than 25 pages). With the exception of reports on Agriculture theses, which averaged almost four pages, there were no significant differences in length between disciplines. The reports give examiners 'voice' – as academic assessors, professionals and supervisors. Examiners are given, or take, some free rein in making their comments. They may judge it important, for example, to comment on institutional process, their expectations about the doctorate, their own research, and their expertise vis-a-vis the thesis work. These elements provide information that can be used to verify and explicate their evaluative comments on the thesis, while at the same time offer insights into the culture of examination and the examiner role.

Only a small number of Australian studies have subjected PhD examiner reports to content analysis (Nightingale, 1984; Pitkethly & Prosser, 1995; Johnston, 1997). These have tended to be one-off studies with reasonably restricted disciplinary coverage. In these studies, as far as the reports show, there was limited or no use of statistical and comparative measures, particularly comparison of examiner reports and ratings on the same thesis, nor were attempts made to correlate categories with each other or with examiner recommendations

## **A COLLABORATIVE MIXED METHODS APPROACH**

In Australia the examiner report, recommendation and final outcome are documented and archived consistently by most institutions. As in the UK the Higher Education Authority (in Australia the Department of Education Science and Training - DEST) does not keep statistics on outcomes other than completions data. At the institutional level access can be gained to the individual examiner reports and the committee decision for each candidate, candidate history (e.g. full- or part-time enrolment, possession of a scholarship, leave taken, any problems notified, time to submission, number and experience of supervisors), candidate demographic data (e.g. age, gender, entry qualification, English proficiency), and some examiner

characteristics (e.g. gender and location). For the researchers this potential access gave rise to the possibility of posing a set of questions that could address many of the issues about process and product raised in the literature, even extending to the rationale embedded in examiner decisions, the difference between a pass and an outstanding thesis, disciplinary differences and a deeper understanding of the nature of the PhD. Only a collaborative and flexible mixed methods model could allow us to reap the full potential of the questions and data in an integrated way.

By collaborative mixed methods research we mean the purposeful application of a multiple person, multiple perspective approach to questions of research and evaluation. Decisions about how methods are combined and how analyses are conducted are grounded in the needs and emerging complexity of each project rather than in preordinate methodological questions. (Shulha & Wilson, 2003)

The team had worked in various ways together but not in this area. Two (Bourke and Holbrook) had worked in collaboration on large-scale empirical studies drawing on their quantitative and qualitative expertise respectively. Lovat brought expertise in the Philosophy of Education, and more specifically a Habermasian perspective on ‘disciplinary knowing’ (Habermas, 1972, 1974, see also Lovat 2004). Holbrook and Bourke also brought a varied background in assessment studies. All the researchers have substantial experience in PhD supervision, examination and peer review, and collectively possess networks that allow them to draw on experts across the full spectrum of disciplines.

The guiding questions for the study were framed in the light of both the literature and collective researcher experience. From the start it was clear that the strength of the collaboration would be located in the variety of perspectives brought to bear on examination and the learning that would ensue. Methods and perspectives were mixed from the outset. This is evident in the design that follows. Shulha and Wilson (2003) flag that the ‘interaction of problem, method and results produces a more comprehensive, internally consistent and ultimately more valid general approach’ (p. 640).

### *Questions about PhD examination process*

Examination process can be very broadly interpreted to encompass the administrative, procedural, personal and academic activities and actions involved in assessing the thesis. From the perspective of the examiner, some of the processes are external. For example, the selection of examiners is essentially an academic matter ratified by the pertinent authority within the institution, as is the provision of guidelines for examining and the administration involved. Another external process is how examiner comments are circulated, read and used. Elements of process that can be deemed internal to the examiner are their interpretation and use of any guidelines, previous experience of process, the form they choose to give to their report, and editorial features essentially invisible to the reader such as using the cut and paste feature in word processing the report, i.e. their editorial processes. Processes more closely connected with the intellectual engagement with the thesis are the disciplinary criteria and standards that examiners apply in examination, the consistency between examiners, and an understanding of how the report will be used, including examiner expectations about the audiences and the procedures that determine the final result.

A series of questions about process were produced on the basis of the known availability of reports, recommendations and candidate and examiner data:

1. How consistent are ratings between examiners on the same thesis? Does consistency differ by discipline area?
2. What types, attributes and characteristics of evaluative comment can be identified in the written examination report? To what extent do these differ for the same thesis? Are different patterns of comment evident by discipline area?

3. What do examiners determine are their main functions or roles as evidenced by the structures and qualities of their reports? How do these reflect thesis, examiner, institution (that is, instructions and procedures) and discipline characteristics?

