Teaching-research Nexus or Mock Research?

Student factors, supervision and the undergraduate thesis in history

STEFAN EKECRANTZ, JENNY PARLIDEN & ULF OLSSON

IN THIS ARTICLE preliminary results from an ongoing study of 88 undergraduate theses in history from five Swedish universities are presented. By matching thesis quality with individual student grades from upper secondary school two inter-related issues are discussed: how student prior knowledge relates to thesis quality in different groups, and how this can be understood in relation to a suggested phenomenon of direct influence on thesis quality independent of student learning.

Narrative competence or competent narrative?

The undergraduate thesis in history and similar disciplines is often seen as something of a pedagogical ideal (Ekecrantz, 2006; Härnqvist, 1999). The thesis is viewed as the most work intensive module in Swedish undergraduate history and is associated with high assessment standards. The finished product is usually original empirical research presented in a 30–50 page thesis written by a single author. As in professional historiography, alternative methodologies and theoretical perspectives from other disciplines are increasingly common but a vast majority still builds on written primary sources. Prior to the undergraduate thesis the students have finished a second semester thesis in a five week module. This work is of a more limited
scope but is in essence a very similar task meant to prepare the students for their subsequent undergraduate thesis.

Founded on concepts like inquiry-based learning and the teaching-research nexus, the thesis is seen both as a superior way to develop student learning and as a highly valid assessment method (Healey, 2005; Kinkead, 2003; Jenkins et al., 2003). Traditionally the relationship between a student’s actual competencies and the finished thesis is often assumed to correspond perfectly. This may not be articulated, but is often implied and embedded in practice. It goes without saying that students sometimes underperform for various reasons and may in fact be more competent than their theses might lead us to believe. The very opposite is also possible, as when supervisors and various support structures may help improve the quality of a text directly, without influencing the author’s independent research abilities to the same degree. When this is the case, a finished thesis may in fact be better than the student’s ability to create such a text, which can lead to potential problems within higher education and beyond. The reasons for such direct influence on theses, we argue, lay in the professional research model traditionally used, where research results and the text are the main intended outcomes. This differs from core principles in genuine research-based learning, where the primary outcome would be student learning. This discrepancy, we argue, may have far reaching consequences for students as well as for quality and research.

Factors influencing thesis quality

In reality there is a vast array of factors influencing thesis quality, including student self-efficacy beliefs, motivation, personality traits, inter-personal relationships, institutional and cultural aspects just to name a few drawn from general research on academic success (Diseth, 2011; Giota, 2010; Poropat, 2009). For the purposes of this study a limited structural perspective is used. These structural factors are divided into: Prior knowledge and abilities (A), student learning during undergraduate studies (B), aspects of supervision that support learning (C), aspects of supervision that only influence thesis quality
(D) and other factors (E) that may influence the thesis without subsequent learning.

The model above is an adaption of the widely used Presage-Process-Product 3P-model of teaching and learning (Biggs, 1979; 1989; Ramsden, 1992; Prosser et al., 1994). The gist of most variations of that model is a partly chronological perspective where students and teachers enter into – and become part of – a learning environment, resulting in some kind of learning outcome. Trigwell & Prosser (1997) give an overview of how the model has come to be used within a wide range of theoretical frameworks, including cognitivist, individual and social constructivist as well as non-dualistic constitutionalist perspectives – all with their varying conceptions of the nature of such relationships. The original model is in itself highly simplistic, which can explain why so many have found use for it. One such use has been to create clarity in scientific debates, where it is sometimes necessary to contrast teaching and learning activities (process) to learning outcomes (product) to avoid misunderstandings. Other
times it is necessary to highlight student factors and teaching context (presage) when the existing discourse assumes a tabula rasa perspective on students, i.e. disregard the importance of prior experiences.

