Course Analytics

Leveraging course assessment through data mining for student satisfaction

ANDREI NECULAU

andrei.neculau@gmail.com

Master's Thesis at Department of Computer and Systems Sciences
Stockholm University (SU) / Royal Institute of Technology (KTH)
Stockholm, SWEDEN

Supervisor: Martin Hassel - xmartin@dsv.su.se
Examiner: Martin Hassel - xmartin@dsv.su.se

TRITA xxx yyyy-nn

September 6, 2010
This master's thesis corresponds to 30 credits — 20 weeks of full-time work.

Abstract

This thesis focuses on combining the perspective of educational service quality with current technological means of assessing the delivery of this type of service. A service quality perspective is chosen over teaching effectiveness, given the latter's proved limitations and trends in educational provision. The Importance-Performance Analysis (IPA), a core model of service quality research, paves the ground for a novel assessment method. An artifact is built to enhance IPA's data collection process by switching from survey-based to data mining techniques, associating importance with usage metrics and performance with opinion metrics. Data collection is instantiated inside an existing academic information system. Both students and instructors are ultimately surveyed on their views on current end-of-course evaluations, on the associative strength between data mining metrics and IPA metrics and on the use of IPA in the academia. Results show that there is a desire for change from end-of-course evaluation, which presents challenges both to students and instructors. Moreover, we observe student interest and confidence into metric associations, while instructors show an undecided approach towards both these associations and to IPA results.

Keywords: academic quality, academic evaluation, service satisfaction, educational data mining, importance-performance analysis, teaching effectiveness

Acknowledgements

I cannot do without bowing to:

Martin Hassel
for his confidence and encouragement, his good spirits and vital analytical support

Aron Henriksson
for his systematic and meticulous comments and for raising the bar of writing skills

the thesis seminar group
for their useful comments, easy and honest discussions, and for the nice ambiance

the survey respondents
for taking the time to go through my questions

Sven Olofsson
for whitelisting the Snowdrop Information System on KTH/SU servers and thus allowing my study to stand a fair chance

Magnus Boman, *Harald Kjellin* and *Martin Hassel*
for allowing me to introduce my thesis, Snowdrop and the short-term study to prospective study participants

Ruxandra Teodoru & *Cătălin Codrean*
for the chit-chats that took my mind off my thesis when dearly in need

Roula Mecherkany
for your love and support

Contents

List of Figures

Chapter 1

Introduction

1.1 Background

Progress entails assessing current status, identifying weaknesses and strengths. Educational progress follows the same path.

Assessment methods vary in concept, in implementation, in time and in assessed elements. Most commonly, both in the academic and in the research medium, end-of-course evaluations are employed with the scope of summarizing course delivery experiences, sometimes both students' and peers' experiences. To a great extent, the subject of the assessment is *teaching effectiveness*, with most of the research endeavors and their practice within institutions having a *product*-oriented approach to quality over time by looking at the end-product—the graduate (see section 2.1). The dimensions of this *teaching effectiveness* are disputed, both in type and in number (Shevlin et al., 2000), leaving institutions and instructors in the dark when aiming for reliability and validity (Feldman, 1977). Evaluation tools follow the same trend of being exhaustive, yet unreliable (O'Neill and Palmer, 2004).

Technology is ubiquitous and it has made its way into the educational environment. A new way of collecting information has emerged as *Data Mining*. In short, this new field entails looking for answers by collecting and processing large amount of data (Romero et al., 2008). Looking at students' actions inside an information system can be an evaluation complement to the student ratings by collecting different types of information during the course delivery. This may help evaluate the instructional design and materials or adjust itself automatically to bring a tailored experience to the learner. *Educational Data Mining* is a new field and its research directions, while numerous, are limited to providing data for the above *product*-oriented evaluation—assessment of the graduate and, implicitly, of *teaching effectiveness*—ignoring alternative emerging dimensions.

With education provision expanding to a large-scale service, *educational quality* is starting to be redefined by *service quality*, and students being redefined as primary customers with rights. It is these customers that define quality as fitness for use and assess it by building expectations, as they enter the service process, and satisfactions, at the end of it. While dependent on many factors and in need of careful handling, data regarding expectations is valuable for creating a unique advantage (Sander et al., 2000).

At this moment, "universities aren't coming close to meeting their own expectations for what should be happening on campus" although they are reacting in response to concerns over course evaluations, which gave birth to a *nonaggression pact* (O'Brien, 2010). Despite challenges ahead, it is thus even more important to address as many perspectives of *educational quality* as possible, and envision data collection and assessment methods beyond conventional thought. The concept of *quality* is ever variable as ever dependent (see chapter 2) and thus the quality of the educational process becomes more of *objet petit a* (Lacan, 1975), an object of desire which we seek in the other. Keeping to only one perspective—*teaching effectiveness*—would not only be synonymous to taking an untrustworthy direction (see section 2.1), but it would also confine the progress of education to a volatile development. Furthermore, in a world of rankings giving a glimpse of the quality level, *educational quality* must be taken outside common practices since research shows that increased educational institutional emphasis on teaching evaluations, and research productivity alike, can be a path to a lower teaching effort, and to a lower qualitative education, inherently (Love and J Kotchen, 2010). Yet addressing alternative perspectives requires adequate assessment, and while the importance of the aforementioned service quality aspects bears little disagreement, identifying and implementing appropriate measurement tools remains a challenge (see section 2.3).

The evaluation of educational service-quality, which translates to the assessment of the graduate's satisfaction, has been approached through various models (see section 3.1) used in service quality research, all within the core *confirmation-disconfirmation paradigm* (Bitner and Hubbert; Churchill and Surprenant; Oliver; Oliver and DeSarbo; Rust and Oliver; Tse and Wilton; Yi in Appleton-Knapp and Krentler, 2006). Confirmation is seen as validation of pre-consumption expectations or standards, which creates satisfaction, while disconfirmation is the fall of performance below the standard, giving in to dissatisfaction (Wells and Prensky; Oliver in O'Neill and Palmer, 2004). Despite these various models, service-quality evaluation has yet to find a framework for making use of data mining techniques within the educational paradigm, since this topic has not been approached by the research community until now.

Education provision is expanding in Sweden as well, slowly but steadily becoming a global market. It is one of the few countries in the world that does not charge fees for foreign students and this plays as a primary attraction for a vast majority of students, seconded by high rankings of Swedish universties (QS Quacquarelli Symonds Limited, 2010; Shanghai Ranking Consultancy, 2009). This will change with 2011 admission rounds, when Swedish institutions of higher education will set tuition fees for students outside the European Union (EU) and the European Economic Area (EEA) (Tranæus, 2010). Tobias Krantz, current Minister for Higher Education and Research, expressed his view on the new law on tuitions fees by saying "Students choose among universities worldwide. On the market, I think Sweden shall compete by having colleges of high quality, and not by free education" (Utbildningsdepartementet, 2009). The view is shared by the Swedish Institute Manager of Study in Sweden, Niklas Tranaeus (2010), and others.

With newly arising business strategies and with new means of data collection becoming available, the IT strategies of the educational institutions, Swedish ones alike, are bound to align themselves and make use of technological resources to fulfill their business goals (Henderson and Venkatraman, 1993). Given the aforementioned developments of education provision in Sweden, implicitly arguing for a higher educational quality, the environment advances as a feasible ground for pursuing alternative perspectives regarding the concept and the assessment of *educational quality.*

1.2 Problem Statement

Current educational data mining research focuses on a concept that lacks a widely recognized definition—*teaching effectiveness*—and does not take into account another concept that is highly recognized—*service quality.*

New means of data collection have not been put to use in an academic environment for the purpose of evaluating student satisfaction, and, in the bigger picture, for the purpose of assessing the quality of the educational service from more than one perspective. Furthermore, new technology can provide the means to move away from the controversy of summative evaluations (Love and J Kotchen, 2010) (see subsection 2.3.1) to formative ones (see subsection 2.3.2).

1.3 Goal and Purpose

This research attempts to put forward a method of assessing service quality in an educational environment. This method has a broad scope and is highly original, providing an approach suitable for many types of educational institutions by combining research studies and concepts from various domains. Based on research in the service quality domain (see section 2.2), and positive results of applications in the educational environment (see chapter 3), a previous assessment method is extended in order to make use of new data collection methods, in the form of usage and opinion mining. The ultimate goal is to address primary aspects of this method, such as exploring its feasibility and its implementation and identifying its benefits and beneficiaries, its obstacles and their scale.

