Reinforcing Learning in an Engineering Master's Degree Program: The Relevance of Research Training

Fredrik Asplund (fasplund@kth.se) and Martin Edin Grimheden (mjg@kth.se)

KTH Royal Institute of Technology, Department of Machine Design, Division of Mechatronics, Brinellvägen 83, 10044 Stockholm, Sweden

Abstract

Master students at our institute were graduating without acceptable research proficiency. We intervened by shifting our research training from teaching-centred to student-centred, and from research-related subject content to research-related processes. We performed a mixed methods study aimed to confirm there was improved research proficiency without a negative trade-off for our students' engineering skills. Results indicated improvements to research proficiency, which our students were able to transfer to engineering-related learning activities to increase their ability to achieve engineering synthesis. This outcome was potentially supported by our courses including several perspectives on scientific knowledge production. This implies that research training, rather than having a negative effect on engineering skills, can be helpful in learning diametrically opposing aspects of thinking required by current engineering. As engineering education evolves towards more cross-disciplinary cooperation, this implies the need to pursue the increased opportunities for students to learn about different perspectives on knowledge production.

Keywords: Engineering education; Master's education; Science education; Research-teaching nexus

1 Introduction

Institutes within the Swedish higher educational system are required by regulation to provide research training at the master's level [1]. How this subject is taught is mostly decided at the institutional level. At our institute, traditional teacher-led and content-focused research training did not achieve acceptable results. Some students graduated with little proficiency in research, and few synergies with other parts of the curriculum were observed. As with many other engineering programs demand from firms and the opening up of new career paths mean that we are currently being tasked with fostering new skills, such as cultural awareness, sustainability, innovativeness and entrepreneurship [2, 3]. Extending the time spent on the subject was thus not an option, since our engineering curricula are already stretched beyond their limits. New student-centred teaching practices, such as inquiry-based learning, have been suggested as solutions to this dilemma [2]. Conceptually related, *inductive teaching and learning* approaches [4, 5] are at times equated with *conceptual*, *epistemic*, *social* and *procedural* aspects of research in learning activities [6].

Unfortunately, the situation is complicated by the fact that the inclusion of research training in engineering curricula has been historically contested. That research is synergetic to learning is only supported by weak evidence [7], and an emphasis on engineering science implies less time spent on practical skills [8]. The popular view of science as providing unambiguous facts might also be problematic: if engineering is posed as an applied science it might result in risky expectations that even complex, highly critical systems can always be reduced to a set of assessable facts [9]. While we wanted to improve our students' research proficiency, we had to acknowledge that research training could have powerful implications for our students' understanding of knowledge production and engineering skills. Using inquiry-based learning might aggravate the situation, since an abstract and often unfamiliar subject such as research training is not optimally matched to this approach [4]. The result could be that our students would, regardless of their ability to independently conduct research, gravitate towards forms of scientific knowledge production that would impair their ability to perform engineering design.

This paper describes the study of our intervention into the research training in one of our engineering master's programs. This intervention shifted the teaching from teaching-centred to student-centred, and the emphasis from research-related subject content to research-related processes, skills and worldviews. Our interest was to understand any causal relationships between this shift and improvements to research proficiency; and whether these improvements would come with a negative trade-off to our students' engineering skills. This involved identifying the nature of any improvements and relating it to fine-grained elements of research and learning. The novelty of this research focus is twofold. Firstly, the graduate level itself is understudied in regard to the relationship between research and teaching [10], and engineering education [11, 12]. Secondly, studies of the research training provided by graduate engineering programs are scarce, even though research training could be seen as the core of the graduate engineering degree [13].

The next two subsections provide a basis for the paper by describing the research discourse closest to the domain of study and the conceptual framework adopted for the study. The background and research design of the study is described in the subsequent section. A mixed methods design was used, primarily due to the many confounding variables that had to be controlled upon identifying a shift in the student population. The results are then presented, analysed and discussed in regard to learning and future implications. Research proficiency was found to have improved without negative trade-offs to our students' engineering skills; in fact, it would seem that our students were able to apply knowledge from the context of research to the benefit of engineering-related learning activities. The paper ends by summarizing the conclusions. We find that research training can be helpful in teaching students the diametrically opposing aspects of thinking required by current engineering processes. We also conclude that teachers should grasp opportunities for students to learn about different perspectives of knowledge production as engineering education evolves towards more cross-disciplinary cooperation.

1.1 The research-teaching nexus

The connection between research and education, the *research-teaching nexus*, is much debated. Researchers take differing standpoints, including that this link supports synergies [14], has no substantial impact [15], or can be harmful [16]. Supporting each standpoint is complicated due to the many opportunities for variations. For instance, the conceptualization of research and teaching varies [17]; the strength of the relationship differs across institutions, disciplines and levels of education [15, 18]; and the entity/activity in focus can vary from teacher/teaching, to student/learning, to policy, to recruitment, and so on [19, 20]. Furthermore, curricula are also affected by occurrences at the societal level [21] – emphasis in engineering education on theory and scientific skills vs hands-on problem-solving and non-technical skills has varied across nations and throughout history. Nevertheless, the idea of a connection between research and teaching remains appealing to many in the academic profession [15].

Several reports have discussed the research-teaching nexus in regard to higher education contexts. In the US, the Boyer Commission [22] propose basing education at research universities on research and inquiry from the first year onwards. In Canada, Halliwell [23] strongly emphasize action towards creating a common vision on the research-teaching nexus among higher education stakeholders. In Australia, Cherastidtham, Sonnemann and Norton [24] down-play the importance of the research-teaching nexus for deciding between teaching practices in higher education. Tight [7], as part of a larger research project, summarize many of the national and international perspectives on the research-teaching nexus. Prince, Felder and Brent [25] identify that the most empirical support for a positive benefit of strengthening the research-teaching nexus comes from interventions where teaching has been shifted towards emulating the research process, rather than conveying research content. Together, these reports highlight how teaching practices, levels of education, institution and geographical location can all combine to complicate the study of the research-teaching nexus in higher education. When it comes to master's programs, even when limiting oneself to Europe and North America, the challenge is further evident in how students can be taken in drastically different directions [26-28]: the underlying intent of a program can range from preparing students for a career in academia to putting emphasis on skills required in professional positions in the industry.

The research-teaching nexus seems especially weak at engineering institutions [29]. Griffiths highlights the attitudes of both teachers and students to explain this phenomenon [30]. In regard to teachers, an explanation is likely the large proportion of academic staff recruited from industries in which orthodox science has little value in day-to-day operations [30]. Academics at engineering institutions are also aware that research in their fields is usually driven by government policy and industry, rather than by research institutions [30]. In regard to engineering students, an explanation is likely that these emphasise hands-on skills rather than methods to recognize and handle complexity [30]. The perspective is often that academia overemphasizes science, generating engineering students with too little experience in the practice of engineering and design [8]. Most studies on the research-teaching nexus are conducted in an undergraduate setting [10], where it is assumed that the case for a relationship is weaker [31]. However, there are exceptions such as the study by Aditomo et al. [32], which provides examples of different types of learning tasks used in disciplines akin to engineering at the undergraduate and graduate levels. These tasks are defined as inquiry-based, and several more or less mimic research activities. When characterizing which of these *"can be regarded as close to the kinds of research that academics engage in"*, one can argue that Aditomo et al. [32] use standards at odds with much of the research conducted in engineering. This suggests that the perceived weakness of the research-teaching nexus in engineering may be based on different conceptualizations of research.

Arguably this indicates that the complex relationship between research and teaching makes it difficult to ignore the influence of the subject content when looking at changes to research training.

Strong opinions of teachers and students in engineering, and implications for knowledge production, suggest research has special implications for various types of teaching and the self-regulation of learning itself. Therefore our aim calls for a theoretical base that can be used to discuss psychological concepts as they relate to a wide range of other factors that affect teaching and learning activities.

1.2 Conceptual framework

To carry out our study we require a conceptual framework that can be used to (a) describe our intervention, (b) analyse the results and (c) discuss the outcomes.

The search for a framework suitable for our purposes started with the intent behind our study as it relates to the *discussion* of the outcome. As mentioned, it will depend on a wide range of factors involving both the individual student and the institutional context. To this end we chose Entwistle's model of the teaching-learning process as a conceptual foundation for discussing our results [33]. In contrast to many other similarly broad frameworks, it has a strong construct validity and has been developed for the context of higher education with an eye towards ecological validity [34, 35].

Entwistle's model is based on the two dimensions of *deep* vs *surface approaches to learning*, and *strategic* vs *apathetic approaches to studying* [33]. A deep approach to learning is trying to understand the underlying ideas of the learning material, while a surface approach to learning is to focus on the learning material and what it explicitly conveys [36]. A strategic approach is to optimize the time spent in a deep vs surface approach to learning to get the highest possible grade for the least effort. A deep approach to learning can be undertaken in a *holist*, *serialist* or *versatile* way [33]. The holist style is broad and personally structured, while the serialist style is critical, cautious and step-by-step structured. Students with a holist approach thus tend to try to get to an understanding of the learning material as a whole, only looking at separate parts based on mood and interest. Students with a serialist style instead tend to try to break down the learning material into a series of logical steps, only arriving at generic conclusions later by combining what has been learnt in isolation. The versatile style is to alternate between the holist and serialist styles to avoid the negative effects of taking either to the extreme.