### *Questions relating to outcome*

There is a potentially wide range of outcomes associated with PhD candidature. Most, but not all, are achieved by the candidate. There are management and resource-based outcomes associated with completion, for example. There are personal outcomes for the candidate that range from personal growth and satisfaction through to skills development, and similarly for the supervisor. In addition there may be longer-term outcomes for the candidate and supervisor in such areas as publication and employment as well as that connected with successful 'partnership'. There are potential outcomes for the nation in terms of research application and innovation. But the outcomes that are most accessible through examiner reports and ratings are those that indicate success, or otherwise, in the production of the thesis.

Candidature detail and examiner reports and ratings can give some limited insight into institutional outcomes (e.g. completions), but primarily they can provide information about what is expected of the research candidate, whether or not they meet expectations, and whether such expectations transcend disciplines. In addition they provide indications about thesis quality. The indications in the written report can be captured in two ways: (1) by specific evaluative content contained in the reports (conceptualised as discrete categories), and (2) by qualities identified through data merging, analysis and the questions emerging from these processes.

4. What characteristics of student, candidature, examiner, institution and discipline predict final rating and category of evaluative comment?
5. How do examiner comments reflect expectations about thesis quality and standards? To what extent are these shared and consistently applied to the same thesis, across theses, institutions and disciplines?
6. What characteristics of student, candidature and thesis examination are related to examiner ratings, final rating, category of evaluative comment and thesis quality? Are there similar patterns across institutions and disciplines.

### *Examination discourse*

In their reports examiners are consciously positioning themselves in relation to knowledge – what it is to know, how they 'know', what it is important to know and why. It can be anticipated that examiners, as members of a particular group, will share a familiar set of common-sense understandings about research at the PhD level and what is acceptable. Such understandings (or at least the interpretative repertoire they draw on to express them) will be captured in what they say about examination in their reports. Multiple perspectives, as well as a range of approaches, need to be brought to bear to elicit what it is that examiners look for, emphasise, and act on, and to determine commonalities that may indicate consensus in practice.

7. In what ways does examiner comment contribute to our understanding of the skills and knowledge specific to PhD study, the role and traditions of the PhD and the features of disciplined inquiry at that level? In what ways can examiner comment inform research pedagogy, specifically thesis supervision?

## ACCESS & ETHICS

Heightened concern about quality assurance, particularly in regard to examination procedures, examiner consistency and high attrition have placed research higher degree examination in the spotlight. Access to information was never going to be an issue for such an important topic, given the acceptance of assurances that the data collection was to be handled responsibly and ethically. It was of interest to find, however, that not all institutions are in a position to access historical information on candidates and tie it in with examiners reports and recommendations even though they collect all three sets of information.

In Australia access by researchers to such information is underpinned by the requirement that the data are gathered and de-identified by the legal custodians (i.e staff in graduate schools or higher degree research administrative units) which is required by Commonwealth Privacy Legislation (Guidelines approved under section 95A of the *Privacy Act 1988, Privacy Amendment (Private Sector) Act 2000*). A further extension is not to use identifying material. In our study quotation has to be screened and where necessary edited to make sure that it will not identify an individual or institution. It would prove difficult to report data by discipline on an institutional case-by-case basis for the same reasons – confidentiality of reports would be compromised in a discipline with few candidates in a year. The solution adopted is to group smaller disciplines into larger fields. For each institutional case, disciplines are grouped by up to ten Broad Field of Study (BFOS) specified by the Australian Government Department of Education, Science and Training (DEST). These are:

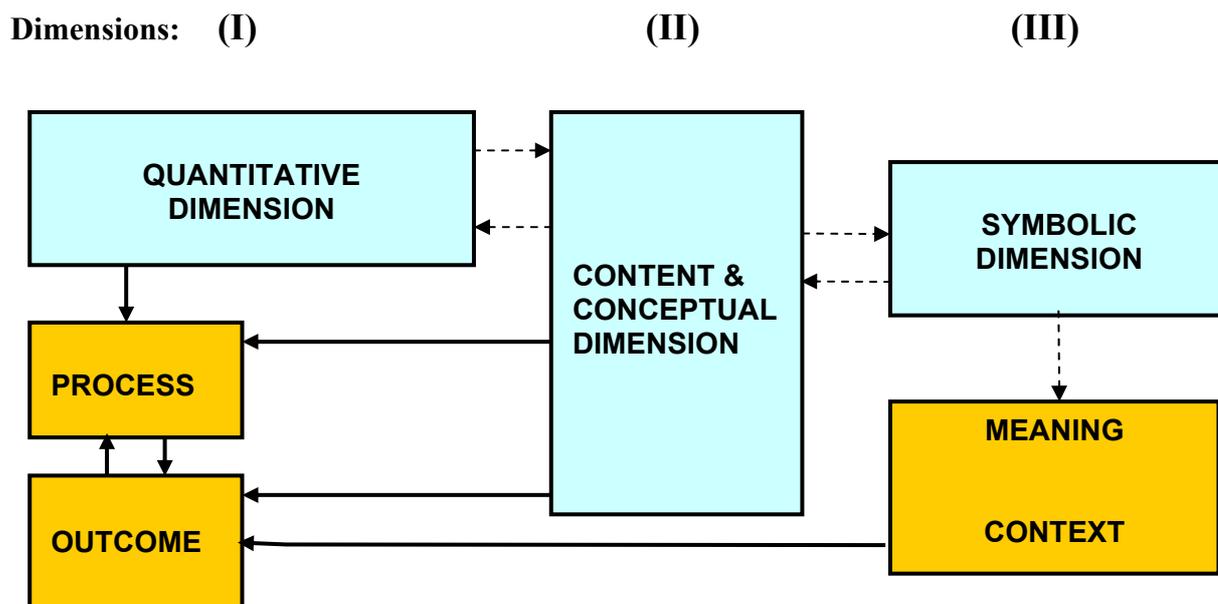
- Agriculture and Animal Husbandry
- Architecture and the Built Environment
- Arts, Humanities and Social Sciences
- Business, Administration and Economics
- Education
- Engineering and Surveying
- Health
- Law and Legal Studies
- Science and Mathematics
- Veterinary Science