Tapping into current technological and service-oriented trends fills in a gap both in the academic and in the research environment, serving as a purpose of this thesis. Traditional face-to-face education will not disappear in the near future, but will be complemented more and more by blended or computer-based courses. This opens up possibilities not only in assisting educational evaluation, but also to help redefine what *educational quality* is.

1.4 Method Description

The present work strives to follow closely the guidelines of design science in information systems research (Hevner et al., 2004) on understanding, executing and evaluating the production of an artifact—the assessment model. This is due to the goal of providing utility by building and evaluating and not through analysis of previous studies, which would classify as behavioral science. This conformation assumes seven guidelines, out of which six are fulfilled, arguably with different magnitudes, by

1. designing an artifact—here, an evaluation method;
2. addressing a relevant problem that transcends technological boundaries—here, an educational problem;
3. evaluating the design—here, by using experimental and descriptive methods;
4. contributing to existing research—here, by an instantiation artifact and by building a methodological foundation;
5. designing as a search process—here, searching for an adequate subset of technological and evaluation means;
6. and communicating the research with an adequate language for both technological-oriented and management-oriented audiences.

As a research method, this study performs as a deductive and quantitative one. The assessment method put forward in this thesis will make use of an instantiation process, which poses as a feasibility and implementation test, and empirical data collected through surveys, which fulfills evaluation purposes related to obstacles and benefits, attitude and availability of the main beneficiaries.

The method is built under the *confirmation-disconfirmation paradigm*. While at the core of service quality research, this defines quality as being the relation between service pre-consumption expectations and post-consumption satisfaction. The artifact constructed by this research is based on a disconfirmation model of this paradigm—the Importance-Performance Analysis (IPA)—which was successfully tested in the educational environment, and acts as an extension of the model by shaping new ways of collecting data for its analysis. Other models have been considered and discussed, before focusing on the IPA model (see section 3.1).

This artifact is then incorporated into an academic information system (see subsection 4.1.1) that allows for the collection of usage and opinion data of each user, thus producing a secondary artifact—an instantiation of the method. This system interfaces with most of the official information systems of the Computer and Systems Science (Swedish abbrev. DSV) department—the educational campus chosen for this study.

The DSV department is shared between two Swedish educational institutions that will also need to align their educational assessment with the new political, business and technological trends. The two institutions, Stockholm University (SU) and Royal Institute of Technology (Swedish abbrev. KTH), assume a great deal of information cohesion and therefore make use of a large number of informational systems, from regular websites and email communications to proprietary administrative systems—Daisy (2010)—, advanced course management systems—Moodle (2010)—, learning management systems—Ping Pong (2010)—, online conferencing systems—FirstClass (2010)—and resource planning systems—TimeEdit (2010). Students interact on a regular basis with these systems, in search of needed academic information. This poses as an appropriate medium for gathering data regarding their educational expectation and satisfaction through data mining means. Furthermore, given the developments within the Swedish education provision (see section 1.1), the DSV department will need to intensify the transition to computer-based distance course delivery in order to alleviate tuition fees. Most importantly, the department, just like any other Swedish educational institution, will need to aim for a higher quality of its teaching process, both under the old paradigm of teaching effectiveness and under new ones, like student satisfaction, if it wants a comprehensive quality progress. Swedish academia already receives the world's highest spending on research and development in relation to its Gross Domestic Product (Organization for Economic Co-operation and Development in Tranæus, 2010). Without a proper quality shift and visibility, international students that were previously attracted by the free tuition will now be severely discouraged when facing tuition fees ranging from 80.000 to 180.000 SEK per year in some institutions (Waara, 2010).

Beyond the aforementioned instantiation, the primary actors of education provision within the DSV department, students and course instructors, will provide final empirical data that will help evaluate the assessment method by confirming some of the theoretical premises, awareness of limitations in current evaluation methods, interest in alternative evaluation methods, comprehension of the suggested method, and by highlighting some of the issues and experiences of interacting with the instantiation itself and with its reports.

1.5 Limitations

This study is based on only one possibility of implementing service quality evaluation through data mining—best suited given the online nature of most course management systems available. Moreover, generalization to different environments than the one observed in this study may not apply (Markus et al. in Hevner et al., 2004). Due to this, the research is far from extensive in its implementation. It is worthwhile mentioning that collected data would be on a larger scale in a cloud-computing information system, where all of the academic events and resources would be available only on demand and where user interactions could be monitored with greater accuracy. The limited time to perform data collection, the limited number of students and instructors, the relatively low number of respondents and the subjective nature of the primary targeted element—satisfaction regarding the educational service quality—also contribute to a reserved thesis outcome.

The above are complemented by general limitations of design science in information systems research (Hevner et al., 2004). One such limitation is that the existing knowledge base is not capacious and that a constructed artifact relies to a great extent on the designer's comprehension of the problem and of the solution. More than others, this research suggests a new perspective and a new method to fit the perspective, extending the limitations even more. Nevertheless, developing and implementing prototype artifacts are of great importance since it helps to learn in depth about the problem, its environment and possible solutions (Newell and Simon in Hevner et al., 2004).

1.6 Disposition

This introductory first chapter is followed by an extensive background of educational quality, its perspectives and its assessment methods. Chapter 3 focuses on a disconfirmation model to address the assessment of educational quality from a satisfaction perspective and suggest new technological means for collecting data to analyze using the model. This primary artifact is then instantiated. The development and the means of evaluation are detailed in Chapter 4, followed by the analysis of collected data in Chapter 5. Finally, this thesis is recapitulated and its produced results are discussed.

Chapter 2

Academic Quality

Quality has been of interest to humanity for millennia now, the term being said to have been coined as early as Cicero's times (Harper, 2010c). The same is valid for *education*, a term residing well before the 16th century, but defined as "systematic schooling and training for work" only in the 1610s (Harper, 2010b), just after Shakespeare associated *to educate* with *to "provide schooling"* in 1588 (Harper, 2010a).

The mixture of *quality* and *education* lead to debate and it only caught interest and priority in the early 1900s. Under the title *Academic Efficiency*, Kent (1912) is arguing for taking on "the great educational experiment which will lead to improving the methods of training our future citizens". This experiment continues even today.

Despite the early origins of the concept of *quality*, its meaning is ever variable as ever dependent. However, at its most basic, quality is *conforming to requirements* (Crosby in O'Neill and Palmer, 2004). This implies defining requirements and specification and then achieving quality by following the specifications. Yet whose requirements and whose specifications (Palmer et al. in O'Neill and Palmer, 2004)?

2.1 Traditional Product-Oriented Educational Quality

Until a couple of decades ago, the term *quality* did not exist in the vocabulary of most universities (O'Neill and Palmer, 2004). And when it did exist, it was an approximate synonym with *teaching effectiveness*. This can be regarded as a product-oriented approach to quality, a true duty of universities as stated by some researchers (Emery et al. in O'Neill and Palmer, 2004). This in turn would hold programs responsible for "producing knowledgeable, effective students who possess skills and talents valued by public and private corporations" (Emery et al.

in O'Neill and Palmer, 2004). Failure to follow this approach, would directly mean that "pedagogy becomes entertainment" (Franz in O'Neill and Palmer, 2004).

2.1.1 Teaching Effectiveness

Teaching encompasses material, methods and instructor, to name a few of the most abstract dimensions. In order to assess *its* effectiveness, the effectiveness of *its dimensions* needs to be defined and clearly distinguished. Yet this is an overwhelming challenge as there is little agreement on the nature and the number of dimensions that represent teaching effectiveness (Patrick and Smart in Shevlin et al., 2000). Since 1976, research has reached non-linearly from 2 dimensions to 9, 22 or even 35 dimensions (Marks, 2000).

To elaborate, this labyrinth is even more intricate as research does not seem to progress by continuing previous work, but rather by providing different perspectives on the matter. For example, Swartz et al. identified (1) clear instructional presentation and (2) management of student behavior in 1990, whereas Lowman and Mathie, 3 years later, identified (1) intellectual excitement, and (2) interpersonal rapport (Shevlin et al., 2000). Agreement on whether the different dimensions are discrete or are representatives of a single higher-order dimension is yet another ongoing debate (Abrami et al.; Marsh and Roche in Shevlin et al., 2000).