However, to *analyse* our results we required a taxonomy that describes student *learning activities* in more detail than Entwistle's model. We chose to use the taxonomy by Vermunt and Verloop [37], since it is student-centred and shares enough background with Entwistle's model to allow the discussion to be related to the analysis [38]. The Vermunt and Verloop [37] taxonomy differs between cognitive, metacognitive and affective learning activities: cognitive activities process subject matter, for instance by structuring or analysing it; metacognitive activities plan the learning process, for instance by orienting the student in regard to what to learn; and affective activities involve dealing with emotions that arise during learning, for instance by actively focusing on learning rather than alternative activities. A student which realises that he has not understood a text although he has read it several times (monitoring, a metacognitive activity), overcomes the frustration related to this (dealing with emotions, an affective activity) and proceeds to focus on distinguishing the main points of the text (selecting, a cognitive activity) has passed through all types of learning activities. All parts of the taxonomy are identified in Table 1, with those important to this study described in further detail in the Results section.

[Table 1 here]

To *describe* our intervention we used the model by Griffiths [30] to conceptualize the links between research and education. It defines four ways to structure the research-teaching nexus: *research-led* which organizes education around state-of-the-art research content; *research-oriented* which emphasizes the teaching of research-related processes, skills and worldviews; *research-based* where learning takes place through inquiry-based activities not necessarily focused on learning subject content; and *research-informed* in which the teaching and learning process itself is inspired by systematic inquiry. Using results from recent research studies as examples during lectures is thus a research-led approach, while involving students in research activities is an example of a research-based approach. Healey [39], as shown in Figure 1, makes the point that these categories differentiate both teacher/student focus and research content/processes emphasis. This captures the essence of our intervention, which involved a shift across both of these scales.

[Figure 1 here]

Figure 2 summarizes the conceptual framework of the study. Using Griffiths [30] model we can describe a shift away from teacher-led and content-focused research training. We expected a student- and content-focused approach to allow students to become more independent and efficient when performing research activities, which should lead to improved outcomes in research-intensive learning activities. However, we feared that this would also lead to negative trade-offs with students adhering to the hypothetical-deductive model even when inappropriate during engineering design [40]. This should be observable using the taxonomy by Vermunt and Verloop [37] if students e.g. showed less consideration of design alternatives (see e.g. Relating/Structuring, Analysis and Selecting), emphasised a non-repetitive process (see e.g. Analysis, Appraising and Orienting/Planning) or ignored uncertainty (see e.g. Selecting and

Orienting/Planning). This does not mean we believe that these problems are intrinsic to the hypothetical-deductive model, but rather that these trade-offs might occur when a *novice* to both research and engineering combines learning about both. This is the reason we need Entwistle's [33] model to discuss the implications of our results in the context of higher education.

[Figure 2 here]

2 Methodology

This section motivates and describes the research design of the study. It starts by establishing the background and studied intervention. Thereafter the methodology is motivated: first the overall choice of approach, and then each method in regard to validity and limitations.

2.1 Background and context of the intervention

By 2007 higher education in Sweden had adapted to the European Bologna process [41], which is based on three degree cycles [42], with a linear progression from bachelor to PhD using the master's as an intermediate step. For Swedish universities, preparatory change started earlier – with a stricter focus on research training already initiated in 2003. At our university, KTH Royal Institute of Technology in Stockholm, this meant the launch of a number of pilot programs. The existing 5-year professional engineering programs were divided in two: bachelor's (3 years) and master's (2 years). At the master's level, course-based programs were formed incorporating both professional and research-related learning goals.

The context of our study is a master's program, more specifically the Mechatronics Track of the Engineering Design master's program. During the first half of the second year a team-based capstone course integrates the knowledge gained throughout the students' engineering education, assessing whether they have the engineering skills required to develop products, processes and systems [43]. The second year then ends with a master's thesis course in the subject of Mechatronics, which assesses the students' research proficiency and *individual* mastery of engineering. It is thus not an option to, as some institutions, allow theses that focus almost exclusively on either research or engineering [26, 44]. Historically the engineering tasks of our theses have mostly come from an industrial context with students physically located at industrial premises. To allow a dual focus and to keep the engineering relevant to industry this practice has continued for our master theses. Recent examples of our theses' engineering tasks thus include the prototyping of a classification system for tracking objects in autonomous trucks, modelling the unwanted pressure oscillations produced by auto-ignition in engine cylinders and designing a control strategy for dampening out vibrations in an active cabin suspension system. The students identify research questions for the theses, ideally supported by finishing the engineering tasks and proven to be valuable by academic literature. As an example, the intention of the thesis prototyping a classification system was to investigate ways of using machine learning to improve object identification accuracy despite signal noise and environmental factors. As during our capstone course there is regularly tension between learning goals and the expectations of industry. However, as in capstone courses [45], these are usually possible to overcome by focusing on communication and defining the responsibilities of all involved. In theory our master's program thus ensures a high level of proficiency both in research and engineering. In practice, earlier external assessments of the program indicated only an acceptable level of research proficiency at graduation [46]. A closer look even revealed large differences between individual students in this regard. This motivated an effort to change the situation by intervening in our context.

2.2 Intervention

Prior to the intervention, a comparison group of students from our division was established, henceforth referred to as Y0 Students, i.e. *Year 0 Students*.

During the first year we replaced the lectures on ongoing research projects at the department, which encompassed 3 European Credit Transfer System (ECTS) credits, i.e. 2 weeks' worth of a semester. To date, these lectures had been *research-led*, i.e. they were traditional lectures and their emphasis was on making students understand research findings. The replacement was *research-based*. The students were divided into groups and presented with a set of questions concerning competing research methods, processes and worldviews. These questions were to be answered based on real examples of research, elicited from self-study and a series of three seminars. In the seminars senior researchers from the Department spent an hour explaining their research and another hour answering students' questions. Rather than passively receiving information in lectures, students had to look actively for knowledge. After submitting a report answering the questions, the students received guidance in the form of feedback on par with that given when reviewing journal publications. Rather unsurprisingly most students had to resubmit the report several times, while continuing to interact with teachers and researchers. Thereby an increased responsibility for regulating the learning was taken by the students. The students that received this type of research training will henceforth be referred to as Y1 Students.

During the second year of the study, we changed the *research-oriented* approach of the remaining lectures of our research training (equivalent to 4.5 ECTS credits) to a *research-based* one. During the previous years these traditional lectures had introduced a number of research methods, processes and worldviews. Students had then studied these concepts in more detail while putting them together to form a master's thesis plan. The new approach still introduced these concepts in a lecture, as a way of putting all students on the same level. However, we then relied on a bottom-up approach whereby the students received most support *after* their thesis plans had been formulated. A random choice of students had to present their plans in front of the entire class at two seminars, receiving detailed critique in the process. All students were expected to consider this feedback before handing in their final report. The students that received this type of research training will henceforth be referred to as Y2 Students.

Table 2 summarizes the treatment of the three cohorts of the study. As can be seen the treatment of the Y2 cohort was an extension of the treatment of the Y1 cohort, and involved a shift towards a more student- and content-focused approach.

[Table 2 here]

2.3 Choice of overall methodology

We wanted to verify that the intervention improved our students' research proficiency. Furthermore, the focus of the study included the causal relationships connecting any improvements to the intervention, especially as related to negative trade-off to our students' engineering skills. This first requires the nature of the improvements to be identified. On the one hand, the nature of improvements could be related solely to the students' grasp of the subject matter; on the other hand, it could be related to the students' way of self-regulating their learning activities. This means that the study had to include both confirmatory and exploratory elements [47]. With our Division being occupied with a multi-disciplinary research field, many of us share a pragmatic worldview [48]. It is therefore not uncommon for us to adapt or mix different types of research approaches, since proving an effect and understanding it better can often be best supported by different methods [49]. A way to build on different types of research approaches to include both confirmatory and exploratory elements is to use a *mixed methods design* employing a *sequential explanatory strategy* [50]: a phase employing quantitative methods precedes a qualitative phase. We decided on this approach since it allows for quantitative results to *direct* qualitative data gathering. Studying the self-regulation of learning solely with a quantitative approach would be difficult considering the many confounding variables related to any dependent variable; however, without first confirming an effect on specific cohorts, it would also be difficult to know which of our cohorts had changed enough to motivate a detailed study. The following three subsections describe the approach of the different phases, and the triangulation of their combination.

2.3.1 First part, Confirming an effect

To confirm an improvement, and allow for a study of its nature, we measured a part of the curriculum with strong opportunities for both self-regulation and research. The choice fell on the master's thesis course, since it is driven by the students themselves and has research-related learning goals. This course is also separate from those that made up the intervention. We decided on the completion time as the dependent variable, as increased research proficiency should translate into more independent and efficient self-regulation of research-related activities and thus a shorter completion time. Self-regulated changes to completion time should also be readily measurable as there is no time limit imposed on finishing the course – each student decides when to submit their thesis. The characteristics of the data and cohorts motivated the use of a Kruskal-Wallis H test [51, 52]. For reasons of brevity, this motivation is given in the next section on validity and limitations.

The design was *quasi-experimental*, given that we intervened on groups that had not been formed through random selection [53].

2.3.2 Second part, Understanding the effect in depth

To explore a phenomena as complex as research proficiency, we followed Creswell's suggestion to use a qualitative analysis of qualitative data [49]. We considered our students' inexperience in research terminology the largest obstacle to analysis. Therefore, we chose to use inductive content analysis as outlined by Cohen et al. [54]. In line with this we each separately read through and coded all master theses, creating codes inductively. The textual definitions provided in the theses were helpful in avoiding misunderstandings: the use of research-related terms was unorthodox in several cases. The final sets of codes were discussed, merged and refined into categories during two work sessions. This ended in the creation of primary categories around learning activities defined by Vermunt and Verloop [37].