In a similar way this adapted model is first and foremost meant to work as a contrast to more simplistic, single factor explanations regarding thesis quality. Presumptions of single factor explanations can be suspected when poor thesis quality is explained by student factors only, or when reforms to improve thesis quality is aimed exclusively at one factor such as quantity of supervision or other. A clear example of a single factor explanation is the Swedish national quality audit system 2011–2014, where external assessment of theses was used as the main indicator of educational quality in all disciplines leading to Bachelor or Master degrees (Swedish National Agency for Higher Education, 2011). The rationale behind this output-system was the idea that the educational quality of e.g. history departments and their degrees corresponded more or less perfectly with the quality of their students’ theses. In Figure 1 above this would be akin to an unobstructed link between undergraduate education and thesis quality, perfectly mediated through individual students’ abilities. A more complex but still limited input-intervention-output system had been suggested by Härnqvist (1999) who in a pilot study included both upper secondary school grades and thesis quality as means to assess higher education quality. In that system a department with students with poor prior grades (input) who produced mid-range quality theses (output) would be seen as having better educational quality (intervention) than if the same theses had been written by students with better grades.

Prior knowledge (A)

In a systematic review of various pitfalls in using theses as simple measurement of educational quality, Hamilton and colleagues (2010) expand on the relationship between prior knowledge and abilities:

"Programs using theses or dissertations to document outcomes must realize that verbal ability among students might not significantly be improved by educational programing, but could well be a pre-existing
asset students bring with them to educational programs. [...] Writing a high-quality thesis or dissertation and adequately defending it orally are both tasks that require sophisticated verbal ability. However, Nie and Golde (2008) provide evidence suggesting that verbal ability, rather than being improved by education, is an attribute that enables students to succeed in schools and universities. Thus, while education and verbal ability have a strong positive relationship, the direction of causality may be counterintuitive. (Hamilton et al., 2010: 570-571)

An individual’s upper secondary grades are no more certain to be perfect representations of such ‘pre-existing assets’ than the undergraduate thesis is of his or her abilities to produce such a thesis several years later. Yet existing research tells us grades are valid predictors of academic success on a group level (Cliffordson, 2008; Martin et al., 2001; Morgaman et al., 2002; Tumen et al., 2008; Urban et al., 1999). Furthermore, the undergraduate thesis is widely believed to be a key indicator of important academic competencies and prior grades can thus be expected to correlate with thesis quality to some degree (Härnqvist, 1999). Another possible indicator of prior knowledge would be the Swedish Scholastic Aptitude Test (SweSAT), albeit research by Cliffordson and others show that this correlates less with academic success than upper secondary grades do (Cliffordson & Askling, 2006; Henriksson & Wolming, 1998; Lyrén, 2008).

Undergraduate studies (B)
That both undergraduate education in general and thesis supervision may influence student ability to conduct independent historical research can be seen as self-evident, although more knowledge about these highly complex and contextualized processes is needed. The overall degree can be expected to support such learning via two basic strategies: through general, research-led higher education and through research oriented modules and interventions specifically aimed at supporting the upcoming thesis work (Brew, 2003; Jenkins et al., 2003). Examples of the latter could be courses in academic writing, research methodology, philosophy of science and similar. Traditionally in Sweden, the undergraduate thesis in history has been
the most significant threshold for students to overcome to complete their degree. For some time there has been an increased focus on the ‘production’ of degrees as part of Swedish adaptations to the Bologna process, pressuring history departments to step up their efforts to prepare students for their theses. In recent years the aforementioned quality audit system has added to this pressure.

**Supervision and student learning (C)**
Another main strategy to support thesis work is of course supervision. In recent years there has been an increasing amount of research published on supervision of undergraduate and graduate theses. Many theoretical concepts and models developed in research on doctoral supervision have been used to develop corresponding research on supervision on lower levels. This research can be divided into three main strands: interpersonal relationships and supervisor models (cf. de Kleijn et al., 2012; Dysthe, 2002; Grant, 2003; Greenbank & Penketh, 2009; Ylijoki, 2001), supervisors’ and students’ perceptions (cf. Anderson et al., 2006; Todd, 2006) and research on so called ‘best practice’ in various settings (cf. Dysthe et al., 2006). This research in many instances address thesis quality or student learning, or the relationship between the developing thesis and the student, but substantially less research address potential discrepancies between the finished product and the student.

