*Lärarlärdom, högskolepedagogisk konferens, 2017*

# LÄRARLÄRDOM, HÖGSKOLEPEDAGOGISK KONFERENS

vid

Blekinge Tekniska Högskola

2017

ISBN 978-91-7295-961-3

Redaktör
Christina Hansson

Enheten för utbildningsutveckling
Blekinge Tekniska Högskola

# Innehåll

# Introduktion

## Lärarlärdom 2017

Högskolepedagogisk utveckling har under senare år fått en tydligare och mer framträdande roll vid universitet och högskolor. Under senare år har också ett allt större fokus riktats mot utbildningars resultat. Frågor som rör kvaliteten i undervisningen blir därmed centrala. Att utveckla och stödja lärares pedagogiska skicklighet samt belysa villkoren för den undervisning som bedrivs inom högre utbildning är angeläget. Behovet av en gemensam samlingspunkt för pedagogiska och didaktiska diskussioner där lärare och andra yrkeskategorier som är intresserade av av ämnet på BTH, Högskolan i Kristianstads och Linnéuniversitetet kan träffas och föra dessa diskussioner tillsammans och över ämnesgränser är stort. Genom att anordna en årlig högskolepedagogisk konferens vill vi främja ett utbyte av erfarenheter från den dagliga undervisningen och insikter kring det lärande som möjliggörs.

Lärarlärdom 2017 gick av stapeln på Blekinge Tekniska Högskola, campus Karlshamn den 16 augusti och samlade ett femtiotal deltagare.

### Huvudtalare

#### Det akademiska lärarskapet som del av den akademiska professionen

*Åsa Lindberg – Sand:* docent vid institutionen för utbildningsvetenskap och avdelningen för högskolepedagogisk utveckling vid Luns universitet.

Trots att alla anställningar i högre utbildning som kräver vetenskapligt kompetens benämns ”lärare” vet alla som arbetar inom området att det som oftast väger tyngst vid befordran är de vetenskapliga meriterna. Lärarbenämningen kan därför uppfattas som lite urholkad inifrån. Men eftersom undervisningens kvalitet är helt beroende av skickliga lärarinsatser och god pedagogisk ledning har det under de senaste decennierna vuxit fram olika kompenserande system för att stödja utveckling av pedagogisk skicklighet på andra sätt: särskild pedagogisk meritering, pedagogiska akademier, konferenser och publikationsmöjligheter. Efter Boyers inspel i början av 1990-talet har merparten av dessa initiativ hämtat sitt perspektiv från en av de fyra delar av hela den akademiska professionen som han pekade ut: The scholarship of teaching (and learning, som lagts till senare – SoTL). Scholarship är omöjligt att översätta, så på svenska föreslogs t.ex. både Lärarlärdom och Akademiskt lärarskap. I min presentation tänker jag resonera kring samspelet mellan det som kan inrymmas i akademiskt lärarskap och vad helheten i den akademiska professionen kräver, med en utgångspunkt i den SUHF-rapport om högskolepedagogisk utbildning som publiceras under våren. Fokus kommer att ligga på hur variationer i det akademiska lärarskapet samspelar med övriga krav i den akademiska professionen. Finns det en risk att vi avgränsar det akademiska lärarskapet för tydligt och för snävt?

### Presentationer

Några presentatörer valde att publicera sina artiklar i full text. Nedan följer abstracten till dessa. I följande kapitel kan den fulla texten läsas.

Under konferensen spelades ett urval av presentationerna in. Inspelningarna, samt alla abstract och presentationer finns på: www.bth.se/lararlardom

### A, B, C – U eller G? Vi får väl se! Om bedömning och betygssättning inom högre utbildning

*Åsa Lindberg, Linnéuniversitetet*

Assessment and examination in higher education takes place in a complex context where many perspectives affect; management, academic developers, students and not least teachers. That assessment strongly influences students' learning is well documented (Gibbs 1999; Marton 2005). The study's starting point is that a better understanding of how teachers really relate to this very important part of the teaching profession, create better prerequisites for working continuously to enhance the quality of higher education. The purpose of this study is to describe university teachers' perceptions on assessment and examination as a basis for creating conditions for the development of grading as being a strong influencing factor in student learning. The method is a combination of questionnaires and interviews, thus both qualitative and quantitative methods was used to collect data of both qualitative and quantitative nature, which was analyzed with both qualitative and quantitative methods. Results show the relationship between the conditions, criteria for what is regarded as assessment and examination of good quality, and the development of the same. The empirical results also identify diverse perceptions to a variety of aspects related to assessment and examination, but also educational development in general. The importance of knowledge transfer is repeated where discussions and explanations between colleagues and students, in order to create meaning and development is required.

### Are finance students over- or under confident – A study on the ability to predict grades

*Emil Numminen och Ola Olsson, Blekinge Tekniska Högskola*

Overconfidence is a cognitive bias that most people suffer from. A person suffers from overconfidence bias when his or her own subjective estimation of an ability is significantly higher than an objective estimation of the same ability. Previous research in pedagogy has established that students suffer from overconfidence when it comes to grade prediction in business and economics. A student suffering from overconfidence bias have a propensity to study less than required since the subjective estimation of comprehension of the subject is higher than it really is when measured objectively. The implication of overconfidence is thus that a student will not fulfill his or her own full potential of learning the subject. This paper adds to the overconfidence research in pedagogy by measuring the level of overconfidence throughout an entire course to analyze the relation between learning and overconfidence. This has not been done in previous research. Students made estimation of their final exam score at five times throughout the course. Results show that students are overconfident and that they do not calibrate their expectations over time, on a general level. as they perhaps should given how they perform in learning the subject. Female students show a lower degree of overconfidence and had a higher tendency to calibrate their expectations. After having taken the exam and making a final estimation of expected grade, overconfidence drastically went down but less so for fourth year students in relation to third year students. In this estimation female third year students even became under-confident.

### Analyzing the Impact of Differences in Academic Cultures on the Learning Experiences of Overseas Master's Students

*Javier Gonzales-Huerta, Blekinge Tekniska Högskola*

*Problem*: As teachers at Blekinge Institute of Technology (BTH), we sometimes observe that students from overseas partner universities experience difficulties in participating with learning and assessment activities on the courses we teach. We hypothesise that a significant cause is the difference in academic cultures between the students' home universities and BTH.
*Intended Outcomes*: Our objective is to understand the challenges and barriers to effective learning

faced by overseas students from partner universities as a result of differences in academic culture.
*Method*: The context of the research work is overseas students from partner universities in China and India taking the Master's in Software Engineering program at BTH. The study was conducted as a series of focus group interviews with students enrolled in this program. The resulting discussions were analysed using constructive grounded theory to identify the main challenges and barriers experienced by the students during their first months at BTH.
*Relevance*: Guidance, supported by empirical evidence, as to the nature of the differences which would help our program in developing resources to help students and teachers to be better prepared to accommodate differences in academic culture at course, program, and school-levels.

#### Assessing Knowledge Through Written Reviews of First-Year Programming Students

*Emil Folino, Blekinge Tekniska Högskola*

In this paper, a method of qualitative assessment of programming students' knowledge and comprehension is investigated. The qualitative assessment is done by reading students' review texts from three subsequent courses' individual programming project. The review texts are analyzed according to the SOLO Taxonomy and the students are awarded a SOLO level of Unistructural, Multistructural or Relational. The SOLO level is compared to the final grade of the three courses and a relation between a student's final grade and the SOLO level is shown. Furthermore, a positive progression in the students' comprehension and understanding of the course material is observed as the students progress through the three subsequent courses. A recommendation is given to complement programming exercises with written assignments where the students get an opportunity to reflect and expand on the completed exercises.

1

# A, B, C – Fail or Pass? We will see! On assessment and examination in higher education

**Åsa Lindström**
***School of Business and Economics, Linnaeus University***

asa.lindstrom@lnu.se

#### Abstract

Assessment and examination in higher education takes place in a complex context where many perspectives affect; management, academic developers, students and not least teachers. That assessment strongly influences students' learning is well documented (Gibbs 1999; Marton 2005). The study's starting point is that a better understanding of how teachers really relate to this very important part of the teaching profession, create better prerequisites for working continuously to enhance the quality of higher education. The purpose of this study is to describe university teachers' perceptions on assessment and examination as a basis for creating conditions for the development of grading as being a strong influencing factor in student learning. The method is a combination of questionnaires and interviews, thus both qualitative and quantitative methods was used to collect data of both qualitative and quantitative nature, which was analyzed with both qualitative and quantitative methods. Results show the relationship between the conditions, criteria for what is regarded as assessment and examination of good quality, and the development of the same. The empirical results also identify diverse perceptions to a variety of aspects related to assessment and examination, but also educational development in general. The importance of knowledge transfer is repeated where discussions and explanations between colleagues and students, in order to create meaning and development is required.

*Key words: academic development, assessment, assessment culture, examination, grading, knowledge transfer*

#### Introduction

The everyday context of teachers, at all levels of education, include assessment, and for students it involves being assessed (Marton 2005). Moreover, he states that assessment, and the grades it result in, is the most important form of influencing the way people learn in general. In other words, young pupils, as well as students of higher education, predominantly learn what they expect to be graded upon. Schools are basically evaluative settings, Lundahl (2006) agrees, and continues stating that it is not only what you do there, but what others think of what you do that is important. Teachers engage in both formative and summative assessment of multiple forms (Taras, 2008). The final result, not rarely, is deciding a grade. In the best of worlds, a grade that in a just way reflects the student's performance. A key assumption in this paper is to acknowledge the role of the teacher to promote student learning. Hence, the focus is on the teacher's perspective on grading, with the intention that an understanding of the university teachers' perceptions can better shape education, including assessment and grading, which supports student learning. This study has posed questions such as; what is their opinion on different forms and scales of grading? And what about different categories

of teachers; are more experiences teachers reluctant to change, more so than their younger colleagues? In the context of grading, the criteria used is an operative factor (Sadler 2005). Therefore, one might wonder if teachers agree and in what way they use criteria and set grades. Understanding of teachers' views of grading is a necessary prerequisite in order to be able to improve the quality if learning as being one of the great challenges for institutions of higher education worldwide (Boud et al. 2010).

The purpose of this study is to describe university teachers' perceptions on assessment and examination as a basis for creating conditions for the development of grading as being a strong influencing factor in student learning.

#### Theoretical framework

If you should choose to single out one measure to improve education quality and enhance student learning, you should consider the assessment system, where grading is a significant component. Rowntree (1987) states that if we wish to discover the truth about an educational system, we must look into its assessment procedures. He further explains the dynamics of assessment and the relation to grading, describing grades being based on assessment of different kind; a result. Sadler (2005) defines that *assessment* refers to the process of forming a judgment about the quality and extent of student achievement or performance, and therefore by inference a judgment about the learning that has taken place. *Grading*, in turn, refers to the evaluation of student achievement on a larger scale, either for a single major piece of work or for an entire course, subject, unit or module within a degree program. It is not a clear-cut distinction between assessment and grading as terms, hence making in challenging to elucidate. There is no doubt that assessment is of the outmost importance when it comes to influencing student learning, their behaviour and results (Gibbs 1999, Biggs 2003, Boud & Falchikov 2007, Harland 2015). The ability to manage the process of assessment and the act of grading, hence aiding students to develop strategies for handling different kinds of academic assignments, is the most significant task being a teacher of higher education (Gillett & Hammond 2009). However, assessment in general and grading in particular, is not undertaken is isolation, but as part of a wider context. Biggs (2010) refers to the correspondence between learning objectives, course content, structure, teaching and assessment, as *'aligned teaching'.* This principle is established as part of a European approach to quality, which includes concepts such as learning outcomes and clear student-centred perspective (Standards and Guidelines for Quality Assurance in the European Higher Education Area, 2015). The standard on student-centred learning, teaching and assessment explicitly underline the importance of assessment for the students' progression and their future careers.

Taras (2008) believes that it is central to everyone in higher education, not just specialists, to understand assessment on a deeper level; its terminology, processes and relationships between them. Her results, however, show the opposite that knowledge about assessment is often fragmented both theoretically and practically. The conclusion is to chart through which processes assessment is done, to be clear about what we do and how. It also enables evaluation of both how the students' learning is influenced and the teachers' assessment (Taras 2008). To covet a deeper understanding and consider assessment and grading in a broader context seems to improve quality (Dahlgren et al. 2009). The overall view of a system of components that strongly dependent on each other, rather than considering assessment and examination as individual phenomena, is illustrated:

> *"Examination in higher education as well as in all part of the education system is a highly interdependent system of grading, assessment tasks, judgement criteria, students' approaches to learning and features of the learning outcomes."* Dahlgren et al. (2009:193)

Aligning learning outcomes with assessment and specifying assessment criteria are frequent objectives where institutional learning and teaching strategies focus on assessment (Gibbs & Simpson 2002). However, they race an alert, underlining that it is not only about the measuring of student performance – it is about learning. With an international perspective, universities face

major changes in the near future (Boud et al. 2010). In the article *'Assessment 2020: Seven propositions for assessment reform in higher education'*, some 50 researchers and academic developers from e.g. Australia and the United Kingdom have made concrete proposals on how assessment in higher education can be reformed (Boud et al. 2010). Rust, Price and O'Donovan (2003) assume that there is increased demand for higher education assessment in a more transparent manner for all parties involved. This demonstrates better reliability in grading and satisfying increasing demands from reviewing agencies to prove what results are being created. Rust et al. (2003) focuses on how to increase understanding of the assessment process and its criteria among students. An important conclusion was the importance of both students and teachers expressing the need to discuss criteria to make its application comprehensible. If not, there is a risk of a *'Hidden Curriculum'* (Snyder 1971) where hidden knowledge, *'tacit knowledge'*, fill the gaps. Rust et al. (2003) warn of an exaggerated belief in *'explicit knowledge'*, the more explicitly expressed form, for example, in an assessment matrix. An understanding of both forms of knowledge of assessment must be developed by both teachers and students. Thus, it is not enough only the explicit formulation of assessment templates or criteria, but a socializing process is required where understanding is created and the explicit knowledge is transmitted (Rust et al. 2003, Sambell, McDowell & Montgomery 2013).

Dahlgren et al. (2009) emphasizes that assessment criteria in higher education are often problematic in formulating and not in itself leading to high quality or promoting student learning. Instead, it is in the discussion between teachers, and between teachers and students, that the criteria become meaningful and can be applied in a manner that gives the positive effects sought. Discussion in this context involves formulation, negotiation, application, rewording and critical reflection (Dahlgren et al. 2009). Even small changes in methods and tasks can give immense impact on student behavior and study results (Gibbs 1999).

> *"Students are tuned in to an extraordinary extent to the demands of the assessment system and even subtle changes to methods and tasks can produce changes in the quality and nature of the student effort and in the nature of the learning outcomes out of all proportion to the scale of the change in assessment."* Gibbs (1999:52)

Gibbs, Hakim and Jessop (2014) have found tremendous dissemination regarding assessment and application of assessment criteria. They emphasize the value of developing a common collegial assessment culture to promote student learning. When the students are constantly assessed based on very different premises, the feedback value becomes marginal and the progression through an education becomes unclear. There is evidence that subjects and corresponding academic groups create distinct assessment environments linked to traditions, rules, perceived demands or myths about how assessment is done (Gibbs, et al. 2014). They address the question of whether this affects students' learning, especially students who meet representatives of different disciplines in their education. They say, among other things, that there are significant variations in quality, quantity and timing of feedback in connection with assessment of examinations. This in turn, resulted in students perceiving feedback as unreliable and not useful for developing their abilities in future tasks or courses. An existing culture, or approach, is that assessment is an intuitive process that cannot be formulated explicitly, as illustrated by the following quote.

> *"These words uncannily echo the normative, 'connoisseur' model of assessment typified by the phrase 'I cannot describe it, but I know a good piece of work when I see it'. A model of assessment judgement most often likened to the skills of wine tasting or tea-blending, and 'pretty much impenetrable to the non-cognoscenti'."* Webster et al. se O'Donovan et al. (2004:328)

Criteria-based strategies for assessment and grading in higher education have increased in use as a consequence of good theoretical motivation and its educational effectiveness (Sadler 2005, O'Donovan et al. 2004). However, there is no consensus on what criteria-based really means. Sadler (2005) has compared 65 universities in Sweden, UK and Australia, all of which have grading policies that claim to be criteria-based. Although those type of policies has a broad desirability, there are dif-

ferent conceptions amongst higher education institutions of what it means in theory and in practice, according to Sadler. To pin-point a common denominator, criteria as a term is explained as attributes or rules that are useful as levers for making judgments. Sadler (2005) also adds that judgments can be made either analytically, built up progressively using criteria, or holistically without using explicit criteria. Furthermore, four different grading models and their connection with criteria are identified; 1) Achievement of course objectives; 2) Overall achievement as measured by score totals; 3) Grades reflecting patterns of achievement; 4) Specified qualitative criteria or attributes. These different models of grading can come on a collision course with each other. For instance, the third model ranks poorly if grading criteria prescribe the second model. There are many situations in which a rule for adding sub-points gives obvious discrepancies in terms of qualities that appear to be contrary to the university's best assessment of a student's achievement. A weak development of a student in certain areas can be compensated by superior performance elsewhere. When a total sum is calculated, patterns of strengths and weaknesses may be lost (Sadler 2005). The fourth type of grading model has grown in use in recent decades. The conclusion that specified qualitative criteria or attributes increased in use summarize challenges; to examine how criteria and standards can be contextualized, to negotiate and reach a consensus on appropriate levels of standards, to allow students to make assessments using standards, and finally to apply the criteria consistently when grading (Sadler 2005).