2.3.3 Triangulating the parts

To further corroborate findings, the quantitative and qualitative phases can be methodologically triangulated [55], i.e. positive/negative results from one method can be corroborated by positive/negative results from another. To allow for this corroboration, subgroups of theses from each cohort were identified by use of completion time and four qualitative variables indicative of an effect.

The four variables were the master theses' grade, research questions, methodological approach and discussion content. The choice of the latter three was based on the Tashakkori and Teddlie [56] framework for describing research studies.

To elicit subgroups the variables were (re)classified as dichotomous variables. The reasoning behind the assessment of the latter three variables was then coded directly into the theses to ease analysis. For completion time, we divided the theses into two groups based on the average completion time.

For grades, we divided the theses based on whether they achieved an A grade according to the ECTS. With one exception, our examiners only handed out A and B grades. It should be noted that grades A to E all signal a pass, meaning that all theses reported in this paper were deemed to be acceptable overall.

For research questions we separated theses with high vs low quality research questions. High quality was defined as providing direct guidance to the direction of the investigation conducted during the thesis. This was in contrast to many research questions, which only indicated which area the study should be conducted in. To show the gist of this classification some examples are given in Table 3.

For methodological approach, we divided the theses according to whether they included a structured empirical investigation beyond the ad hoc development of engineering artefacts. Examples primarily included case studies, but there were also questionnaires and interviews. To show the gist of this classification some examples are given in Table 4.

For discussion content, we divided the theses according to whether the discussion in them reflected a serious attempt at critical inquiry. This was defined as going beyond addressing the research questions by simply stating the capabilities of any system engineered as part of the thesis. While perhaps not a problem in the context of many other countries, this is a real risk in Sweden: as mentioned, our master theses are almost exclusively performed with students physically located at industrial premises, where hands-on engineering is emphasised. To show the gist of this classification some examples are given in Table 5.

2.3.4 Summary

To ease the understanding of the relationship between phases the important points from previous subsections are visualized in Figure 3.

[Figure 3 here]

To ease the understanding of the relationship between data sets the important points from previous subsections are visualized in Figure 4.

[Figure 4 here]

2.4 Validity and limitations

As Creswell and Miller did, for validity, we "most closely align ourselves with the use of systematic procedures, employing rigorous standards and clearly identified procedures" [57]. In the following subsections we discuss our approach to validity, important validity concerns and associated limitations.

2.4.1 The complete study

Due to factors out of our control it was not possible to use a true experimental or stronger quasi-experimental design [58]. These factors included gaps in previous data sets, that the curriculum could not differ within year groups, and that practically comparable control groups were not available. Therefore, the methodological triangulation was an important measure to ensure validity, given the difficulties in ruling out alternative explanations in quasi-experimental designs [59]. However, several alternative explanations still merit a discussion in the following paragraphs.

Prior to this discussion a reminder on statistical significance and power is valuable [60]. A required sample size is calculated a priori using the statistical significance, statistical power and effect size that make sense for each test at hand. We have no control over the size of our cohorts, and we have therefore identified underpowered tests in our study. We proceeded anyway, arguing that the triangulation allows for this, but it *still* has two key implications: firstly, when identifying significant results in underpowered tests, we have to discuss the associated effect size; secondly, when identifying non-significant results in underpowered tests, these cannot be used to accept the null hypothesis, due to the large probability of a Type 2 error.

The first alternative explanation considered was that the groups differed in some other aspect than the treatment they received [59]. At a cursory glance, the recruitment of women (5%, 11% and 4%, respectively) and students with a bachelor's degree from another university than KTH (18%, 18% and 16%, respectively) are similar across the cohorts. All students were full-time students. For a more detailed inspection of the differences between cohorts, we turned to their grades and ages. We essentially considered these as a *rough* indicator for large differences in maturity and capability.

For the grades we conducted a one-way ANOVA [61]. Grade averages were based on the time spent in the master's program and official calculations used when deciding scholarships/grants. Test data is summarized in Table 6. Student grades increased from Y0 Students (n=22, mean=3.8, SD=0.33), to Y1 Students (n=28, mean=4.0, SD=0.33), to Y2 Students (n=25, mean=4.1, SD=0.33), but the differences between the student groups were not statistically significant. We argue that we can thus be acceptably sure that the groups do not differ significantly in regard to grades.

[Table 6 here]

The age test data is summarized in Table 7. Outliers in, and the distribution of, the data indicated that the Kruskal-Wallis H test [51, 52], a nonparametric method, would be appropriate. It is difficult to establish a required sample size for this method [62]. However, an estimate based on rule-of-thumb and the one-way ANOVA sample size calculation show that we are within bounds [62]. The mean rank of ages was not statistically significantly different between groups. We argue that we can thus be acceptably sure that the groups do not differ significantly in regard to age.

[Table 7 here]

We can also assert that there were no substantial changes to the acceptance criteria for the different student groups. Furthermore, the examiners at our Division didn't notice any large differences in regard to student capability between the cohorts, and the findings presented in this paper regarding the Y0 Students fit well with our and the examiners' impression of the state of earlier year groups. Therefore we argue that we have covered the most plausible indicators for large differences.

The second alternative explanation considered was mortality, i.e. selective drop-out of participants [63]. A retrospective check shows that two Y0 Students never finished their master's thesis, whereas all Y1 and Y2 Students managed to complete theirs. We therefore argue that mortality is not a substantial biasing factor in this study.

The third and fourth alternative explanations considered were those of history and maturation, e.g. the influence of significant events other than the intervention. We note that there were no substantial changes to the curricula for the different cohorts outside the intervention. Furthermore, the master theses were conducted in similar contexts. However, two possible concerns along these lines merit closer examination.

Firstly, using completion time as a measure of self-regulation relies on all students and teachers in the study perceiving the same ideal completion time. However, the start of a few of our master theses was delayed by a whole semester. There is no natural deadline for these theses, while the normal cases are generally perceived by students as ideally ending before the summer vacations. Even if the delayed theses were considered in the quantitative phase, the use of the aforementioned categories could not be relied on during qualitative analysis. To avoid confusing the analysis, eight such theses, roughly evenly distributed, were therefore removed from the data sets. To err on the side of caution separate tests have been carried out to ensure that, had the eight theses been included, they would not have changed the statistical significance of any results.

Secondly, the behaviour of members of our faculty is important, since regulation of learning is driven by both students and teachers. In regard to supervision substantial differences across the cohorts are unlikely: the supervisor group was stable, and the supervision of students administrative and focused on technical expertise. Structural aspects of the thesis course are rather addressed by texts available via the Department’s website. Furthermore, there is a substantial resistance to emphasizing research across the supervisor group, due to reasons outlined by Griffiths [30]. However, for full disclosure, we note that one teacher involved in changing the curriculum supervised one thesis from the Y0 Student cohort and one from the Y2 Student cohort.

The question of a uniform assessment is of greater concern, since examiners at our Division might not interpret the assessment guidelines in the same way. One examiner was also involved in changing the curriculum, and might therefore evaluate theses from later cohorts differently. Therefore a Kruskal-Wallis H test was used to identify differences between examiners in regard to completion time. One examiner had only handled one thesis, and since this was not an outlier we decided to exclude it from the analysis. This resulted in six groups (n=12, 14, 20, 13, 7 and 8). Test data is summarized in Table 8. Mean ranks increased across the groups (27.58, to 36.57, to 37.18, to 37.42, to 44.36, to 48.94), but the differences were not statistically significant. With an underpowered test we cannot reject the null hypothesis based on these results. We therefore interviewed the examiners. Based on the interviews we could not identify any substantial differences in their understanding or application of the learning goals of the master thesis course.

[Table 8 here]

2.4.2 The quantitative part

Quantitative research considers data gathering a separate activity from inferences and therefore raises special validity concerns [64].

A quantitative concern was the diligence needed over a long period to avoid errors entering the data set. All quantitative data was therefore checked against external use. As an example, the completion time was gathered internally and checked against announcements for end seminars.

Another concern was the way some theses may appear to take longer to complete because they span several semesters, with varying vacation time in between. To avoid this effect, official vacations and weekends were deducted from relevant completion time data points. This is acceptable, since our students were all full-time students and it penalizes our statistical tests for significant differences.

2.4.3 The qualitative part

Using the framework by Creswell and Miller we can identify three procedures for establishing validity in qualitative research that are in line with our paradigm worldviews. These are triangulation, member checking and an audit trail [57]. The use of triangulation is, as previously explained, a cornerstone in our study. As alluded to in previous subsections we have also made use of member checking: we interviewed the examiners at our Division to establish whether our understanding of the master's program, results and conclusions were credible and trustworthy [64]. We believe examiners are in a position to correctly evaluate self-regulation of learning, since they are the other half of said regulation. We also had our study audited by a professor external to our Division. He was provided with the data, results and analysis of the study. Feedback was provided in written form.

Feedback from the member check and audit has been incorporated into the study and this paper.

2.4.4 Limitations

We believe cognitive and metacognitive learning activities were the most important considering our intervention, and that the research design was suitable for studying them. Furthermore, none of our students seemed particularly weak in affective learning activities, and our member check did not reveal any specific concerns in that direction. Pressure to complete early or difficulties in the students' private lives should thus not have biased the study. However, it should be noted that the research design does not allow us to say whether our results translate to cohorts with an overall different capability in this regard. As an example, a cohort of very motivated students might look for more information earlier when faced with research-related learning goals, and vice versa.

3 Results

This section describes the results from the two phases of the study in preparation for the discussion. Associated data is found in Appendices A and B.

