





# What makes a good PhD thesis? Norms of science as reflected in written assessments of PhD theses

Sofie Kobayashi  and Claus Emmeche 

Department of Science Education, University of Copenhagen, Copenhagen, Denmark

## ABSTRACT

This study looks at assessment of PhD theses from two perspectives: criteria in use in assessment reports at a science faculty and norms of science. Fifty assessment reports were analysed inductively, resulting in thirteen categories that examiners consider when assessing a thesis. These categories were compared with norms of science as described in the sociology of science. The study shows a high congruence between the two perspectives, but also new categories worthy of further discussion and research. Relevance of the thesis research and quality by proxy (that publication is an indicator of quality) stand out very clearly in this study compared with earlier assessment research. These two categories are both relatively new categories in assessment research and indicate that the classical norms of science are changing with an increasing influence of post-academic norms in academia.

## KEYWORDS

Research evaluation;  
doctoral examination;  
PhD dissertations;  
assessment criteria;  
sociology of science

## Introduction

Norms of science are collective expectations for appropriate behaviour within the scholarly community. They concern scientists' motivations, how to relate to other scientists, standards for and evaluation of scientific work (Anderson et al. 2010). Scientists may, to varying degrees, subscribe to or resist these norms. Examiners are gatekeepers for membership of the scholarly community, governed by norms of science, while they also provide feedback that can support candidates in their further academic development (Stracke and Kumar 2010). Norms of science have been studied in the sociology of science, while assessment of PhD theses has been studied in the field of higher education and assessment. How is assessment of PhD theses connected with the norms of science? We explore this obvious question that seems to have escaped earlier research.

Assessment of PhD theses is assigned to experienced researchers in the field in question, as only they can judge the quality of a thesis. Assessment guidelines provided by universities are very general, to accommodate disciplinary differences and to leave room for interpretation and negotiation among committee members. However, there is an ongoing discussion and a call for more specific assessment criteria (e.g. Sharmini et al. 2015; Krumsvik, Øfstegaard, and Jones 2016). Research that explores assessment of PhD theses can qualify this discussion and eventually help develop a globally recognised understanding of the standards required for a PhD degree.

Understanding the criteria is a cornerstone in developing autonomy and self-regulated learning (Stracke and Kumar 2010). Gurr (2001, 85) describes the aim of research education as producing graduates with 'competent autonomy' who 'are cognisant of the norms, expectations and standards within their discipline and are able to assess their own plans and actions to ensure compliance with these'. A deeper understanding of the norms and assessment criteria can inform examiners, supervisors and PhD students about the demands for a PhD degree. In this study we analyse assessment reports in a science faculty at a Danish university to identify which norms come into play when criticism or praise is given in assessments of PhD theses in the natural sciences. Our research questions are:

1. Which evaluative themes emerge from an analysis of examiners' comments?
2. Can such themes be related to the norms described by the sociology of science?

In the Danish context, the committee makes a preliminary assessment of the thesis and concludes whether it is acceptable for defence and the candidate has made a 'significant contribution to the development of new knowledge', as stated in the Danish Qualifications Framework ([www.ufm.dk/en](http://www.ufm.dk/en)). Critique given in a positive assessment can be addressed at the defence. The conclusion must ideally be based on the committee's deliberations, the evaluative details and argument. Theses in the natural sciences are typically article-based with three to five manuscripts or amended published papers, and a synopsis to introduce and connect these. The assessment committee has three members, one of whom is recruited internationally (as in Norway, see Kyvik and Thune 2015) and the committee writes one common report.

## Background

We intend to bridge the scholarly domains of assessment of PhD theses and the sociology of science.

### *Assessment of PhD theses*

Assessment of PhD theses has been studied for decades. An early study was conducted by Johnston (1997), who made a qualitative analysis of 51 examiners' reports. She found two overall types of comments: one directed at the significance of the study, originality and scholarship, and one concerning communication. Golding, Sharmini, and Lazarovitch (2014) contributed with a thorough literature review that we draw on here.

### *Presentation*

Good use of language, style and terminology enhances readability, and this also reflects good communication skills (Holbrook et al. 2004). When a thesis is not well presented, examiners get irritated, and may even lose confidence in the quality of the research (Golding, Sharmini, and Lazarovitch 2014).

### *Coherence*

A thesis should be coherent, with a common thread and logical structure linking a clear research problem with an answer, linking the findings with the literature, and aligning the use of theory, research design and research question (Golding, Sharmini, and Lazarovitch 2014; Holbrook, Bourke, and Fairbairn 2015).

## Literature

The review of literature should be adequate, up to date and relevant, without omissions of central references, but not necessarily exhaustive (Holbrook et al. 2007; Chetcuti, Cacciottolo, and Vella 2022). Literature should be critically appraised to reflect a comprehensive understanding, and it should be used with critical engagement throughout the thesis.