**Supervision of ‘text’ regardless of learning (D)**
In various stages of the writing process, the supervisor is expected to suggest changes and improvements of the text. In the latter stages one might tell the students how to rephrase the research problem to better fit the actual results, add a series of spot-on references and suggest changes in the overall disposition. Earlier on the supervisor might have identified a series of themes in the material that fundamentally alter the presentation from mere description to a proper analysis. Archival problems may be circumvented by the creative use of alternative sources suggested by the supervisor, and so on. Ideally, discussing and making such changes is a learning experience. Other times the student may make changes more or less mechanically,
without genuine understanding. When this is the case, the supervisor can be said to have influenced thesis quality more so than student ability. In a similar vein, Hamilton and colleagues have argued against a simplistic, single factor explanation in such cases and that ‘the resulting theses or dissertations can reflect more about the capability of the supervisor than of the student’ (Hamilton et al., 2010: 569). The same can be said about changes suggested by the examiner and peers, for example when the author may learn that the text needs to be altered in some very specific ways to get a passing grade.

**Bureaucracy (E)**

Another type of direct input is here labeled ‘Bureaucracy’. Timeframes and variations in ways to organize the assessment and revision processes may affect thesis quality directly. In some history departments the work is to be completed within ten weeks of full time study with no extensions. In others the undergraduate thesis may be scheduled as a part time module during a whole semester with the opportunity to finish the thesis the following semester, prolonging the process up to a year. Sometimes a final thesis presentation is the absolute end point, while in other institutions the student gets substantial room to revise the thesis in light of criticism put forth during the final presentation. The latter might significantly affect the quality of the texts archived for external audits, making for a direct ‘bureaucratic’ input on thesis quality.

In addition, various types of support material, exemplars, templates, checklists et cetera are often developed over time to address problems with retention and specific, recurring shortcomings in student work. Again, this material will ideally help students learn but sometimes it does not. Then the overall effect might be an environment where some students are able to complete their work partly in a paint-by-numbers type process.
history majors and their undergraduate theses

Following in Härnqvist's footsteps...

In Härnqvist's (1999) original study theses in history and economics were assessed by external examiners engaged specifically for that particular research project. The two disciplines were chosen to represent large, traditional disciplines in humanities and social sciences respectively. In history a randomized sample of 35 undergraduate theses from five different institutions were analyzed. Each thesis was rated on a scale from one to five along six quality dimensions: prior research; definition of problem; theory, methodology; procedure and conclusions; and language and formalities. Thereafter, the overall quality of each thesis was rated holistically on the same scale.

By using an existing database all upper secondary grades in Sweden from students born in 1972–1979 were available. This data was cross referenced with students registered on second and third semester undergraduate history in 1995–1997. Hereby a prior grade average for each institution that was part of the study was obtained. The grade averages correlated strongly with thesis quality averages. One outlier was a history department with students with good grades but with relatively poor theses. Among the other departments, upper secondary grade averages were strong predictors of subsequent thesis quality, supporting the hypothesis of a relationship between prior knowledge and higher education outcomes (Härnqvist, 1999: 97).

The initial motivation for the present study was to attempt to replicate and refine Härnqvist's results using the qualitative data collected in the quality audit system, where a randomized sample from all undergraduate theses in history with passing grades in 2012 had been assessed by external examiners in a similar fashion. By identifying individual authors in retrospect and matching each student with prior grades through a national database the aim was to create a dataset that was both larger and qualitatively superior to the one Härnqvist had to work with – with individual grades and individual theses being matched, as opposed to grade averages per institution. By using this new data, the goal was to better quantify
the explanatory power of prior grades in relation to thesis quality. This, in turn, could make for a more informed view of the potential room for influence by the overall degree and from supervision. All transcripts from the randomized, double blind external assessment exercise in 2013 were obtained from the governing body, along with the original anonymized theses. Thereafter the five institutions with the most theses were contacted and asked to submit the identity of each author previously selected for the exercise. Subsequently each author was cross referenced with a national grades database, making for a dataset with a total of 88 history theses and individual upper secondary grades, creating data including external assessment of thesis quality, original thesis grade and prior grades.

The external assessment had been done in relation to the intended learning outcomes specified in the degree specifications in the Swedish Higher Education Ordinance. Each Bachelor thesis was assessed in relation to four quality dimensions consisting of a total of eleven sub-dimensions. The four dimensions were selected from a wider range of outcomes in the Bachelor degree specifications:

1. Knowledge and understanding in the main field of study, including knowledge of the disciplinary foundation of the field, knowledge of applicable methodologies in the field, specialized study in some aspect of the field as well as awareness of current research issues.