3.1 Quantitative results

Outliers in, and the distribution of, data indicated the Kruskal-Wallis H test as appropriate. The test held no assumptions on the similarity of the shapes of the involved distributions, and the comparison had already been established as underpowered. Comparing the cohorts reveals that the distributions of student completion time were statistically significantly different between groups. Test data is summarized in Table 9. Subsequently, pairwise comparisons were performed using Dunn's procedure with a Bonferroni correction for multiple comparisons. Adjusted p-values are presented. This post hoc analysis revealed statistically significant differences between Y0 (n=22, mean rank=46.30) and Y2 (n=25, mean rank=28.48) (p=.015), but not in any group combination involving Y1 (n=28, mean rank=39.98). Estimating an effect size can be done by using the Hodges-Lehmann estimator (HLΔ) on the cohorts in question [65]. HLΔ is originally only intended to be used for distributions with similar shapes. However, it has been shown that HLΔ can be used in the case of symmetric distributions [66]. Inspection of a boxplot and comparing medians to means indicate that apart from a few outliers the completion time is fairly symmetrical. HLΔ is estimated to 20.9 (95%: 6.4, 45.9) for the completion times of Y0 and Y2. We thus conclude that we can be acceptably sure that our intervention has had a significant effect on master's thesis completion time.

[Table 9 here]

3.2 Triangulation

This subsection provides the distributions of theses based on the dichotomous variables used for triangulation. This gives eight subgroups in each table laid out according to completion time (from low to high on the y-axis), grade (from low to high on the x-axis), and the quality of research questions, methodological approaches and discussion content (low quality to the left and high quality to the right in each cell). To facilitate interpretation the optimal subgroup (low completion time, high grade and high quality) is highlighted in grey. The frequencies of all subgroups add up to 100% of the theses for Y0 and Y2 Students respectively, as directed by the quantitative result.

[Table 10 to 15 here]

Results indicate that the improvement to completion time is strongly tied to improvements of aspects of our students' research proficiency: as completion time improves from Y0 to Y2, the groups with improved research aspects grow strongly. This growth is especially noticeable for the optimal groups. As an example, as seen in Table 10 only 18% of the theses by Y0 Students both finished prior to the average completion time and achieved an A grade. Of these theses none (0%) had a research question of high quality. As seen in Table 11 56% of the theses by Y2 Students both finished prior to the average completion time and achieved an A grade. When dividing these theses further one can see that 32% of all theses by Y2 Students belonged to the optimal group – they were optimal in regard to completion time, grade *and* research question quality.

The aim of the study was to understand any causal relationships between the intervention and improvements, and to identify any associated negative trade-off to our students' engineering skills. It is then beneficial to compare each type of subgroup in regard to the qualitative results. The causal relationships could for instance be straight-forward, i.e. that an improved research proficiency meant students could more easily fulfil the learning goals of the master thesis course. Any negative trade-offs due to the intervention should then be most obvious in the optimal groups. Trade-offs could also be contingent on the abilities of the student, in which case the middle subgroups should give indications of the nature of these dependencies. Even the worst subgroups (quantitatively speaking) are interesting, since these students seem to be the least affected by the intervention. This can for instance help in identifying differences between Y0 and Y2 Students that are unlikely to be related to our intervention.

3.3 Qualitative results

As previously noted, primary categories were formed around some of the learning activities defined by Vermunt and Verloop [37]. Below we list the four of these which differ substantially between Y0 and Y2 Students, as directed by the quantitative result. Apart from this, no substantial differences were found when comparing different combinations of subgroups. Furthermore, these primary categories do not indicate any negative trade-offs to our students' engineering skills. Indeed, almost all theses indicated students were strong in concretizing/applying learning activities [37], and even if the research aspects of the theses by Y2 Students had improved substantially the theses remained strongly focused on engineering. Therefore, while the learning activities described below were framed as applied research, they can more accurately be described as engineering influenced by aspects of research. The difference between Y0 and Y2 Students thus appear to be that increased research proficiency made Y2 Students able to handle engineering tasks in ways Y0 Students could not. Unfortunately this means that our results mostly limit us to discussing Y2 Students, as they do not provide a way of discussing Y0 Students in isolation. However, in the Discussion section this will allow for a focused explanation of the causal relationship between effect and intervention, as well as point to a troubling limitation of our intervention that will require future research.

3.3.1 Adjusting

Adjusting involves changing learning plans or goals on the basis of monitoring one's observations [37]. A primary category for Adjusting was seen when analysing the Y2 Students in regard to the quality of research questions (Table 11). 6 out of 12 students in the groups with low quality research questions had in fact started out with high quality research questions.

This was not only connected to students with a higher than average completion time. 3 out of 6 seem to rather have used it as a strategy to de-emphasize critical inquiry. In other words, by removing the direct guidance on the direction of the investigation, the discussion in the thesis could be kept generic. As indicated in Appendix B, Table 17, this for instance meant removing parts of research questions that directed the investigation towards identifying optimal solutions. As there is no course requirement to actually succeed in engineering such an optimal solution, the only real difference was that students could thus limit the investigation to an ad hoc engineered prototype – avoiding discussing the implications of other engineering choices. Instead the mechanical aspects of applying scientific methods seem to have been stressed, allowing students to refer to these to motivate a more narrow investigation. As an example, 2 of these theses utilized unstructured interviews to support their case, which was otherwise quite uncommon.

3.3.2 Analysing

Analysing involves breaking down a problem into steps highlighting important aspects [37]. A primary category for Analysing was seen in 4 out of 7 in the Y2 optimal group (7 out of 12 also counting theses sub-optimal in regard to grade) in regard to a discussion reflecting critical inquiry (Table 15).

The difference between Y0 and Y2 Students in regard to Analysing was connected to students breaking down the field context when deploying an engineered system into parts discussable in separation. As indicated in Appendix B, Table 17, this for instance meant discussing what adding or removing different parts of the engineered system implied, and discussing each part of the engineering process in relation to the end result. As an example, one thesis discussed the engineered system from the perspective of each type of sensor that could be attached to it. Many students rather simply referred to the capabilities of their complete system, which engineering had only been limited by the components available at the time. Another discussed engineered artefacts on a scale going from simulations to prototype, highlighting what each form indicated in regard to the use of a real system. This differed from many theses that went through the steps of an engineering process, but never challenged the initial assumptions regarding the system to be engineered formed at the start of the process.

3.3.3 Processing Critically

Processing Critically (PC) involves arriving at one's own conclusions based on facts and arguments [37]. A primary category for PC was seen in 4 out of 7 in the Y2 optimal group (7 out of 12 also counting theses sub-optimal in regard to grade) in regard to a discussion reflecting critical inquiry (Table 15).

The difference between Y0 and Y2 Students in regard to Processing Critically was connected to students raising validity concerns in regard to their study, or arriving at the limitations of it. As indicated in Appendix B, Table 17, this for instance meant that they challenged their own attempts to verify that their system worked as specified, and suggested tests that would validate that they had built the right system rather than simply built a system according to a specification. As an example, one thesis analysed the installation of a system that measured vehicles passing an intersection, rather than, as most, accepting the associated statistics and guidance from industrial supervisors directly. Another discussed the effect of loose clothing, rather than user experience, when measuring the effect of different prototypes on body awareness. This came about due to user tests with people from many different backgrounds, which highlighted difficulties with using the specified sensors that had not been identified when researchers had tested the prototypes themselves.

3.3.4 Relating/Structuring

Relating/Structuring (R/S) involves connecting different parts of the learning experience, e.g. by imposing a structure on the main concepts of an article [37]. A primary category for R/S was seen in 4 out of 11 in the Y2 optimal group (8 out of 17 also counting theses sub-optimal in regard to grade) in regard to the structure of the empirical investigations (Table 13).

The difference between Y0 and Y2 Students in regard to Relating/Structuring was connected to students structuring their empirical investigations into inter-relatable stages, or realizing several engineering concepts and comparing them to each other. As indicated in Appendix B, Table 17, this for instance meant that students identified what could not be verified using simulations and proceeded to test this using prototypes, and built several prototypes to investigate the limitations of different design concepts. As an example, during the writing of one thesis two fall-detection systems were built and the different aspects of these realizations compared. Another realized simulation models for vehicle steering and braking and then related these to field tests.

4 Discussion

To organise our discussion we divide it into three parts, relating our results to theory and practice, as well as discussing the need for further research.

4.1 Theory

In regard to the learning goals of our master thesis course, a surface approach to learning is to focus on building engineering prototypes without reflecting on the overall purpose or strategy. This was the approach of most of the Y0 Students. However, Y2 Students showed strength in (a) systematic planning at the process level (see R/S), and (b) breaking down the learning experience into separate logically debatable steps (see Analysis and PC). This is evidence of a serialist style of a deep approach to learning. At the same time Y2 Students showed an increased ability to analyse a situation from different perspectives, while organizing parts into a meaningful whole. This evidence of a holist style was evident when Y2 Students brought together aspects of their field context and discussed validity. It seems Y2 Students were more versatile, able to alternate between both styles of deep approaches to learning. Furthermore, Y2 Students utilizing a strategic approach seem better at directing their efforts to engage in a deep approach to learning (see Adjusting).

Learning approaches in engineering education have been discussed with regard to the research-teaching nexus, mostly in terms of the tension between convergent and divergent learning approaches [40, 67]. For the purpose of the discussion in this paper, the convergent vs divergent distinction can be equated with the aforementioned serialist vs holist learning styles. Convergent learning approaches are thought to be promoted by traditional “engineering science” curricula, which emphasize engineering analysis skills. These curricula are thus thought to be detrimental to skills in engineering design, which would instead benefit from fostering the ability for divergent inquiry.