## Methodology

The approach should be appropriate, well-explained and justified, with motivated choice of methods (Golding, Sharmini, and Lazarovitch 2014). Holbrook, Bourke, and Fairbairn (2015) emphasise that theory should be used in alignment with research design and research questions, and Chetcuti, Cacciottolo, and Vella (2022) confirm this.

## Results

Results should be accurately presented (Chetcuti, Cacciottolo, and Vella 2022) and critically discussed (Ramlall, Singaram, and Sommerville 2020). Golding, Sharmini, and Lazarovitch (2014) advise that results should be interpreted, analysed and critically appraised and connected with literature, conclusions drawn from them, and implications shown.

## Contribution

While other aspects are important, the contribution (original or significant) of new knowledge is a requirement. Golding, Sharmini, and Lazarovitch (2014) state that a thesis must be publishable, and that 'the most convincing way to show examiners that a thesis is publishable is to publish from it' (571). When discussing what a contribution to new knowledge entails, the terms 'novelty' and 'originality' are often used (Clarke and Lunt 2014; Poole 2015; Chetcuti, Cacciottolo, and Vella 2022). Poole (2015) argues that publishability is a better term, since originality may mean anything from 'not copied' to 'ground-breaking', and examiners have extensive experience in judging whether research is publishable.

There are of course other ways to 'cut the cake', and Kyvik and Thune (2015) use 'external relevance' as a category, referring to applied, societal, cultural and industrial relevance. Chetcuti, Cacciottolo, and Vella (2022) found that a contribution was significant when it had a utilitarian purpose, making the category of relevance an aspect of the significant contribution.

## Sociology of science

Founded by Merton in the mid-twentieth century, the study of norms by the sociology of science moved from initial tentative formulations (Merton 1942) to more empirical investigations. The CUDOS norms (*universalism, communalism, disinterestedness, originality and organised scepticism*) originally articulated by Merton became much discussed in the sociology of science. Additional norms and counter-norms were found (Mitroff 1974; Ziman 2000; Anderson et al. 2010; Kim and Kim 2018), partly based on interviews and surveys in which scientists self-reported their opinions of what 'good science' means. As the norms of science can be thought of both as ideals to strive for and as socio-cognitive structures of socialisation that guide real actions (also in supervision, Kobayashi and Berge 2022), it may be advantageous to investigate evidence of such norms *via* actual assessments that researchers make of each other (an expression of the norm of organised scepticism) in peer-reviews of draft papers, funding applications and assessments of PhD theses. There are important additional norms about publications, authorship and proper use of citations, as summarised in the prescription 'to give credit where credit is due' (Hoffmann et al. 2016). The study of citation norms took its point of departure in Merton's work (Small 2004).

A focus for Anderson's group was not only to give a fuller, evidence-based picture of the complete set of norms governing science, but also to investigate the extent to which scientists deviate from these norms *via* 'questionable research practices' (QRP), denoting a grey zone between 'responsible conduct of research' (RCR) and outright breaks on scientific integrity by *falsification* in the sense of data forgery and fraudulent research, *fabrication* (of data), and *plagiarism* (FFP). Anderson's group (Martinson, Anderson, and de Vries 2005) found a surprisingly high share of researchers (33%) who anonymously self-reported that they had engaged in at least one of the top ten behaviours of QRP, e.g. 'failing to present data that contradict one's own previous research' or 'changing the design, methodology or results of a study in response to pressure from a funding source'. Now, research training and supervision of PhD students aim at socialising the student into RCR and avoiding practices of especially FFP but also QRP. Given the nature of PhD supervision, we expected to find no assessment comments on FFP, but some comments on QRP that include both deliberate forms of slight dishonesty in data handling and unintentional forms of bias, negligence, sloppiness and common errors (Allchin 2001).

PhD theses are the result of an intersection of original research and research education, and we did not expect to find the full range of norms/counternorms governing the whole of science (Anderson et al. 2010) in PhD thesis assessments. Neither was Ziman's (2000) work of extending the normative taxonomies to cover research in industrial settings (the PLACE-norms, i.e. research as *proprietary*, *local*, *authoritarian*, *commissioned* and *expert*) expected to be pertinent in the evaluations, but his norm of *utility* (for solving societal challenges) might play a role.

## Methodology

This project obtained ethical approval from the university legal unit. A third party anonymised 230 assessment reports from 2016 by deleting names and addresses of persons and departments. Each assessment report is 2–5 pages. All theses in this study were article-based, and all assessments recommended the thesis to be accepted for defence (in practice a pass). All assessment reports were written in English.

Based upon a first close reading of 50 randomly selected assessment reports, we identified excerpts that contained a judgment rather than being purely descriptive. The excerpts were thematized, following procedures from Braun and Clarke (2006), aiming at inductive identification. Based on the judgments, we constructed a list of evaluative categories. We continued analysing assessment reports beyond saturation that we reached after 18 reports (no new categories emerged). The rest of the reports contributed more nuance to make thicker descriptions of the categories. Each category is exemplified by one or more quotations from the reports to provide a richer description of the assessors' use of criteria. The categories that emerged were then compared with the norms described within the sociology of science.