2. Ability to search for, gather, evaluate and critically interpret the relevant information for a formulated problem and also discuss phenomena, issues and situations critically.

3. Ability to identify, formulate and solve problems autonomously and to complete tasks within predetermined time frames.

4. Ability to make assessments in the main field of study informed by relevant disciplinary, social and ethical issues.

No holistic judgment was made of each thesis in the original audit but each dimension and sub-dimension were graded on a 1–3 scale allowing for a quantified total assessment of thesis quality post facto in three levels of similar size. Furthermore, several school and grading
reforms had taken place during the time period in which the different students had attended upper secondary school, creating a highly complex picture. By translating all passing grades in seven core subjects, a unified grade average could be constructed for each individual student.

Conflicting results
Surprisingly, prior grades and thesis quality did not match up as expected and did not mirror Härnqvist’s previous results. Only weak correlations between prior grades and thesis quality could be found in the combined population of students, a result also at odds with general research on academic success. As a combined category, the students with the best grades produced the best theses, but poor prior grades did not correlate with theses of poor quality as had been expected. Various reasons could be assumed causing these differing results.

The methodologies used in the two studies were similar but not identical. There had been changes in the grading system due to school reforms, resulting in grade inflation (Skolverket, 2012). In the fifteen years between the two studies there had also been some changes in Swedish undergraduate history that might at least account for some of these differences, such as an increased focus on degrees and completion within ten weeks of study. How these differences might relate to the conflicting results is unclear but none can be disregarded completely.

At closer inspection a major difference could be found in the two populations. It turned out that as many as 20 out of the 35 theses in Härnqvist’s study were originally graded as ‘Pass with distinction’ (57%), making for a sample with unusually strong theses. How this came to be is unknown, but in the randomized sample used in the present study only 21 out of 88 theses were originally graded as ‘Pass with distinction’ (24%), a number much closer to actual grading practices at this level. In trying to understand these results we first replicated Härnqvist’s sample by a random selection of 20 ‘Pass with distinction’ and 15 ‘Pass’ theses. In doing so a significant correlation between prior grades and thesis quality could be shown. Looking at
all 'Pass with distinction' in isolation this phenomenon was enhanced even further.

Consequently, that part of the present study does corroborate Härnqvist’s results – if the latter is redefined as in fact having been a study of mostly high quality theses. In sum, as a group, the theses with ‘Pass with distinction’ had better thesis quality and better prior grades than the others, but in the variations within that category quality and prior grades correlated significantly, while this was not the case among the others. This raises several important questions. Could the lack of a strong link between prior grades and thesis quality among theses with merely passing grades be indicative of:

- Stronger students receiving less attention from the system, allowing for more relative influence from student background?
- The system as a whole focuses on thesis quality rather than student ability?
- Better theses correspond with students’ true competencies?
- Mid-range theses correspond poorly with students’ true competencies due to different interventions directed at the texts rather than student learning?

At this point it is important to underscore that this discussion only relates to potential processes on a group level. Among the history students with relatively poor grades entering higher education there are numerous individuals in this study who did very well at university level and produced theses of high quality – and vice versa. The list of possible reasons for this is long, some of the most obvious ones being that people develop differently, learn differently and that prior grades often say little about an individual’s future abilities, ambitions and interests years ahead.

That being said, the lack of correlation between prior grades and thesis quality in the group that received a passing grade without distinction on their theses deserves further attention. An optimistic explanation could be that more resources might have been channeled to weaker students, e.g. through redistribution of supervisors’ time and engagement, leading to improved learning and thus nullifying
the expected impact of student presage in this particular group. This would not have to be the result of explicit policy, but may simply be the result of supervisors wanting all their students to succeed. A more pessimistic and possibly more plausible explanation would be that in some cases the theses are better than the authors’ abilities to produce them. Or, in other words, the highly valid assessment method of assessing student learning in these modules mainly through their theses might be lacking in reliability for this group. This would be consistent with the hypothesis that the ‘system’ is focused on producing theses rather than student learning, and that when a student is struggling tradition may lead faculty and bureaucracy to help improve the text directly without necessarily improving student ability to the same degree. And, developments in the Swedish higher education landscape in the last couple of years – and likely years ahead – create even stronger economic and other incitements to do so.