Our results initially seem to point to an outcome in line with stressing convergent thinking. Theses converged quicker and in more of a step-by-step analytical fashion. However, rather than being detrimental to divergent thinking, our results point to a more versatile and strategic learning approach. This can be best understood as an increased ability to combine analysis and design to successfully create a working system, i.e. *engineering synthesis*. While Y2 Students structured their entire learning experience in a more convergent way, they did not approach each part in a more single-minded search for *the* truth.

It is therefore an important observation that these improvements to learning activities were primarily evident in the support for engineering tasks, and not in pursuit of independent research. The underlying mechanism for the improvements was not a simple re-enactment of research methods learnt during research training. The mechanism rather appears to be a transfer of key insights of the production of knowledge from the context of research to the context of engineering. In *general,* transferability of skills and knowledge is an often highlighted benefit of inquiry-based approaches [4]. In regard to research, this can *specifically* be contrasted with traditional lectures and reading material, which are often based in one worldview. Accounts of how to apply methods therefore often leave out underlying assumptions. These assumptions can, on the contrary, be concrete in discussions where students are allowed to independently assess competing worldviews and question researchers.

The important mechanism to support our students' understanding of how to integrate engineering design and analysis would seem to be two-fold: to not only stress top-down structuring in research training, but also the ability to think freely and creatively about confounding factors. The result was not restricted to the isolated examples used to describe the differences between Y0 and Y2 Students in the previous section. Several students that structured their engineering as a series of steps used their discussions on validity to bound their investigations. As an example, a student that performed a sensitivity analysis on a vehicle model provided reasons for why the model could for instance be invalid for a car with a driver, but then stopped at noting that this was a limitation of the results. In other words, these students were able to see their engineering as a means to answer a question reasonably well within identifiable limitations. Students unable to adapt this perspective often extended their engineering needlessly into efforts of unreasonable size or limited value. This dual emphasis is most likely supported by a pragmatic worldview, which routinely weighs the weaknesses and strengths of research methods against each other. As this skill takes time to learn the students were probably mostly affected by the idea that both quantitative and qualitative methodologies can be acceptable as long as the situation merits it. Interviews were used as an example to put this message across to Y1 and Y2 Students. Both structured and semi-structured interviews were motivated in a fictitious study presented to the students, but based on different goals and perspectives of the involved researchers. Although still uncommon, interviews were the most frequently used method by Y1 and Y2 Students in combination with otherwise quantitative case studies.

The perspective that research training is harmful to aspects of engineering associated with divergent thinking might thus be too simplistic. If the research training introduces students to the motivations of a variety of quantitative and qualitative methodologies, it might actually help them to integrate aspects associated with divergent thinking with aspects dependent on diametrically opposed thinking. In addition to the aforementioned differences in how Y2 Students handled validity and accepted methodological combinations involving qualitative methods, there were a few weaker indicators of this ability. Firstly, Y2 Students seemed to have systematically queried more stakeholders for information on the context of the theses. Ultimately, Y2 Students seemed more willing to search for information that contradicted their initial plans, and make changes to the steps in their engineering process if they seemed unlikely to succeed. Secondly, all theses by Y2 Students that achieved an A grade solely based on their engineering achievements seem to have started with a high quality research question. This might indicate that these students started by thinking through the many aspects of their engineering problem more thoroughly than other students.

4.2 Practice

If our students' understanding of competing research methods, processes and worldviews are key to the observed outcomes, then arguably the most difficult challenge to successfully applying our intervention is that most researchers are proficient only in a narrow set of these. We extended the invitation to take part in the intervention to researchers from across several different research fields at our Department, which helped us to largely avoid this problem. However, we appreciate that this might not be possible at other departments. Therefore, the main implication of our results to the practice of engineering education is the need for more cooperation between engineering disciplines.

Traditional learning environments in engineering education include laboratories, cooperative education and research [2]. While a single perspective on knowledge production can permeate these learning environments at any one program, division or department, there are probably differences between them. Current engineering practices include cross-disciplinary projects with a requirement for interpersonal and creative skills [2]. In the future, engineering students are thus likely to see more interaction with research and development at firms other than those traditionally affiliated with their engineering discipline, and with the traditional learning environments of departments teaching other engineering disciplines, or even non-engineering institutions. As engineering education develops in this direction we urge teachers to take the opportunity to include learning goals that involve understanding the knowledge production of these other disciplines. As our results indicate, understanding the production of knowledge in other professions and affiliated sciences might not only improve interactions across disciplinary boundaries, but also an engineering student's own engineering processes.

4.3 Further research

Transferability of skills and knowledge has been mentioned as a benefit of our intervention being inquiry-based. However, the discussion has been focused on the shift from research-related subject content to research-related processes, rather than the teaching-centred to student-centred shift. This might seem like downplaying the latter shift, given the aforementioned close resemblance between inquiry-based learning and research activities. Could not the *primary* reason for the observed improvements be the additional requirements on our students to carry out inquiries?

We do not make this claim, since we found the inductive categories related to learning activities to be connected to the optimal groups in regard to research aspects, completion time and grade. Had the improvement mostly been related to our students' critical thinking skills, we would for several reasons have expected to see the same type of change to learning activities across the whole cohort. Firstly, breaking down the field context into parts discussable in separation does not require that more than technical issues are discussed. Secondly, a longer than average completion time gives students more opportunities to discuss validity, especially if the engineering outcomes are not optimal. Thirdly, while a lack of connection to field tests might be related to completion time, these tests can be expected in an engineering process regardless of the structure of any overlaying investigation. The inquiry-based approach thus seems to be more of a vehicle for achieving our results than the mechanism behind them.

With this in mind, our results point to the disturbing issue that even though we did not see any negative trade-offs, there appears to be a group of students that are left behind by our intervention. This suggests that further research needs to be performed to identify the limitations of our intervention in regard to which students are affected by it and how. The relationship between our results and the form of inductive learning employed should then be a good starting point. It would be interesting to see whether inquiry-based teaching that more strongly mimics research activities could lead to the same results. Case-based teaching utilizing examples from previous years could also be used to make learning less abstract and thus potentially more easily relatable to our students' experience of engineering. How much this is a motivational issue is also an open question – does the lesser extent of constraint in the teaching activities associated with our intervention specifically encourage those students which favour hands-on skills to approach these activities with a surface learning approach?

5 Conclusions

Our intervention affected our students' way of self-regulating certain learning activities. This effect seems to be linked to both our context and our use of inquiry-based research training. Our pragmatic context ensured the research training encompassed competing worldviews and methodologies. Our inquiry-based approach enabled students to transfer research knowledge to engineering practice. In this way our research training did not manifest itself simply as an increased ability to conduct research independently, but rather as an ability to self-regulate learning activities towards achieving engineering synthesis. This suggests research training can be helpful in teaching students the diametrically opposed aspects of thinking required by current engineering processes. It also implies that teachers should use the fact that engineering education is evolving towards more cross-disciplinary cooperation to ensure students learn about different perspectives on knowledge production.

Acknowledgements

Special thanks go to Vicki Derbyshire for her help in proofreading; Professor Mats Magnusson for auditing this study; and Professor Norm Vaughan for his helpful comments on inquiry-based learning.

References

[1] Swedish Council for Higher Education, The Higher Education Ordinance, in, 2016.
[2] L.H. Jamieson, J.R. Lohmann, Innovation with impact: Creating a culture for scholarly and systematic innovation in engineering education, American Society for Engineering Education, Washington, DC, 2012.
[3] National Academy of Engineering, Getting to 2020: Guiding Strategies, in: Educating the Engineer of 2020: Adapting Engineering Education to the New Century, The National Academies Press, Washington, DC, 2005, pp. 17-32.
[4] M.J. Prince, R.M. Felder, Inductive teaching and learning methods: Definitions, comparisons, and research bases, Journal of engineering education, 95 (2006) 123-138.
[5] V. Lee, D. Greene, J. Odom, E. Schechter, R. Slatta, What is Inquiry-Guided Learning?, in: Teching and Learning Through Inquiry, Stylus Publishing, LLC, Sterling, Virginia, 2004, pp. 3-16.