## Results

Our inductive thematic analysis resulted in the following 13 evaluative categories and subcategories that examiners consider:

1. Contribution to research
2. Relevance
3. Quality – overall
  - 3.1. Empirical dimension
  - 3.2. Theoretical dimension

- 3.3. Methodological dimension
- 3.4. Connecting theory and experimental work
- 4. Proxy for quality
- 5. Quantity
- 6. Focus and depth
- 7. Contribution in collaboration
- 8. Overview of the earlier research in the field
- 9. Own contribution in relation to previous work
- 10. Ability to discuss and judge own results
- 11. Communication (structure, coherence, clarity, illustrations, language, editing)
- 12. Engagement
- 13. Conditions

These categories can all be seen as describing aspects of the overall quality of the thesis work – it is difficult to imagine a thesis making no contribution whatsoever still having high overall research quality. Yet, from our analysis, quality (3) emerged as an evaluative category with four subcategories (3.1 to 3.4), and in general, different aspects of the categories can overlap. Many of the assessors' comments are complex and obviously cut across different aspects of the judgement of whether the thesis is ready for defence.

We found more positive comments than negative, perhaps because all assessment reports recommended the thesis to be accepted for defence. In the descriptions of categories, we have included both praise and criticism, because criticism often points more precisely to assessment criteria, while positive comments tend to be more general.

### **Contribution to research**

This category concerns assessments of achievements that bring science forward. Statements like *The most significant contributions of the thesis are...* (002) signify an overall assessment, while the category also includes subcategories pertaining to four identified dimensions: the empirical, theoretical, methodological and theory-experiment connection. Excerpts related to the overall assessment for instance read:

All the presented results represent novel research which can help advance the field (004)

This might very well become a classic study of... (042)

Some results were assessed as important because they go against common understanding:

This challenges the long-held view that... (016)

### **Relevance**

Judgements assigned to this category commend the research for its relevance for society in contexts like climate change, public health, agriculture or environment. Assessment may concern research objectives as well as results:

The research objective of the thesis is important, especially as the reduction of greenhouse gases is recognised as significant research question for society. (007)

The data show that... might lead to important consideration for human health. (027)

Relevance can refer to multiple dimensions, as in this case *'forest monitoring using the wealth and detailed [type of] data is a very important and 'hot' topic in this field with relevance for both science, services and policy development'* (040).

## Quality

When the assessment committees make statements about the overall thesis and its papers, they judge how well the work lives up to the standards of good science, its degree of excellence, etc. Statements like:

The quality of the results presented in this thesis, and by extension, the papers that comprise it, is high and notable (002)

exemplify an overall norm of quality, without actually defining quality. A compliment like:

the work deserves commendation for approaching some old topics from a rather new and original perspective (017)

both touches upon originality as a central norm of science and the candidate's capacity for creative thinking. It also shows that quality pertains to many dimensions of a scientific work. These are not graded, but some committee members interpose a comparative aspect in their evaluation:

Compared to other PhD dissertations that I have examined over the past ~10 years this is one of the best (top 10%). (047)

## Empirical dimension

This subcategory concerns field work, experiments, data collection and analysis, like *his experimental results do a commendable job* (002). Data shared among scientists have an important role, not only for judging the validity:

The study is of extraordinary quality and the morphological studies revealed a wealth of new data which may be important far beyond the current thesis itself. (009)

Yet, when data are shared, they should be annotated correctly:

However, it remains unclear whether the individuals providing the stock were collected in the field or come from a lab culture. (009)

From the assessments it is evident that there exist a variety of field-specific technical norms, standards and protocols which, if not followed correctly, raise criticism:

It would be interesting to know if the candidate completed negative controls. (016)

In general the presented data are of high quality and include relevant controls. (041)

## Theoretical dimension

This subcategory reflects evaluation of how the thesis contributes to theoretical development or applicability of theory to new areas. Development of specific models is also covered in this dimension:

a very useful simple model is formulated... which give an intuitive microscopic insight. (017)

The theoretical dimension also involves identifying gaps of knowledge and improving the connections between different fields:

[NN] has identified major gaps in understanding of... significant... historical events and has constructed and carried out novel studies that directly address them (026);

we are impressed by the work the candidate has done to merge these disciplines (026).

The latter comment reminds us about unification and coherence between theories as a general norm, especially in the physical sciences.

Testing theoretical hypotheses is a common, though challenging, component of research that may be absent:

There is a lack of hypothesis testing although some thesis formats do not lend themselves easily to hypothesis construction. (028)

The theoretical dimension *per se* generates relatively fewer judgements than the comments on data, methods and theory-data connection. This may reflect the brief, focused and modest character of thesis work in the natural sciences, often located within larger collaborative research projects.