## THE STUDY DESIGN

To obtain as direct grasp as possible of examiner consistency, thesis qualities, examiner execution of role and application of standards, a multi-dimensional study was called for, not only to allow us to draw as much as possible from a complex data set, but to provide a firm structure for validation, theory building and testing. The model developed has three dimensions that arise from the nature of the raw data, its treatment and how the information elicited (including contextually embedded information) contributes to an understanding of process and outcome. Process is also hypothesised as a factor contributing to outcome. The linearity of the diagram depicting the study in Figure 1 suggests a continuum of methods – statistical at one end (I) and interpretative at the other (III). The arrows suggest sequence but also an integrated information flow. This design has sequential and concurrent elements. Data collection is concurrent, some of the key analyses are independent and can also occur together but some are dependent, for example the qualitative coding of the examiner reports precedes the analysis linking candidate and outcome data to examiner comment. Moreover there is overlap between analysis of data in one case and data collection and analysis in the next.

The first of the dimensions in Figure 1 focuses on the quantitative information obtained from university records, and from the quantifiable elements of the text in the examiner reports (core coding categories). The second dimension focuses on the core features and attributes of the reports. The reports are subjected to text analyses that identify content but which also categorise patterns, emphases, and discursive and other communicative qualities. The third dimension moves the study into the interpretative realm.

The study of the symbolic (III) contributes a unique critical line of questioning that, juxtaposed with the other dimensions, assists to refine and question findings emerging from the core statistical and text analyses. In particular it keeps to the fore the language and traditions of the disciplines, the intentions behind examination, and acts as a brake against oversimplification. The aim of the analysis is to locate, contextualise and appraise comments that are clearly related to different apprehensions and uses of knowledge. Such an approach will further assist in the identification and elaboration of features consistently associated with high rating PhD theses across disciplines.



**Figure 1. Methodological dimensions and links**

In Figure 1 the broken lines and arrows signify paths of analysis, and also suggest how the methods are integrated. The unbroken lines signify the flow of interpretation that will contribute to understanding involving collaborative engagement.

The quantitative dimension draws on two forms of raw data, text counts generated through content and conceptual analysis of examiner reports, and pre-coded data drawn from student records including the thesis recommendation given by each examiner and the final or institutional decision. There is a reciprocal flow in the analysis as findings from one dimension contribute to extended and more theoretically-driven questioning and analysis in other dimensions. Similarly the content and conceptual analyses provide both a navigational aid and source of questions to assist in extended forms of analysis (symbolic dimension). These in turn may further inform or refine the content and conceptual dimension of the text analysis, occasionally leading to the development of a text coding category that may be applicable to quantification.

The culture and language of the doctorate, what it is to become accepted as ‘Doctor’, and the disciplinary knowing that this assumes, all contribute layers to the examination process that range from clearly articulated expectation to assumption and myth. The examiner report is a limited window on the latter, yet the symbolic layer evident in the organisation and language of the reports allows us to explore the situated and self-referential nature of thesis evaluation. Extended analysis of the text has already played a role in the interpretation of examiner comment to support the development of core coding categories (a contribution represented by the second horizontal dotted arrow on the far right of Figure 1).

The research questions are addressed by the following methods singly or in combination:

- Cross tabulation (Questions 2, 3, 5; Dimensions I & II)
- Correlational analysis (Question 1; Dimension I)
- Correlational/causal analysis (Questions 4, 6; Dimensions I & II)
- Content & conceptual analyses (Questions 2, 3, 4, 5, 6; Dimensions I, II, & III)
- Semiotic analyses (Questions 3, 5, 7; Dimensions II & III)

Ultimately it may be possible to predict thesis recommendations based on the categories identified in the reports, and so obtain stable and consistent representation of examiner report content to guide new examiners and to allow for international and inter-institutional comparison of thesis quality.