[6] E.M. Furtak, T. Seidel, H. Iverson, D.C. Briggs, Experimental and quasi-experimental studies of inquiry-based science teaching: A meta-analysis, Review of educational research, 82 (2012) 300-329.
[7] M. Tight, Examining the research/teaching nexus, European Journal of Higher Education, 6 (2016) 293-311.
[8] A.J. Dutson, R.H. Todd, S.P. Magleby, C.D. Sorensen, A Review of Literature on Teaching Engineering Design Through Project-Oriented Capstone Courses, Journal of Engineering Education, 86 (1997) 17-28.
[9] M. Downey, Is There More to Engineering that Applied Science?, Lund University, Lund, 2012.
[10] M. Elken, S. Wollscheid, The relationship between research and education: typologies and indicators. A literature review, Nordic Institute for Studies in Innovation, Research and Education (NIFU), Lysaker, 2016.
[11] M.T. Nettles, C.M. Millett, A Map from the Past to the Present, in: Three magic letters: Getting to Ph.D., The Johns Hopkins University Press, Baltimore, 2006, pp. 1-6.
[12] M.C. Loui, Editor's Page: Board Changes and Neglected Research Topics, Journal of Engineering Education, 105 (2016) 3-5.
[13] S.W. Rogers, R.K. Goktas, Exploring engineering graduate student research proficiency with student surveys, Journal of Engineering Education, 99 (2010) 263-278.
[14] A. Brew, E. Jewell, Enhancing quality learning through experiences of research-based learning: implications for academic development, International Journal for Academic Development, 17 (2012) 47-58.
[15] J. Hattie, H.W. Marsh, The relationship between research and teaching: A meta-analysis, Review of educational research, 66 (1996) 507-542.
[16] T. Pocklington, A. Tupper, Teaching and Research at Canadian Universities, in: No Place to Learn, UBC Press, Vancouver, 2002, pp. 105-120.
[17] J. Robertson, C.H. Bond, Experiences of the relation between teaching and research: What do academics value?, Higher Education Research & Development, 20 (2001) 5-19.
[18] M.Q. uz Zaman, Review of the academic evidence on the relationship between teaching and research in higher education, Research Report RR506, Department for Education and Skills, London, 2004.
[19] M. Healey, Linking research and teaching to benefit student learning, Journal of Geography in Higher Education, 29 (2005) 183-201.
[20] R.S. Hathaway, B.A. Nagda, S.R. Gregerman, The relationship of undergraduate research participation to graduate and professional education pursuit: an empirical study, Journal of College Student Development, 43 (2002) 614-631.
[21] B. Seely, Patterns in the History of Engineering Education Reform: A Brief Essay, in: Educating the Engineer of 2020: Adapting Engineering Education to the New Century, The National Academies Press, Washington, DC, 2005, pp. 114-130.
[22] The Boyer Commission on Educating Undergraduates in the Research University, Reinventing Undergraduate Education: A blueprint for America\'s research universities, Stony Brook, New York, 1998.
[23] J. Halliwell, The nexus of teaching and research: Evidence and insights from the literature, Higher Education Quality Council of Ontario Toronto, Toronto, 2008.
[24] I. Cherastidtham, J. Sonnemann, A. Norton, The teaching-research nexus in higher education, Grattan Institute, Melbourne, Australia, 2013.
[25] M.J. Prince, R.M. Felder, R. Brent, Does faculty research improve undergraduate teaching? An analysis of existing and potential synergies, Journal of Engineering Education, 96 (2007) 283-294.
[26] H. Davies, Survey of master degrees in Europe, European University Association, Brussels, 2009.
[27] S. Francis, L. Goodwin, C. Lynch, Professional Science Master´s, in: Challenges to Sustaining Professional Science Master's Programs, Council of Graduate Schools, Washington, DC, 2011, pp. 109-116.
[28] C. Kasworm, L. Hemmingsen, Preparing professionals for lifelong learning: Comparative examination of master's education programs, Higher education, 54 (2007) 449-468.
[29] J. Gilmore, D.M. Lewis, M. Maher, D. Feldon, B.E. Timmerman, Feeding Two Birds with One Scone? The Relationship between Teaching and Research for Graduate Students across the Disciplines, International Journal of Teaching and Learning in Higher Education, 27 (2015) 25-41.
[30] R. Griffiths, Knowledge production and the research–teaching nexus: The case of the built environment disciplines, Studies in Higher education, 29 (2004) 709-726.
[31] A. Jenkins, M. Healey, R. Zetter, Linking teaching and research in disciplines and departments, in, York, 2007, pp. 96.

[32] A. Aditomo, P. Goodyear, A.-M. Bliuc, R.A. Ellis, Inquiry-based learning in higher education: principal forms, educational objectives, and disciplinary variations, Studies in Higher Education, 38 (2013) 1239-1258.
[33] N.J. Entwistle, Improving teaching through research on student learning, in: University teaching: International perspectives, Routledge, New York, 1998, pp. 73-112.
[34] F. Coffield, D. Moseley, E. Hall, K. Ecclestone, Should we be using learning styles? What research has to say to practice, Learning and skills research centre, London, 2004.
[35] F. Coffield, D. Moseley, E. Hall, K. Ecclestone, Learning styles and pedagogy in post-16 learning: A systematic and critical review, Learning and Skills Research Centre, London, 2004.
[36] F. Marton, R. Säljö, On qualitative differences in learning: I—Outcome and process, British journal of educational psychology, 46 (1976) 4-11.
[37] J.D. Vermunt, N. Verloop, Congruence and friction between learning and teaching, Learning and instruction, 9 (1999) 257-280.
[38] N. Entwistle, V. McCune, The conceptual bases of study strategy inventories, Educational Psychology Review, 16 (2004) 325-345.
[39] M. Healey, Linking research and teaching exploring disciplinary spaces and the role of inquiry-based learning, in: Reshaping the university: New relationships between research, scholarship and teaching, Open University Press, Maidenhead, England, 2005, pp. 67-78.
[40] C.L. Dym, A.M. Agogino, O. Eris, D.D. Frey, L.J. Leifer, Engineering design thinking, teaching, and learning, Journal of Engineering Education, 94 (2005) 103-120.
[41] Å. Lindberg-Sand, The Embedding of the European Higher Education Reform at the Institutional Level: Development of Outcome-Based and Flexible Curricula?, in: European Higher Education at the Crossroads, Springer, Dordrecht, 2012, pp. 191-207.
[42] G. Heitmann, Challenges of engineering education and curriculum development in the context of the Bologna process, European Journal of Engineering Education, 30 (2005) 447-458.
[43] M. Grimheden, M. Hanson, Collaborative learning in mechatronics with globally distributed teams, International journal of engineering education, 19 (2003) 569-574.
[44] A. Atrens, R. Truss, A. Dahle, G. Schaffer, D. St John, C. Caceres, J. Gates, Graduate attributes in relation to curriculum design and delivery in a Bachelor of Materials Engineering programme, International Journal of Engineering Education, 20 (2004) 834-848.
[45] M.G. Rasul, J.J. Lawson, P. Howard, F. Martin, R. Hadgraft, Learning and teaching approaches and methodologies of capstone final year engineering projects, International Journal of Engineering Education, (2015).
[46] KTH Royal Institute of Technology, Education Assessment Exercise (EAE) 2011: External panel of assessors report, Ineko, Stockholm, 2011.
[47] A. Tashakkori, C. Teddlie, Mixed methodology: Combining qualitative and quantitative approaches, Sage Publications, Thousands Oaks, California, 1998.
[48] J.W. Creswell, The Selection of a Research Approach, in: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage Publications, London, 2013, pp. 3-24.
[49] J.W. Creswell, The Introduction, in: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, Sage Publications, London, 2013, pp. 107-122.
[50] J.W. Creswell, Mixed Methods Procedures, in: Research Design: Qualitative, Quantitative, and Mixed Methods Approaches, 2013, pp. 215-240.
[51] N. Draper, H. Smith, Checking the Straight Line Fit, in: Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1998, pp. 47-77.
[52] D. Montgomery, Experiments with a Single Factor, in: Design and Analysis of Experiments, John Wiley & Sons, Inc., New York, 2000, pp. 60-125.
[53] L. Cohen, L. Manion, K. Morrison, Experiments, quasi-experiments, single-case research and meta-analysis, in: Research Methods in Education, Routledge, Abingdon, 2007, pp. 272-296.
[54] L. Cohen, L. Manion, K. Morrison, Content analysis and grounded theory, in: Research Methods in Education, Routledge, Abingdon, 2007, pp. 475-500.
[55] L. Cohen, L. Manion, K. Morrison, Validity and reliability, in: Research Methods in Education, Routledge, Abingdon, 2007, pp. 133-164.
[56] A. Tashakkori, C. Teddlie, Research Designs Issues for Mixed Method and Mixed Model Studies, in: Mixed methodology: Combining qualitative and quantitative approaches, Sage Publications, Thousands Oaks, California, 1998, pp. 40-58.
[57] J.W. Creswell, D.L. Miller, Determining validity in qualitative inquiry, Theory into practice, 39 (2000) 124-130.

[58] P. Steiner, A. Wroblewski, T. Cook, Randomized Experiments and Quasi-Experimental Designs in Educational Research, in: The SAGE International Handbook of Educational Evaluation, SAGE Publications, Thousand Oaks, California, 2009, pp. 75-96.
[59] W. Shadish, T. Cook, D. Campbell, Experiments and Generalized Causal Inference, in: Experimental and Quasi-Experimental Designs for Generalized Causal Inference, Houghton-Mifflin, Boston, 2002, pp. 1-32.
[60] D. Montgomery, Simple Comparative Experiments, in: Design and Analysis of Experiments, John Wiley & Sons, Inc., New York, 2000, pp. 21-59.
[61] N. Draper, H. Smith, Multiple Regression Applied to Analysis of Variance Problems, in: Applied Regression Analysis, John Wiley & Sons, Inc., New York, 1998, pp. 473-504.
[62] T. Ryan, Nonparametric Methods, in: Sample Size Determination and Power, John Wiley & Sons, Inc., Hoboken, New Jersey., 2013, pp. 323-340.
[63] D. Campbell, J. Stanley, Experimental and Quasi-Experimental Designs for Research, in: Handbook on Research on Teaching, Rand McNally & Co., Chicago, 1963, pp. n.p.
[64] A. Tashakkori, C. Teddlie, Sampling, Measurement and Quality of Inferences, in: Mixed methodology: Combining qualitative and quantitative approaches, Sage Publications, Thousands Oaks, California, 1998, pp. 61-94.
[65] M. Hollander, D. Wolfe, E. Chicken, The Two-Sample Location Problem, in: Nonparametric Statistical Methods, John Wiley & Sons, Inc., 2014, pp. 115-150.
[66] G.K. Rosenkranz, A note on the Hodges–Lehmann estimator, Pharmaceutical statistics, 9 (2010) 162-167.
[67] W.H. Wood, Decision-based design: a vehicle for curriculum integration, International Journal of Engineering Education, 20 (2004) 433-439.

Biography

Fredrik Asplund received his PhD at the KTH Royal Institute of Technology in 2014, and since then has been a Postdoctoral Researcher at Rolls-Royce plc in Derby, UK. Research interests include system safety, empirical software engineering and engineering education in the Cyber-Physical Systems domains. Dr Asplund is a Marie Cure Fellow and a VINNMER Fellow.