### **Methodological dimension**

The use of method(s) also requires quality assessment. In the biological fields it can be the development of specially adapted model organisms:

she now suggests a new system which is comparatively simple, easy to access, easy to culture, and with a short generation time. (009)

In applied statistics concerns of computational or analytical efficiency are important:

The efficiency of the new method is demonstrated by application to both simulated and real life data. (011)

Efficiency and suitability are also pertinent concerns when alternative methods can be applied:

Some more reflection on the choice of methodology... as well as discussing more thoroughly the implications of applying the proposed methodologies to situations with fewer data constraints would have been recommendable. (007)

Learning a field's methodology and applying it to a broader range of phenomena can be commented:

[NN] has clearly obtained a very solid grasp of the key methods in the field of molecular magnetism, and here it is applied to a range of interesting compounds. (021)

### **Connecting theory and experimental work**

The theoretical and empirical aspects of research should be well connected in a thesis:

This PhD work contains strong experimental and theoretical parts, which is quite unique and it is a strong added value of the thesis. (017)

When this is not achieved, assessors may comment on lack of coherence that also touches upon norms like objectivity, scepticism and avoidance of confirmation bias:

the evaluation and discussion of results were not always objective and at places the candidate seemed to ignore data that did not conform with her a priori hypothesis and instead focused on data that pointed towards supporting ideas she had from the outset. (036)

The very character of the connection between theory and experimental data is rarely made as explicit as in this quote, in which this connection is deficient, and we assume that for most fields 'well-connectedness' is central.

### **Proxy for quality**

When a thesis includes published papers, the assessors frequently rely on the journal's peer-review system as a proxy for quality, and in that sense delegate the judgement of the PhD work to

other peers, with comments like *published in high quality journals* (002). This excerpt spells it out: *such placement of this contribution in a journal like this speaks in favor for its quality* (009). However, the committee does sometimes assess published articles independently:

Four of the included papers are already published or accepted for publication in some of the highest ranking peer-reviewed journals. The quality of the published papers and the manuscripts is considered excellent and they certainly fulfil the aim of the thesis. (014)

For co-authored papers, it is often mentioned whether the candidate is the lead author or had less status as a contributor, and more rarely the impact factor as a proxy for a journal's quality is mentioned:

NN's main work is published in a paper in the prominent journal [name mentioned] (Impact Factor = 6.767), in which he is joint-first author. (032)

The proxy for quality may be taken even further to citations, as witnessed in this assessment:

Paper 1 has been cited 4 times, and Paper 2 has been cited 11 times according to Google Scholar'. (018)

This is a little surprising as the validity of Google Scholar citation counts (*per se* and as a proxy for quality) is debatable.

### Quantity

This category concerning judgements of the amount of work completed within the time limit of three years in the Danish system, prompts comments like:

The work documented in the thesis is of high technical standard and the volume of the work is what is expected for a PhD degree to be awarded. (031)

The number of manuscripts or publications is also considered:

Moreover, [NN] has published another two papers (one of them as first author), which report on other projects not included in the thesis. This very impressive publication record is a testament to the substance and scope of the candidate's work as well as her productivity. (038)

We found that there are fewer explicit comments about the volume of the work than its qualities.

### Focus and depth

This category reflects an expectation that a PhD thesis should have one coherent narrative, rather than being a compilation of unconnected contributions, like *a series of well-written but disjointed sections that do not create a coherent narrative* (028).

The thesis does an admirable job of tying [5 papers] together into one 'research story'... but the mapping of research area to paper is a bit more one-to-one than the opening portion of the thesis suggests. (018)

However, not all assessors may subscribe to this expectation, as this excerpt reveals

The PhD study includes an impressive range of approaches and methods... The PhD student did an excellent job as evidenced by the novel, relevant and interesting results, the amount of work and dedication, and the fact that the thesis covers such a wide range of topics related to [topic]. (019)

Here, the 'lack of focus' was counter-balanced by the relevance of the work to meet a great societal challenge.



### **Contribution in collaborative research**

Research is increasingly practiced within larger teams, and apart from checking that the candidate did 'enough' work (quantity), the assessment committee may also assess the collaboration. As hinted at under *proxy for quality*, lead authorship in several papers may indicate a certain prowess at coordinating a collaboration:

All of the peer-reviewed papers have the candidate as the first author. (010)

Although evaluation committees are provided with 'co-author statements' explaining the candidate's specific contribution to co-authored papers, it seems sometimes challenging to assess the real distribution of work roles, and this can lead to attenuated criticism:

The candidate is first author on three of the manuscripts and third and fourth author on the two published papers. The contribution of the candidate to these manuscripts is not evident from the thesis, but it is recognized that research is and should be a collaborative effort... it was not possible to assess the contribution of the candidate to the work, but the thesis is evidence that she understands all aspects of the project. (027)

Clearly, the evaluation committee should not evaluate a whole research programme, but must judge the extent of the special contribution the individual PhD candidate made.