Harland et al. (2015) has studied university teachers' experiences and the results showed that, among other things, teachers had no idea how many examinations the students were exposed to, and there was weak communication between teachers, departments and programs. Assessment can be regarded as a very difficult task. Barnett (2007) problematizes assessment in higher education and asks whether it is even an impossible task. He believes that in our complex and constantly changing contemporary setting, it is almost a mission impossible to assess knowledge, understanding and skills at the level of higher education. Reimann and Wilson (2012) argue that in order to improve student learning, teachers' perceptions of teaching and assessment must change. Certainly, perceptions may differ from actual behavior, but Reimann and Wilson (2012) still point to perceptions as a prerequisite for achieving change. This, in turn, motivates the focus of this study teachers' perceptions. The research question was formulated as follows; how do university teachers perceive assessment and examination in higher education?

## Methods

Prior to this study, many perceptions were observed among colleagues who were quantitative in nature, for example; *'Most teachers think grading criteria are meaningless and impossible to formulate.'* Was that really the case? These observations led to the selection of a quantitative first part of the study; a survey. The population consisted of all teachers at Linnaeus University, located in southern Sweden. The education and research at Linnaeus University's is conducted within a wide range of subjects within the Faculty of Engineering, Business and Economics, Humanities and Art, Health and Life Sciences and Teacher Education (Linnaeus University 2015). The data collection generated 202 responses with good dissemination between the five faculties. The results were distributed with a spread of 13% response from the Technical Faculty, to 23% at the School of Business and Economics. The data types that were collected were of quantifiable type, e.g. number of active service years in higher education, as well as qualitative type, e.g. questions about perceptions. Regarding how long respondents have been working as teachers in higher education; the results were categorized into four segments; 0-5 years, 24%; 5-10 years, 21%; 10-15 years, 23%; > 15 years, 33%. The distribution shows a good spread between different scientific disciplines, as well as between more experienced teachers and those who are newer in their role as university teachers.

The results of the survey identified interesting observations, which were sought explanation and understanding of through a second part of the study; qualitative interviews with a targeted selection (Bryman 2011). The ambition was that the chosen respondents could contribute interesting explanations and interpretations of the survey results. The distinction between qualitative and quantitative methods does not need to discriminate in perspective or approach. Qualitative perspectives

are not synonymous with qualitative methods. It is thus fully possible to apply quantitative methodology in the context of qualitative perspectives. Data compilation and analysis was carried out using descriptive statistics, significant tests through X2 analysis, and thematization depicted by a constructed framework matrix (Richie et al. 2003, in Bryman 2011). To categorize the results on how grades usually are decided on a specific examination, the grading models of Sadler (2005) was used.

The content of this paper was presented at the pedagogical conference *Lärarlärdom* in August 2017 at Blekinge Institute of Technology, Sweden, and constitutes a selection of results from a larger study where a more thorough description of methods can be found (Lindström 2016).

## Results

University teachers' perceptions on assessment and examination, resulting in setting grades, was examined through questions concerning; 1) forms of assessment, 2) methods of grading, and 3) scales of grading. Firstly, the study shows the most frequently used forms of assessment on the courses they normally teach. The intention of the question was to give a picture of how the teachers work with different forms of assessment, without locking them into predefined alternatives or entering a certain number of possible forms of assessment, which justifies the open character. In data processing, the answers have been categorized into four types: written exam, various forms of written reports, oral presentations/seminars, and practical forms such as laboratory work, clinical examination and design. How common these different types are is illustrated in the table below.

| | |
|---|---|
| Written exams | 39% |
| Written reports | 33% |
| Oral presentations/seminars | 19% |
| Practical forms | 9% |

*Table 1. Most commonly used forms of assessment*

Combination response that indicated more than one form was frequent on this question. The most common form of use as the sole form of assessment was the written exam, as indicated by 46% of respondents. Thus, there are just over half that use written exams in combination with any other form of assessment. The second most common form is written reports, which includes written work that may be termed PM, paper, essay or the like, is used in combination with other forms in 73% of the answers given. The form that is combined to the highest extent was oral examinations; 94%. While practical examinations are combined in 85% of cases.

The results reflect the teachers' perception of the most common forms of examination, which is not equivalent to the actual forms. However, the purpose of the question was, in addition to giving this image, to serve as an input to the following question of any desire to change them. The tendency to change methods of grading was depicted by inquiring *'Would you like to change the forms of examination you mostly use today if it was practically and in terms of resources possible?'* This hypothetical question showed that 69% would *not* like to change their examinations. The remaining 31% answered the supplementary question as to what they would like to change. This resulted in a highly diversified response, as shown by the following diagram.

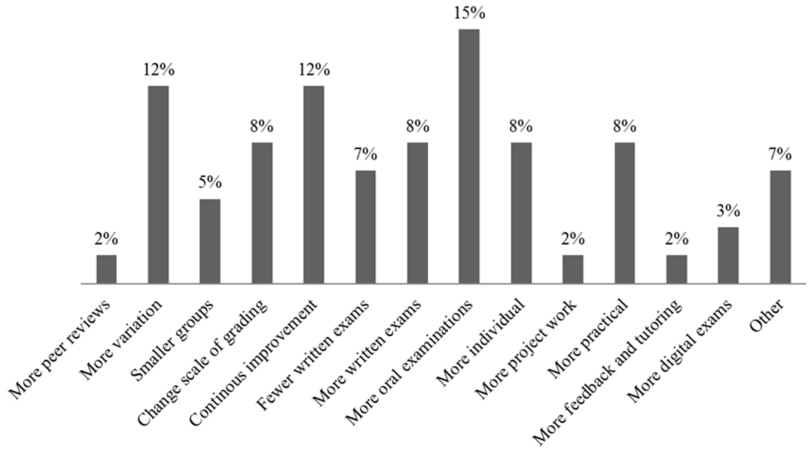


*Diagram 1. Desired changes of forms assessment*

The highest frequency was given to requests for more oral examinations; 15%. It is supported by respondent views on how to make the most effective examination. The next size largest group consists of requests for more variety in the different forms of examination used, and the desire to work more with continuous improvement; 12% on each proposal. One respondent argues: *"Most of all, that change is good, both to elevate different skills of the students, but also that I, as a teacher, will be developed and inspired."* Variation was also highlighted in several comments, for example: *"A variation in forms of assessment is preferable for both students and teachers. A variation favors the ability to reflect and learn in different ways."* In addition, there were comments that included more oral elements, individual examinations to a greater extent, and more practical elements. It can also be noted that about as many, 7% and 8%, respectively, want *fewer* written exams versus *more* written exams. It can be noted that all examinations does not serve the same purpose and balance the often massive criticism of written exams. It's not as simple as saying that written exams are always bad. The category *"Other"* contains, for example, comments such as the forms of examination can easily be changed with unchanged resource allocation, and suggestions like: *"More 'open' exams where I grade students ability to argue for their ideas, in an on-going conversation."*

The number of active years as a teacher in higher education had no significance for change propensity. The result showed very little dispersion; between 67-71% for *No* and 29-33% for *Yes*, and no significance for the differences with an X2 of only 0.11. Among the scarce thirds who indicated that they would like to change the types of assessment that are mostly used today if it was practical and resource-friendly, there was the opportunity to comment. The categorization above of these answers shows a diversified image. However, it is not only what answers are given that are interesting, but also what is not commented at all, or to a very low degree. For example, there was a very low degree of comment regarding student engagement; 2%, formulated in an answer: *"Have a higher degree of student assessment; both of their own achievements and of other students."* Compilation of interviews also support the presence of different models of grading, as illustrated by the following quotation from one respondent:

> *"Here we put numbers on things. If you have 60%, you have actually passed. Other departments have their culture. Comparing the student's overall performance throughout the course, including seminars, discussions, reflections and submitted exam papers and possibly practical examinations, which makes it extremely difficult to put numbers. These are quite different cultures!"*

The argument posed suggests that different forms of assessment involve different degrees of subjectivity, but that numbers are subjective too. There are definitely different cultures at different departments and it is about adapting, is a frequent view. One respondent describes a similar type of differences within departments and subjects. They distinguish between more technically oriented

and business-oriented courses within the same subject, which have completely different assessment practices.

Secondly, to learn the methods of grading used, how the respondents usually decide which grade a student gets on a specific examination (regardless of scale), was investigated. When the university teachers in the sample of this study commonly set grades the complied results show that specified qualitative criteria or attributes dominate; 55% of respondents were placed in the fourth category using Sadler (2005). Even though the question was designed with open-ended response, the answers were easily categorized based on answers like: *"Using given criteria compiled with colleagues according to the requirements"*. The table below shows the overall division.

| | |
|---|---|
| 1. Achievement of course objectives | 20% |
| 2. Overall achievement as measured by score totals | 12% |
| 3. Grades reflecting patterns of achievement | 13% |
| 4. Specified qualitative criteria or attributes | 55% |

*Table 2. Grading Models (Sadler 2005)*

One fifth indicates that the course objectives, category 1, determines the grade a student receives, while just over ten percent gives answers categorized within 2. Overall achievement measured by total score, as illustrated by answers as *"According to a percentage of marks on the written exam."* Category 3. The grade reflects patterns of achievement, is approximately the same extent and can be exemplified with explanations like *"I set grades based on knowledge learning; i.e. what has the student understood and how can the student use the knowledge."*

Thirdly, perceptions on scales of grading was investigated; for two reasons. Due to the fact that Linnaeus University, where data was collected, recently changed scales of grading, but even so still uses four different scales across the university. This implies that different teachers must adapt to several scales, as must many students. The second motive was present critique against moving from a scale with few grading levels to a more finely divided one, such as is the case of Linnaeus University. The argument being that more grade levels risk controlling students towards more shallow learning. A current question regarding grading concerns the grading scale used; a scale with few levels, such as Pass/Fail only, or a multi-level scale such as A-F. Finding out what experience is available from different scales enables analysis of different perceptions based on experience. Thus, a survey question was posed: *"The Rector has decided that a seven-point grading scale, A-F, is introduced at Linnaeus University from the autumn term 2015. The A-F scale will be used in all courses where international students can attend. What best describes you?"*

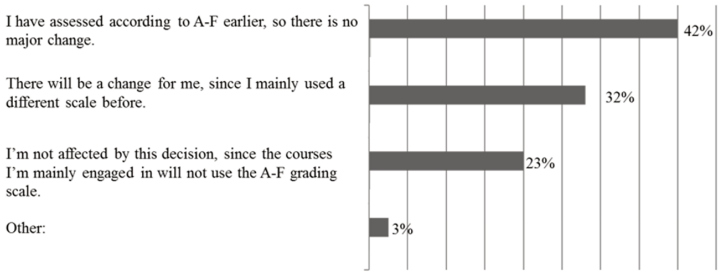


*Diagram 2. Experience of assessment according to A-F grading scale*

The majority of the respondents, 42%, have experience of assessing according to A-F. A relatively large proportion, 32%, has mainly used a different scale and therefore will have to change the way they assess and grade. Among the few who stated *'Other',* and used the free text option, there were

some comments that all courses should have A-F, but also those that articulate risks. One comment describes the risk of the level of approval being lowered: *"I have done both, evaluated, and saw that there is a need for discussions concerning the risk of A-F lowering what is regarded to be a pass grade."* Another comment highlights the time perspective: *"It will require more time by both teachers and students before the scale will provide measurable difference, is my spontaneous insight."* Perceptions differ and some argue that a multi-level scale does not have any positive effects, while others describe: *"A-F, a scale that gives both the student and the teacher a good educational tool in their learning and understanding of their own and the others learning processes."*

In order to investigate the perception of switching to A-F is predominantly positive or negative, the question was asked; *"Whether switching to the A-F scale affects you or not – what is your opinion?"* The result showed that a majority, 44%, found it positive, and 31% indicated that it was predominantly negative. 25% had no opinion. A comparison of the answers to the outcome of this question and the number of active years as a teacher in higher education, showed that the number of years of service did not significantly affect positive or negative perception, or no opinion. A weak correlation could be identified between the 0-5 year of service with 29% responding to *"no opinion"*, while 21% in the group >15 years had no opinion. In other respects, the result showed great coherence in this regard. When clustering the two groups with the shortest time as active teachers in higher education and the two with the longest experience, it also showed a weak connection. In the 0-10 year group, 38% were positive and in the group >10 years, 48% were predominantly positive.

The result showed stronger correlation with a comparison of predominant positive versus negative perception, compared with previous experience of assessing A-F. The most positive group were those who stated that they assessed according to A-F earlier and therefore did not experience any major change. In this group, 68% were predominantly positive, 13% predominantly negative, while 19% responded no opinion. This difference is significant at the 0.1% level ($X2 = 20.7$) compared with those who used a different scale earlier and therefore experience a change. The analysis thus showed that the experience of seven-grade assessment creates a more positive perception. The qualitative results also provide confirmation of the relevance of experience, for instance, one respondent testifies on how to develop their assessment ability in line with the number of active years.

To summarize; a profound consensus was identified regarding the significance of assessment and grading, as illustrated by comments such as: *"Important, as it controls students' learning!"*; *"We need to get better at it."* And; *"There's a need for many discussions about this, to ensure just assessment of good quality".*

## Discussion

The result shows a diverse picture of perceptions; which types of assessment that are preferred, what grading scale or what actions that should be prioritized. Even current assessment cultures showed rich variations. Patterns that can be identified are that assessment questions are perceived as important and there is a demand from teachers for recognition of educational work in general. The perception is that teachers who engage in student learning and take time to develop, for example, assessment forms and grading are a bit strange; *"It's not illegal to be interested in development of teaching and assessment, but it's a bit odd"*, is a comment that has been left. Clear signals from management are required; not only in words but in action. Empirical results and theory show a large dispersion regarding assessment and application of grading criteria, which gives reasons for forming common collegial assessment cultures. A department or academic discipline is part of an organization, that is also characterized by a culture; an organizational culture, which the assessment culture is a part of.

Confidence and commitment are words that are repeated both empirically and theoretically. Management is one variable, but the teachers' attitudes and actions are also culture-creating. Confidence enables sharing of knowledge and experience, as well as creating the discussion described in the theoretical framework regarding assessment. For instance, explicit grading criteria in courses are not a quality-creating universal solution as such, but the *process* between teachers and students to create, implement and revise, is promoting student learning and quality-driven education. Higher education consists of a great number of academic disciplines and degrees. Respect for diversity is required; that different subjects have different assessment practices. Nonetheless, common

denominators, such as assessment, clearly guide students' learning, and that its design needs to be noted. One question that engages is the seven-grade scale A-F and whether its application is predominantly positive or negative. In theory there is criticism; so even among the respondents. However, the results showed that the majority thought it to be predominantly positive, a proportion that increased significantly with experience. The manner in which grades are usually set, has been categorized and shows a clear overweight for specified qualitative criteria, which is consistent with trends described in the theoretical framework. Assessment criteria together with the requirement to achieve course goals represented ¾ of respondents' responses and thus dominate. That finding also reinforces arguments for working consciously and systematically throughout the assessment process.

Transparency, in the sense of clarifying to all parties involved, how assessment and grading is conducted, is imperative. Theory strengthens both this result and demonstrates a tendency where demands for increased transparency can be observed both nationally and internationally. There are perceptions present, that one would rather want to proceed by themselves and do as they please, without interference and definitely not be subjected to scrutinizing eyes of any kind. The dominant perception though, is that transparency is desirable and something we will have to get accustomed to, if we have not already done so. Conversely, seeking transparency is not unproblematic, since the use of concepts and international comparability can create confusion. Transparency is not just a requirement for clear processes, but also clear conceptual use. For instance, it may counteract complications with international comparability, as well as misperception concerning application of terms and concepts, if we can demonstrate clearly both the assessment process and its criteria. A conclusion is to emphasize the benefit of sharing and valuing colleagues' experiences.

Propensity to change was identified as a category of quality in the empirical thematization. The result recognized which forms of examination that are most common in occurrence. On the basis of that, it was found that most, more than two thirds, would *not* like to change their forms of examination, even if it was practically and in terms of resources possible. For this cluster, it can therefore be ruled out, that lack of time and resources is considered a barrier to changing forms of examination. Among the scarce third that would like to change, the findings show a wide spread of desired changes, such as more oral examinations and greater variety. The question was hypothetically posed, and how many respondents think they actually have sufficient resources and other conditions to fulfill their wishes, does not appear. Even though, the results show that the number of active years as a teacher did not have any significance for propensity to change. This conclusion contradicts some of the respondents' perceptions. It does not therefore have to be the case, that teachers with more active years become comfortable and not willing to change, instead experience can create a momentum and ability to see opportunities that less experienced colleagues do not acknowledge. The reflection underlines the added benefit of getting acquainted with colleagues' understanding of the various aspects of assessment.

Knowledge transfer is a recurring quality criterion both empirically and theoretically. Discussions and explanations are required between colleagues, and with students, in order to create meaning and development. It is also intimately associated with the requirement that grades be argued for in agreement with set goals and criteria. To realize this necessary knowledge transfer, also requires that conditions in terms of time and resources exist. Since assessment is not performed in isolation, an integrated approach is required that also involves goals, teaching methods and feedback. Combining explicitly formulated assessment criteria with the socialization process of the more implicit parts, need not be that resource-intensive. The main insight is that this is actually needed at all, as well as conversations with colleagues and students.

*"How do university teachers perceive assessment and examination in higher education?"* was the research question posed, with the purpose of describing the very same and thus creating conditions for the development of grading as being a strong influencing factor in student learning. With a comprehensive interpretation, one might not say that this was fulfilled. Still, within the limits of the study, a contribution has been given, which may hopefully be inspiring, and perhaps even influential. Given the importance of assessment for student learning, it is a prerequisite for achieving good quality, or

even excellence, to work consciously and strategically with educational development in general and assessment in particular.