Martin Edin Grimheden, associate professor at KTH Royal Institute of Technology, is head of the Mechatronics Lab and Program Director of the Mechanical Engineering Program. Dr Grimheden has a combination of degrees in Engineering and Education. He has been heavily involved in the transition into the BSc/MSc-system at KTH. His current research interest includes studies of learning in higher education, of learning in Mechatronics, and of internationalization of higher education.

Appendix A, Quantitative data

Table 16 includes the quantitative data on which this paper is based.

[Table 16 here]

Appendix B, Qualitative data

Table 17 includes examples from qualitative data on which this paper is based.

[Table 17 here]

Table 1 Taxonomy by Vermunt and Verloop [37]

Cognitive	Metacognitive	Affective
Relating/Structuring	Orienting/Planning	Motivating/Expecting
Analysing	Monitoring/Testing/Diagnosing	Concentrating/Exerting Effort
Concretizing/Applying	Adjusting	Attributing/Judging Oneself
Memorizing/Rehearsing	Evaluating/Reflecting	Appraising
Critical Processing		Dealing with Emotions
Selecting		

Table 2 Study cohorts and their treatment

Course Theme	Y0 Students (22 Students)	Y1 Students (28 Students)	Y2 Students (25 Students)
Research at the Department, 3 ECTS credits	Research-led	Research-based	Research-based
Research methods, processes and worldviews, 4.5 ECTS credits	Research-oriented	Research-oriented	Research-based

Table 3 Examples of research questions using quotes from theses

Low Quality	High Quality
“The purpose of this thesis is to investigate if the estimation of vehicle mass of an HDV can be improved if the road grade is retrieved from a map database instead of not using it.”	“Sub questions that arise are: what are the advantages and disadvantages with an automated environment, in terms of effectiveness, safety and time? Does it add uncertainties into the testing process?”
“Develop a control strategy for two electrically actuated bypass valves that operates the exhaust gas into two separate TEGs with the condition that the exhaust gas do not overheat and damage the TEGs.”	“How does the new media affect the stability and responsiveness when applying the multicopter techniques' under water?”
“What robust control strategy can be designed and implemented on an active damping test rig in order to reduce vibrations on a forwarder cabin?”	“Can the system given a reasonable guess of initial system settings optimize the process with regards to robustness, capacity and efficiency?”

Table 4 Examples of methodological approaches using quotes from theses

Low Quality	High Quality
“The project started with an extensive research on solutions to this problem and by delving into the current system used at [Firm]. After the initial research, the development process took the shape of an iterative methodology although no textbook procedure was applied.”	“To be able to present such a result, data will be gathered by interviews and literature analysis.”
“The first part of the thesis consisted of learning about EMG measuring, and the demands and limitations that could relate to the thesis ... Once the information had been stripped down to its most basic fundamentals that related to the thesis, focus was changed to the technical aspect, meaning the specific components and their technical data that would be used in the project. ... For a project of this size and organization complexity consisting of only one person, a macro cycle version of the V-model was used since it would provide a systematic and logical approach to the different areas of interest during the project ...”	“The idea was, through a case study, to explore the possibility to transfer the stabilizing and manoeuvring platform techniques' from airborne multi-copter vehicles (multi-copters, quadcopters etc.) to a new medium.”
[No methodology discussed or used to structure the investigation of the research questions.]	“Once the model strategy has been chosen and the model has been built in MATLAB/Simulink, measurements from a [Firm] LNG truck will be used to verify the model through simulation by supplying the model with the same input as the real tank in the measurement. Furthermore the hold time of the tank model will be simulated and verified against indicative data provided by the

	tank manufacturer. Once the model is verified, the computational time and the processor load of the developed model will be analysed, together with an observer solution in the form of an extended Kalman filter, that will be tested on the model and evaluated, both with respect to performance and processor load.”

Table 5 Examples of discussion content using quotes from theses

Low Quality	High Quality
"The goal of this project was to research scheduling algorithms for multi-core embedded systems … several algorithms have been studied and compared. The linear clustering algorithm was chosen to be implemented. In the practical phase a toolchain for programming parallel applications was implemented. The target platform was the Epiphany E16 development board. Different software modules for that target had to be implemented … Several experiments were carried out in order to help evaluate the performance of the system. Parallel computing should only be used when there is enough computation to be parallelized. If there is little parallelization to be done, the mailbox system is not worthy to use. However, the execution can still be sped up by using algorithms based on task duplication. A mailbox system is worthy to use in applications with a lot of parallelization because the communication overhead can then be neglected."	"Among the three measuring procedure concepts of the DOC, the HC-slip test seems to have the highest potential to measure the oxidation performance of the DOC. In comparison to the NOx transient test, it has the ability to measure the performance of the DOC alone. Also, it is not dependent on the condition of neither the SCR nor the NOx-sensors. In comparison to the comparative test, it show tendency to be able to measure the oxidation performance and has fewer model dependencies. It also has a higher potential of measuring the light-off temperature. Since Concept 1 show tendency to be able to measure the oxidation performance of the DOC, the used exhaust mass flow in the tests seems to be sufficient to stress the DOC to obtain a measurement of the performance. Since the resolution of the results is still undetermined, it is not possible to decide whether if it could be lowered or must to be increased further. The HC-slip concept included some drawbacks, such as long duration time and troublesome temperature regulation. These drawbacks have to be investigated further, to find potential improvements of increasing the efficiency of this test."
"In the implementation in this thesis of the troubleshooting application, the main areas for	"The system in an application will have multiple benefits compared to a traditional static system.

improvements are the correctness of the Bayesian model, and a more complex troubleshooting algorithm. However, since the purpose of this thesis was to demonstrate how an integrated troubleshooting system that uses Bayesian network models for preparation of an action plan, it's natural that these were not as optimal as they could've been. The troubleshooting algorithm, as mentioned earlier, only looked one step ahead in time and never considered the possibility to conduct a test later in time. This limited the efficiency of the algorithm heavily, but the efficiency was sufficient enough for the implementation in this thesis. Since the troubleshooting algorithm depended heavily on the outcomes of the Bayesian network model and its probability distribution, it's concluded that for a successful troubleshooting (in the sense of minimizing repair cost and minimize downtime of the vehicle), both the model and algorithm need to be as optimal as possible. Hence, both are an area of focus for future work."	One of the main being the reliability of operation to be expected after the initialisation phase has passed. This reliability of operation is due to the level sensors, emergency mechanism and evolutionary learning from earlier cycles. For the traditional conveyor a source of machine downtime is failure of the filter caused by overfilling with material … Another large benefit of the optimisation is that due to that the system can without risk be operated closer to maximum performance, the system can therefore be used more efficiently or in new applications … A mayor question regarding this thesis is how to relate the measured performance improvements to what could be expected to be achieved by an operator? This is of course a question without a definitive answer since it will depend on the operators' level of skill and time assigned for the task. A skilled operator that has long experience will of course use that experience and achieve good performance of the system within a relatively short time-span with high probability. If the operator instead is a novice the time required will probably increase significantly. The novice operator will also face the problem of identifying behaviour that may lead to problems in the long run such as robustness issues due to too heavy plugs forming that he or she has not encountered before …

	Another large issue that needs to be addressed in the process in creating a product of this technology is how it should be implemented. One vision is to create an optimisation system that is add on to the conveying system and runs the optimisation until the operator is content with the results and then aborts and removes the extra equipment. That equipment can therefor consist of high quality components and be expensive as it will be able to service a large number of machines. Another approach is to have complete system distributed on all locations and create a database of solutions that have been proven that can be distributed to benefit all. The step of taking this technology from the laboratory to the factory will raise some ethical aspects on if such technology should be released on all markets and applications …”

Table 6 Grade test data summary

Sample Size Calculation		Test Results	
Minimum detectable difference	0.5 grade step (considered minor)	Outliers	No (assessed through boxplot)
Standard deviation	0.5 grade step (expected based on previous year groups)	Data normally distributed	Yes (Shapiro-Wilk's test ($p>.05$))
Power	0.8 (standard)	Homogeneity of variances	Yes (Levene's test for equality of variances ($p=.945$))
Calculated sample size (one-way ANOVA)	21 students/cohort	Test statistics	$F(2,72)=2.315$, $p=.106$

Table 7 Age test data summary

Sample Size Calculation		Test Results	
Minimum detectable difference	3 years (length of Swedish education cycles)	Outliers	Yes (assessed through boxplot)
Standard deviation	2 years (expected based on previous year groups)	Data normally distributed	No (Shapiro-Wilk’s test ($p<.05$))
Power	0.8 (standard)	Distributions of ages similar	No (assessed through boxplot)
Calculated sample size (one-way ANOVA)	10 students/cohort	Test statistics	$\chi^2(2)=.906$, $p=.636$

Table 8 Examiners test data summary

Sample Size Calculation		Test Results	
Minimum detectable difference	15 days (considered minor)	Outliers	Yes (assessed through boxplot)
Standard deviation	35 days (expected based on previous year groups)	Data normally distributed	No (Shapiro-Wilk's test ($p<.05$))
Power	0.8 (standard)	Distributions of completion times similar	No (assessed through boxplot)
Calculated sample size (one-way ANOVA)	106 students/cohort	Test statistics	$\chi^2(5)=5.570$, $p=.350$

Table 9 Cohort test data summary

Sample Size Calculation	Test Results	
Underpowered, see Table 8.	Outliers	Yes (assessed through boxplot)
	Data normally distributed	No (Shapiro-Wilk's test ($p<.05$))
	Distributions of completion times similar	No (assessed through boxplot)
	Test statistics	$\chi^2(2)=8.207$, $p=.017$