### **Overview of the earlier research in the field**

When a committee writes that the *various parts are well-referenced, citing many of the most relevant works in the various sub-fields* (003), it is referring to the way a PhD thesis should clarify its contribution in relation to the international 'state-of-the-art' within a discipline (this demand is explicit in the university guidelines).

Some committees express frustration when a literature review has too few details, but they recognise the constraints set by theses that are a collection of papers:

Given the nature of this compiled paper style of thesis and the space constraints given the publication format of each included paper, the review of related work has less depth and breadth than a traditional monograph-style thesis. (018)

An alternative is to include the state-of-art review in the introduction section before the papers:

the Introduction could better place the study in the context of current state-of-the-art. (019)

### **Own contribution in relation to previous work**

The thesis should also specify the candidate's own contribution to the advancement of knowledge:

especially chapter two is really interesting reading, because [name] weaves past research with his own reflections (015)

and a lack of such specification elicits critique:

in this connection it would have been appreciated with a clear statement of what is own work and what is taken from there. (003)

This theme can relate to the theme of focus and the need for a narrative giving a clear aim of the research:

calls for a clearer sense of direction: 'where is all this going', and 'what can we understand now, which we couldn't before?' (003)

This, again, is related to the theme of presentation.

### **Ability to discuss and judge own results**

This ability is commented on when lacking, *some more discussion of potential biases of the dataset would have been valuable* (034), or when examiners lack a discussion of an apparent contradiction:

It could have been discussed in the synopsis part why NN induced increase of [substance] promotes growth in paper II while in paper III the opposite... also seems to promote growth. (008)

This category also covers critical thinking:

there may be room for further discussion of certain assumptions underpinning the modelling exercises. (019)

The judgement relies on the committee's expertise as the word 'appropriate' indicates:

The papers/manuscripts to elucidate these areas are generally very well structured and the results are presented and evaluated a precise way with appropriate critical analysis. (031)

### **Presentation (structure, coherence, clarity, illustrations, language, editing)**

Most of the assessments comment on this category. The introduction (synopsis) should explain the role of the individual articles or manuscripts for the overall work. This comment:

Chapter 1, which represents the introduction to the thesis, could have benefitted from extra work to attend to grammatical and typographical errors and to reduce the extensive use of colloquial non-scientific language. Some paragraphs could have been more precise in order to give the reader a clearer picture of the subject... Arguments are not always clearly developed... a more structured writing could significantly improve the manuscript. It is a shame that often there is less time to get the introduction up to date making sure that all relevant references are read and discussed. (031)

indicates the importance of the synopsis, an aspect many PhD students (or supervisors) do not prioritise, as the manuscripts and papers get priority.

### **Engagement**

Examiners use positive words in their comments like dedication, enthusiasm, engagement or diligence. With comments like *These chapters give an impression of someone who is enthusiastic about the topic* (015) or *The incorporation of six supplementary papers is evidence of a very high level of engagement and diligence* (014) the examiners attempt to judge the candidates themselves through the thesis. If examiners find the work interesting to read, this reflects their own engagement in science: *We are very much interested in this research work and we congratulate NN for coming up with such a lively discussion* (011).

### **Conditions**

Judgements often depend on the conditions in which the work is carried out, and some committees try to explicate the context of the research:

It should be noted that a part of the presented results are from a brief three months visit to University of... Being able to produce experimental results from a new setup in such a short period is an achievement. (004)

this is still a highly competitive field of research. (009)

Criticism regarding the low quality of specific findings may be moderated with a comment like:

It should be noted that this branch of research is almost untouched, and therefore any contribution here is very important. (043)

The local conditions of doing a PhD in Denmark are also considered, *Given the three-years term*, (004) by the international assessment committee.

## Comparing assessment criteria with norms of science

The categories and subcategories that emerged from the analysis invite a comparison with the norms and counternorms known from the sociology of science. We find instances of congruence, new norms and absent norms.

### Congruence

Category 1, 'Contribution to research' is in complete agreement with the norm of originality (Clarke and Lunt 2014); mere replications of previous experiments, though important for the validity of research, do not count as a genuine contribution to the advancement of the research frontier, and many assessments stress that the candidate's work produced new knowledge of importance to the field of study.

Category 2, 'Relevance' is often commended in the applied research theses and corresponds well with the norms of industrial research, where focus is on providing solutions to complex technical, environmental or health problems. Given that much research in applied science is funded by external sources, it is not surprising that relevance can be highlighted as a normative dimension of the assessment of a piece of research (there are no criticisms of lack of relevance of theses in the basic fields). That science should be justified by its utility is contextualised by Ziman (2000, 73f) in relation to post-academic settings (inspired by the PLACE norms of industrial research that is commissioned to solve practical problems).