The core analyses are designed to be replicated across institutions, and the data sets merged and compared. Given the volume of data involved, sequences and processes have been devised to facilitate validation procedures, expedite methodological integration and allow regular reporting.

### **SAMPLING AND DATA COLLECTION**

The initial study design called for the sampling of nine universities from the 35 Australian universities that were judged to have sufficient numbers of PhD students across a range of disciplines. Each university would be asked to provide information for the most recent 100 candidates who had submitted a PhD thesis for examination across all their discipline areas, and for whom the examination process was complete. Such a sample of 900 candidates would provide between 2400 and 2700 examiner reports, depending on the number of examiners required by each university.

Selection of the nine universities was based on research quantum (i.e. the research income of institutions including income generated by PhD candidature) which provides a stable basis of institutional classification over time. Universities were divided into three categories on the basis of research quantum: high (consisting of 8 universities), medium (14) and low (13). The intent was to sample three universities from each category to provide sufficient numbers of students for each of the major BFOS for stability in estimates of candidate and examination variables by type of institution to be made. Although there was no evidence to hand that State of location or institutional size is related to the work of research students, care was also taken to select universities to ensure representativeness by both geographic area and size.

When a university declined to participate in the study, its replacement was another university from the same category that most closely matched it on the other criteria of location and size. The three universities that declined did so for different reasons. One university indicated that their research higher degree candidature records were inadequate for the data collection. Another replied that they had no centralised file system for research higher degrees in place, and the dispersed nature of their administrative arrangements made their participation impracticable. The other university initially agreed to participate but subsequently decided not to proceed because of workload pressure. The participating institutions were offered funding to support the appointment of casual staff to undertake the data collection.

The initial pilot university was from the medium research quantum group and these data were used to develop the core coding categories and key procedures for the examiner reports. The categories were subsequently checked with the reports from the second university in the sample, also from the medium research quantum group, and a few minor adjustments made. Eventually data were collected from a total of eight universities, three from each of the high and medium research quantum groups, and two from the low group. There were three reasons for reducing the intended sample from nine to eight universities. First, sampling of the major discipline areas was adequate for stability of results with eight universities involved – approximately 800 candidates. Second, the proportions of text categories and relationships between them and examiner recommendations did not alter significantly as we progressively

analysed the reports for the first four universities – approximately 400 candidates and 1100 examiner reports. We judged that when the total of 2100 examiner reports (approximately) were coded and analysed, the stability of text category estimates made would be high. Finally, the funding received for the study was not all that was requested, which would have necessitated a reduction in the sample in any case.

Two main sources were used to obtain data from university records. They are the pre-coded candidature information and the full text of the examiner reports.

*University records*

Not all institutions keep records in exactly the same form. The table below indicates what proved possible to obtain uniformly. Each case requires, in pre-coded form, the doctoral candidate personal, enrolment and supervision history and examination information. Students and examiners are given paired numeric identification and all identifying information is removed before it is entered in coded form in *eXcel* format.

<i>Candidate information:</i>	<i>Candidature information:</i>	<i>Examination information:</i>
entry qualification	length of candidature; duration in equivalent full-time semesters and ratio of full-time to part time semesters	Examiner recommendation
citizenship	discipline area; scholarship and fee details; any upgrade from another research degree	examiner gender
age	change of supervisor and reason,	examiner location
gender	supervision type (for example, sole supervisor, co-supervisor, etc.),	final 'institutional' decision on the thesis
English proficiency	supervisor experience (choice of 3 designations, i.e. inexperienced =1 student, some experience = 2-5 students, very experienced = more than 5 students),  student leave of absence - type and reasons given, and  when and if a problem in candidature is flagged by a supervisor or candidate	

**Figure 2. Information collected from University records**

*Examiner reports*

Photocopies of the original reports (and re-examination reports where pertinent) on each thesis were obtained from each university with identifying information removed. They are scanned electronically (or re-typed if necessary). The scanned data are archived, and an electronic copy is formatted using a standard set of procedures that allow text unit comparability. A text unit in this study is a typed line with a standard number of characters. The scanned copy is checked against the original for errors caused by scanning. Some other minor typographic errors or abbreviations that may have existed in the originals are also corrected in so far as they do not change the sense of the report. The latter step is necessary because typographical errors can impede text string searches. The reports are then prepared for N6 software in Courier font with 80 characters per line and single spacing. Where there are sub-headings these are collapsed into the text followed by a colon. If a diagram, picture or equation is featured in the original report the equivalent line count is estimated and inserted

(the characteristics of the special feature are described by the researcher entered in brackets within that line allowance). Diagrams and pictures are relatively rare, equations are more common in certain disciplines. It takes on average 3 hours to prepare and code data for each candidate.