## Concluding remarks

Teachers' perceptions exist in a context of conditions that cannot be ignored. These are the basis of criteria for what is considered to be good quality and what is possible to implement. Institutions of higher education must balance commitment and prerequisites. In general, what you want to achieve, the driving force, motivation or even passion, to actually work with and against the ambitions of both management and teachers. Those who show dedication want to experience recognition for what they achieve and that their experience and knowledge are given value. This goes for both teaching and research, which in an ironic sense, often are perceived as conflicting activities. One argument that could possibly help to reduce the gap between research and educational activity is the view that it is basically one and the same. These are both processes where knowledge will be problematized, measured and generated. Certainly, epistemological assumptions may differ, but if such a supposition is rooted, it could inspire a new way of considering assessment questions.

The creation of academic development in general and assessment in particular is based on the criteria identified. One reflection is whose interpretation and prioritization of criteria that take precedence; teacher, student, management, administration or maybe any other stakeholders such as external auditors or professional academic developers. Illuminating possible different perceptions makes it possible to ascertain common basic criteria as a starting point in development work. Is it a goal to create an examination that demands students to spend 40 hours a week on studies? Or is it most important to distinguish students by differentiated grades? What will create the desired outcome? Academic development can have many goals and be achieved in many ways, hence a generous approach is preferable with a shared view regarding what actually are desired outcomes.

## References

Barnett, R. (2007). "Assessment in Higher Education: An Impossible Mission?" In *Rethinking Assessment in Higher Education: Learning for the Longer Term*, edited by D. Boud and N. Falchikov, pp. 29-40. London: Routledge.

Biggs. J. (2003). *Teaching for Quality Learning at University – What the Student Does.* Buckingham: Open University Press.

Biggs, J. (2010). *Aligning teaching for constructing learning*. The higher education academy. Available at: http://www.bangor.ac.uk/adu/the_scheme/documents/Biggs.pdf. [2013-02-05].

Boud, D. and Associates. (2010). *Assessment 2020: Seven propositions for assessment reform in higher education*. Australian Learning and Teaching Council. Available at: https://www.uts.edu.au/sites/default/files/Assessment-2020_propositions_final.pdf. [2015-09-23].

Boud, D. & Falchikov, N. (2007). *Rethinking Assessment in Higher Education: Learning for the Longer Term*. London: Routledge.

Bryman, A. (2011). *Social Research Methods*. Malmö: Liber AB.

Dahlgren, L-O., Fejes, A., Abrandt-Dahlgren, M. and Trowald, N. (2009). Grading systems, features of assessment and students approaches to learning. Teaching in Higher Education. Vol. 14:2, pp. 185-194.

Gibbs, G. (1999). *Using assessment strategically to change the way students learn*. I Brown, S. & Glasner, A. (red.). Assessment matters in higher education: Choosing and Using Diverse Approaches. Buchinghamshire: SRHE and Open University Press.

Gibbs, G., Hakim, Y., Jessop, T. (2014). The whole is greater than the sum of its parts: a large-scale study of students' learning in response to different programme assessment patterns. *Assessment & Evaluation in Higher Education*. Vol. 39, no. 1, pp. 73–88.

Gillett, A & Hammond, A. (2009). Mapping the maze of assessment: An investigation into practice. *Active Learning in Higher Education.* Vol. 10:2, pp. 120-137.

Harland, T., McLean, A., Wass, R., Miller, E. & Nui Sim, K. (2015). An assessment arms race and its fallout: high-stakes grading and the case for slow scholarship. *Assessment & Evaluation in Higher Education*. Vol. 40:4, pp. 528-541.

Lindström, Å. (2016). *A, B, C – U eller G? Vi får väl se!: Om bedömning och examination inom högre utbildning*. Available at: http://lnu.diva-portal.org/

Linnaeus University. (2015). *Linnéuniversitetets organisation*. Available at: http://lnu.se/om-lnu/organisation. [2015-09-03].

Lundahl, C. (2006). *Viljan att veta vad andra vet: kunskapsbedömning i tidigmodern, modern och senmodern skola.* Diss., Uppsala universitet.

Marton, F. (2005). *Inlärning och omvärldsuppfattning: en bok om den studerande människan*. Stockholm: Norstedts akademiska förlag.

O'Donovan, B., Price, M. & Rust, C. (2004). Know what I mean? Enhancing student understanding of assessment standards and criteria. *Teaching in Higher Education*. Vol. 9:3, pp. 325-335.

Reimann, N & Wilson, A. (2012). Academic development in 'assessment for learning': the value of a concept and communities of assessment practice. *International Journal for Academic Development*. Vol. 17:1, pp. 71-83.

Rowntree, D. (1987). *Assessing Students: How Shall We Know Them?* London: Penguin.

Rust, C., Price, M. & O'Donovan, B. (2003). Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*. Vol. 28:2, pp. 147–164.

Sadler, R.D. (2005). Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*. Vol. 30:2, pp. 175-194.

Sambell, K., McDowell, L. & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.

Snyder, B.R. (1971). *The Hidden Curriculum*. Cambridge, MA: MIT Press.

Standards and Guidelines for Quality Assurance in the European Higher Education Area (ESG). (2015). Brussels, Belgium. Available at: http://www.enqa.eu/wp-content/uploads/2015/11/ESG_2015.pdf. [2017-01-23].

Taras, M. (2008). Summative and formative assessment: Perceptions and realities. *Active Learning in Higher Education.* Vol. 9:2, pp. 172-192.

2

# Analysing the Impact of Differences in Academic Cultures on the Learning Experiences of Overseas Students

Javier González-Huerta, Simon Poulding
*Blekinge Institute of Technology*

{javier.gonzalez.huerta, simon.poulding}@bth.se

**Abstract**

*Problem:* As teachers at BTH, we observe that students from overseas partner universities can experience difficulties in participating with learning and assessment activities. We hypothesise that one cause is the difference in academic cultures between the students' home universities and BTH.

*Outcomes:* Our objective is to understand the challenges and barriers to effective learning faced by overseas students from partner universities as a result of differences in academic culture.

*Relevance***:** Understanding the impact of differences in academic culture, both positive and negative, will assist students and teachers to be better prepared to accommodate these difference at course, programme, and institution levels.

*Context:* The context of the study is overseas students from partner universities in China and India taking the MSc in Software Engineering programme at BTH.

## Introduction

Many universities actively recruit relatively large number of international students to their study programmes, but courses in which the participants come from different academic cultures can create a number of challenges for both teachers and students. Several studies have identified the issues affecting these international students: prominent examples include a series of studies in the 2000s that focused on students from East Asian countries who study on programmes in Australia [1]–[3].

The majority of these studies acknowledge that the problems encountered are not just simply because of general cultural differences, e.g. between Confucian Heritage and Western cultures but because of the difference in academic cultures.

Swedish universities have also acknowledged these type of issues and some — such as Kristianstad University [4] — have published guidelines to assist both teachers and students.

At BTH, the MSc in Software Engineering is an example of such a programme: participants include students from partner universities in China and India; students from Sweden; students from Europe on Erasmus programmes; and a small number of students from other countries. Our experience as teachers and examiners on this programme is that students from partner universities in China and India appear to experience more difficulty in participating in some learning activities and assessments than, for example, students from Sweden and other European countries. Our desire to understand the reasons for these difficulties is the motivation for this study.

Our hypothesis is that a significant cause of the difficulties experienced by these students are differences in the academic cultures between BTH (and in general any Swedish universities) and the

partner universities the students attend before taking the programme at BTH, and it is this hypothesis we wish to explore in this study

We note that the context at BTH is different from many existing studies. The students from Chinese and Indian partner universities form the vast majority (sometimes over 80\%) of participants in many of the course in the MSc in Software Engineering programme. In contrast, the proportion of overseas students in the studies mentioned above at Australian universities was 20-30%. These studies showed that the effect of differences in academic culture were mitigated relatively quickly – over a few months- through the knowledge gained from local students. We are concerned therefore that the opportunity from learning from Swedish and European students is diminished on the MSc at BTH and thus the impact of any differences in academic culture may remain longer into the programme.

#### Purpose

The main goal of this study is to analyse the problems faced by students coming to BTH from partner universities that limit the ability of students to engage in the MSc in Software Engineering programme. Specific areas of interest include:

- course content, e.g. are examples and case studies culturally relevant, do the students have the prerequisite knowledge;
- pedagogy, e.g. can the students effectively engage in teaching activities;
- assessment, e.g. the emphasis on demonstrating learning objectives.

Our intention is that the outcomes of this study are feedback to programme and course responsible on the MSc in Software Engineering programme, with the aim of giving us, as teachers, a better understanding of the challenges faced by this cohort of students, and enabling us to modify our teaching practice to mitigate these challenges.

#### Research Questions

We formalise this goal as the following research questions:

- *RQ1:* What are the differences in academic culture between BTH and partner universities in China and India that are experienced by students from these partner universities taking the MSc in Software Engineering programme at BTH?
- *RQ2:* How do students perceive the impact – either positively or negatively — of these differences on their learning at BTH?

### Research Method

#### Data Collection: Focus Group Interviews

The data was collected by interviewing MSc students using focus groups. There are three main reasons why focus groups were chosen instead of alternative methods such as interviews and surveys.

Firstly, the objective of this study is a broad exploration of challenges faced by overseas students at BTH. Although we, as researchers, will have prompting questions related to possible activities in which challenges could arise, a focus group offers the opportunity for the discussion to move into new areas that we had not initially considered. The same could occur in a semi-structured interview, but the interaction and discussion between participants in the focus group is likely to lead to greater breadth than would arise in an interview.

Secondly, we wish to collect input from as wide a number of students as possible. Within the

time constraints in which the study was carried out (i.e., cfthe last weeks of the last semester of the course), the use of focus groups enables us to gather data from a broader sample of students than would be possible using single participant interviews.

Thirdly, we believe that a focus group consisting of fellow students from similar backgrounds will be more comfortable for the participants than an interview setting. This is particularly important as both of us may be lecturers and/or examiners on courses that some participants have taken or will take. Our concern was that this relationship could inhibit discussion in general because of any perceived difference in status between teacher and student; and of some particular topics, such as quality of lecture materials or usefulness of feedback from examiners, that might involve us personally. By having such discussions in a group setting, our aim was to reduce the reluctance of the participants to offer their opinions.

#### *Participants*

The pool of potential participants in the focus group were drawn from the Master of Science in Software Engineering programme. This programme has a high proportion of students who take the programme as part of an agreement between BTH and partner universities in China and India. Both authors teach courses on this programme.

Our pool of participants consisted students who have been at BTH between six and eighteen months on this programme on the basis that they will have experience of adapting to the academic culture at BTH that is both relatively recent and extensive. Students from partner universities in China will have been at BTH since August 2016, while students from partner universities in India will have begun their studies at BTH in January 2016.

Within each focus group, the participants were recruited so that they came to BTH from the same partner university. The intention was to facilitate discussion amongst participants by ensuring they had a similar experience at their partner universities against which they could compare their experiences at BTH.

Given these constraints, the numbers of students from which we could potentially recruit participants, categorised by the programme and partner university, is shown in Table 1.

| Start Date at BTH | Partner University | Cohort Size | Group Id |
|---|---|---|---|
| Aug 2016 | *University of Science and Technology, Beijing* | *10* | *C1* |
| | Qingdao University | 11 | |
| | Zhejiang University of Technology | 7 | |
| Jan 2016 | *Jawaharlal Nehru Tech. Univ. Hyderabad* | *11* | *I1* |
| | Jawaharlal Nehru Tech. Univ. Kakinada | 6 | |

*Table 1. Cohorts and Potential Focus Group Participants*

Our original intention was to hold a series of focus groups, ideally one for each combination and university and programme. However, after initial analysis of the data collected from the first focus group, we decided to limit the number of focus groups for the purpose of this study to two[1]. We realised that we were unlikely to be able to complete the data collection and analysis of four focus groups before the semester finished. However, the quality and accessibility of the data from the first group was sufficiently good that we felt that we could obtain answers to our research questions using only two focus groups.

The two focus group from which data was collected for this study were:

- *Group C1:* 4 students from the University of Science and Technology, Beijing, China.
- *Group I1:* 5 students from the Jawaharlal Nehru Technological University, Hyderabad, India.

In addition, the focus group interview was piloted with 4 PhD students from the Department of

Software Engineering whose studies prior to their PhD took place in a non-Swedish academic system.

#### *Method*

*Participant Recruitment:* Participants were recruited via a short presentation at the end of one lecture; by personal emails from the researchers; and in the case of Group I1, by a reminder email from the programme coordinator responsible for the partner university. Participants were asked to confirm their attendance — and in the case of Group C1, their choice of two possible dates — using a Google Form.

*Introduction and Provision of Informed Consent:*

At the beginning of each focus group session, the interviewer explained:

- the general purpose of the focus group and how the data will be used;
- that the session would be video- and audio-recorded;
- that the recordings would be used only by us as researchers in order to transcribe the discussion and then destroyed;
- that transcript of the discussion would be available to only us as researchers for the purpose of analysis;
- the analysis presented in any report or publication would not make it possible to identify individuals who participated;
- that the transcript and any analysis would be shared with the participants;
- that participants would be free to withdraw their participation and consent to use their data at any time (including before, during, and any time after the focus group);
- participants were asked to speak in English wherever possible, but also permitted to briefly clarify, for example the interviewer's use of specific terminology, with their fellow participants in a different language should this be necessary;
- that the focus group was a voluntary activity that was entirely independent of their study programme;
- participants were encouraged to feel able to give critical feedback, both positive and negative, on courses on which the researchers had been their lecturers and/or examiners.

Participants were asked to sign a form to show that they were giving their informed consent to participate in the focus group and regarding our use of the data we would collect from them. This form is shown in Appendix A.

*Logistics:* Once consent had been given by the participants, one researcher began the video and audio[2] recording equipment, and monitored this equipment during the session, while the other researcher led the discussion with the participants. Participants were supplied with drinks (water, soda) and snacks during the session[3].

*Focus Group Discussion:* Each focus group discussion was timed to last approximately 60 minutes. During this time, the researcher leading the discussion would initiate new topics using a series of prompting questions. The questions used for Group C1 are listed in Appendix B. For the second group, I1, we extended the set of questions with new questions to prompt discussion of new topics that emerged in the discussion during first session. Specifically, we added questions regarding the teaching and learning, the communication patterns with teachers, how the students cope with the different roles in courses, the structure of the programme and the courses. The revised set of the questions is listed in Appendix D.

The questions were prioritised so that if there was insufficient time to cover all topics, the information that we were most interested in was collected first. However, in both focus groups, all the

topics were covered. In addition, some prompting word associations were prepared in case it was necessary to stimulate discussion (see Appendix C), but it was not necessary to use them. The researcher also followed and encouraged discussion around other relevant topics when they occurred in the conversation, returning to the prompting questions only once the discussion had been exhausted.

*Transcription:* After each focus group, the discussion was transcribed by the researchers. Each researcher independently transcribed approximately half of each video. The video (rather than audio) recording was used for this purpose since the sound quality was sufficiently good; it was easy to identify which participant was speaking; and, it enabled us to note in the transcript any relevant non-verbal communication.

#### *Ethical Considerations*

We identified the following ethical considerations in the context of data collection and strategies to address them:

*Subsequent Student Assessment:* Both researchers may be lecturers and/or examiners for some of the participants in the future. The ethical consideration is whether we will act differently to such students based on their participation in and their input to the focus group; and whether we may be privileging participants over non-participants by discussing (including revealing our opinions on) the nature of assessment at BTH. To some extent, the use of focus groups rather than interviews, is likely to minimize any subsequent bias towards individual students. In terms of our privileging participants over non-participating students, we believe the risk to be minor.

*Data Privacy:* The data we obtained (the video recording, audio recording, notes, transcripts etc.) is kept securely; we will not make it available to others (without subsequent consent of the participants); and it will be deleted when no longer required. Any report or other publication will not use the data in a manner that will enable individual participants to be identified. At any stage, the participant may withdraw from the experiment, and the participant's data will be deleted. We have made these policies clear to the participants, and ask them to acknowledge that are participating by signing the informed consent shown in Appendix A.

#### Data Analysis

Our original intention was to use *Grounded Theory* [5] to analyse the data collected from the focus groups. Grounded Theory is a research methodology that provides a systematic framework for conducting qualitative studies with a systematic, inductive and comparative basis, with the purpose of constructing theory [6], [7].

However, after an initial analysis of the data from the first focus group (C1) we re-considered our analysis approach and decided instead to apply *Thematic Analysis* following the guidelines by Braun and Clarke [8]. Thematic Analysis is a widely-used approach for identifying recurring patterns (termed *"themes"*) across qualitative data but without the need—in the words of Braun and Clarke—*"to fully subscribe to the theoretical commitments of a 'full-fat' grounded theory which requires analysis to be directed towards theory development"*. This argument in favour of thematic analysis is consistent with our context: to answer our research questions it is sufficient to identify and evaluate the effect of academic cultural differences, i.e. providing evidence in support of a theory rather than necessarily developing a novel theory from the data.

Braun and Clarke recommend six phases, which we applied as follows. The first three phases were applied after first focus group (C1) data before collecting data for second focus group (I1). After the second focus group, the first two phases were applied to the data collected from that group, before continuing with phases 3 to 6 on the combined data collected from both groups. The organization of the data collection and analysis is shown schematically in Figure 1.

- *Phase 1* In this phase the objective is to familiarise ourselves with the data. We achieved this through sharing the transcription of videos of the focus group, and then each researcher reading the entire transcript.
- *Phase 2* The objective of this phase is to generate and apply initial *"codes"* that label relevant aspects of the data. We re-read the transcript of the focus group and applied *"sticky-notes"* with codes (labels) to a printed copy in order to identify parts of the transcript that were relevant to our research questions.
- *Phase 3* In this phase the process of organising the coded data into broader "themes" begins. For the purposes of our study, we used a tree-like mind-map to organise the coded data derived from the first focus group, C1. After applying phases 1 and 2 to data from the second focus group that took place after the interim study, we supplemented this mind map with additional codes and themes arising from the second focus group.
- *Phase 4 and 5* During these phases, the themes are refined and more clearly defined. This was achieved through discussions involving both researchers, during which we also identified the need to represent the data visually as a more flexible graph-like *"thematic map"* in place of the tree-like mind-map so that themes and codes could be related to more than one other theme. This re-organisation also facilitated the consolidation of themes and the clearer identification of sub-themes directly related to the research questions grouped into broader topic-related main themes. We validated this refinement by cross-checking the sub-themes that emerged against our intuitive interpretation of the key topics that arose in the focus groups.
- *Phase 6* The final phase is the communication of the final thematic analysis — in this case, as Section Analysis below. We have followed the recommendation of Braun and Clarke to present the themes, as a thematic map, supported by selected examples from the data.