Not surprisingly, category 3, 'Overall quality', as an evaluative category, is an expression in this context of PhD theses of the general norm of quality, explicated by Anderson et al. (2010) as belonging to the list of norms of science and balanced by the emergent norm of quantity (as exemplified in this advice from a young scientist's mentor: 'Early in your career, you need to publish a lot. Later in your career, you have the luxury to publish good things', *ibid.*: 380). Thesis evaluators often comment positively upon the quantity (our category 5) of the work, but that there are fewer such comments than about quality may reflect the finding in the sociology of science that more researchers subscribe to the norm of quality than to the counternorm of quantity. The category 'quality' also reflects the examiners' awareness of time constraints in a PhD project.

Theory is not commonly commented on, except for praise in relation to development of models. This is in line with the observation made by Holbrook, Bourke, and Fairbairn (2015) that examiners' reference to theory was common in education, humanities, arts and social science but less so in science and health.

Quality is a special norm in the sense of being difficult to explicate. Assessment of quality relies on the expertise of the contributors to a given field (the assessors) and their expert knowledge of state-of-the-art in the field. The judgements about 'quality' represent a dimension of the assessments concerned with the entire character of the candidate's work. Of course, 'quality' is something all-pervasive and comes into play also when judging the degree of perfection regarding many detailed and technical aspects of the research methods, like data quality and representativeness, how convincing the evidence is for confirming a hypothesis, etc. Two aspects of quality stand out in particular; first, the craftsmanship of doing good, valid science that can withstand the judgment of peers in the peer review system, and second, the creativity in contributing with new, innovative research that brings science forward. Craftmanship does not relate directly to any of the original CUDOS norms, while the creativity aspect relates to the amended Mertonian norm of originality (Merton 1957; Ziman 2000).

The categories 'Overview of the earlier research' (8), and 'Own contribution in relation to previous work' (9) are both consistent with the overall aim of the Mertonian norms to ensure a disinterested stance of intersubjectivity and to give due credit to the research one builds on.

We also see a congruence between the norm of calling and our category 'Engagement' (12). Furthermore, engagement is reflected in the way assessors voice their own enthusiasm or calling, as they use terms like 'beautiful', 'mature' and 'elegantly'.

### **New norms**

The four sub-categories—addressing the empirical (3.1), theoretical (3.2), methodological (3.3) and theory-experiment connectedness (3.4)—flesh out more concretely what is meant by 'quality'. They have only been treated superficially in the sociological literature as 'technical norms' relating to empirical evidence, reliability, consistency and predictability (Merton 1942; Anderson et al. 2010, 368). It is debatable whether category 4, 'Proxy for quality' – as often used as a tool for the evaluation, allowing the committee members to rely on, e.g. journal status (cf. Sharmini et al. 2015) – is a 'new' norm of science, but it relates to the discussion about quality (norm) versus quantity (counternorm).

Category 6, 'Depth and focus' is related to the genre of a PhD thesis and reflects an expectation that a thesis should have one coherent narrative. While it is a well-known assessment criterion in previous studies of thesis assessment (Chetcuti, Cacciottolo, and Vella 2022; Golding, Sharmini, and Lazarovitch 2014; Ramlall, Singaram, and Sommerville 2020), it seems to be a new norm in the sociology of science that should be included in a description of the normative structure of science.

The set of assessment criteria that we have summarised under category 11, 'Communication' (quality of language, stylistic elegance, degree of clarity, etc.) also emerges as a hitherto overlooked normative element in scientific work. It is likely that it has simply been taken for granted.

### **Absent norms**

Not all the norms described by the sociology of science were expected to be useful for the examiners. For the original CUDOS norms/counternorms, organised scepticism was expressed on two levels: the general critical stance of the examiners, and their specific comments on the candidates' capacity for discussing and judging their own results (category 10). Regarding Anderson et al. (2010) governance/administration and breath/narrowness, these were not found, and neither was employment (the counternorm of calling). This is not surprising, as these norms concern the context for doing research rather than the research itself.

### **Other categories**

The categories 'Contribution in collaborative research' (7) and 'Conditions' (13) are not assessment criteria *per se*, but things that assessors consider when judging the candidate's competence. If it is unclear how much collaborators contributed, or the candidate may have received too much help, the conditions for doing the research may make the examiners less harsh in their judgement.

### **Discussion**

The results show, for the first time, how it is possible to identify the norms of science in written PhD thesis assessments. This has implications for sociology of science, as it supports that norms

of science are used in practice as assessment criteria, and it makes the very concept of norms of science more concrete. Regarding the field of assessment, we point to a new framework for understanding research assessment, an explication of which can support examiners, supervisors and doctoral students. Some of the categories we identified also contribute to the ongoing research and discussion of assessment criteria. Relevance and proxy for quality are particularly interesting for further discussion.