### *Statistical analyses*

The core statistical analyses take four forms.

- Statistical analyses of candidate information
- Statistical analyses of the coded text of examiners reports (the measures include instances of occurrence in the report, and proportion of total text units, using standard line lengths, in the report)
- Relational analyses between candidate information and text coded by category
- Comparison between core data sets

The quantitative analyses undertaken in addressing six of the seven main research questions are:

- Correlation of ratings between examiners for each discipline area and comparison of coefficients by discipline (Q.1).
- Tabulation and comparison of categories of evaluative comment by examiner, and by discipline (Q.2).
- Tabulation of structures and qualities of reports and descriptive comparison. Report content is quantifiable for most attributes in terms of each category text unit counts as a proportion of total text units, or alternatively by counts of occurrence, regardless of length (Q.3).
- Factor analysis of examiner report categories to develop five constructs, including two based on evaluative comment. Multiple regression analyses with final examiner recommendation and evaluative comment constructs as dependent variables with characteristics of student, candidature, examiner, institution and discipline as independent variables. These analyses are multi-level, with examiner report data at level 1, candidate data at level 2, and either institutions or disciplines as alternative level 3 variables (Q.4).
- Tabulation and descriptive comparison of findings relating to standards and qualities in theses across examiners, candidates, institutions and disciplines (Q.5).
- Repeating the analyses as for Q.4 but with constructs related to thesis quality as dependent variables (Q.6).

### *Text analyses*

The qualitative data analysis software *QSR N6* supports a mixed method approach (Bazeley, 1999, 2003). Searching and retrieval are based on Boolean and contextual operators. Hence each text unit can be coded more than once to reflect interconnecting layers of information. Because it is possible to standardise the examiner reports to a particular format, comparable measures (based on line counts) are possible for a range of features of the reports where such an approach is helpful and meaningful, including proportions of text units coded by category, and the number, pattern and sequence of instances of coded text.

All of the text units associated with the 'examination' of the thesis are coded at least at one node (i.e. coding category) in *QSR N6* software. The core categories were arrived at after coding trials. They emerged from the pilot study report text, however, the fine definition in some of the categories (particularly between types of evaluative comment) was arrived at through close study of the assessment literature and doctoral studies literature and close discussions with colleagues in the area of assessment. In their final form these core categories constitute the most stable and replicable patterns emerging from the data.

The core content categories were tested in the course of establishing inter-rater agreement, through peer review, iteration across the pilot data and data for second institution coded, and through confirmatory factor analysis of relationships between categories. Because high inter-rater agreement is required during the core coding stage (we aim for 90 per cent agreement between two coders) detailed coding notes with examples from the reports were developed, tested and progressively refined.

Each of the core categories tells us something fundamental about the report, from its structure to the nature of examiner judgment. The text they capture is quantified but also explored for further patterns, contrasting and irregular information and deeper thematic structures. These further explorations then form the basis for probing questioning and lines of inquiry against which existing theories and emerging ideas can be tested. Given the breadth of the data-set it was always envisaged we could draw on additional expertise, that is to involve colleagues as well as discipline experts as required during the course of the main project.

The core codes are built on a hierarchical structure of ‘parent’ or primary coding categories. Each of them has first and level sub-categories. The act of coding occurs at the ‘child’ or 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 are:

**CATEGORY 1: REPORT ORGANISATION**

How the examiner structures their report

**CATEGORY 2: EXAMINER AND PROCESS**

The elements of the report where examiners give us information about themselves, what they know of examination and the processes they are using

**CATEGORY 3: ASSESSABLE AREAS COVERED**

All comment about the possible outcomes, subject matter and presentation of the thesis under examination

**CATEGORY 4: DIALOGIC ELEMENTS**

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).

**CATEGORY 5: EVALUATIVE ELEMENTS**

All comment that contains evaluation and judgement

The sub-categories are explicated and illustrated in Holbrook, Bourke, Lovat & Dally (2004b).

Examiner reports constitute a complex discursive terrain. While the two coders are immersed in the core coding they identify promising segments of text for later closer inspection. The category *Dialogic Elements* is particularly useful for this. The language of examiners exhibits situatedness, and nowhere as clearly as when they engage intensively with the subject matter, or in direct conversation with the reader. They may move into reflective mode, or draw on meta-narratives that reflect discipline or sub-culture. Segments, or the whole report, may exude a certain tone, for example, apology, frustration enthusiasm, etc. (See also Holbrook et.al 2004a for the tone of examiner reports in cases where examiners require revision and resubmission of a thesis). Tone may be indicated by repetition of a word or phrase, or sequence of remarks. Such features provide opportunities to investigate the dynamic of examination, conceptually and contextually (such as examiner role, and examiner expectations of thesis ‘readiness’), extending into the deeper layers of meaning about knowledge and discipline. Much of the analysis concerned with such elements will be prompted by questions and findings that emerge as the project moves beyond the core coding. Extended coding and analysis are primarily connected to Dimension III.