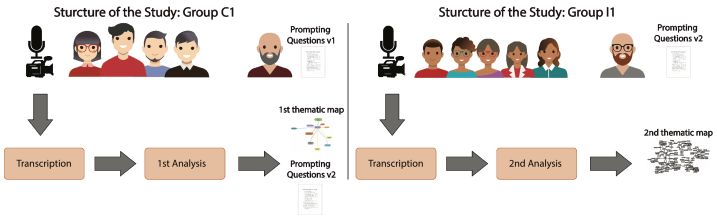


*Figure 1. Procedure followed for the Data Collection and Analysis*

### Results of the Thematic Analysis

In this section, we summarise the results of the thematic analysis of the data collected in the two focus group interviews. Since the whole thematic map is too complex to be visualized in the paper[4], in the sub-sections below, we identify cohesive clusters of themes for discussion. We will present the corresponding section of the thematic map in a readable form, and discuss the themes identified and the evidence in the data that supports these themes.

All the thematic maps shown in the following sections share the following legend, whose graphical syntax is described in Figure 2:

- Blue rounded-rectangles**:** main themes that are general pedagogical topic areas (*e.g.,* teaching and learning activities or assessment); the network of main themes may be thought of as a scaffold on which sub-themes are organised.
- *Green ellipses, yellow rectangles, red rectangles:* sub-themes that typically correspond to academic differences between BTH and partner universities. The colour and shape notation indicates the impact: green ellipses indicate positive impact, red rectangles negative, and yellow rectangles neutral impact (or where the impact is not clear from the data).
- *White "document" icons:* codes derived from the transcript of the focus groups. It is not typical to show the codes within a thematic map, but we do so here to provide evidence for both our identification of sub-themes and the positive or negative impact of them.
- *Labels C1 and I1:* indicate which focus group was the source of the data. Sub-themes that are derived from the data of only one focus group contain one of these labels as a prefix to the description of the sub-theme. Sub-themes without such a prefix are derived from the data collected from both focus groups.
- *Notations @BTH and @host:* within themes and code descriptions, these indicate the academic culture experienced by students when studying at BTH, and at the host (partner) university, respectively.

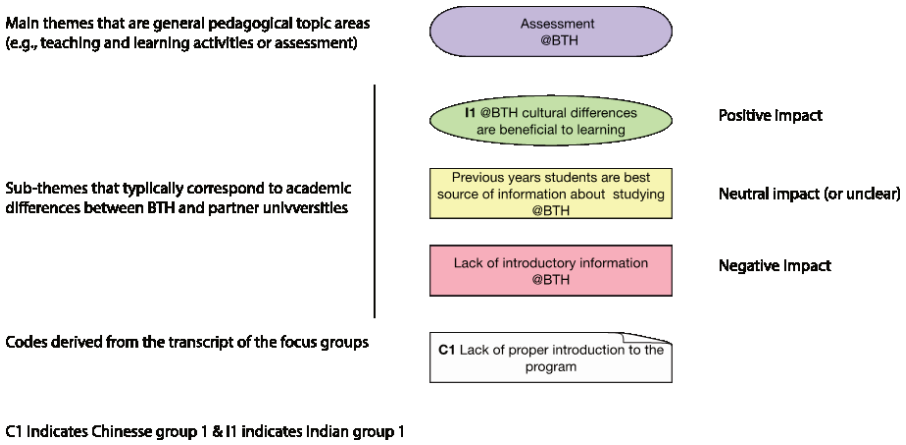


*Figure 2. Thematic Maps Graphical Syntax*

#### Teaching and Learning Activities

One of the biggest theme clusters that arose is the differences and challenges regarding teaching and learning activities. Figure 3 summarises the most relevant themes, sub-themes and codes discussed around this theme-cluster.

One of the main themes in this cluster was the switch to team work-centric courses, both for Chinese and Indian students. In their previous experiences as bachelor students in their home universities, students were more used to mainly work individually. When the students arrive at BTH, they have to first adapt to work in teams, but also understand and get used to be assessed for their performance as a team.

Another main theme was the report-writing skills, highlighted as a challenge both by Chinese and Indian students. Indian students discussed their lack of knowledge on how to write reports, and

how to adhere to a specific formatting templates (e.g., IEEE). The Chinese students acknowledged that they are not familiar with writing reflective essays and reports, since this is only included in some optional courses in the bachelor degree. Another interesting theme, which is aligned with the findings by [4], is the fact that Chinese students are used to include lots of contextual information in their reports, before discussing the main topic. This contrasts with their perceptions about the intended structure of reports in Sweden, where they perceive that they should be brief and go to the point as quick as possible.

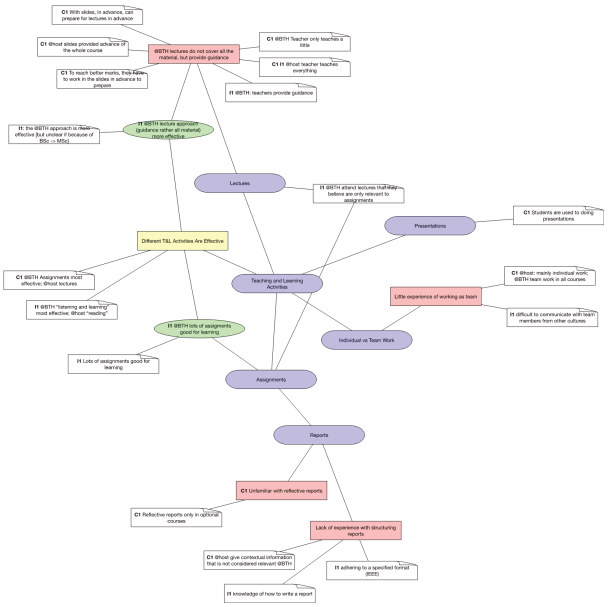


*Figure 3. Subcategories regarding the Learning and Teaching Activities*

Their perception of the lectures was another big theme in this theme-cluster. Both Chinese and Indian Students perceive that in their home university the lectures cover all the material, whereas in Sweden provide *"only"* guidance. For example, one Chinese student said that *"In China the teacher teaches us everything"* whereas here at BTH *"teacher only teaches us a little"*. Chinese students perceive that they have to dig in the topic by themselves while preparing their assignments. They also perceive as a challenge the fact that they do not have access to the slides in advance. In their home university they have access to the slides for the whole course, and they can prepare themselves for the lectures in advance. They acknowledge that this was the main vehicle to get higher marks in China *i.e.,* working in the slides beforehand to prepare for the lectures. Similarly, the Indian students perceive that the teachers provides mainly guidance, although they also acknowledged that the BTH approach is more effective in terms of learning[5].

There is also a theme covering the students' perceptions regarding the most effective teaching and learning activities. Chinese students perceive that the most effective learning activity in Sweden are assignments and that this is what helps them to understand the topics. Similarly, for some Indian students, having several assignments that spread all along the course was good for their learning. Other Indian students expressed that they find more effective *"listening and learning"* referring to the fact that in Sweden attending to the lectures is the most effective way of learning for them . This contrasts with their perception about their Indian university, where they perceive that they learn more by reading the materials while preparing for their exams.

### Language and Cultural Aspects

There is also a theme cluster around language and cultural aspects and their effect on the students' learning. Figure 4 depicts the main themes, sub-themes and codes in this topic cluster.

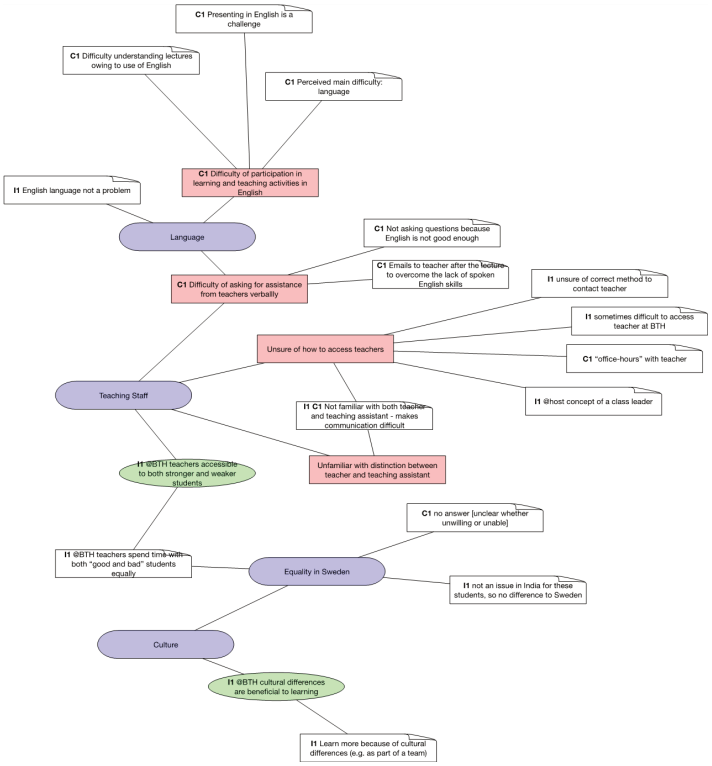


*Figure 4. Subcategories regarding the Language and Communication with the Teacher*

One big theme in this cluster were the language difficulties. Language was the main topic mentioned in the interview with the Chinese students when asked about the challenges they face when arriving to Sweden. They perceive their main challenge / problem was the use of English as the language of instruction. They perceive that it is hard for them to participate in some of the learning and teaching activities due to their lack of language skills. However, during the interview they pointed out that this is something that has improved over time (a perception that we can confirm after discussing with them 6 months after having met them in our courses). According to them, the main problem they have are their listening skills and they acknowledged having problems understanding the lectures. They also pointed out that the different accents of the lecturers represent a problem for them.

Another sub-theme in this topic is the difficulties to communicate with teachers (particularly, verbally). Since the speaking skills of the Chinese students were limiting them, the communication with teachers also represents a challenge. They explicitly mentioned that sometimes they do not ask questions "because our English is not good enough". They solved this problem by communicating with their teachers by email. They pointed out that communication through email make them lose some of the interaction face-to-face provides, but also highlighted as beneficial that they can ask questions at any time.

The access to teaching staff is another big theme in this theme-cluster. Having different roles involved in the courses i.e., teachers and teaching assistants, and the differences between pedagogic cultures make students being unsure how to access the teachers. Indian students explicitly mentioned that sometimes they feel unsure about the correct method to contact the teachers. Chinese students pointed out that in China teachers have *"office hours"* just after their lectures, and they usu-

ally interact face-to-face with the teachers during these office hours. Here they perceive that the teacher is not as available as in China. Similarly, Indian students mentioned that sometimes they find difficult to access teachers at BTH, and they find the procedure of making appointments not very effective. In India they have the concept of class leader to help other students and ease the communication flow.

Interestingly, Indian students highlighted the fact that in Sweden teachers spend the same amount of time both with *"good and bad"* students, and this was perceived as a positive aspect of the education at BTH.

As part of the cultural aspects we wanted also to investigate a theme around equality in Sweden, which was suggested by the Programme Manager of the MSc in Computer Science. With this theme we wanted to investigate whether the Swedish emphasis on equality had any impact on their education. The Chinese students deflected the question, no answers were given, and was unclear whether they were unwilling or unable to answer this questions. On the other hand, Indian students claimed that they do not recognise equality as an issue in their home cities in India either.

The last cultural-related theme was the fact that cultural differences are beneficial for learning. Indian students highlighted the fact that having this mixture of cultures at BTH, as part of their teams allowed them to learn *"more"*.

#### Course and Programme Organization

Course and programme organization is another theme-cluster highlighted during the analysis, and whose decomposition in themes, sub-themes and codes is shown in Figure 5.

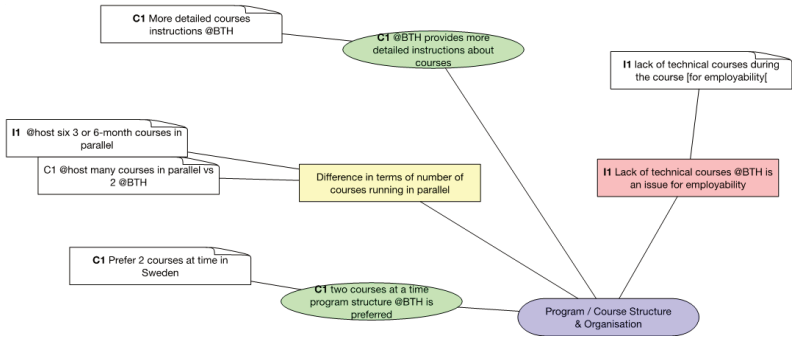


*Figure 5. Subcategories regarding the Course Structure and Programme Organization*

The structure of the programs at BTH, with two courses in parallel and two study periods is one of the main sub-themes that arose. Indian students mentioned that at their home university they have six 3-month or 6-months courses running in parallel. Similarly, Chinese students mentioned that they have many courses running in parallel as opposed to the two at BTH. Chinese students highlighted that the BTH structure with only two courses in parallel per study period, and two study periods per semester, helps them focus on the courses and keeping the pace. They perceive this structure supports their learning better.

The quality of the instruction and information was a sub-theme that was found doing the thematic analysis. The Chinese students highlighted the fact that there is much more detailed information and instructions here at BTH, and they perceive this as a positive difference. They highlighted that in general, here in Sweden they have more clearer instructions for the course organization and for the different assignments. They said literally: *"In China when we have to write a report, the teacher gives just us a topic and the number of words e.g., 2000 or 3000"*, and they perceive having clear instructions, as they do here, helps them to succeed in their assignments.

Finally, the lack of technical courses and it is impact on the students' future employability is the

last sub-theme related to programme and course organization. Indian students agreed that the lack of technical, hands-on courses during the programme can be hindering their future employability, since they will not be able to show the technical skills during the years they were studying their Masters. However, this could be perhaps a misunderstanding regarding the required skills for a Software Engineer in industry.

#### Assessment, Plagiarism and Collusion

Assessment, plagiarism and collusion is a big theme-cluster that emerged during the analysis. In this theme cluster, shown in Figure 6 we categorise the main themes around problems and challenges regarding assessment and the students perceptions regarding plagiarism and collusion.

One of the main sub-themes, that can have a big impact on the initial performance of the overseas students when arriving to Sweden is the fact that both in India and China they hardly ever have mandatory written assignments. In China, if they have written assignments, those will be always optional and these written assignments only represent a small proportion of the final grade (if any). In India written assignments are simple and ungraded, whereas at BTH the assignments are mandatory, most of the times with hard deadlines and always graded. Another interesting finding is the fact that Chinese students are used to only attending lectures, and then demonstrate their learning and understanding through a written exam. The presence of written assignments (also referred to as "Homework" during the interview) is anecdotal, and therefore the main component of the final grade is the written exam.

A sub-theme with impact on their performance when arriving is the different qualities perceived as important in assessment i.e., how to show learning and understanding (surface *vs.* deep-learning). Indian students highlighted the fact that in their home university their responses to the exam should be as verbatim to the book as possible, whereas here in Sweden they perceive that we ask for more reflective, deep and applied practical knowledge. Chinese students discussed that here they perceive that what is important here is *"doing things well"* and cooperation within the team. However, this latter point might be affected by their impressions around the courses taught by the researchers conducting the interviews.

Another important sub-theme was the different grading and whether it is unclear what is required to get each grade. Indian students highlighted the fact that the grading scale is unclear, and they do not know what is required for a high grade, and in any case more effort is required to get an A. The Chinese students discussed the differences in the effort required to pass and to get a high grade. Initially some of the participants were arguing that they perceive the amount of work to pass the course was higher in Sweden as compared to China. But then one of the participants made an interesting comment, and all the participants agreed: to pass the course you must work harder here in Sweden, but if you want a high grade they perceive the amount of work required is the same (everywhere they said). The unfair marking was highlighted as a main difficulty by the Chinese students, who suggested that the teachers here in the introduction of the courses explain the assessment schema.