The relevance criterion (where the purpose of the PhD is extended to include making a contribution to society) is new in assessment research. In our inductive analysis, external relevance clearly stood out as a separate category. Relevance is not a classical ideal (Merton 1942, 1957) in the sociology of science. It indicates that Ziman's (2000) proposed norms for mode 2 research have become an integral part of assessment. It will be interesting to follow future research on the assessment of PhD theses to see trends in the frequency of relevance as an assessment criterion in applied research compared to academic science.

Proxy for quality has emerged in assessment literature over the past two decades. It started from examiners encouraging students to publish based on their thesis (Holbrook et al. 2004). Later research investigated how examiners evaluate research that has already been published, either as work in progress (Sharmini and Kumar 2017) or as a 'quality stamp', especially if the journal is high ranking, as we find in the present study. This also reflects an emerging change in the normative structure of academic (and post-academic, cf. Ziman 2000) science, in which scientists are measured and governed by such proxies for quality as publication and citation numbers. An interesting question for further research would be whether these are more often accepted following publication, either because the PhD student has already learned from the peer review system, or because examiners see publication as a proxy for quality.

The analysis reveals new norms that have not (or only superficially) been described in the sociology of science, such as the four dimensions of overall quality (empirical, theoretical, methodological and theory-experiment connection), deepening our understanding of co-called technical norms.

### **Limitations**

The method we used to identify criteria was to look for judgments in the form of praise or criticism. This method worked well, but it has the limitation that we may have overlooked criteria that emerge from more descriptive passages. Often the examiners describe the research under scrutiny in some detail, and such detail may reflect praise. The examiners find it worthwhile to describe the work done within the specific scientific domain, so other researchers (with a contributory expertise) can recognise the novelty in the descriptions. Reading many assessments, it is inevitable to observe a great variety in the degree of depth, detail and constructive criticism made by the examiners. For a minority of the assessments, we were puzzled to see the brevity and lack of engagement from the committee. This brings to mind the remark made by Chetcuti, Cacciottolo, and Vella (2022, 14) that development of an evaluative expertise within communities of practice 'was not observed'. Such brief descriptions do not show why the thesis is ready for defence, and hence the assessment report lacks a transparent argument to support the decision. Yet, the written assessments were in general of good and detailed quality.

### **Conclusion**

This study extends assessment research with stronger evidence for the importance of relevance to society as an assessment criterion, and with new perspectives on how examiners assess already published research. From the sociology of science perspective, both relevance and quantitative proxies for quality indicate an increasing influence of post-academic norms in

academia. The four subcategories of quality deepen our understanding of originality as a norm. Our explication of the normative categories in PhD thesis assessments may help supervisors and students discuss the expectations and criteria involved in the assessments.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Notes on contributors

**Sofie Kobayashi** is lecturer at the University of Copenhagen and independent consultant. Her research interests cover doctoral education and supervision, currently with a focus on how doctoral students develop independence, learn to do research, and navigate the norms of science. Her second strand of interest covers research and teaching in the development of teaching competences in higher education.

**Claus Emmeche** is associate professor at the University of Copenhagen. His research interests include the philosophy and sociology of science, and philosophy of interdisciplinarity. He has also contributed to theoretical biology and biosemiotics. He is part of the History and Philosophy of Science group at the Department of Science Education, where he teaches philosophy of science to students in the biology, chemistry, biochemistry, and nanoscience programmes.