Given the nature and range of questions and the methods employed in this study, not to mention the heavy emphases on documentary data, the approach to validity has to be guided by complementarity in intent and clarity of position. Maxwell’s typology built on five

categories of qualitative ‘understanding’ underpinned by what qualitative researchers actually ‘do’ (1992, p. 281) and corresponding validity ‘types’ (descriptive, interpretative, generalisable, theoretical and evaluative) has general applicability in this mixed method approach, but specifically for Dimension II where qualitative and quantitative approaches merge. The typology provides a checklist against which the ‘kinds of threats to validity’ can be considered (p. 296). For the emergent textual analyses in Dimension III where the quality of the interpretation rests on its explanatory power, the positions adopted by Strauss and Corbin (1998) and Mishler (1990) have greater salience wherein validation constitutes a continuing dialectic between theory and analysis (p. 438). Concerns to ensure accuracy, visibility, attention to discrepant data, and cross validation are evident in every feature of the study. These concerns are also evident in the attempt to achieve the right balance of expertise and integration of methods. For example, the comprehensive sample is sufficient to provide some stability and hence generalisability of results for candidature information, such as gender and the major BFOS, required for Dimension I but also the purposeful case selection necessary for Dimension III.

#### *Data entry and analysis sequence and management*

One of the strategies we use to make sure that collaboration occurs and analysis is sustained is sequenced and continuous writing. Members meet informally face-to-face almost weekly, there is a project meeting monthly and emails and exchange of written material is frequent. Given the study involves eight institutional cases there also has to be a sequence of data collection, entry and analysis activities overseen by a project officer. The dual sequencing facilitates:

1. routine feedback to promote refinement of methods, analysis and interpretation of data
2. continuous reporting and publication
3. informed questioning and cross checking
4. theory building and testing
5. extension of analysis, finer-grained analysis
6. replication of core coding and analyses
7. ongoing case comparison

There are various phases of data entry and analyses. In the first phase, all written reports are coded but analysis is restricted to the initial examination of the thesis. The second phase adds the re-examination reports, and compares the re-examination reports with the initial reports. The third phase adds a new layer of data to the analysis, that of tone. This is one way we capture all evidence of either positive or negative orientation of evaluative comment. Extended coding builds on the core coding but it may be taking place while the core coding is occurring, as explained above.

Each case affords the opportunity to check the accuracy of the coding, to undertake inter-rater reliability checks (when data are being entered) and to refine the coding notes. The process of reporting of data by case and node provides a further check on coding consistency between cases. The latter is important because each set of reports may be structured differently, based on institutional guidelines, and there is a different mix of disciplines per institution.

## **DISCUSSION**

This article has detailed the mixed methods design of a study that addresses questions about PhD thesis examination process in Australia, examiner consistency, the nature of examination and the quality of thesis outcomes. It has also provided some insights into the management of the project which involves eight Australian universities, a core part of which is intended to be replicable, and the procedures adopted to facilitate validation processes.

Mixed method designs are often employed in educational research, but the detail behind the design decisions is rarely provided in the literature. Rocco et al. (2002) see this as a serious flaw, particularly in relation to judgements about research quality and its usefulness. One way to present the detail is to draw on the various frames of reference and typologies described in the emerging body of mixed methods literature (Teddlie & Tashakkori, 2003; Tashakkori & Teddlie, 2003), however, for us it was not a matter of choosing a mixed methods design of a certain type in order to do the research, but a shared personal and professional interest in the area of PhD Examination, and a predilection for the productive challenges posed by paradigmatic diversity. In short we had identified a very interesting and topical area tied to a difficult-to-access and extraordinarily rich source of information, and we wanted to:

- address the major questions raised about PhD assessment – about consistency of examination, quality of outcome, and the nature of the degree—that had not been answered, but needed to be in the contemporary context of higher education
- do justice to the quality and scope of the information generated by the PhD examination process, and
- make the most of collaboration

Our stance was ‘pragmatic’, that is question driven, insofar as the literature and our personal experience of the phenomenon guided the study. The questions are essentially about assessment – outcomes and processes. However the phenomenon of assessment of the PhD (see question 7 above) captures the quintessential complexities of higher level learning, academic culture and research activity in the 21<sup>st</sup> century – what is known, how knowledge comes about, and how its value is determined. As researchers we are galvanised by the opportunity to explain, as far as we can, the complexity behind examination, and in so doing better understand our own practice as supervisors and examiners.