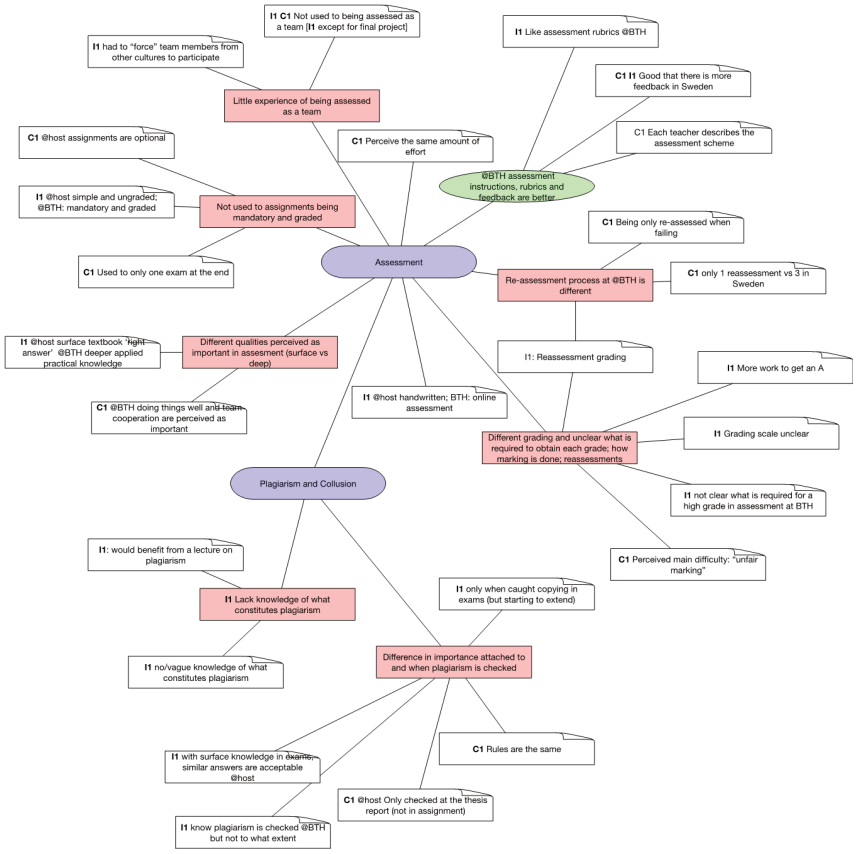


*Figure 6. Subcategories regarding the Assessment and Plagiarism and Collusio*

The number of reassessments and their grading was a main difference that emerged during the analysis. Chinese students highlighted as positive having multiple reassessment opportunities (3 here in Sweden vs only 1 in China). On the other hand, they perceive the Swedish policy "reassessment only when fail" is not totally fair (although is a common practice in most countries in Europe). Similarly, Indian students felt unfair only being able to get an E after an FX (complementation required) whereas you can get an A after an F. All these issues show that perhaps there is a lack of information or understanding on the grading.

On the other hand, a related sub-theme that emerged was the amount of feedback and its role in their learning and the use of rubrics. Chinese and Indian students both perceived that here in Sweden they get more feedback. Indeed, Chinese students said that here they get feedback, as compared to China where they do not receive feedback at all. They highlighted the fact that they believe this feedback is important and helps them in their learning process. Indian students pointed out the use of rubrics as positive for their learning, since helps them recognizing what is important and what they should concentrate on.

Finally, plagiarism and collusion is the last theme in this theme-cluster. During the discussion it was clear that, among the Indian students, there was a lack of deep knowledge regarding what constitutes plagiarism. The Chinese students acknowledged that this was not new for them: that in their home university the concept and the regulations regarding plagiarism are essentially the same. However, in the home university of the Chinese students the anti-plagiarism checking is only applied in the final thesis, and not at the assignment level. They do not expect that we do checks in every single submission, and this could be the cause of the high rate of plagiarism cases among Chinese students at BTH. Indian students were aware that plagiarism was checked at BTH but not to what extent and this might constitute a problem on itself. In general, it seems that there is a difference in the importance given to plagiarism and in their expectations regarding when plagiarism is checked at BTH.

#### Sources of Information

The sources of information are the next theme-cluster from the analysis. Although this was not one of the main hypotheses of the research, it became a theme that might explain some of the issues we experience as teachers, and that can be solved by discussing and including more information in the introductory lectures. This theme-cluster contains themes related to how and from whom overseas students get information about the Swedish system. The summary of its structure, and its organisation in themes, sub-themes and codes is shown in Figure 7.

The first main theme was the previous year's students as source of information about studying at BTH. Both Chinese and Indian students pointed out that the main source of information are their seniors (previous year's students). Chinese students get practical information about day-to-day living in Sweden, but also specific information about the different courses and the main difficulties they will face. Indian students get information about the course structure, which courses are difficult, which classes can they skip, and how to write reports and presentations. Both Indian and Chinese students believe this information to be accurate, although we believe this represents a problem since if a course is changed it will take time for them to notice these changes.

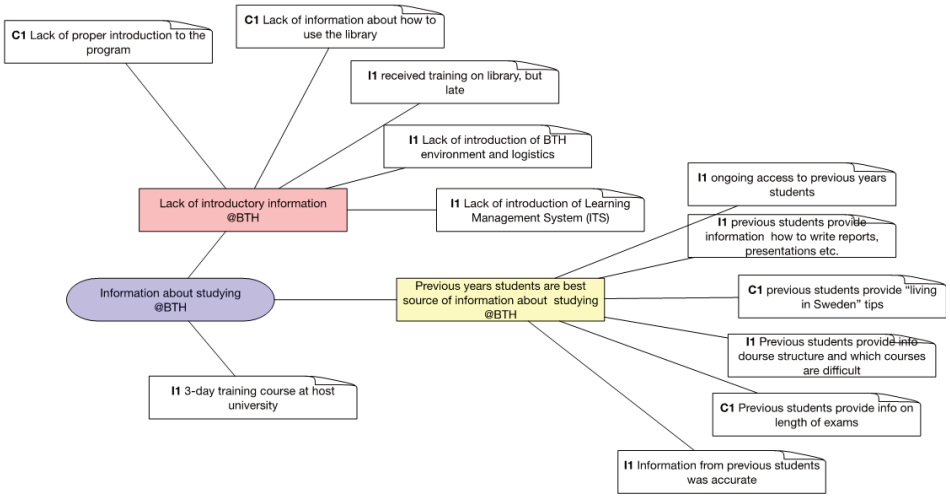


*Figure 7. Subcategories regarding the Sources of Information*

The second main theme was the lack of introductory lectures about the particularities of studying at BTH. Both Chinese and Indian students pointed out the lack of a proper introduction to the program. The Indian students also highlighted the need of an introductory lecture to the ITSLearning management system.

The Chinese students emphasised the lack of an introduction about how to use the library. However, this information might not be accurate since the librarians at BTH offer this type of introductory talks. Perhaps they are not always aware of these events. Indian students acknowledged that they had an introductory lecture to the use of the library, although this was too late for them, since they start using in the first course they take just after their arrival.

#### Motivation

The last theme-cluster that resulted from the analysis of the responses to some of the prompting

questions regarding their motivation to come to Sweden and why to study Software Engineering. The categorization of themes, sub-themes and code is shown in Figure 8.

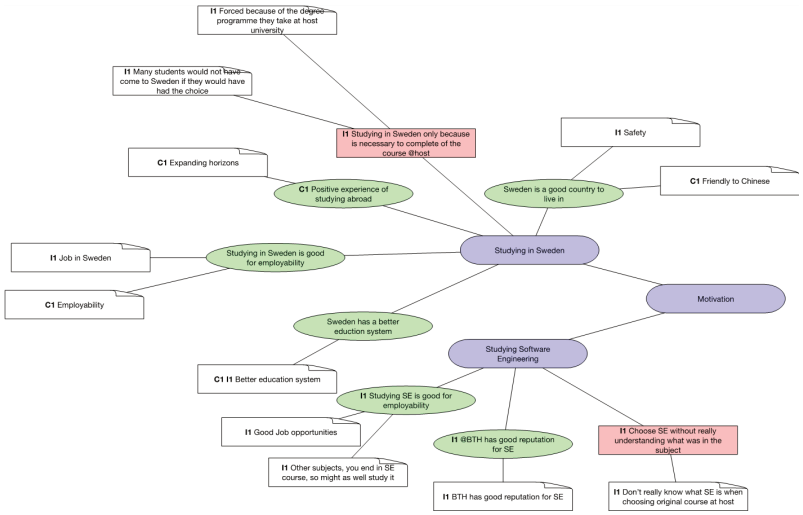


*Figure 8. Motivation Thematic Map*

There were two main themes: the motivation to study in Sweden, and the motivation to study software engineering.

As regards to the why choosing Sweden, both the Chinese and the Indian students highlighted as main motivation to come to Sweden their future employability. Chinese students believe that having an international experience might help them getting a job back in China (but they also mentioned finding a job here in Sweden). Indian students mentioned finding a job in Sweden as motivation to come.

Another sub-theme that arose is the fact that some students are at BTH only because this is a requisite to complete their degree in their home university. Some Indian students acknowledged that they are here only because this was part of their BSc/MSc integrated degrees in their home university. The host university is one of the best universities in their country and the only way for some of these students to access to that university is to be enrolled in these integrated degrees. Many students acknowledged that they would not have come to Sweden if they have had the choice.

The Chinese participants highlighted the fact that they perceive that Sweden has a better education system and that Sweden is friendly and respectful towards the Chinese community as factors to choose Sweden as the country to finish their studies.

Regarding the motivations to study software engineering, being good for employability turned out being one of the main motivations for choosing this topic for their MSc degree. One of the main motivators for Indian students to choose MSc in Software Engineering is the reputation that BTH has in the area. They also made an interesting reflection: chances are that you will end working in a software-related work, regardless of what is your formal background, so might turn out being a good idea to study it as well. However, one sub-theme arose that might explain some of the knowledge and skill gaps: Indian students exhibited certain lack of knowledge on what is software engineering when choosing their degree at the host university.

## Discussion

In this section, we discuss the thematic analysis results presented in Section 3 in the context of our two research questions:

*RQ1:* What are the differences in academic culture between BTH and partner universities in China

and India that are experienced by students from these partner universities on the MSc in Software Engineering (and related programmes) at BTH?

*RQ2:* How do students perceive the impact — both positively and negatively — of these differences on their learning at BTH?

The thematic maps presented in Section 3 was organised in such a way as to facilitate answering these two research questions. The main themes (blue rounded-rectangles in the figures) are related to the general topic areas that arose during the focus groups (and in turn, are influenced by the prompting questions used during data-collection). But many of the sub-themes (red and yellow rectangles, green ellipses) identify specific differences in academic culture between BTH and the partner universities in China and India (addressing RQ1). Moreover, the notation identifies the impact as positive (green ellipses), negative (red rectangles), or neutral/unknown (yellow rectangles).

We consolidate the relevant sub-themes that provide an answer to RQ1 in Table 2. The main cluster of sub-themes that are omitted from this table are those related to motivation for studying software engineering and in Sweden, since although valuable, is not directly relevant to the research questions.

| Academic Difference | Impact |
|---|---|
| Unfamiliar with distinction between teacher and teaching assistant | - |
| Difficulty of asking for assistance from teachers verbally (C1 only) | - |
| Teachers are accessible equally to both stronger and weaker students (I1 only) | + |
| Difficulty of participation in learning and teaching activities in English (C1 only) | - |
| At BTH, lectures do not cover all the material exhaustively, but instead provide guidance | - |
| Different teaching and learning activities are regarded as the most effective | neutral |
| At BTH often required to work as a team, but little experience from host university | - |
| At BTH often assessed as a team, but little experience from host university | - |
| Many more assignments at BTH - and this is good for learning (I1 only) | + |
| Unfamiliar with how to write reflective reports (C1 only) | - |
| Lack of experience with structuring report assignments | - |
| At BTH, assignments throughout the courses are mandatory and are graded | - |
| At BTH, assessment instructions, rubrics, and feedback are much clearer | + |
| Re-assessment process at BTH is different, and unclear to students | - |
| Marking and grading process at BTH is different, and unclear to students | - |
| Different qualities are perceived as being important in assessment: show deeper learning at BTH | - |
| Difference in importance attached to plagiarism and when it is checked | - |
| Lack knowledge of what constitutes plagiarism (I1 only) | - |
| Cultural differences at BTH are beneficial to learning (I1 only) | + |
| BTH provides more detailed instructions about course structure and organisation (C1 only) | + |
| Difference in the number of courses running in parallel | neutral |
| Two courses at a time is better structure for the programme at BTH (C1 only) | + |
| Lack of introductory information at BTH | - |
| Previous years' students are the best source of information about studying at BTH | neutral |

*Table 2. Sub-themes from the thematic analysis that relate to academic differences (RQ1), and their impact (RQ2)*

We also indicate how the students perceive the impact in the column *Impact* in this table, taking this information directly from colour/shape notation used in thematic map. The data in this column addresses RQ2.

We note that we speculate the existence of two *meta-themes* that could be used to group the listed academic differences: those that could be resolved by providing information to the students, e.g. about course and programme organisation; about the re-assessment process; and those that require additional skills, knowledge, or experience on the part of the students, e.g. how to work in teams; language skills. However, at this stage of our investigation, we do not feel that we are able to make this classification confidentially. In future focus groups, we will discuss possible solutions with participants in order to provide data to support this classification: the two focus groups conducted so far did not explicitly discuss solutions. In addition, we would like to conduct interviews with programme managers, course responsible, project coordinators, teachers, and other stakeholders regarding potential solutions to those academic differences with a negative impact.

### Threats to the Validity

In this section, we discuss the main issues that may threaten the validity of the study. We are going to focus in the main three issues discussed in [9] for qualitative studies: validity, reliability and generalisability. We also discuss with more detail the main threats that might have challenged the focus group interviews.

Regarding the general threats for qualitative research we identify the following challenges:

- *Validity:* Validity in qualitative research refers to the appropriateness of the method, data and tools [9]. The main threat to the validity of this study is the appropriateness of the method. We have chosen a method well suited to study the unique phenomena that we are interested in observing. We decided to apply thematic analysis since the data gathered already revealed the main information we were aiming to analyze, and no more profound grounded-theory analysis was needed. Regarding the sampling schema, we contacted the whole population as described in section 2, and encouraged their participation through emails, but their participation was voluntary, although reaching a high response rate (e.g., 40% for C1 and 45% for I1).
- *Reliability:* The main threat to the reliability of qualitative studies refers to the replicability of the process and the results [9]. A margin of variability on the results is accepted when talking about qualitative research [9], since the subjectiveness of the researcher is embedded in the roots of the analysis itself, and also the contexts analyzed are unique and not repeatable. Therefore, the main threat to the validity of this type of studies relies on the consistency. To mitigate these threats, we have consistently followed the analysis guidelines and documented all the different stages of the data collection and data analysis. To mitigate the threats regarding the reliability of the data, we have "manually" transcribed the recordings of the full focus group interviews. Each researcher transcribed roughly half of the material available. We also did a cross validation, by reading the other part of the interview the other researcher have transcribed. We have also performed cross validations during the analysis process, by making sure that each researcher coded the data the other researcher transcribed and by cross-validating and discussing the different categorizations of the thematic map.
- *Generalisability:* We do not aim for any sort of generalisability of the results, since we are analyzing one phenomena in a very specific context, and our aim is to gain an understanding of the differences on academic cultures and its impact on the teaching and learning at BTH. Therefore, generalisability do not represent a threat to the conclusions drafted from this study.

Regarding the main threats that might have challenged the focus group interviews we identify the following threats to the validity:

- *Reluctance to express opinions in front of peers***:** Some participants might have reluctance to express their opinion in front of peers. We tried to alleviate this threat by having small focus groups (4-5 participants) from the same university, which increases the chances of them having more close relationships and being more prone to talk openly. We believe that this strategy worked well during both interviews: the students talked honestly and discussed many issues regarding their education at BTH, however it is not possible to assess this in practice.
- *Reluctance to talk about certain problems in front of the professors:* This turned out being the

other way around: since we have had direct interaction with them, they apparently were more prone to talk and participate, and express their opinions, sometimes regarding our courses or our own performance as teachers.

- *Participants biased by their opinions regarding the courses taught by the researchers conducting the interview:* Some of the opinions might be more prone to focus on the course taught by the researchers or their opinions biased by their experiences on these courses. We believe the participants were talking in general about the program, and sometimes, when talking about these courses they mentioned it explicitly.
- *Reluctance to talk about unethical or compromising problems:* The fact that some topics are related to compromising or unethical behaviour might represent a big threat to the validity. However, the students were discussing openly compromising problems as plagiarism, collusion, fairness of the grading, support to bad-performing students without any visible reluctance. This might be influenced by the previous factors, the relationship with the participants.
- *Language barriers:* This was one of the main challenges, the misunderstandings that might appear due to the language. We mitigated this threat by trying to speak using more simple constructions and by rephrasing our questions in case there was any trace of misunderstanding. Although no visible misunderstanding occurred, nothing prevents the participants to have misunderstood any of our questions or the interviewers not fully understanding their responses. The other big challenge in this was the interchange of information between the participants in their first language. We mitigated this threat by explicitly asking the participants not to talk in their mother tongue (only for small clarifications). Only in the first interview there were some small word clarifications in Chinese, except for the question regarding equality, that required a long clarification in Chinese by one of the participants.
- *Class leader:* The presence of the "class leader" in the Indian focus group (I1) might have influenced the participants' opinions, towards a more positive discourse regarding the teaching and the BTH system. However, we believe that the participants were in general very honest and they were in any case also criticizing some aspect of the teaching and what they do not see as fair of the BTH system, but again, it is not posible to assess the effect in practice.

## Conclusions and Further Work

Our main objective in this study was to understand the academic differences between BTH and the partner universities and the impact, both positive and negative, on the overseas students who come from these partner universities to study for a MSc in Software Engineering at BTH.

We chose to collect data from students using focus groups, and these groups proved to be a rich source of relevant and valuable data. Using thematic analysis, we were able to identify from this data a relatively large list of academic differences, understand the impact of these difference on the students' study at BTH, and to start figuring out causes of this impact and potential mitigating solutions.

As teachers on the MSc in Software Engineering programme at Blekinge Institute of Technology, we have been pleasantly surprised as to how much this study has improved our understanding of the issues experienced by our students — not only from results of the data analysis that addressed the specific research questions we had, but from the entire process of meeting and discussing these

topics with students. Our intention is to disseminate the information we have gathered to our fellow teachers, programme managers, project coordinators, and other stakeholders so that may similarly improve their understanding of the challenges faced by our overseas students and to participate in solving the issues raised.

As future work, we plan to continue this work by conducting additional focus groups where the participants are students from other partner universities of the MSc in Software Engineering and Computer Science programs. The objective will be to consolidate the evidence we have gathered so far as well as to identify new academic differences (which may be specific to one partner university). We would also like to strengthen the reliability of the conclusions by interviewing programme managers, project coordinators, and teachers on courses taken by students on these programs to validate the issues expressed by students, better categorising the root causes, and identifying mitigating solutions. We aim also at triangulate this data with focus group interviews with Swedish students having exchange experiences either in China or India.