Using abstract dimensions separately is not a step for the better as even the instructor alone does not have a definition for its effectiveness. A good instructor could be linked to (1) the ability to affect personal change and development, to (2) the facilitation of satisfactory academic results or, more obviously, to (3) the opinion of the student. While the first option assumes a long-term process and it is hardly quantifiable, the second one is questionable in terms of incorporating unequivocal elements of educational quality while the instructors get benefits according to the grades which they themselves give to the students. This leaves student ratings, the third option, as one of the most widely accepted and used method (Shevlin et al., 2000) with consequences in decisions concerning faculty status even from the early 70s (Bejar in Feldman, 1977). This is not to say though that their results have a high degree of confidence.

2.1.2 Reliability and Validity

Widely accepted for the apparent ease of application, reliability—the attribute of a consistent and stable measurement—and validity—the ability to measure what we claim to measure—of student ratings alone (Feldman, 1977) are scarce, especially when rating concerns *teaching effectiveness*.

There are readily available statistics to measure reliability. Even so, a number of extraneous variables may confound the results. Some are already proven to give unreliable measurements of teaching effectiveness—expected grades, prior subject interest and motivation, grading leniency, class size, instructor reputation and appearance, student workload and instructors from minority groups. (Marsh; Feldman; Marsh and Roche; Greenwald and Gillmore; Fernandez et al.; Griffin; Gurung and Vespia; Carle in Stieger and Burger, 2009)

Moreover, students are not trained in rating nor in psychometrics—the measurement of knowledge, abilities, attitudes, and personality traits. It is also argued that they rate specific features of teaching on the basis of a global evaluation, like instructor charisma (Shevlin et al., 2000).

Validity is questionable as well. That is due to the low level of statistical training of the personnel that uses the collected information, suggesting caution when interpreting this data. (McKeachie in Shevlin et al., 2000)

In addition to that, one cannot analyze an item when the item is not readily defined. And this is the case with teaching effectiveness, when there is no single criterion of it (Marsh in Kember and Leung, 2008) although there is a set of common principles regarding excellence in teaching (Kember and McNaught in Kember and Leung, 2008). A clear verdict in favor of discriminant validity was not reached even with a representation of suggested aspects of teaching effectiveness as exhaustive as possible (Marks, 2000).

All in all, lack of well founded and widely agreed answers with respect to student ratings of teaching effectiveness and to *teaching effectiveness* alone leaves little to argue for a trustworthy evaluation of educational quality.

2.2 Service-Oriented Educational Quality

Education provision has generally developed from a centrally planned service to a large-scale service with resources allocated by market forces. Especially when coupled with international students or tuition fees in general, providing service excellence became a factor of prime importance between business success and failure (Slade et al. in O'Neill and Palmer, 2004). When talking about state education, countries like the United Kingdom are linking funding levels to quality levels, making service excellence no less important (Gilroy et al., 1999). While some argue for a business-oriented focus only on student recruitment, that poses like a short-term solution at best (Simpson and Tan, 2008).

Debate has covered various topics including quality management frameworks, quality dimensions, quality implementation problems, customer satisfaction and even whether customer focus is appropriate for the educational sector. (Owlia and Aspinwall; Crawford and Shutler; Rowley; Rodgers and Ghosh; Roffe; Aldridge and Rowley; Emery et al. in O'Neill and Palmer, 2004)

Studies suggest, though, that the current climate in higher education puts more and more students in the position of primary customers that are aware and exercise their customer rights (Hill; Thorne and Cuthbert in Sander et al., 2000). Among other countries, Germany, offering both tuition-based and tuition-free education, reacted promptly already by introducing "student satisfaction guarantees", which not only set performance standards but also help increase accountability in both students and lecturers (Gremler and McCollough in Voss et al., 2007).

2.2.1 Education as a Service

Putting the previous question *Whose requirements and whose specifications?* into the context of a service industry, with service providers and service consumers, quality starts to be defined as *fitness for use*, which is at its basic *satisfaction of customers' needs* (Juran in O'Neill and Palmer, 2004).

Gaps have been identified between *what it is expected* and *what it is actually provided* (Parsuraman et al. in Gilroy et al., 1999) in the service industry, and parallels have been drawn to the gaps in higher education (Gilroy et al., 1999). More importantly, they fall under *technical* and *functional* dimensions of service provision. "*Technical quality* refers to the result of the service" (i.e. the graduate), whereas *functional quality* "refers to the way the service has been provided" (i.e. the satisfaction of the graduate). It is the latter that can be used to create a competitive edge (Grönroos in O'Neill and Palmer, 2004). And this competitive edge can differ dramatically from one institution to another, with one emphasizing academic values and another—social values (Clark and Trow in Marks, 2000).

Customers' *expectations* build up as they enter the service process and they need to be elicited at that point in time (Zeithaml et al. in Sander et al., 2000) in order to ensure provision of quality service. This "is the first, and possibly most critical, step in delivering service quality" (Zeithaml et al. in Hill, 1995) otherwise customers' memories regarding their initial expectations may be biased after the service transaction (Appleton-Knapp and Krentler, 2006).

2.2.2 Expectations as Desired Quality

Educational research has recognized the importance of *expectations* with more focus on positively affecting retention and performance (Steel in Sander et al., 2000), rather than on meeting the remodeling of the educational process in order to meet them. An example of the latter would be the Expectation Led Planned Organization (ELPO) model, which specifies that teaching and learning programs should be designed using pre-course expectations (Stevenson et al. in Sander et al., 2000).

Expectations build up on a considerable large number of factors like culture (Shank et al.; Tale et al. in Sander et al., 2000), gender (Walker et al. in Sander et al., 2000), age (Levine in Sander et al., 2000), university type (Shank et al. in Marks, 2000) and mode of study (Stevenson and Sander in Sander et al., 2000). They also build on less personal factors like word of mouth communications or external communications from the service (i.e. printed advertisement) (Zeithaml et al. in Hill, 1995).

While this implies careful handling of data regarding *expectations*, it is this data that may offer distinct advantages (McElwee and Redman in Sander et al., 2000) and paves the premises for an educational service that is both effective and pleasing (Sander and Stevenson in Sander et al., 2000), or at least to detect unrealistic and inappropriate expectations and to address them (Hill in Sander et al., 2000).

It is interesting though that while expectations of service quality change over time (Boulding et al. in Sander et al., 2000), they are formed early in the course with a low probability of changing during the course (Hewett, Chastain, and Thurber; Ortinau and Bush; Sauber and Ludlow in Sander et al., 2000). Some argue that they are formed even prior to entering the education cycle as a whole (Hill, 1995). Furthermore, it was concluded that expectations regarding course content, teaching quality, teaching methods, personal contact and feedback were constant throughout a 3 year long study (Hill, 1995).

Expectations regarding an educational service and a regular industry service differ though, diminishing the role and the benefit of pre-transaction elicitation of service expectations. One can even argue that the necessary condition of being a student is that they lack the ability to comprehend the course, both its content and its rationale. Otherwise, if they had such an ability, what would be the point of studying the course in the first place? (Gilroy et al., 1999)

2.2.3 Satisfaction as Perceived Quality

Satisfaction has been defined as "the perception of pleasurable fulfillment of a service" (Oliver in Appleton-Knapp and Krentler, 2006). While operationally similar to *expectation*, it is a postdecision construct.

Along with the personal factors that influence *expectations*, others have been identified to influence *satisfaction*: employment (Fredericksen et al. in Appleton-Knapp and Krentler, 2006) or grade point average (Porter and Umbach in Appleton-Knapp and Krentler, 2006). Since satisfaction is post-factual, influencers can also be institutional—instructor teaching style (Dana et al. in Appleton-Knapp and Krentler, 2006), quality of instruction (DeBourgh; Lado et al. in Appleton-Knapp and Krentler, 2006), research emphasis of the school (Porter and Umbach in Appleton-Knapp and Krentler, 2006), quality and promptness of feedback from instructor, clarity of expectations from instructor (Fredericksen et al. in Appleton-Knapp and Krentler, 2006), and class size (Krentler and Grudnitski in Appleton-Knapp and Krentler, 2006)—or a mix of institutional and personal—participation and interaction with the instructor and classmates (Fredericksen et al. in Appleton-Knapp and Krentler, 2006).

And above all, *satisfaction* is influenced profoundly by *expectations* (Murray et al. in Appleton-Knapp and Krentler, 2006) which, in turn, are positively associated with perceptions of the *service quality* (Claycomb et al. in Telford and Masson, 2005) and of the organizational climate (Kelley et al. in Telford and Masson, 2005), along with the level of beneficial outcomes (Lengnick-Hall et al. in Telford and Masson, 2005).