Our approach is as much driven by a dialectic position as a pragmatic one. There is a degree of freedom in this position that is signified in the device of both ‘core’ and ‘extended’ coding and analysis (see above) and in the multiple lines of questioning that reflect a range of purposes for the study. Specifically, these include strengthening the knowledge base, predicting outcomes and also understanding and exploring the meanings embedded in assessment activity at the pinnacle of university study, particularly in assessment discourse (see also Newman et al., 2003). The design has sequential and concurrent elements, as well as the fixed (e.g. the replicable core) and flexible features, across multiple institutional cases and there is a multi-dimensionality - temporal and collaborative - across all the components of the study (purposes, conceptual framework, questions, methods, validation and the drawing of inferences) (see also Maxwell & Loomis 2003). There is complementarity in the methodological elements in our study, including in our ideas about what constitutes ‘good’ research. We see the different methods and paradigms we bring to bear as both strategic and supportive, offering the opportunity to verify ‘design quality and interpretive rigour’ (Teddlie & Tashakkori, 2003, p. 37) in both specifiable and holistic overarching ways. We employ and pool our range of distinctive talents and knowledge to this end and have identified literature on validation that intersects (Maxwell, 1992; Mishler, 1990; Strauss & Corbin, 1998) where the researchers’ knowledge and understandings also intersect.

It is both remarkable and intriguing that the assessment of the doctoral thesis, unlike assessment at other levels of education, has attracted so little large-scale empirical research. In designing such a study, however, it becomes quickly apparent that there are many factors that should be taken into account even in a higher education system which is reasonably uniform in its approach. In Australia there are minor variations in the options offered for examiner recommendation, and in institutional procedures and instructions. Other variations emerge from the source and background of examiners, number of examiners, and variations in their communication and reporting styles, and possible disciplinary differences. Yet others emerge from the relative ‘flexibility’ of the process in conjunction with the substantial nature

of the thesis. Examiner reports contain many strands of information, and among these there are various types of evaluation employed by examiners, multiple examiners, multiple foci and multiple purposes. It is this level of complexity, and the unique and unexplored qualities of PhD thesis assessment in relation to outcomes that call for a collaborative mixed method approach.

The coding of examiner reports is at the heart of the study, and the coding occurs in core and extended phases. The reports generally concentrate on specific components of a thesis and are organized in a range of ways attend to them. By dint of academic culture, history, and disciplinary and individual differences, their emphases as well as the arrangement, and substance of what they say can tell us a great deal about process, expectation and outcome. When the texts of 303 examiner reports on 101 theses were examined in the pilot phase, it was found that four broad categories of information were present in most and also that the way the examiner structured the report reflected both process and emphasis. Firstly, examiners tended to provide information about themselves and their expectations of the thesis (organization). Secondly, and very importantly from an assessment perspective, was what aspects of the thesis examiners emphasized in their assessment (examiner and process). A third category of information in the reports was concerned with the style of engagement with the thesis and the audience (dialogic elements). The common fourth layer of information was concerned with evaluative comment – how the thesis was judged and in what terms (evaluative elements).

The findings of the study will provide information that is directly applicable to postgraduate pedagogy and supervision practices, administration and examination. It will render visible the expectations of examiners and so should directly flow through to informing processes and procedure to the benefit of students, supervisors and examiners. By clarifying thesis standards, the study can provide new information relevant to the field of learning theory, contribute to the constructive critique and development of disciplined inquiry, and facilitate comparison, mapping and strategic planning for research training nationally and internationally, with the ultimate aim of enhancing research performance.

## REFERENCES

- Bazeley, P. (1999). The Bricoleur with a computer: Piecing together qualitative and quantitative data. *Qualitative Health Research*, 9 (2), 279-287.
- Bazeley, P. (2003). Computerized data analysis for mixed methods research. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioural Research* (pp. 385-422). Thousand Oaks: Sage Publications
- Becher T (1993). Graduate education in Britain: the view from the ground, In B.R. Clark (Ed.) *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan* (pp.115-53). Berkeley, University of California Press.
- Bourke, S., Hattie, J. & Anderson, L. (2004). Predicting examiner recommendations on PhD theses. Paper presented at the EARLI, Higher Education SIG Conference, Stockholm, June.
- Bourner, T., Bowden, R. & Laing, S. (2001). Professional Doctorates in England. *Studies in Higher Education*, 26 (1), 65-83.
- Bowen, W.G. & Rudenstine, N.L. (1992). *In Pursuit of the Ph. D.*. Princeton NJ: Princeton University Press.
- Burgio-Ficca, C. (2001). The contribution of the higher education sector to Australian research and development. *Australian Journal of Education*, 45 (3), 263-283.
- Clark, B. R. (ed.) (1993). *The Research Foundations of Graduate Education: Germany, Britain, France, United States, Japan*. Oxford, England: The University of California Press.
- Denicolo, P. (2003). Assessing the PhD: A constructive view of criteria. *Quality Assurance in Education*, 11 (2), 84-91.
- Gilbert, R. (2004). A framework for evaluating the doctoral curriculum. *Assessment & Evaluation in Higher Education*, 29 (3), 299-309.