## References

[1] P. E. Barron and C. Arcodia, "Linking learning style preferences and ethnicity: International students studying hospitality and tourism management in Australia," *J. Hosp. Leis. Sport Tour. Educ.*, vol. 1, no. 2, pp. 1–13, 2002.

[2] J. Ryan and S. Hellmundt, "Excellence through diversity: Internationalisation of curriculum and pedagogy," in *17th IDP Australian International Education ...*, 2003, pp. 1–10.

[3] J. Ryan and K. Louie, "False Dichotomy? 'Western' and 'Confucian' concepts of scholarship and learning," *Educ. Philos. Theory*, vol. 39, no. December, pp. 404–417, 2007.

[4] J. Mattisson Ekstam, L. Ahlin, and A. Fjelkner, "A Guide for Teachers and Students in a Cross-Cultural Context," Kristianstad, Sweden, 2014.

[5] B. G. Glaser and A. L. Strauss, *The discovery of grounded theory : strategies for qualitative research*. Aldine Pub. Co, 1967.

[6] K. Charmaz, *Constructing grounded theory : a practical guide through qualitative analysis*. Sage Publications, 2006.

[7] A. Bryant and K. Charmaz, *The SAGE handbook of grounded theory*. Sage Publications, 2007.

[8] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qual. Res. Psychol.*, vol. 3, no. 2, pp. 77–101, Jan. 2006.

[9] L. Leung, "Validity, reliability, and generalizability in qualitative research.," *J. Fam. Med. Prim. care*, vol. 4, no. 3, pp. 324–7, 2015.

[1] We do, nevertheless, intend to hold additional focus groups in the next few months, as discussed in Section 6.

[2] The separate audio recording using a high-quality omni-directional microphone acted as a backup in case of problems with the video-recording, or poor quality in the sound track of the video.

[3] Our original intent was to also offer gift cards, e.g. for the local cinema, as both an inducement for participation, and to acknowledge the time the participants have given us voluntarily. However, our thesis supervisees from China — who would not be asked themselves to participate in the focus groups — advised us that this could be interpreted as an encouragement for the participants to provide only positive feedback.

[4] The whole thematic map is available for download at:
http://www.gonzalez-huerta.net/wp-content/uploads/2017/10/Thematic_Map.pdf

[5] It was unclear if the difference was due to the fact that now they are in MSc level.

3

# Are finance students over- or under confident – A study on the ability to predict grades

Emil Numminen
*Blekinge Institute of Technology and Kristianstad Univeristy*

Ola Olsson
*Kristianstad University*

emil.numminen@bth.se, ola.olsson@hkr.se

**Abstract**

Overconfidence is a cognitive bias that most people suffer from. A person suffers from overconfidence bias when his or her own subjective estimation of an ability is significantly higher than an objective estimation of the same ability. Previous research in pedagogy has established that students suffer from overconfidence when it comes to grade prediction in business and economics. A student suffering from overconfidence bias have a propensity to study less than required since the subjective estimation of comprehension of the subject is higher than it really is when measured objectively. The implication of overconfidence is thus that a student will not fulfill his or her own full potential of learning the subject. This paper adds to the overconfidence research in pedagogy by measuring the level of overconfidence throughout an entire course to analyze the relation between learning and overconfidence. This has not been done in previous research. Students made estimation of their final exam score at five times throughout the course. Results show that students are overconfident and that they do not calibrate their expectations over time, on a general level. as they perhaps should given how they perform in learning the subject. Female students show a lower degree of overconfidence and had a higher tendency to calibrate their expectations. After having taken the exam and making a final estimation of expected grade, overconfidence drastically went down for less academic experienced students but instead increased for more academic experienced students. In this estimation less academic experienced female students even became underconfident.

## Background

There are several studies (e.g. Tversky and Kahnemann, 1974; Kahnemann and Tversky, 1979) conducted on various biases and heuristics that many people are affected by. A heuristics can be defined as a mental shortcut based on one-handed information to simplify a complex setting (Lewis, 2008). The resulting cognitive systematic error is called a cognitive bias (Tversky and Kahnemann, 1974) or simply put; a bias. The problem of suffering from biases is that people do not make informative decisions since they base their decisions on one-sided information and tend not to pay attention to contradicting information (Plous, 1983). There are many types of heuristics and resulting biases, see e.g. Tversky and Kahnemann (1974), Baron (2008) or Hillbert (2012) for an extensive review, but one of the most common one is overconfidence bias (Brenner et al., 1996). A person suffers from

overconfidence bias when her own subjective estimation of an ability is significantly higher than an objective estimation of the same ability (Pallier et al., 2002).

There is an extensive literature on overconfidence bias, see e.g. Skata (2008) for a review of the literature, and it has been studied in various academic disciplines as well as in various empirical settings. Hribar and Yang (2015) studied how various dimensions of overconfidence affect management forecasting, Malmandier and Tate (2005) studied the effect of overconfidence in timing decisions of when to exercise options and Ho et al. (2016) studied how overconfidence may explain the 2007-2009 financial crises. It is not only in economics psychology where overconfidence is used as an explanatory variable; Cassam (2017) argues that overconfidence is a major factor explaining diagnostic errors in medicine, Andreou et al. (2014) analyzed the effect of overconfidence for dopaminergic activity and found a positive effect. Overconfidence has also been reported in pedagogy in various studies when it comes to grade expectations. Grimes (2002) studied overconfidence measured as the difference between expected and actual score on a midterm exam and found students to suffer from overconfidence. Koku and Qureshi (2004) studied the effect of overconfidence in a setting of multiple-choice examination. Nowell and Alston (2007) studied the relation between overconfidence, measured again as the difference between expected and actual grade, and pedagogies. It is relevant to study overconfidence in a pedagogical setting of teaching and learning given the implication of the bias. A student suffering from overconfidence bias have a propensity to study less than required since the subjective estimation of comprehension of the subject is higher than it really is when measured objectively. The implication of overconfidence is thus that a student will not fulfill his or her own full potential of learning the subject. This study contributes to previous studies of overconfidence in pedagogy by studying how overconfidence evolves during a course taken by the students. Whereas previous research has measured overconfidence at one time and most often during or after the exam (e.g. Koku and Qureshi, 2004; Novell and Alston, 2007) we measure the expected grade five times during the course and after the exam. This enables us to study if there is a relation between overconfidence and how much of the course has been covered, i.e. is there a relation between overconfidence and learning.

The rest of the paper is structured as follows. Next, we will review the related literature of the areas we are analyzing. After that we will discuss how data was collected and analyzed. Finally, we will discuss and draw conclusions based on the empirical study conducted and make suggestions for further research.

## Related literature

Previous studies in pedagogy has revealed that business and economics students are overconfident in their expectations of final grades for courses they have taken (Grimes, 2002; Koku and Qureshi, 2004; Novel and Alston, 2007). Research has tried to explore this overconfidence in various ways. Armstrong (2013) experimented using grades from previous year as a base for present years to predict their final scores as an assessment of their progress in an undergraduate business course. 29% of the respondents stated that grades were lower than expected whereas only 6% stated that the score was higher than expected. Further, 47% stated that they were studying more than expected for this prediction whereas only 3% were studying less. A possible implication of this is that students were overconfident given the relation between expected grades and predicted grades using previous year data versus the time spent on studies by the students. Other studies have focused on students studying intentions and the gap between expected and realized grades. Armstrong and MacKenzie (2017) found that students with high academic abilities have a smaller gap between grades and goals whereas students with high personal control have a larger gap between the same. However, the study also reports that these gaps narrows over time since students with a larger gap increased their study efforts. The effect of feedback to students on study effort and performance in pedagogy has been confirmed in several previous studies, see e.g. Hattie and Timberley (2007), Shute (2008) and Schartel (2012) for further discussions and results.

The effect of students abilities on performance has been widely studied and confirmed before.

Andersson, Benjamin and Fuss (1994) and Straton and Arnold (2012) have all found positive effects on students performance in economics based previous education in math. Other factors that research have found positively related to ability and overconfidence include age, gender and previous grades. Gustavson and Hall (2011) found positive correlation between previous grades and overconfidence in library research skills. Lundeberg, Fox and Puncohar (1994) and Geffert and Christensen (1998) report gender effects when it comes to overconfidence. Male students are more inclined to overestimate their work and performance whereas female students on the other hand are more likely to underestimate their ability. Other effects of gender on grades also include choice of subject to major in. Owen (2011) found relation between grades in introductory economics courses and the propensity to major in economics for female students whereas for male students the same relation could not be observed. However, the gender effect on grade prediction and thus overconfidence is not univocal; nor Ballard and Johnson (2005) or Hossain and Tsigaris (2015) could see any significant gender effects in grade estimation in economics nor business and economics statistics courses analyzed.

Studies on the relation between age and overconfidence has revealed that age has a negative effect on overconfidence when it comes to grade prediction. Grimes (2002) found less overconfidence in older business students. This result is also confirmed by Novell and Alston (2007) in their study of economics students. Age seams to have a negative effect on overconfidence in general as suggested by several psychological studies as e.g. Hanson et al. (2008) and studies in behavioral decision making e.g. de Bruin, Parker and Fischoff (2012).

The literature review show that there is a support in the literature of overconfidence among students. Varoious research has also shown that there are factors that can explain differences in overconfidence among group of students. Our analysis will add to this literature by measuring how overconfidence evolve over an entire course and if gender and academic experience among students can explain overconfidence among students.

### Method and data collection

Data for the study was collected using students taking a course in finance at Kristianstad University during the autumn 2016. This is not an introduction course in finance but designed for Bank and Finance (BF) students in their last year on the bachelor level as well as for Accounting (RR) student in their first year on the master level. Hence, students in this course have been trained for academic work during three and four years respectively. The class included 46 participants and was outnumbered by B&F students but more balances when it came to gender. The average student was 25 years old.

| | | Academic focus | | |
|---|---|---|---|---|
| | | BF | RR | Total |
| Gender | Men | 22 | 4 | 26 |
| | Women | 13 | 7 | 20 |
| | Total | 35 | 11 | 46 |

*Table 1. Sample distribution*

Students were asked to predict their expected score on the final examination at pre-specified times during the course on the scale 1 to 100. In this paper we call these measurement occasions (MO:s). MO1 took place during the first day after the course introduction where students were told what to expect about learning outcomes, course activities and examinations. After the second and the fourth week of the course MO2 and MO3 were done. During the last week of the course students studied on their own. MO4 took place by the end of the fifth week and only a few minutes ahead of the final exam. This was after a week of self-studies where students were given the possi-

bility to repeat key elements of the course at home. MO5 took place a few minutes after the final exam. The data collected from MO1 to MO5 were based on student expectations of total score of the final exam measured on scale from 1 to 100 percent. This data was then compared with the data from the actual performance (MO6) as measured by e.g. Grimes (2002). All students were informed about the scoop of this study and that participation was on a voluntary basis with full confidentiality guaranteed. All data of expectations and actual performance of scores were collected in Excel and all statistical tests were done in SPSS. Given the aim of the paper and the literature review the test of overconfidence was done using t-tests and ANOVA with post doc analyze.

| | Expectations | | | | | Performance |
|---|---|---|---|---|---|---|
| MO schedule for GPA | (1) Start | (2) 2 weeks | (3) 4 weeks | (4) pre-test | (5) post-test | (6) Actual |
| Means | 69,8 | 67,4 | 66,2 | 69,1 | 59,4 | 56,2 |
| Standard deviation | 9,7 | 9,9 | 8,5 | 8,5 | 14,3 | 15,9 |
| Participants | 44,0 | 38,0 | 30,0 | 44,0 | 44,0 | 45,0 |
| Diff of means, expected and actual performance | 13,6 | | | | | |
| Diff of means, expected and actual performance | | 11,1 | | | | |
| Diff of means, expected and actual performance | | | 10,0 | | | |
| Diff of means, expected and actual performance | | | | 12,8 | | |
| Diff of means, expected and actual performance | | | | | 3,1 | |
| Diff of means, expected and actual performance | | | | | | 0,0 |

*Table 2. Student expectations and performance: means and standard deviations*

The data was translated into means or class grade point averages (GPA) at any given point of observations over time. The number of participants varied but this is considered in all GPA calculations as well as in all statistics below. Positive differences of means based on expected (MO 1-5) and actual performance (MO 6) indicate overconfidence. Hence, if an expected GPA was 80,0 and the actually performed GPA was 70,0 the subject would be overconfident. Table 1 above show the descriptive statistics for the GPA over time during the MO:s used in experiment. The difference in means is positive throughout the entire period but decreases in MO5. The standard deviation on the other hand increases in MO5.

### Statistics and results

In this section we will present descriptive statistics of the collected data and the statistical tests for overconfidence.

#### Test of mean expectations

The means are tested in Table 3 to Table 6 and the results are displayed in Figure 1. All students did not participate in every observation. This will influence the mean value when you compare expectations for the same individuals over time. Our focus was the development of mean values of the expected final score on the examination. Hence, we ran paired sample t-tests to see the implications of student absence and to be able to suggest if the adjusted changes were statistically significant.

**Paired Samples Statistics**

| MO comparison | | Mean | N | Change of mean | Sig. (2-tailed) |
|---|---|---|---|---|---|
| Pair 1 | MO1 | 69,9 | 37 | | |
| | MO2 | 67,3 | 37 | -2,5 | 0,04 |
| Pair 2 | MO2 | 65,6 | 29 | | |
| | MO3 | 66,3 | 29 | 0,6 | 0,68 |
| Pair 3 | MO3 | 67,2 | 28 | | |
| | MO4 | 70,6 | 28 | 3,4 | 0,00 |
| Pair 4 | MO4 | 69,1 | 44 | | |
| | MO5 | 59,4 | 44 | -9,7 | 0,00 |
| Pair 5 | MO5 | 59,4 | 44 | | |
| | MO6 | 56,2 | 44 | -3,2 | 0,18 |

*Table 3. Change of student expectations over time – statistical significance of means*

The changes of expected performance means are statistically significant with 95 probability in all cases but one (Pair 2). The initial minor loss of expectation (Pair 1) was, however, recovered during the last week of the course (Pair 3) and ahead of the final examination. In addition, the dramatic change of mean values (Pair 4) was confirmed between MO4 and MO5 (-9,7). Hence, the class substantially lost momentum when it came to expectations of performance during the final examination and this change is statistically significant on a 95% level. These results suggest that GPA expectations remain on a high level during the entire course but drops significantly during the final examination. At that point standard deviation increased. This imply that the class uncertainty over the actual performance is almost the same during the course, see Table 2, but increases when the final exam is taken.

To further explore the material Table 4 and Table 5 below presents the material when it is separated based on gender and academic experience.

| | Expectations | | | | | Performance |
|---|---|---|---|---|---|---|
| MO schdule for gender | (1) Start | (2) 2 weeks | (3) 4 weeks | (4) pre-test | (5) post-test | (6) Actual |
| Men - number | 25,0 | 19,0 | 18,0 | 26,0 | 26,0 | 26,0 |
| Women - number | 19,0 | 19,0 | 12,0 | 18,0 | 18,0 | 19,0 |
| Men - mean | 72,8 | 69,7 | 69,3 | 71,5 | 61,9 | 58,6 |
| Women - mean | 65,9 | 65,1 | 61,6 | 65,7 | 55,7 | 53,1 |
| Men - SD | 10,5 | 11,5 | 8,5 | 8,6 | 14,1 | 16,5 |
| Women - SD | 7,1 | 7,6 | 6,4 | 7,4 | 14,2 | 14,8 |

*Table 4. Development of expectations and gender – mean and standard deviation*

In general men expected higher scores than women in all observations and this was statistically significant on the 95 percent level in MO1, MO3 and MO4. In all MO:s but one (MO5) standard deviation was lower for women. Both groups kept their initial level of expectation a few minutes ahead of the final examination (compared MO1 and MO4) but made similar plunges of means in absolute numbers (expectations) after the examination was finalized (compared MO4 and MO5). Standard deviation rapidly increased for both groups after M04. The results in Table 4 indicate that both groups are overconfident. However, men indicate a higher tendency for overconfidence in relation to women throughout the course.

| | Expectations | | | | | Performance |
|---|---|---|---|---|---|---|
| MO schedule for academic experience | (1) Start | (2) 2 weeks | (3) 4 weeks | (4) pre-test | (5) post-test | (6) Actual |
| BF (3-years) - number | 34,0 | 27,0 | 21,0 | 33,0 | 33,0 | 34,0 |
| RR (4-years) - number | 10,0 | 11,0 | 9,0 | 11,0 | 11,0 | 11,0 |
| BF (3 years) - mean | 69,2 | 66,7 | 67,2 | 69,8 | 56,7 | 54,4 |
| RR (4-years) - mean | 71,8 | 69,1 | 63,9 | 67,0 | 67,5 | 61,9 |
| BF (3 years) - SD | 10,2 | 10,0 | 9,6 | 8,7 | 13,1 | 16,4 |
| RR (4 years) - SD | 7,8 | 10,0 | 4,9 | 8,0 | 15,4 | 13,3 |

*Table 5. Development of expectations and experience – mean and standard deviation*

More academic experienced students participated to a higher extend in the MOs. Less experienced students kept their initial level of expectations until a few minutes ahead of the final examination (compared MO1 and MO4) but significantly lost momentum after the examination was finalized. More experienced students lost momentum already after the first MO and the mean expectations were actually the same before and after the final examination. Standard deviation did, however, rapidly increased for both groups after MO4. The results in Table 5 suggests that both groups are overconfident. These results were, however, not significant on a 95% level.