In an educational context, *satisfaction* is bound to be influenced by the active participation of the customer himself. Higher education is seen as "an ongoing transformation of the participant" (Harvey and Knight in Telford and Masson, 2005), through an "analytical and critical development" (Harvey and Green in Telford and Masson, 2005). It is in this perspective that assessment of *educational quality* through the lens of *satisfaction* should be sought, and not as a unitary assessment machine that would lessen the path to pedagogy as entertainment (Franz in O'Neill and Palmer, 2004). Research has shown that congruence of expectations and values of different actors involved in the educational process has little to do with student satisfaction (Telford and Masson, 2005). Therefore, it is far from impossible to have both an act of *teaching effectiveness* using high-quality material and methods and an act that leads to *satisfaction*.

2.3 Quality Assessment

Options are discretely available for today's educational manager when it comes to measuring student perceptions of *educational quality*. Due to the cost, the complexity and their inappropriate usage, choosing the right option is highly discouraging though (Ford and Bach in O'Neill and Palmer, 2004).

Moreover it is the validity and the reliability of these tools that is under current debate (Joseph and Joseph; Rowly; Oldfield and Baron; Sureshchandar et al.; Welsh and Dey; Hill et al. in O'Neill and Palmer, 2004). Along with the question of replicable measures of service quality, assessing *quality* in an academic environment is a vicious circle with no clearly confined concepts as a number of extraneous variables may interfere (e.g. instructor charisma), and obviously with no reliable tools to evaluate the application of these concepts (Shevlin et al., 2000).

2.3.1 Traditional Evaluation Methods

Traditionally, universities would employ a mix of *qualitative* and *quantitative* methods using observation and/or communication techniques.

Qualitative methods can go from interviews and focus groups to role-play and observation research. They carry a high degree of subjectivity and a high cost due to the need of specialist training but they do provide a comprehensive view of the students' perspectives.

Quantitative methods, on the other hand, are thought to be much more objective and they are not expensive to pursue. These methods are also very flexible as they can be administered face-to-face (i.e. exit surveys), indirectly (i.e. by phone) or as take-away surveys, either paper or computer based and they can be enacted on regular and much shorter intervals (O'Neill and Palmer, 2004).

Depending on the interval, evaluations can be *formative* or *summative*. One performs the evaluation after each course unit in order to improve the next unit, while the other at the end of the course in order to improve the next course delivery. The latter is the most widely spread in the academia and most studies are based on it (Cave et al.; d'Apollonia and Abrami; Marsh and Roche; Rice in Appleton-Knapp and Krentler, 2006), although it is "questionable whether students are able to evaluate a course as a whole" especially when trying to keep the evaluation representative for each course unit (Stieger and Burger, 2009).

Moreover, when applying this method as an evaluator of satisfaction, it implies recalling expectations as well and retrospective expectations do not match with their actual expectations prior to the course because students are inevitably influenced by their satisfaction level (Appleton-Knapp and Krentler, 2006).

However, quantitative methods are popular because they can be standardized to a greater extent, especially when dealing with summative evaluation. One of these standards is Student Evaluation of Teaching (SET), nowadays ubiquitous in the United Kingdom and the United States of America, where research has shown that expectations and preferences of students are valuable to collect and consider (Hill; Harrop and Douglas; Booth; Naramsimhan in Sander et al., 2000). Yet despite the importance given to SET and other student rating alike, aided by ignoring the substantial contaminating influences (Greenwald in Marks, 2000), theoretical and psychometric aspects of *teaching effectiveness* remain pending (see subsection 2.1.2).

Student ratings are no more than perceptions and impressions. The basic conundrum is "How can students evaluate whether they learned what they should have learned when they do not know what they should have learned? In rhetorical response, they may rely on perceptual factors, such as instructor liking/concern, workload/difficulty, expected/fairness of grading, and organization." (Marks, 2000)

All of that pose more as satisfaction rating influencers, rather than linking to *teaching effectiveness*. And because higher education requires a high level of human interaction and labor intensity in the delivery of its education service, the whole process becomes heterogeneous with each service being unique. This in turn leads to a low level of standardization when it comes to evaluating *service quality* through the pinhole of *satisfaction* (Hill, 1995).

2.3.2 Educational Data Mining

In 2007, one fifth of all students in the United States of America were participating in online courses (Sload Consortium in Hislop, 2009).

Technology became ubiquitous and, while being the most important innovation in educational systems, it started to change expectations, means of communication (Ha et al. in Romero and Ventura, 2007), means of education delivery and means of educational evaluations (Hislop, 2009). Traditional, face-to-face education, criticized heavily (Johnson et al. in Romero and Ventura, 2007), is now complemented by blended or fully computer-based courses, whether online or offline. The information systems associated with these courses gather a large amount of information that can create valuable educational data (Mostow and Beck in Romero et al., 2008).

Data Mining is a field that focuses on collecting and processing large amounts of data in order to bring meaning to it through statistics, visualization, clustering, classification and association rule mining (Romero et al., 2008). While Data Mining is widely used in the business industry, *Educational Data Mining* is a new field, with new specific requirements (i.e. consider teaching aspects related to the system and the student). Its applications can provide a more effective learning environment (Ingram in Romero and Ventura, 2007), but this goal is more subjective and more subtle to measure (Romero and Ventura, 2007) (see also subsection 2.1.1).

Educational Data Mining commonly acts like a formative evaluation technique (Arruabarrena in Romero and Ventura, 2007), thus complementing the most widely used evaluation technique—summative student ratings. Looking at how students use an information system may help evaluate the instructional design and materials (Ingram in Romero and Ventura, 2007). Furthermore, this field is not restrictively oriented only towards educators. It can also be used for the benefit of the students directly. By using student models based on previously completed tasks and their outcome or on similarities with other users, the information system can adapt itself in order to recommend activities, resources and tasks to improve students' learning experience (Heraud et al.; Farzan; Lu; Tang and McCalla; Zaïane Romero and Ventura, 2007).

Choosing the appropriate evaluation method is a daunting task on its own. Adding the complexity of data mining, it can only become worse. Data mining tools are increasingly available (Klosgen and Zytkow in Romero and Ventura, 2007) but many of them are not designed within an educational paradigm (Zaïane et al. in Romero and Ventura, 2007).

However, *what is the educational paradigm?* It seems that Educational Data Mining is focusing solely on the traditional view that quality education implies *teaching effectiveness*, and it has been previously highlighted in this thesis where that path is going. Research has been scarce on how data mining can be used to offer a complementary and radical view on education as a service, and not only to improve several types of issues with traditional evaluation methods.

Chapter 3

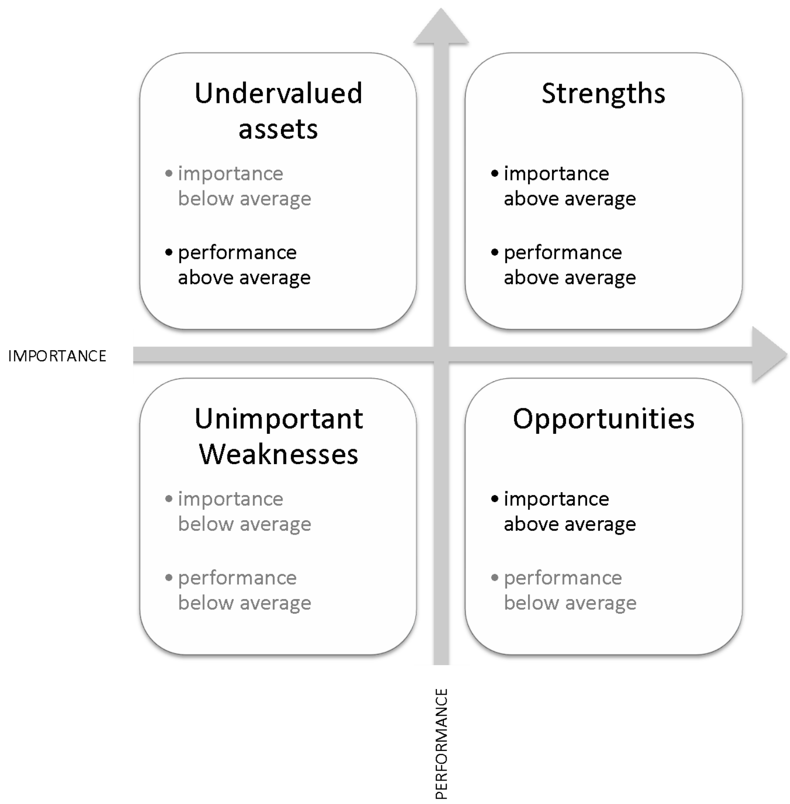
Importance-Performance Analysis of Pure Satisfaction