Discussion and implications for practice

The results of this study do not at all undermine the picture of the undergraduate thesis as one of the most profound learning experiences during the entire undergraduate degree in history. Decades after completion the former student can be expected to be able to describe it and share insights from the work, which is a standard few other parts of the curriculum can live up to. It is an authentic assessment method in itself and it fulfills many established principles for successful undergraduate education, such as active learning, iterative feedback, time on task, and high expectations to name a few (Chickering & Gamson, 1987). Equally clear is the fact that traditional thesis supervision is an effective teaching method that often supports this learning. What we do suggest is that sometimes there are problems that might remain obscured when student learning and the quality of the finished text is assumed to be synonymous. And, that this might have consequences for research, development and supervision.
A most common supervision process is to let the student submit proposals and drafts followed by feedback and suggestions for revisions and continued work. This tradition allows for a range of different supervising models (cf. Dysthe, 2002; Dysthe et al., 2006). Some supervisors prefer to use a more discussion based format before student work, while others prefer to work exclusively with submitted texts and build their supervision as more summative feedback on texts. When either model is successful iterative discussions about texts function as formative feedback, allowing the supervisor to gauge the student’s present level of understanding and make adaptations in light of this information. Other times supervision does not work out as intended. Feedback on drafts is typically aimed at influencing learning through the text, sometimes blurring the line between student and text in the eyes of the supervisor. In many ways this would be similar to the way doctoral students submit thesis chapters for review over a period of four years or more. The question is if that research ideal really is possible to scale down to a ten week format at undergraduate level, and what might be lost in the process?

A developing text written by a novice researcher under time constraints while learning-by-doing will most likely be a poor presentation of the student’s current understanding, which in turn demands quite a lot of the historian turned supervisor. Much would be gained if valid and reliable assessment methods independent of the thesis could be developed. In teaching, such instruments could be used as designated formative assessment methods, allowing for more effective supervision and learning. In research and development, such instruments could be used as an independent point of reference more closely related to the theses on an individual level than prior grades as has been done in this study.

1 So called ‘narrative competence’ can be seen as the synthesis of historical mindedness – the understanding of history expected of professional historians (Rüsen, 1987; Wertsch, 1998). In this article it is argued that student competency needs to be distinguished from their texts, i.e. their narratives.

2 The 2011–2014 system was highly contested and is being revised at the time this article is written. An expected development is that institutions rather
than departments will be audited through a selection of disciplines. Furthermore, external assessment of undergraduate and graduate theses will most likely be part of that process, but will be given less relative weight compared to other data. Thus, it is plausible that history departments will continue to experience external pressure to focus on the quality of finished written ‘products’ on all levels.

Härnqvist’s work was a main inspiration for the system that was eventually developed but due to political and practical concerns within the Swedish Ministry of Education and Research a strict outcome-based system was opted for – in effect disregarding all other factors with regards to thesis quality.

One major difference from the official system 2011–2014 was that Härnqvist meant to ultimately develop methods to assess whole institutions, rather than separate disciplines and departments.

There was also some variation in systems used when the theses were originally graded at the five different institutions in 2012, and the more detailed ECTS-scale was translated into ‘Pass with distinction’ for A-B and ‘Pass’ for C-E.

Spearman’s rho = 0.236, (p < 0.05), n=85. Three students did not have records in the national grades database. One student did not have a thesis grade recorded.

‘Pass’: rho = 0.430, (p > 0.50), n = 64. ‘Pass with distinction’: rho = 0.514, (p < 0.05), n=20.
References


Cliffordson, Christina (2008) 'Differential prediction of study success across academic programs in the Swedish context: The validity of grades and tests as selection instruments for higher education', *Educational Assessment*, vol. 13, no. 2: 56-75.


Diseth, Åge (2011) 'Self-efficacy, goal orientations and learning strategies as mediators between preceding and subsequent academic achievement', *Learning and Individual Differences*, vol. 21, no. 2: 191-195.

Dysthe, Olga (2002) 'Professors as mediators of academic text cultures – an interview study with advisors and master’s degree students in three
disciplines in a Norwegian university’, *Written communication*, vol. 19, no. 4: 493-544.


Henriksson, Widar & Wolming, Simon (1998) ’Academic performance in four study programmes: a comparison of students admitted on the basis of GPA and SweSAT-scores with and without credits for work