#### Test of overconfidence

After testing mean values of expectations Table 6, 7 and 8 present the test of overconfidence over the course for students, gender and academic experience. To do this, expected scores are compared to actual performance and the final and differences are tested through paired sampled t-tests.

| Expectations - actual performance | | | Diff of means | Sig. (2-tailed) |
|---|---|---|---|---|
| Pair 1 | MO1 | MO6 | 14,03 | 0,000 |
| Pair 2 | MO2 | MO6 | 9,39 | 0,002 |
| Pair 3 | MO3 | MO6 | 8,67 | 0,015 |
| Pair 4 | MO4 | MO6 | 12,93 | 0,000 |
| Pair 5 | MO5 | MO6 | 3,20 | 0,183 |

*Table 6. Overconfidence over time, students – significance of means*

The results in Table 6 shows that students suffers from overconfidence during the entire course and these results are statistically significant on a 95% level apart from when they have taken the exam. Overconfidence is highest in the beginning of the course and slowly decrease during the course. However, the bias remains and after the last week of preparation for the written exam the overconfidence once again rise almost to the same level as the initial level. The bias does, however, decrease directly after the final exam as expectations starts to calibrate with the actual level of performance. This result is, however, not significant on a 95% level.

| | | | Diff of means - gender | | |
|---|---|---|---|---|---|
| Expectations - actual performance | | | Men | Women | Sig. (2-tailed) |
| Pair 1 | MO1 | MO6 | 13,86 | 17,05 | 0,568 |
| Pair 2 | MO2 | MO6 | 15,66 | 11,11 | 0,159 |
| Pair 3 | MO3 | MO6 | 15,31 | 7,58 | 0,012 |
| Pair 4 | MO4 | MO6 | 17,46 | 11,67 | 0,024 |
| Pair 5 | MO5 | MO6 | 7,88 | 1,72 | 0,163 |

*Table 7. Overconfidence over time, gender – significance of means*

The results in Table 7 indicate that men and women suffers from overconfidence during the entire course. However, women show that they calibrate downwards after the last lecture of the course (pair 3). In addition, results show that men increase their overconfidence a few minutes ahead of the final exam (M04). These results are statistically significant on the 95 percent level.

| | | | Diff of means - academic exp | | |
|---|---|---|---|---|---|
| Expectations - actual performance | | | Less | More | Sig. (2-tailed) |
| Pair 1 | MO1 | MO6 | 16,27 | 11,70 | 0,488 |
| Pair 2 | MO2 | MO6 | 12,69 | 15,09 | 0,504 |
| Pair 3 | MO3 | MO6 | 13,21 | 9,89 | 0,335 |
| Pair 4 | MO4 | MO6 | 15,79 | 13,00 | 0,353 |
| Pair 5 | MO5 | MO6 | 2,67 | 13,45 | 0,029 |

*Table 8. Overconfidence over time, academic experience – significance of means*

When analyzing the results from Table 8 both less and more academic experienced students seem to suffer from overconfidence during the entire course. The results indicate that less academic experienced students calibrate downwards after the final exam. In addition, the results show that more experienced students instead increase their level of overconfidence after their final exam and this result is statistically significant on the 95 percent level.

To further analyze the development of overconfidence, Figure 1 below depicts the results from an ANOVA test with post-hoc analysis for the four groups.

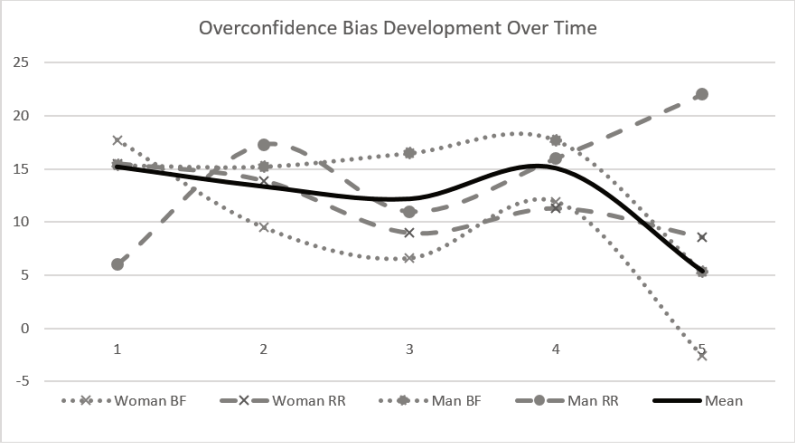


*Figure 1. The development of overconfidence over time*

Once again the result indicate that student overconfidence is high at start and only slowly reduce but gain new momentum ahead of the final examination as shown in MO4 in Figure 1 above. After this, overconfidence remains but on a much lower level then before as shown in MO5.

The results indicate that less academic experienced women, i.e. Woman BF in Figure 1, start off with the highest level of overconfidence but the bias fades over time. However, gain new momentum ahead of the final exam. After the final exam less academic experienced women actually indicate under-confidence as shown in MO5. Less academic experienced men, i.e. Man BF in Figure 1, start with an overconfidence level in-line with the class mean. During the course this bias grows stronger and stronger and indicate the highest level of strength just prior to the final exam. After the examination the level plunges to the level of the class mean as shown in MO5.

More experienced women, i.e. RR Women in Figure 1, start with an overconfidence level in-line with the class mean. However, the bias reduces over the course and converge to the less academic experienced women just prior to final exam. After the final exam, the overconfidence remains but on a lower level as shown in MO5. More academic experienced men, i.e. RR Men in Figure 1, start with a lowest level of overconfidence but ends up with the highest level of overconfidence as shown in Figure 1 when MO1 is compared to M05.

According to the post hoc analysis the difference between less academic experienced men and

women in MO3 was statistically significant on a 95% level. This show that men were more overconfident than women, with the same academic experience, after the last lecture or one week ahead of the final exam. The post hoc analysis also showed statistical significant on the 95% level when it came to the difference between men and women with different academic experience in MO5. This show that after the final exam, men with more academic experience showed overconfidence versus women with less academic experience that showed underconfidence.

## Discussion and suggestions for further research

This paper have reported the analysis of whether finance students suffer from overconfidence or under-confidence when it comes the ability to predict grades. To add to previous research the study was conducted so that level of overconfidence could be measured over an entire course. In-line with previous research, we also controlled for gender and academic experience.

The results of the analysis show that the students did exhibit overconfidence. This result is in-line with previous studies by e.g. Grimes (2002), Koku and Qureshi (2004) and Nowell and Olston (2007). Further, the results show that students on a general level do not calibrate their expectations during the course as a response to their learning since overconfidence remains significant throughout all occations of measment. Results could also to some extent confirm a gender difference in overconfidence as in previous studies by Lundeberg, Fox and Puncohar (1994) and Geffert and Christensen (1998). Male students showed increased overconfidence after the last lecture in the course and before going to take the final exam, compared to initial estimation. In addition, female students showed reduced overconfidence after the final lecture. However, results could not show any other gender effects in the beginning of the course, in the middle of the course nor after the exam was taken. Results also controlled for the academic experience effect reported by e.g. Andersson, Benjamin and Fuss (1994) and Straton and Arnold (2012). There was only a difference between more and less academic experienced students in predicting their exam score after the exam was taken. More academic experienced students did increase their expectations more than the less academic experienced students and ended up further away to the actual final score. This would imply that self-regulated learning feedback is less effective when it comes to more academic experienced students which differs from previous research. The study also found that less academic experienced women even suffered from under-confidence when given self-regulated feedback.

Implications and suggestions for further research based on this study are as follows. Given the effects of self-regulated feedback on overconfidence reported in study, it is key to implement these measures in courses continuously to control the over- or underconfidence of students. If students have a more realistic view of their performance they hopefully will also calibrate the effort needed to meet their potential. It is up to future research to analyze how these feedback measures should be designed given different various degrees of academic experience and other dimensions that might matter.

## References

Andersson, G., Benjamin, D. and Fuss, M. A. (1994) "The Determinants of Success in University Introductory Economics Courses", *Journal of Economic Education*, Vol. 25, No 2, pp. 99-119

Andreou, C., Moritz, S., Veith, K., Veckenstedt, R. and Naber, D. (2014) "Dopaminergic Modulation of Probabilistic Reasoning and Overconfidence in Errors: A Double-Blind Study", *Scizophrenia Bulletin*, Vol. 40, No. 3, pp. 558-565

Armstrong, M. J. (2013) "A Preliminary Study on Grade Forecasting by Students", *Decision Sciences Journal of Innovative Education*, Vol. 11, No. 2, pp. 193- 210

Armstrong, M. J., MacKenzie, H. F. (2017) "Influence of Anticipated and Actual Grades on Studying Intentions", *The International Journal of Management Education*, Vol, 15, No. 1, pp. 49-49

Ballard, C., Johnson, M. (2005) "Gender, Expectations, and Grades in Introductory in Microeconomics in at a US University", *Femininst Economics*, Vol. 11, No. 1, pp. 95-122

Baron, J. (2007) "Thinking and Deciding", *Cambridge University Press*, New York

Brenner, L. A., Koehler, D. J., Liberman, V. and Tversky, A. (1996) "Overconfidence in Probability and Frequence Judgments: A Critical Examination", *Organizational Behavior and Human Decision Processes*, Vol. 65, No. 3, pp. 212-219

de Bruin, W. B., Parker, A. W. and Fischoff, B. (2012) "Explaining Age Differences in Decision-Making Competence", *Journal of Behavioral Decision Making*, Vol. 25, No. 4, pp. 352-360

Cassam, Q. (2017) "Diagnostic Errors, Overconfidence and Self-Knowledge", Palgrave Communications, Vol. 3, pp. 1-8.

Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2951662 2017-05-25

Geffert, B., Christiensen, B. (1998) "Things They Carry: Attitudes Towards, Opinions About, and Knowledge of Library and Research Among Incoming College Students", *References and User Services Quarterly*, Vol. 37, No. 3, pp. 279-289

Grimes, P. (2002) "The Overconfident Priciples of Economics Student: An Examination of Metocognitive Skills", *The Journal of Economic Education*, Vol. 33, No. 1, pp. 15-30

Gustavson, A., Nall, H. C. (2011) "Freshman Overconfidence and Library Research Skills: A Troubling Relationship?", *Collage and Undergraduate Libraries*, Vol. 18, No. 4, pp. 291-306

Hansson, P., Rönnlund, M., Juslin, P. and Nilsson, L.-G. (2008) "Adult age Differences in the Realm of Confidence Judgments: Overconfidence, Format Dependence, and Cognitive Predictors", *Psychology and Aging*, Vol. 23, No. 3, pp. 532-544

Hattie, J., Timberley, H. (2007) "The Power of Feedback", *Review of Educational Research*, Vol. 77, No. 1, pp. 88-112

Hilbert, M. (2012) "Towards a Synthesis of Cognitive Biases: How Noicy Information Processing can Bias Human Decision Making", *Psychological Bullitin*, Vol. 138, No. 2, pp. 211-237

Ho, P-H., Huang, C-W., Lin, C-Y. and Yen, J-F. (2016) "CEO Overconfidence and Financieal Crises: Evidence From Bank Lending and Leverage", *Journal of Financial Economics*, Vol. 120, No. 1, pp. 194-209

Hossain, B. Tsigaris, P. (2012) "Are Grade Expectations Rational?: A Classroom Experiment", *Education Economics*, Vol. 23, No. 2, pp. 199-212

Hribar, P., Yang, P. (2016) "CEO Overconfidence and Management Forecasting", *Contemporary Accounting Research*, Vol. 33, No. 1, pp. 204-227

Kahnemann, D., Tversky, A. (1979) "Prospect Theory: An Analysis of Decision Under Risk", *Econometrica*, Vol. 47, No. 2, pp. 263-292

Koku, P. S., Qureshi, A. A. (2004) "Overconfidence and the Performance of Business Students on Examinations", *Journal of Education for Business*, Vol. 79, No. 4, pp. 217-224

Lewis, A. (2008) "The Cambridge Handbook on Psychology and Economic Behaviour", *Cambridge University Press*, New York

Lundeberg, M. A., Fox, P. W. and Puncohar, J. (1994) "Highly Confident but Wrong: Gender Differences and Simililarities in Confidence Judgements", *Journal of Educational Psychology*, Vol. 86, No. 1, pp. 114-121

Malmendier, U., Tate, G. (2005) "CEO Overconfidence and Corporate Investment", *Journal Finance*, Vol. 60, No. 6, pp. 2660-2670

Nowell, C., Alston, R. M. (2007) "I Thought I got an A! overconfidence Across the Economics Cirriculum", *The Journal of Economics Education*, Vol. 38, No. 2, pp. 131-142

Owen, A. L. (2001) "Grades, Gender, and Encouragement: A Regression Discontinuity Analysis", *The Journal of Economic Education*, Vol. 41, No. 3, pp. 214-234

Pallier, G., Wilkinon, R., Danthiir, V., Kleitman, S., Knezevik, G., Stankow, L. and Roberts, R. (2002) "The Role of Individual Differences in the Accuracy of Confidence Judgments", T*he Journal of General Psychology*, Vol. 129, No. 3, pp. 257-299

Plous, S. (1983) "The Psychology of Judgment and Decision Making", *McGraw Hill*, New York

Schartel, S. A. (2012) "Giving Feedback – An Integral Part of Education", *Best Practice and Research Clinical Anesthesiology*, Vol. 26, No. 1, pp. 77-87

Shute, V. (2008) "Focus on Formative Feedback", *Review of Educational Research*, Vol. 78, No. 1, pp. 153-189

Skata, D. (2008) "Overconfidence in Psychology and Finance: An Interdiciplinary Literature Review", *Bank I Kredyt*, No. 4, pp. 33-50.

Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1261907 2017-05-25

Straton, I J. M., Arnold, J. T. (2012) "Motivation and Math Skills as Determinants of First-Year Performance in Economics", *The Journal of Economic Education*, Vol. 43, No, 1, pp. 33-47

Tversky, A., Kahnemann, D. (1974) "Judgement under Uncertainty: Heuristics and Biases", *Science*, Vol. 185, No. 4157, pp. 1124-1131

4

# Assessing Knowledge Through Written Reviews of First-Year Programming Students

Emil Folino
*Blekinge Tekniska Högskola*

emil.folino@bth.se

**Abstract**

In this paper, a method of qualitative assessment of programming students' knowledge and comprehension is investigated. The qualitative assessment is done by reading students' review texts from three subsequent courses' individual programming project. The review texts are analyzed according to the SOLO Taxonomy and the students are awarded a SOLO level of Unistructural, Multistructural or Relational. The SOLO level is compared to the final grade of the three courses and a relation between a student's final grade and the SOLO level is shown. Furthermore, a positive progression in the students' comprehension and understanding of the course material is observed as the students progress through the three subsequent courses. A recommendation is given to complement programming exercises with written assignments where the students get an opportunity to reflect and expand on the completed exercises.

## Introduction

It is often easy to quantitatively assess students in programming and computer science courses. Have the students completed the assigned tasks? Does the application work as it is supposed to? Does any software tests fail? These questions are easily answered by automated procedures or a simple yes or no. However assessing the students' comprehension and understanding of the completed assignments and carry out a qualitative evaluation is harder [1]. The SOLO Taxonomy was proposed by Biggs and Collis in 1982. SOLO is abbreviated from Structure of the Observed Learning Outcome and the SOLO Taxonomy is used to qualitatively assess students' assignments. The SOLO Taxonomy consists of five levels of understanding: *Prestructural*, *Unistructural*, *Multistructural*, *Relational* and *Extended Abstract*. In the five SOLO levels are exemplified in section III.

In [2] McCracken et al. evaluated first year Computer Science students' programming competency. Several universities took part in the study that showed disappointing results. To reverse the disappointing results a framework outlining the expectations of first year Computer Science students is proposed. However, the assessment is carried out in a quantitative way and there are no recommendations for any qualitative assessment methods.

Paulo Blikstein [3] uses snapshots of students' code during programming assignments together with data mining to automate a technique to assess and analyze students learning programming. It is concluded that together with other data sources (Blikstein propose interviews, tests and surveys) the automated assessment can give insights into the understanding of the learning of programming. The assessment of the students' code is done entirely with quantitative methods. However, the proposed interviews, tests and surveys can give qualitative insights into the students' learning.

In [4] Lister et al. study written and think-aloud responses to exam questions from computer science exams. The authors analyze the responses according to the SOLO Taxonomy and conclude that experienced programmers more frequently answered with SOLO Relational responses compared to the Multistructural responses given by novice programmers. Lister et al. recommend that the students are given written assignments together with programming assignments. The written assignments make it easier to evaluate the understanding and comprehension the students obtain in programming courses.

The study in this paper builds upon the results outlined by Lister et al. in [4] by analyzing review texts written by first-year programming students. The review texts are handed in as a complement to the source code of programming projects. The review texts together with weekly assignments and the project at the end of the course forms the basis for the final grade given in the course.

### Method

In the daily teaching at Webbprogramming (dbwebb.se) at Blekinge Tekniska Högskola (BTH) the students hand in programming assignments every week. Together with the exercises they hand in a review text, answering 3-5 questions centered around the topics and assignments of the week. At the end of each 10 week study period the students hand in an individual project together with an extended review text. The students are graded both with regards to the completed assignments and the review texts. The following web pages [5] and [6] from the educational program's website explains how the students are graded on their review texts according to the SOLO Taxonomy.

In this paper the review texts of three subsequent courses' projects are analyzed according to the SOLO Taxonomy. The three courses are given in three consequtive study periods during the first year of studies. The three courses are "Databases, HTML, CSS and script based PHP-programming" (htmlphp), "Programming with JavaScript, HTML and CSS" (javascript1) and "Programming webservices in Linux" (linux)[1]. The htmlphp course is a beginners course and the course starts at a very basic level. The students' programming skills range from complete novices to experienced programmers looking to broaden their skills. The javascript1 and linux courses builds upon the knowledge and practical skills obtained in prior courses.