- Gilbert, R., Balatti, J., Turner, P. & Whitehouse, H. (2004). The generic skills debate in research higher degrees. *Higher Education Research & Development*, 23 (3), 375-88.
- 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.
- Hoddell, S., Street, & Wildblood, H. (2002). Doctorates – converging or diverging patterns of provision. *Quality Assurance in Education*, 10 (2), 61-70.
- 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.
- 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. viva. *Assessment & Evaluation in Higher Education*, 26 (4), 355-366.
- Johnston, S. (1997). Examining the examiners: an analysis of examiners' report on doctoral theses. *Studies in Higher Education*, 22 (3), 333-347.
- Johnston, B. & Murray, R. (2004). New routes to the PhD: cause for concern? *Higher Education Quarterly*, 58 (1), 31-42.
- 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.
- Kouptsov, O. (1994). *The Doctorate in the Europe Region*. Bucharest: CEPES UNESCO.
- 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.
- Maxwell, J.A. (1992). Understanding and Validity in Qualitative Research, *Harvard Educational Review*, 62 (3), 279-300.
- Maxwell, J. A. & Loomis, D. M. (2003). Mixed methods design: an alternative approach. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioral Research* (pp. 241-271). Thousand Oaks: Sage Publications.
- Mishler, E.G. (1990). Validation in Inquiry-Guided Research: The Role of Exemplars in Narrative Studies., *Harvard Educational Review*, 60 (4), 415-442.
- 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.
- Newman, I., Ridenour, C.S., Newman, C. DeMarco Jr, G.M.P. (2003). A typology of research purposes and its relationship to mixed methods. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioural Research* (pp. 167-188 ). Thousand Oaks: Sage Publications.
- Nightingale, P. (1984). Examination of research theses. *Higher Education Research and Development*, 3 (2), 137-150.
- Noble, K. (1994). *Changing Doctoral Degrees: An International Perspective*. Buckingham: Society for Research into Higher Education and Open University Press.

- O'Brien, P. (1995). The reform of doctoral dissertations in humanities and social studies. *Higher Education Review*, 28 (1), 3-19.
- 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), 104-115
- Powell, S. & McCauley, C. (2003). The process of examining research degrees: some issues of quality. *Quality Assurance in Education*, 11 (2), 73-83.
- Rocco, T. S., Bliss, L. A., Gallagher, S. Perez-Prado, A., Alacaci, c, Dwyer, E.S., Fine, J.C. & Pappamihel, N. E. (2003). In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioural Research* (pp. 595-615). Thousand Oaks: Sage Publications.
- 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 Open University Press.
- Shaw, M. & Green, D. (2002). Benchmarking the Ph. D. – a tentative beginning. *Quality Assurance in Education*, 10 (2), 116-124.
- Shulha L.M. & Wilson R. J. (2003). Collaborative Mixed Methods Research. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioural Research* (pp.639-669). Thousand Oaks: Sage Publications.
- Strauss, A. & Corbin, J. (1998). *Basics of Qualitative Research*. Thousand Oaks, CA: Sage.
- Tashakkori, A. & Teddlie, C. (2003). The past and future of mixed methods research: from data triangulation to mixed model designs. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioural Research* (pp.671-701). Thousand Oaks: Sage Publications.
- Teddlie, C. & Tashakkori, A. (2003). Major Issues and controversies in the use of mixed methods in the social and behavioral sciences. In A. Tashakkori & C. Teddlie (eds). *Handbook of Mixed Methods in Social and Behavioral Research* (pp.3-50 ). Thousand Oaks: Sage Publications.
- Tinkler, P. & Jackson, C. (2000). Examining the Doctorate: institutional policy and the Ph. D. examination process in Britain. *Studies in Higher Education*, 25 (2), 67-180.
- Tinkler, P. & Jackson, C. (2004). *The Doctoral Examination Process: A Handbook for Students, Examiners and Supervisor*. 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.
- Winter, R., Griffiths, M. & Green, K. (2000). The 'academic' qualities of practice: what are the criteria for a practice-based Ph. D.? *Studies in Higher Education*, 25 (1), 25-37.
- Wright, T. & Cochrane, R. (2000). Factors influencing successful submission of Ph. D. theses. *Studies in Higher Education*, 25 (2), 181-196.