Before Educational Data Mining can assist, a definition of the framework and of the items that it needs to focus the evaluation on is needed. Quality of education provision, when taken as a service per se, follows the two basic tenets of *service quality*: (1) the service customers are active participants in the delivery process (Lovelock and Young; Mills and Morris; Kelly et al.; Schneider and Bowen; Zeithaml and Bitner; Rodie and Kleine; Bateson in Telford and Masson, 2005) and (2) their perception of quality regarding the service experience is dependent on how much their expectations of the service are met (Tolman; Berry et al.; Gronroos; Parasuraman et al.; Haywood-Farmer and Nollet in Telford and Masson, 2005). Frameworks that define customer expectations and values have been widely developed and applied in the service sector, but it is important to note that there are a number of differences that prioritize one over the other when taking into account the specific context of education provision (Dickson et al. Telford and Masson, 2005).

3.1 Overview

Confirmation-disconfirmation paradigm is at the core of most of the service quality research, studied and tested by many satisfaction researches (Bitner and Hubbert; Churchill and Surprenant; Oliver; Oliver and DeSarbo; Rust and Oliver; Tse and Wilton; Yi in Appleton-Knapp and Krentler, 2006). The paradigm seeks to observe pre-consumption expectations in relation to perception of actual service performance. This creates the notion of *satisfaction*, where the performance exceeds some form of standard, and the notion of *dissatisfaction*, where the performance falls below the standard (Wells and Prensky; Oliver in O'Neill and Palmer, 2004).

Figure 3.1. Quandrants of a IPA graphic used for classification



A variety of disconfirmation models exist, using both inferred or direct approaches. Among these, pre-eminent are the studies of Parasuraman et al. (1985, 1988), Cronin and Taylor (1992) and Martilla and James (1977) which had as output the SERVQUAL instrument, the SERVPERF technique and, respectively, the Importance-Performance Analysis (IPA) technique. The latter gained popularity in marketing strategies due to its simplicity, ease of application and diagnostic value (O'Neill and Palmer, 2004). While all three of them measure absolute performance, the IPA seeks to identify the importance associated to various quality criteria. "Higher importance ratings are likely to play a more critical role in determining satisfaction" (Barsky in O'Neill and Palmer, 2004), correlating to higher expectations. In order to communicate the outcomes, a matrix is plotted using each item's ratings following importance and performance axes. Depending on its quadrant, each item will then be properly categorized as an undervalued asset, a strength, an opportunity or, respectively, as an unimportant weakness.

In March 2002, IPA was applied in an academic environment, testing the educational quality (process, empathy and tangibles) through 368 respondents—out of a tentative number of 500. A thorough analysis of its reliability and validity shows that the IPA holds a value both in assessing and in directing continuous quality improvement efforts in the educational sector—which are of demonstrated importance to the students—due to its psychometric and diagnostic performance. This technique is valuable also due to the ability to gather relevant information regarding students' perceptual processing and satisfaction level and to present it in a format that can be used to target specific service improvements, both at core and perception levels. (O'Neill and Palmer, 2004)

The IPA technique is chosen for this thesis among the aforementioned three, not only for its appraisals and for its flexibility (see section 3.2) and for the ease of associating it with educational data mining techniques (see section 3.3 and section 3.4), but also for the critiques brought to the other two. The SERVQUAL technique attracts evidence of psychometric problems and raises questions regarding whether the expectations, as defined by it, are a clear benchmark to use against evaluated perceptions. Maybe most importantly is the fact that there is a debate around the fatique created by the SERVQUAL technique since it assumes filling in 22-items questionnaires immediately before and immediately after the service consumption (Brown et al; Andersson, Iacobucci et al. in O'Neill and Palmer, 2004). These critiques gave birth to the SERVPERF technique, which assumes only performance evaluation. While it does overcome some of the problems raised by its predecessor, much of the advantageous information around the service quality is dis-considered by the lack of an overview on expectations (O'Neill and Palmer, 2004). It is finally the lack of means to undergo formative evaluation and to associate technological means that make the above two techniques, and possibly others, more appropriate for a later investigation, succeeding the present exploration of the IPA technique.

3.2 Attributes

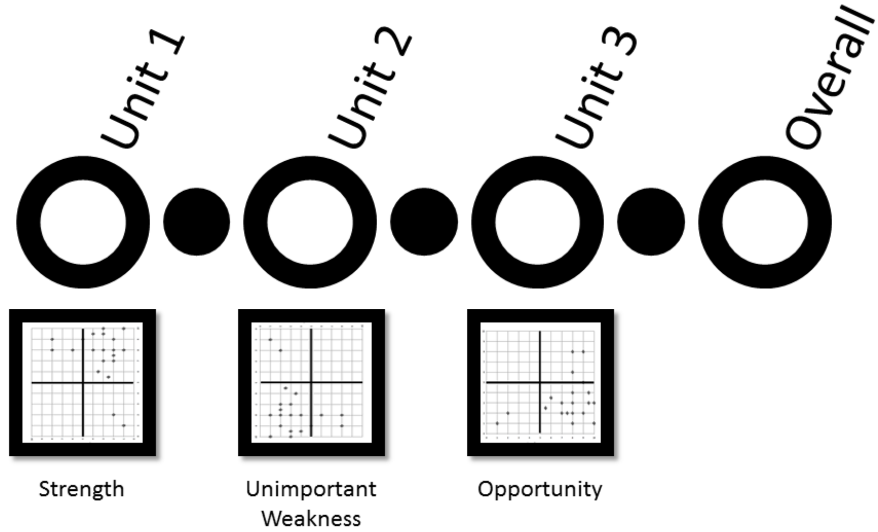
Generally, prior to the application of this technique, two issues need to be addressed: "the determination of the actual attributes to be assessed" (Joseph and Joseph in O'Neill and Palmer, 2004) and the bias towards separation of the performance and importance.

This thesis focuses on a limited area of educational quality—teaching or course delivery quality—, making some of the 22 items used by SERVQUAL not applicable as they were in O'Neill and Palmer's (2004) study on educational quality. That is not to say that within the area of teaching quality, attributes related to the process, the empathy and the tangibles of a course cannot be distinguished for IPA usage.

However, in chapter 2 it was made clear that there is no clear boundary between attributes of the teaching process, and that subjectivity encompasses all of them. Instead of arguing for a subset of attributes and adding to a vicious debate, the author takes a birds-eye-view approach and focuses on the known, rather than the unknown.

Each course unit (i.e. lecture, seminar, exam) must meet some form of standard, defined by expectations, and thus it can act as an attribute. An extra attribute will be the overall teaching process (see Figure 3.2). The two types of attributes, per unit and per course, can be regarded as part of both formative and summative evaluation, depending on the interest of the evaluator. It is important to note that (1) formative evaluation does not yield the same results as the summative one and that (2) the formative evaluation does not influence on the summative ratings (Stieger and Burger, 2009).

Figure 3.2. Illustration of intended outcome



3.3 Usage as Importance

In a traditional assessment, the IPA process would take place by surveying students through the use of Likert-type scale (Likert, 1932) to gauge the importance, and respectively, the performance of several attributes at stake. The rating process is questionable due to a number of factors (see subsection 2.3.1), letting the two seek evaluation through different means and methods.