Table 10 Y0 Completion time, grade and research questions quality

Completion Time	18% / 5%	0% / 0%
	50% / 9%	18% / 0%
	Grade	

Table 11 Y2 Completion time, grade and research questions quality

Completion Time		
	8% / 0%	8% / 4%
	8% / 16%	24% / 32%
	Grade	

Table 12 Y0 Completion time, grade and structure of empirical investigation

Completion Time	18% / 5%	0% / 0%
	45% / 14%	9% / 9%
	Grade	

Table 13 Y2 Completion time, grade and structure of empirical investigation

Completion Time		
	4% / 4%	12% / 0%
	0% / 24%	12% / 44%
	Grade	

Table 14 Y0 Completion time, grade and discussion reflecting critical inquiry

Completion Time		
	18% / 5%	0% / 0%
	50% / 9%	18% / 0%
	Grade	

Table 15 Y2 Completion time, grade and discussion reflecting critical inquiry

Completion Time	8% / 0%	12% / 0%
	4% / 20%	28% / 28%
	Grade	

Table 16 Quantitative data, completion time (days)

Student Number	Y0	Y1	Y2
1	398	106	103
2	100	195	149
3	147	108	106
4	154	108	101
5	119	108	101
6	149	108	195
7	100	101	101
8	114	101	106
9	114	108	112
10	100	113	102
11	100	95	105
12	101	95	100
13	147	109	93
14	147	118	102
15	92	102	102
16	280	102	93
17	111	179	93
18	136	159	92
19	126	132	92
20	175	96	95
21	159	170	95
22	342	92	130
23		92	130
24		292	202
25		100	97
26		212	
27		216	

28		269	

Table 17 Examples of qualitative data

Associated Learning Activity	Examples of Inductive Categories	Text Examples (Coder notes in square brackets)
Adjusting	“Left out agreed learning goals”	“Identify the optimal … in regard to robustness, cost, length of service and impact on the environment.” [Not found in thesis] “How can the performance of … be verified? [Against field behaviour]” [Not found in thesis]
Analysing	“Field is heterogeneous”	“Materials:[Long list of materials to be analysed in separation]” “… highlighting the importance with interaction between different disciplines, such as the [Different important disciplines]”
	“Critical concepts”	“In this section the effect of adding additional sensors [of different types] are discussed.” “The faults that have been tested are the following: [List of separate fault modes to be considered]”
Processing Critically	“Unsuitable verification?”	“More preferable would be to have done the tests in a pool, inside a house. To prevent

		disturbances the tests were performed … Because of this it is possible to expect that the tests presented here is repeatable with the same results … when performing tests in a reactor tank." "During the tests it was noticed that the cable did not disturb the [system] as much as expected." "When tests were conducted between ... there were some issues that could affect the outcome of the tests. The main issue was the lack of time for prolonged testing."
	"Complex validation"	"To improve the validity of these conclusions global methods can be considered and more simulations should be done in different speeds and with different parameter values." "The final prototype was tested with a group of users that had no previous experience with [intent of system] and with different background."
Relating/Structuring	"Models compared to field tests"	"… evaluating the accuracy of two different vehicle models,

		comprising the steering system, against real car measurements." "… the prototype was used to perform tests that was not possible to simulate with help of the developed model."
	"Field to field comparisons"	"Three prototypes will be created to examine the behaviour … The outcome of the tests with this prototype will contribute to decisions made when ... for a second and third prototype." "To test which … panel would work best … both panels were ..."

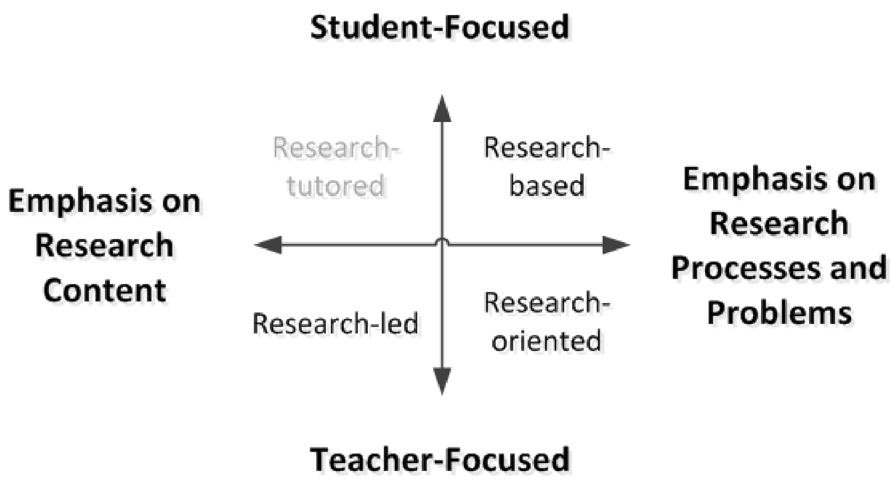

Figure 1 Healey’s [39] elaboration of the model by Griffiths [30]

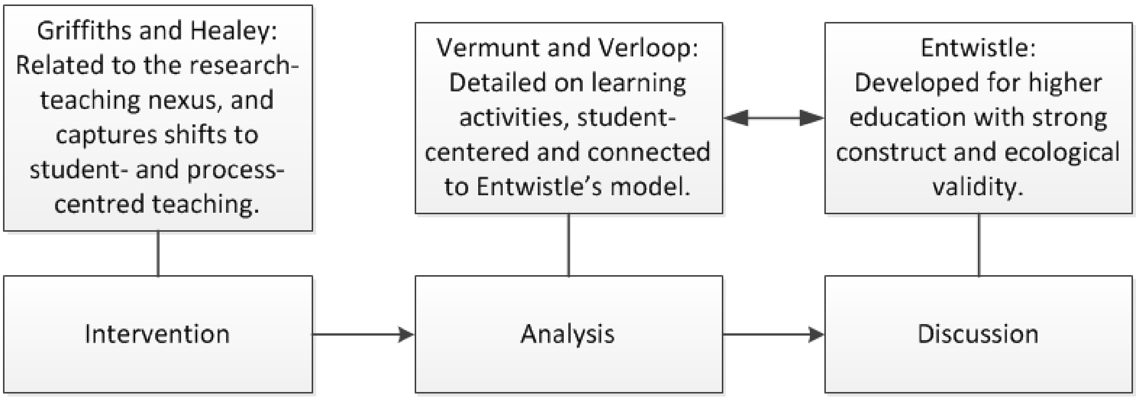

Figure 2 Summary of conceptual framework

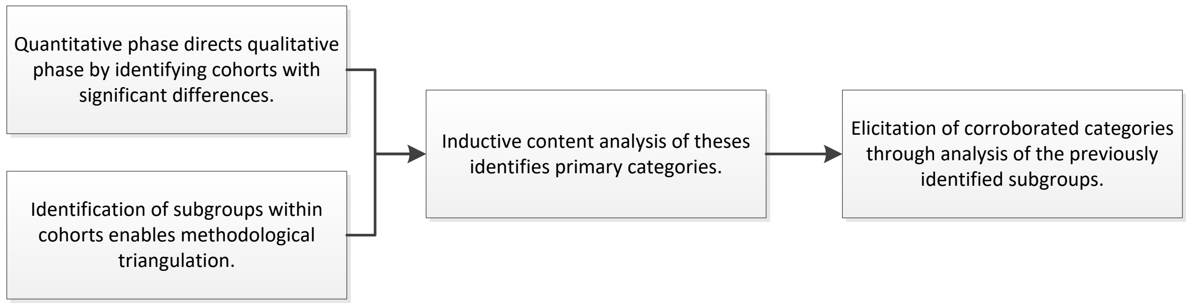

Figure 3 The relationship between phases

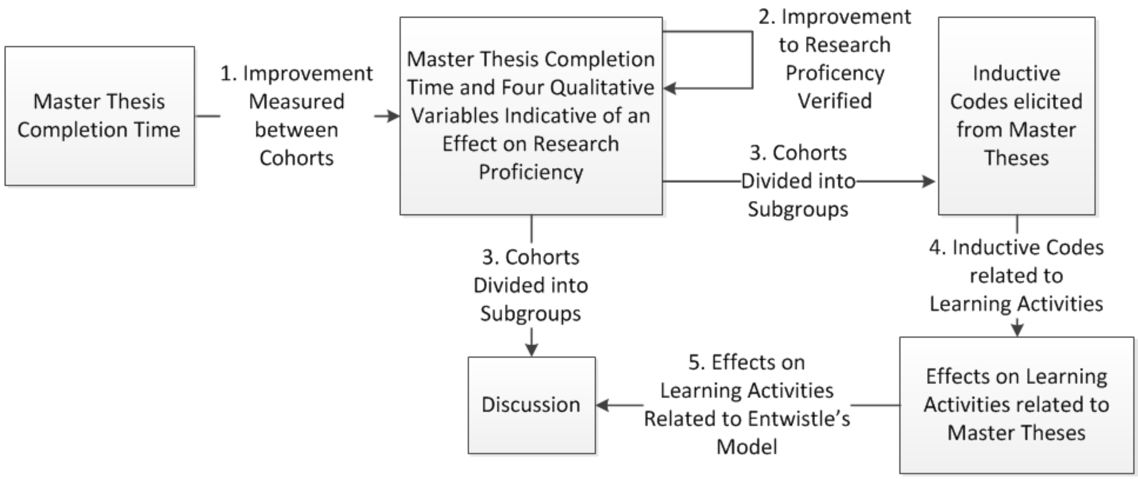

Figure 4 The relationship between data sets