## ORCID

Sofie Kobayashi  <http://orcid.org/0000-0002-3001-9496>

Claus Emmeche  <http://orcid.org/0000-0001-6141-7860>

## References

- Allchin, D. 2001. "Error Types." *Perspectives on Science* 9 (1): 38–58. doi:10.1162/10636140152947786.
- Anderson, M., E. Ronning, R. De Vries, and B. Martinson. 2010. "Extending the Mertonian Norms: Scientists' Subscription to Norms of Research." *The Journal of Higher Education* 81 (3): 366–393. doi:10.1080/00221546.2010.11779057.
- Braun, V., and V. Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77–101. doi:10.1191/1478088706qp063oa.
- Chetcuti, D., J. Cacciottolo, and N. Vella. 2022. "What Do Examiners Look for in a PhD Thesis? Explicit and Implicit Criteria Used by Examiners across Disciplines." *Assessment & Evaluation in Higher Education* 47 (8): 1358–1373. doi:10.1080/02602938.2022.2048293.
- Clarke, G., and I. Lunt. 2014. "The Concept of 'Originality' in the Ph.D.: How is It Interpreted by Examiners?" *Assessment & Evaluation in Higher Education* 39 (7): 803–820. doi:10.1080/02602938.2013.870970.
- Golding, C., S. Sharmini, and A. Lazarovitch. 2014. "What Examiners Do: What Thesis Students Should Know." *Assessment & Evaluation in Higher Education* 39 (5): 563–576. doi:10.1080/02602938.2013.859230.
- Gurr, G. M. 2001. "Negotiating the 'Rackety Bridge' - A Dynamic Model for Aligning Supervisory Style with Research Student Development." *Higher Education Research & Development* 20 (1): 81–92. doi:10.1080/07924360120043882.
- Hoffmann, R., A. A. Kabanov, A. A. Golov, and D. M. Proserpio. 2016. "Homo Citans and Carbon Allotropes: For an Ethics of Citation." *Angewandte Chemie (International ed. in English)* 55 (37): 10962–10976. doi:10.1002/anie.201600655.
- Holbrook, A., S. Bourke, and H. Fairbairn. 2015. "Examiner Reference to Theory in PhD Theses." *Innovations in Education and Teaching International* 52 (1): 75–85. doi:10.1080/14703297.2014.981842.
- Holbrook, A., S. Bourke, H. Fairbairn, and T. Lovat. 2007. "Examiner Comment on the Literature Review in Ph.D. theses." *Studies in Higher Education* 32 (3): 337–356. doi:10.1080/03075070701346899.
- Holbrook, A., S. Bourke, T. Lovat, and K. Dally. 2004. "Investigating PhD Thesis Examination Reports." *International Journal of Educational Research* 41 (2): 98–120. doi:10.1016/j.ijer.2005.04.008.
- Johnston, S. 1997. "Examining the Examiners: An Analysis of Examiners' Reports on Doctoral Theses." *Studies in Higher Education* 22 (3): 333–347. doi:10.1080/03075079712331380936.
- Kim, S. Y., and Y. Kim, Y. 2018. "The Ethos of Science and Its Correlates: An Empirical Analysis of Scientists' Endorsement of Mertonian Norms." *Science, Technology & Society* 23 (1): 1–24. doi:10.1177/0971721817744438.

- Kobayashi, S., and M. Berge. 2022. "Learning Norms of Science through Laughter: A Study of Humour in Life Science Supervision." *International Journal of Science Education* 44 (10): 1680–1699. doi:10.1080/09500693.2022.2090634.
- Krumsvik, R. J., M. Øfstegaard, and L. Ø. Jones. 2016. "Retningslinjer og Vurderingskriterier for Den Artikkelbaserte ph. d-Avhandlingen." *Uniped* 39 (1): 78–93. doi:10.18261/issn.1893-8981-2016-01-07.
- Kyvik, S., and T. Thune. 2015. "Assessing the Quality of PhD Dissertations. A Survey of External Committee Members." *Assessment & Evaluation in Higher Education* 40 (5): 768–782. doi:10.1080/02602938.2014.956283.
- Merton, R. K. 1942. "A Note on Science and Democracy." *Journal of Legal and Political Sociology* 1 (1/2): 115–126.
- Merton, R. K. 1957. "Priorities in Scientific Discovery: A Chapter in the Sociology of Science." *American Sociological Review* 22 (6): 635–659. doi:10.2307/2089193.
- Martinson, B. C., M. S. Anderson, and R. de Vries. 2005. "Scientists Behaving Badly." *Nature* 435 (7043): 737–738. doi:10.1038/435737a.
- Mitroff, I. 1974. "Norms and Counter-Norms in a Select Group of the Apollo Moon Scientists: A Case Study of the Ambivalence of Scientists." *American Sociological Review* 39 (4): 579–595. doi:10.2307/2094423.
- Poole, B. 2015. "The Rather Elusive Concept of 'Doctorateness': A Reaction to Wellington." *Studies in Higher Education* 40 (9): 1507–1522. doi:10.1080/03075079.2013.873026.
- Ramlall, S., V. S. Singaram, and T. E. Sommerville. 2020. "Doctorates by Thesis and Publication in Clinical Medicine: An Analysis of Examiners' Reports." *Perspectives in Education* 37 (1): 130–147. doi:10.18820/2519593X/pie.v37i1.10.
- Sharmini, S., and V. Kumar. 2017. "Examiners' Commentary on Thesis with Publications." *Innovations in Education and Teaching International* 55 (6): 1–11. doi:10.1080/14703297.2017.1294491.
- Sharmini, S., R. Spronken-Smith, C. Golding, and T. Harland. 2015. "Assessing the Doctoral Thesis When It Includes Published Work." *Assessment & Evaluation in Higher Education* 40 (1): 89–102. doi:10.1080/02602938.2014.888535.
- Small, H. 2004. "On the Shoulders of Robert Merton: Towards a Normative Theory of Citation." *Scientometrics* 60 (1): 71–79. doi:10.1023/B:SCIE.0000027310.68393.bc.
- Stracke, E., and V. Kumar. 2010. "Feedback and Self-Regulated Learning: Insights from Supervisors' and Examiners' Reports." *Reflective Practice* 11 (1): 19–32. doi:10.1080/14623940903525140.
- Ziman, J. 2000. *Real Science – What It is and What It means*. Cambridge University Press.