Figure 3.3. Usage of data mining means within the IPA process

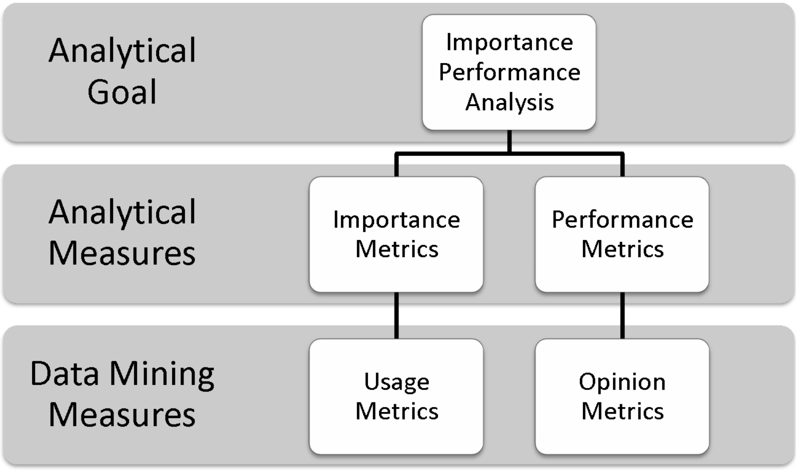


Figure 3.3 depicts the approach towards collecting data to feed a IPA process using data mining and acts as the core of the proposed primary artifact of this research.

With the continued development of web-based information systems, collected user data is ever-growing. Analyzing such data can assist in assessing effectiveness, optimizing functionality, among others (Liu, 2007).

Usage statistics are often a starting point when working with assessment inside an educational information system (Zaïane et al. in Romero et al., 2008). With most of the educational systems available with a web interface, data for these statistics can be extracted using basic tools that feed on different types of web server logs, or using tools specialized in statistics for e-learning systems. Number of website visits, number of visits per page, number of downloads for a specific resource, duration of the visit (Pahl and Donnellan in Romero et al., 2008), educational focus distribution and most frequently accessed courses (Nilakant and Mitrovic in Romero et al., 2008) are all basic measurements that can give insights into students' usage trend. Advanced statistics though can result in predicting whether a student will give the correct answer to an exam question (Beck and Wolf in Romero et al., 2008), in predicting a student's knowledge overall (Feng et al. in Romero et al., 2008) or their final exam score (Pritchard and Warnakulasooriya in Romero et al., 2008). In the greater perspective, web usage mining can assist in automating discovery and analysis of patterns of web actions—either caught in web server logs or in specific applications logs. The latter type of logs act as enhanced logs, recording comprehensive user interactions with online resources. (Liu, 2007)

This is the case with search engines as well. Their fundamental approach helps to build the concept that *usage metrics* can be treated as *importance metrics* in the IPA process. The ubiquitous search engine named Google uses PageRank, a ranking algorithm that is based on academic citation literature. At its core, the algorithm states that a page X with a higher PageRank weight—given by the sum of weight divided by the number of outbound links of each inbound link of page X—has a relatively higher importance or quality, just like a scientific paper is given more credibility if it is cited by other credible papers with a relatively small amount of citations. Coupled with more than 200 different indicators, PageRank is thought to correlate well with human concepts of importance (Brin and Page, 1998; Google, 2010).

Set in the context of Educational Data Mining, importance of a specific course unit can be correlated in the simplest form with the number of interactions relating to it or associated resources. Checking the schedule for a course event or giving feedback on it are two basic examples of interactions. In a cloud-computing environment, more advanced interactions could be tracked like reading the course material. This can identify the course unit by the section or the page number that is being visualized, and can give importance weight by the time spent actively reading the text. Nevertheless, interaction for a long time with an item implies a high importance but not necessarily a high performance as well. Concepts that are hard to grasps need more interaction in order to be assimilated for instance. Thus, performance assessment requires a different type of analysis.

3.4 Sentiment as Performance

Non-topical text analysis poses a lot of interest in the recent years, focusing not on *what* but on *how* people talk about a certain item. One way to address this is through *sentiment analysis*. Important applications of this type of analysis are in evaluating the positivity and subjectivity level that addresses the targeted item—*sentiment classification*—and in extracting relevant correlations between writers' opinions—*opinion mining*. (Argamon et al., 2009)

Both are suitable for addressing performance levels, yet the latter entails that the target item is well defined and that there is a need for cognitive process of the results in order to bring meaning. In an academic setting, this would translate to looking for an attribute, like course material, in comments of students, and providing a subjective phrase as output, like *The books were hard to read*. This would leave the reviewer to assess the output with few objective, summative and timely fashioned means of evaluation. Moreover, previous chapters showed that teaching effectiveness is anything but defined. Sentiment classification on the other hand uses predefined knowledge to produce readily available meaning in one—positivity/orientation—, two—plus subjectivity/force—or three dimensions—plus attitude type. It is this application that can act as *a performance indicator*. (Argamon et al., 2009)

Each course unit is a mixture of instructor, methods, material, topic, interactions, technological resources and others. Attitude towards a course unit can be translated into positivity and ultimately quantified, both horizontally, between all the students per course unit, and vertically, between all the course units per student. For instance, given the previous phrase, sentiment classification in two dimensions would output *weakly negative*, an objective and aggregating output in its nature.

Chapter 4

Data Collection and Processing

Given the primary artifact described in chapter 3, an instantiation is needed in order to learn more on the possible benefits, on the environment and on the actors that would interact with entities similar to the one proposed.

This instantiation, which poses as a secondary artifact, will be put to use inside the Computer and Systems Science (Swedish abbrev. DSV) department of Stockholm University (SU) and Royal Institute of Technology (Swedish abbrev. KTH). This academic environment poses advantages (see section 1.4) that make the department appropriate for the study of this thesis.

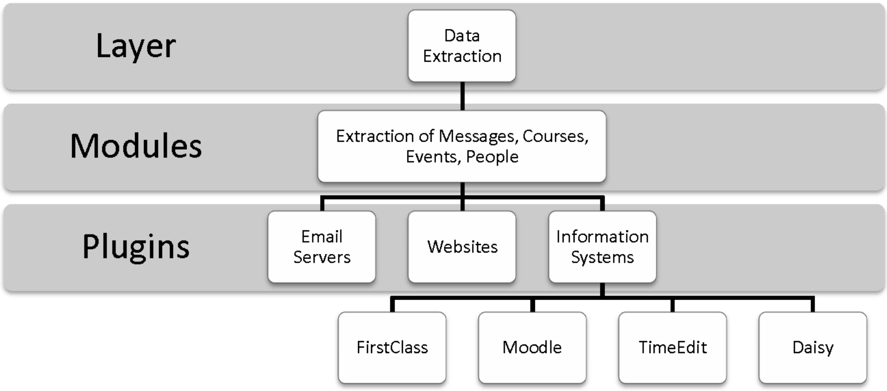
In the spring of 2010, students of the DSV department were invited to take part in a short-term study, through a series of three live and two recorded presentations (Neculau, 2010b; Neculau, 2010a) and one email announcement, reaching approximately 500 students. The reasons for limiting the study's reach were of technical and logistical nature (see subsection 4.1.1). During April $5^{th}$and April 25^{th}, students that previously volunteered were given the possibility to interact with an academic information system that recorded usage and opinion metrics. The study requirement was to provide comprehensive input with the intention to provide a real experience to students before surveying them with respect to the evaluation model described in chapter 3. Moreover, a series of students, both study participants and students without any direct access to the study's system, were surveyed by email over a period of a week, starting April 27^{th}, in order to gain insight on the attitude and the availability towards using this type of course evaluation. A survey approach was taken to assess other actors part of the educational experience—course instructors—who were introduced to potential evaluation reports, that were to be produced by the study's system.

4.1 Data Mining

4.1.1 Snowdrop IS

The model in chapter 3 describes two requirements—*importance* (see section 3.3) and *performance metrics* (see section 3.4)—and two concepts of data collection—*usage* and *opinion mining*. Implementing the latter two in an academic information systems can take numerous paths.

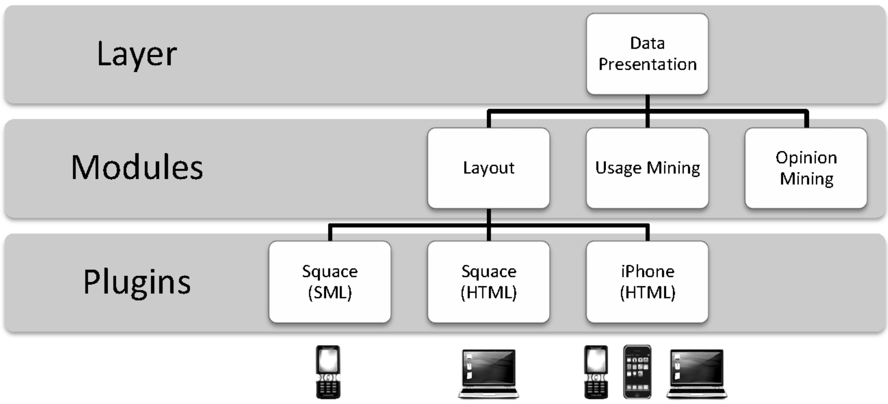
Figure 4.1. Snowdrop - Data Extraction layer



The decision for this study was to make use of an academic information system that interfaced with almost all of the information systems available at the DSV department. The system, entitled *Snowdrop*, is an information system developed as a personal project by the author, while the theoretical foundation is part of a team effort of KTH students, including the author, for an academic course assignment (Neculau et al., 2010). This research decision shares a lot with the research project of developing the system in the first place.