The students publish the completed assignments and review texts on a webserver provided by Blekinge Tekniska Högskola. The author fetched the review texts by using an automated tool programmed by the author. The automated tool fetch the review texts from the webserver. The review texts are stored in a database together with the website URL of the published project and an anonymous, but traceable, reference to the student. The review texts are fetched in a manner that removes the name and student acronym from the review texts to ensure anonymity in most cases.

The SOLO analysis is done manually by the author by reading the review texts. The students are given a grade of 1-5 according to the five levels of the SOLO Taxonomy, *Prestructural* (1), *Unistructural* (2), *Multistructural* (3), *Relational* (4), and *Extended Abstract* (5).

In [1] Biggs and Collis give recommendations for applying the SOLO Taxonomy to different teaching subjects. Programming is not mentioned among the subjects but the authors give examples of applications in technical subjects for example elementary mathematics. For the study reported in this paper the SOLO analysis was adapted to the technical language used in the students' review texts. In section III an explanation of each SOLO level is shown together with examples from the students' review texts.

After the analysis of the review texts the SOLO level is compared to the final grade of the course. The students' final grade is fetched and stored in a database table together with the same traceable reference to the students used in the database table where the SOLO level is stored. The final grade for the course and the SOLO level is compared to evaluate if there is a relation between the SOLO level and the final grade given to the student.

The courses are graded based on the following criteria with a maximum of 100 points in total. 30 points for the completion of the six weekly assignments, 10 points for weekly extra assignments and extraordinary weekly review texts. 10 points for each mandatory requirement of the project and 10

points for each optional requirement of the project. The students are awarded credits on the ECTS-credit scale where more than 90 points equals an A, more than 80 equals B etc.

As the review texts are taken from three subsequent courses the progression of the students' understanding of the course material and programming in general can be investigated.

The automated tool can be found at the author's Github page [2].

### Examples of SOLO levels

In this section a short explanation of each SOLO level is given and examples of the review texts are shown for each level of the SOLO Taxonomy. The review texts have been translated from Swedish to English by the author. The original Swedish texts are found in appendix A.

#### Prestructural

No answer more than repeating the question. The student is failed based on the text. The Pre-structural SOLO level is used for students that have not handed in a project.

#### Unistructural

The answer contains no technical description of how the solutions have been implemented.

*Example:* "The search feature has it's own page (accessible from the navbar) where you can search for articles and object descriptions with a word consisting of letters(a-ö) and numbers(0-9). The search results are presented in a list and the first result in an article/object description is marked with and part of the text."

#### Multistructural

The student gives a technical explanation of the implementation, more or less line of code by line of code, but does not relate the implementation to other parts of the code or prior exercises in the courses.

*Examples:* "For every iteration in the loop an object is appended to slar.json. So when the loop is done it was just adding the last parenthesis and clean the file so it is valid JSON. I thought this requirement was kind of complicated and my solution is absolutely not the fastest but it does what it is supposed to do and that should suffice."

"I have tried to use a lot of built-in functions like 'map', 'reduce', 'filter', etc. to get the correct information from the arrays I use."

#### Relational

The student gives a technical explanation of the implementation and justifies their choices through related course material or real-world examples.

*Examples:* "I chose to not split my client as it is done in the Gomoku assignment [3]. I know that the reason is to separate general and domain-specific code, but I am not going to extend on this client so I have decided to put all the code in the same file."

"With earlier assignments' clients as a base I did a client that tests the server."

#### Extended Abstract

No examples of Extended Abstract texts were found in the review texts. The students are first-year students and are not asked to produce novel material.

#### Results

In table 1 the SOLO level distribution is shown for the three courses: htmlphp, javascript1 and linux. In table 2 a relative distribution of the SOLO levels is shown.

Table 1: Number of each SOLO level for the Courses

| Course | Relational | Multi | Uni | Total |
|---|---|---|---|---|
| htmlphp | 4 | 13 | 5 | 22 |
| javascript1 | 6 | 8 | 2 | 16 |
| linux | 9 | 9 | 0 | 18 |

*Table 1: Number of each SOLO level for the Courses*

| Course | Relational | Multi | Uni |
|---|---|---|---|
| htmlphp | 18% | 59% | 23% |
| javascript1 | 37.5% | 50% | 12.5% |
| linux | 50% | 50% | 0% |

*Table 2: Relative distribution of SOLO levels for the Courses*

The total number of students in figures 1, 2 and 3 and tables 3 and 4 does not correspond with the number of students for each course in table 1. Not all students have received a grade in the courses because the students have not completed the mandatory oral presentation of their project.

Figure 1 shows the SOLO level compared to the grade that the student obtained in the htmlphp-course.

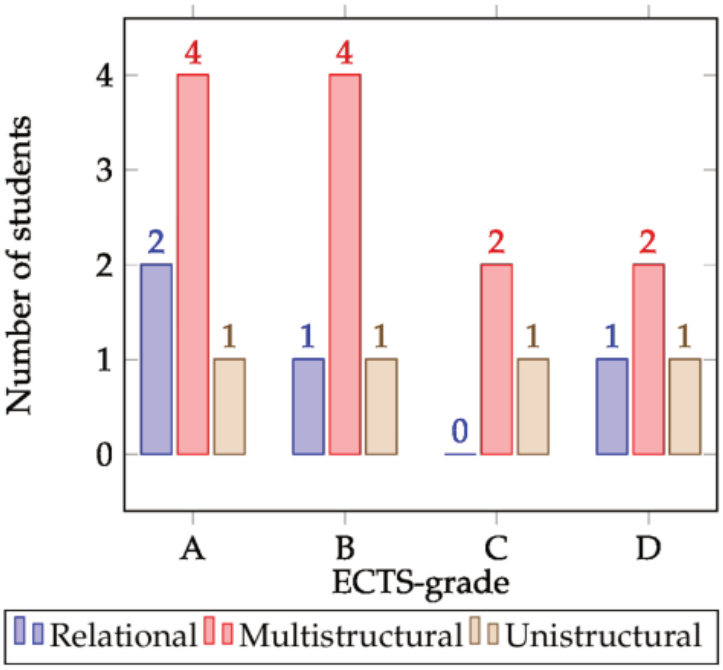


*Figure 1: SOLO level compared to final grade in htmlphp.*

Figure 2 shows the SOLO level compared to the grade that the student obtained in the javascript1-course.

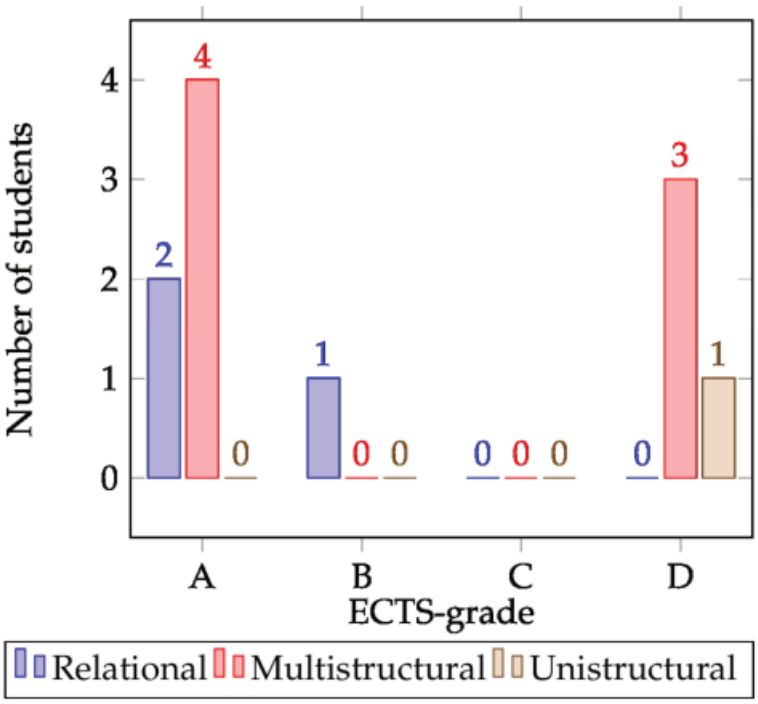

*Figure 2: SOLO level compared to final grade in javascript1.*

Figure 3 shows the SOLO level compared to the grade that the student obtained in the linux-course.

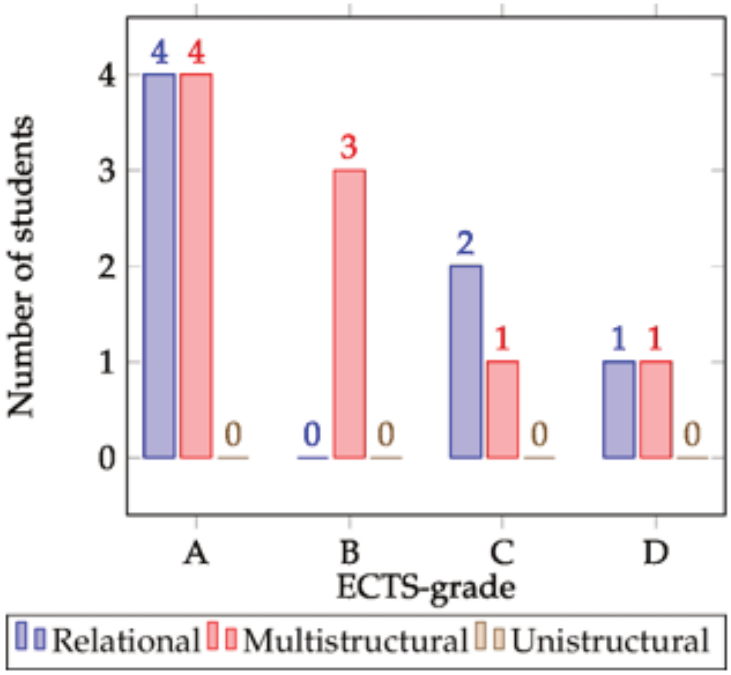


*Figure 3: SOLO levels compared to final grade in linux.*

A comparison of SOLO levels and final grades are summarized in table 3 and a relative comparison of SOLO levels and final grades is shown in 4.

| Grade | Relational | Multi | Uni |
|---|---|---|---|
| A | 8 | 12 | 1 |
| B | 2 | 7 | 1 |
| C | 2 | 3 | 1 |
| D | 2 | 6 | 2 |

*Table 3: Comparison of SOLO levels and final grades.*

| Grade | Relational | Multi | Uni |
|---|---|---|---|
| A | 38% | 57% | 5% |
| B | 20% | 70% | 10% |
| C | 33% | 50% | 17% |
| D | 20% | 60% | 20% |

*Table 4: Comparison of SOLO levels and final grades shown relative to the number of students receiving the grade.*

### Discussion

In table 3 and 4 it can be observed that students receiving a final grade of A in the courses are more likely to hand in a Relational review text than students with a final grade of B or D. Relatively few students received a final grade of C and therefore the grades are skewed towards a higher number of Relational review texts. Despite the skewed distribution of grades students receiving a C there is relatively more A students handing in Relational review texts. Furthermore it can be observed that the students with a final grade of A to a lower degree than any other final grade answers with a Unistructural review text. This confirms the relation between a higher grade and a deeper understanding and comprehension of the course material. The correlation between the SOLO level and the final grade is not statistically significant, but a relation between final grade and SOLO level is shown. The more easily analyzed and classifiable dataset discussed above is similar to the material used in [4]. The correlation shown in [4] between SOLO responses and the level of programming experience is statistically significant.

As a teacher you hope to see a positive trend in your students comprehension of the course material. The SOLO Taxonomy can be used to evaluate the comprehension and understanding that the students have obtained [1]. The htmlphp course given in study period 1 is a beginners course in programming, but the actual level of the students differs from complete novices to more advanced programmers. At the end of the linux course in study period 3 the students have studied 45 ECTS-credits worth of programming courses. The students are now familiar with at least 10 programming languages and technologies and have moved from novices to more advanced programmers. According to Lister et al. in [4] it is expected that advanced programmers answer with Relational answers and the results of this study confirms the results of the study in [4]. In tables 1 and 2 it is observed that the number of Relational review texts increase as the students progress through the three subsequent courses. Furthermore in tables 1 and 2 it is shown that the relative number of students answering with Multistructural review texts is constant and it is the Unistructural review texts that decreases as the Relational review texts increase. This further strengthens the observation that the students have a higher comprehension and understanding of the course material and programming in general.

During the study three of the author's colleagues read a subset of 20 % of the review texts and did their own analysis. The SOLO level for three of the 12 analyzed review texts differed by one SOLO level. In all three cases the author had given the students a higher SOLO level than the colleagues. The colleagues highlighted that the analysis were difficult as the written texts are long form and parts of the texts is a Multistructural review text and other parts is a Relational review texts. The author experienced the same classification problems during the analysis of the entire dataset. Shorter and more concise answers to specific problems would yield more easily analyzed and classifiable material.

## Conclusion

In [4] the authors Lister et al. conclude that as a complement to programming exercises students should be assessed on written or think-aloud responses to programming assignments. The study described in this paper confirms that the final grade in programming courses relate to the level of understanding of the course material. Therefore written assignments or review texts together with programming assignments can be used as a qualitative method in the assessment process. A recommendation is to complement programming exercises with written assignments where the students get an opportunity to reflect and expand on the completed exercises. This introduces a qualitative complement to the conventional quantitative assessment methods used in programming courses and computer science in general.

Furthermore it is observed that the general level of comprehension and understanding of programming increases during the three subsequent courses that were analyzed in this paper. At the beginning of the educational program and the htmlphp course the level of programming skills varies a lot in the student cohort. At the end of the study period 3 and the linux course we see a more homogenized student cohort which further strengthens the notion that the general level of understanding and comprehension of the course material and programming in general is higher.

## Future Work

Based on the observations of the both the author and the author's three colleagues a more suitable dataset would be shorter more concise answers to specific questions. Another study conducted with this type of dataset would be more easily analyzed and probably yield clearer results.

## Acknowledgements

Thanks to Andreas Arnesson, Kenneth Lewenhagen and Mikael Roos at Blekinge Tekniska Högskola for helping to ensure even SOLO levels by reading and analyzing a subset of the review texts.

## References

[1] Biggs JB, Collis KF. Evaluation the Quality of Learning: The SOLO Taxonomy (structure of the Observed Learning Outcome). Educational psychology. Academic Press; 1982.

[2] McCracken M, Almstrum V, Diaz D, Guzdial M, Hagan D, Kolikant YBD, et al. A multi-national, multi-institutional study of assessment of programming skills of first-year CS students. ACM SIGCSE Bulletin. 2001;33(4):125–180.

[3] Blikstein P. Using Learning Analytics to Assess Students' Behavior in Open-ended Programming Tasks. In: Proceedings of the 1st International Conference on Learning Analytics and Knowledge. LAK '11. New York, NY, USA: ACM; 2011. p. 110–116.

[4] Lister R, Simon B, Thompson E, Whalley JL, Prasad C. Not seeing the forest for the trees: novice programmers and the SOLO taxonomy. ACM SIGCSE Bulletin. 2006;38(3):118–122.

[5] Roos M. Att skriva en bra redovisningstext;. Accessed: 2017-02-10. https://tinyurl.com/ldk7jon.

[6] Roos M. Varför regnar det på bergssidan – ett exempel på SOLO taxonomin;. Accessed: 2017-02-10. https://tinyurl.com/mv3uewr.

### A Original Swedish Texts

The original Swedish review texts and their corresponding English translations is shown below.

*Original Swedish Text:* Sökfunktionen finns som egen sida(via navbar) där man kan söka i artiklar och i objektsbeskrivningar med ett ord som utgörs av bokstäver(a-ö) och siffror(0-9). Träffarna presenteras i en lista och första träffen i en artikel/objekt markeras med gult och del av texten.

*Translation:* The search feature has it's own page (accessible from the navbar) where you can search for articles and object descriptions with a word consisting of letters(a-ö) and numbers(0-9). The search results are presented in a list and the first result in an article/object description is marked with and part of the text.

*Original Swedish Text:* För varje varv i loopen appendas ett object till salar.json. Så när loopen är färdig så var det bara att lägga till de sista paranteserna och städa upp filen så att det skulle validera som JSON. Jag tyckte detta krav var ganska krångligt och min läsning är absolut inte den snabbaste men den gjorde vad den skulle och det fick vara bra nog.

*Translation:* For every iteration in the loop an object is appended to slar.json. So when the loop is done it was just adding the last parenthesis and clean the file so it is valid JSON. I thought this requirement was kind of complicated and my solution is absolutely not the fastest but it does what it is supposed to do and that should suffice.

*Original Swedish Text:* Jag har försökt att använda mycket inbyggda metoder som 'map', 'reduce', 'filter', etc. för att få ut rätt information från de arrays jag använder.

*Translation:* I have tried to use a lot of built-in functions like 'map', 'reduce', 'filter', etc. to get the correct information from the arrays I use.

*Original Swedish Text:* Valde att inte dela upp min klient som det är gjort i Gomoku. Jag vet att anledningen var för att kunna hålla isär generell och domänspecifik kod, men eftersom jag inte tänkt bygga vidare på den här klienten så lägger jag allt i samma.

*Translation:* I chose to not split my client as it is done in the Gomoku assignment [4]. I know that the reason is to separate general and domain-specific code, but I am not going to extend on this client so I have decided to put all the code in the same file.

*Original Swedish Text:* När jag bestämde stil för sidan kollade jag runt lite på andra webbplatser som har en koppling till begravningar.

*Translation:* When i decided on the style for the page i looked at other websites with a connection to funerals.

*Original Swedish Text:* Med tidigare uppgifters klienter som grund gjorde jag en klient som kan testa servern.

*Translation:* With earlier assignments' clients as a base I did a client that tests the server.

[1] The course names have been translated by the author from "Databaser, HTML, CSS och skriptbaserad PHP-programmering", "Programmering med JavaScript, HTML och CSS" and "Programmera webbtjänster i Linux".

[2] https://github.com/emilfolino/pedagogy

[3] A programming assignment earlier in the course

[4] A programming assignment earlier in the course