Snowdrop, initially built with a focus on mobile browsers, specifically the Squace browser (Squace AB, 2010), attempts to build a Mobile Virtual Community (MVC). The attention given to mobile development and availability is due to the big potential in countering the "passive nature" of the academic forums (Dawson, 2006), which in turn have the potential to form close-knit student communities that foster learning. This effect is due to a positive correlation between cognitive learning and a sense of community and due to the theory that institutionalized education is not valuable without an affiliation to a social group (Hill, 2003). Snowdrop is also a system that focuses on data integration. Numerous plugins have been developed in order to extract SU and KTH academic information, both public and private (i.e. student specific), from DSV's online repositories—websites, email servers, Daisy, TimeEdit, Moodle, FirstClass. The information is then cleaned and normalized, in order to allow an integrated academic environment tailored for DSV's students. By covering both types of terminals, small and big screen mobile—Squace Markup Language (SML) and/or HyperText Markup Language (HTML) output—and desktop—HTML output only—, keeping in mind that the mobile environment nurtures a great potential for academic communication, by interacting with the majority of the academic information systems available in the department—missing Ping Pong—and, last but not least, having access to the source code of the application for study-related adjustments, the Snowdrop information system makes an easy and rational choice as the appropriate tool for this research.

Figure 4.2. Snowdrop - Data Presentation layer



The system lacked modules for logging user actions and user input. Modules for usage and opinion mining were thus developed and implemented directly into the application's source code.

4.1.2 Usage Mining

Web server logs require data to be prepared in order to be pushed into database records that can later be used for usage mining (Liu, 2007). An alternative is to have application logs that insert specific information straight into the database. Snowdrop's usage module keeps track of user visits, also called hits. Each hit is associated with a database record that retains data to identify the user, to identify and to describe the hit: the date and time, the resource (e.g. information regarding a lecture) that was requested, the referrer—the previous resource that linked to the current one—, the Internet Protocol (IP) address, among others.

There are also cases when a hit cannot be associated with only one resource. For instance, the Snowdrop's homepage acts like a dashboard that presents the latest information on events, recent and forthcoming, and messages, read and unread. When hovering a link to a specific resource for a specified time, partial resource information becomes available. This type of interaction is important because it shows interest for that specific resource, and therefore the user is giving higher importance that will influence our study. In order to record this event, the usage mining module is making use of Asynchronous Javascript and XML (AJAX) technology to send information that describes this event. The specified time after which it is considered that the user is genuinely interested in the resource has been set to 1,5 seconds and is based on research findings regarding reading speed. Factors like age—between 20 and 30 years—and reading goal—between comprehension and memorization—have been taken into account. Research shows that within these boundaries, the reading speed is between 100 and 400 words per minute (Hunziker, 2006). Given that the partial information made available while hovering has on average 6 words of interest, the delay between the hover event and the AJAX call event is well set.

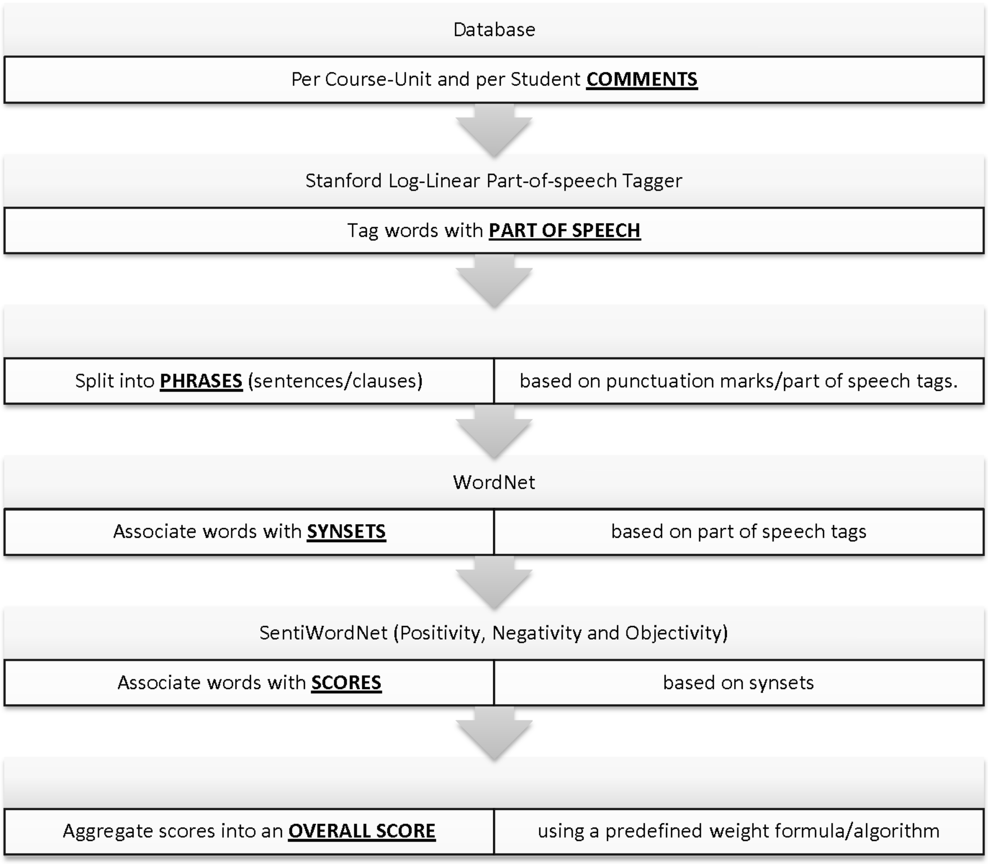
Since mobile terminals do not have hovering events, interest in a specific resource is monitored through a different technique. This entails presenting a low number of resource attributes that identify it, but which do not act as attributes of high interest. To exemplify, a course unit can be shown and uniquely identified by course title and date, while time and location are only made available after a click.

This usage information is not extensive but does provide a good start to assess user's interest in a specific resource. At the same time, the level of complexity is rather low comparing to contemporary usage mining, but it does serve the purpose of providing usage metrics for the importance component of the Importance-Performance Analysis (IPA) artifact.

4.1.3 Opinion Mining/Sentiment Analysis

The module for opinion mining has two main pillars: a tri-state rating and free text commenting. The first one records information regarding the attitude that the user has towards the resource by using two button-like links named *Like* and *Dislike*. The third state is a neutral one, when the user does not click on any of the two. The second pillar allows the user to comment freely on the resource, with no restrictions whatsoever.

Figure 4.3. Outlined process of associating performance ratings to comments



While the first pillar provides information easy to quantify, the second needs advanced processing even if the goal is only to assess the level of positive attitude regarding the resource. One possible approach is outlined in Figure 4.3. One could make use of an English lexical database called WordNet (Miller et al., 1990). This would be applicable only for English-based academia, but semantic networks for several other European languages, including Swedish (Viberg et al., 2002; Borin et al., 2010), are the focus of multiple projects emerging around this project. Wordnet, started in 1985, groups words into sets of synonyms, called *synsets*, with the purpose of creating a more intuitive mix between a dictionary and a thesaurus, while at the same time paving the path for automatic text analysis and artificial intelligence applications. Lexical information is thus grouped in terms of word meanings, rather than word forms, making it an appropriate ground for sentiment analysis. Parts of speech play an important role as Wordnet categorizes words as nouns, verbs, adjectives, adverbs and function words, and builds synsets considering the category. In order to make use of the WordNet database, one needs to tag each word in every phrase with its part of speech. Several techniques and applications are available on a variety of software platforms making this process a matter of choice (Manning, 2010). For the purpose of this research, the author has considered the Stanford Log-linear Part-Of-Speech Tagger (Toutanova et al., 2010), freeware for research purposes, with availability for commercial licensing, which can output its analysis in plain text or Extensible Markup Language (XML) format. The sentence used in section 3.4 would be tagged as "The_DT books_NNS were_VBD hard_JJ to_TO read_VB ._."—determiner, followed by plural noun, verb in past tense, adjective, infinitive marker *to*, and verb in base form.

Having the words associated to parts of speech, and ultimately to synsets, an extension of the WordNet can then be used to associate words with sentiment information. SentiWordNet (Esuli and Sebastiani, 2006) is a database of synsets, where each set is associated with three numerical scores, ranging from 0.0 to 1.0: Positivity, Negativity and Objectivity. Parsing the database for the words in the above example sentence would yield the following associations: books [P=0,N=0,O=1], hard [P=0,N=0.75,O=0.25] and read [P=0,N=0,O=1] (function words carry no weight). The sentence—after a simple mean average—is thus 0 positive, 0.25 negative and 0.75 objective, which translates to a weakly negative performance. Other and most probably more complex arithmetic formulas can surely be used to calculate the overall score. While the afore-mentioned process, outlined in Figure 4.3, is simplistic in its nature, and some of the challenges are easy to spot—i.e. the use of the negator *not* which should switch the positivity and negativity scores—, it suffices as a base for feasible approach to produce objective and quantifiable performance ratings.

4.2 Surveys

4.2.1 Student Survey

Beyond the short-term study that implemented data mining techniques, surveys were used as a method to assess attitude and availability towards this type of implementation for evaluating quality of the educational service. The subjects were both students taking part in the study and students that had no interaction with the information system, and course instructors.

The invitation to the student survey (see Appendix C), carried out between April $27^{th}$and June 3^{rd}, was sent out by email to 650 students, enrolled at KTH or SU in (5) international master programmes and/or registered for (5) individual courses ongoing at the DSV department during the study period (see list of programmes and courses in Appendix D). Due to the limitations of the Google Docs platform

(Google Inc., 2010)—which was used to host the surveys—and the goal of zero tolerance to faulty data, some security policies were put in action. For example, each student was required to fill in the email address where the survey invitation was received, thus allowing the author to filter out responses from unknown email addresses.

The 13 items of the survey were introduced as affirmations. Respondents were asked to rate their agreement with the affirmation on a four-point Likert scale anchored at *(1) Strongly Disagree* and *(4) Strongly Agree* for 7 of these items, while the rest of 6 items required a *Yes* or *No* choice. Only 8 of them were made compulsory, while the rest were intended only for those that took part in the study. The items targeted to confirm the premises of this thesis, to confirm the interest for evaluation methods based on data mining and alternative or complementary to the current end-of-course survey, and, optionally, to evaluate the participation in the study. The structure and the survey items have been scrutinized for errors, vagueness, misleading and other unsuitable factors by a sample of 4 individuals—master students at Swedish universities. They provided general feedback and specific improvements that shaped the final version of the survey.

4.2.2 Instructor Survey

The instructor survey (see Appendix F) followed a similar pattern. Carried out during the same period, 57 invitations were sent out to course instructors working at the DSV department by using the internal communication system FirstClass. The selection of the individuals was based on the list of courses available at the department during the 2010 spring semester (Möller, 2010), and respectively their responsibles. Lecturers, assistants and invited guests have not been included due to their lower interest and responsibility for course assessment.

In the same manner as with the student survey, affirmative items were assessed with the same Likert scale, except for one with explicit multiple choice. The survey went deeper into explaining the primary artifact by offering some brief explanations and visual elements, in order to address more of the reasoning, and less of the attitude towards it. All items were compulsory, and targeted the same categories of the student survey, except for the participation in the study. An added category was the assessment of the IPA process, in order to position the instructors with respect to the graphical outcome of the process. This survey was also scrutinized the same way as the student survey.

Chapter 5

Analysis

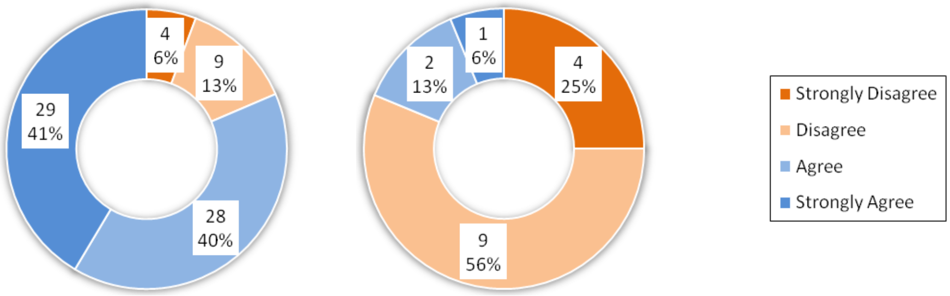
The instantiation of the primary artifact has not produced widespread data collection unfortunately. On average, only 3 out of 12 registered participants have been active in using the system, and only one of these has been regularly providing feedback with respect to each course unit that has taken place during the study. This low turn out has forced the author to use this initiative only as a proof-of-concept and as a mean to provide a real experience to the participants. They were later surveyed about the possible shortcomings of the current instantiation that have led to this outcome.

5.1 Surveys

The student survey had 70 respondents, while the instructor survey gathered 16, reaching a response ratio of 11%, respectively 28%. Extensive results including charts are available as Appendix E (students) and Appendix G (instructors). A call for a wider scale survey and a higher response rate is needed in order to consolidate current findings.

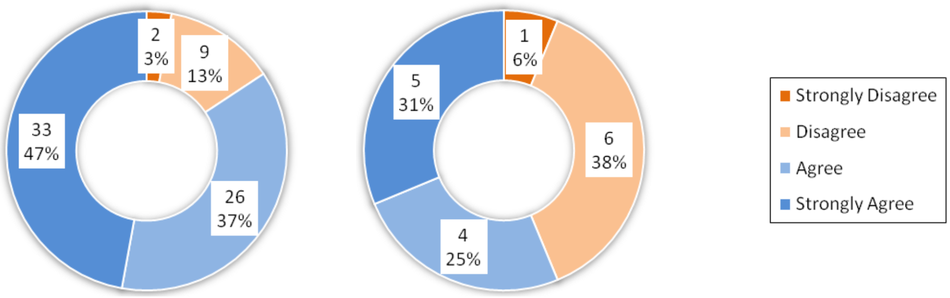
A first conclusion and an important variable of these surveys is that student responses seem to mainly come from those that are actively engaged into providing feedback through end-of-course evaluations. While this is true for a strong majority of the respondents, it is inadequate to conclude that this is true for a strong majority of all Computer and Systems Science (Swedish abbrev. DSV) (DSV) students, given current studies show a lower return rate (Donovan et al., 2006) and, moreover, given the assessment of the instructors on the same matter—81% disagree on different levels that all students provide end-of-course feedback for the department's courses (see Figure 5.1). This is also a strong reason to look for alternative evaluation methods that would increase response rate and provide a more homogeneous perspective, slightly less or not dependent at all to students' availability and attitude towards actively providing feedback. This conclusion can be the root of a perception that the results of the survey are biased, but, on the contrary, it only means that the outcomes are inferred from the response of those highly concerned with the quality of the academic experience.

Figure 5.1. Students vs. Instructors: Response Rate for End-of-course Evaluations



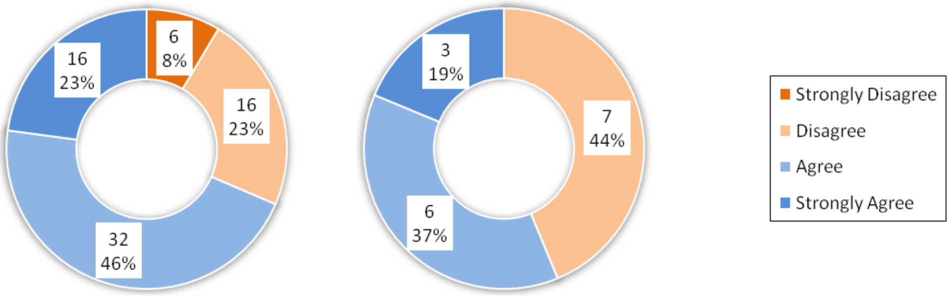
It is not a surprise to observe that these students try to provide the most accurate feedback, but overall, since they do not constitute the majority, the instructors disagree on a high scale that students in general follow the same practice as the student respondents (see Figure 5.2). At the same time, half of the instructors question their own end-of-course evaluations due to their accuracy and the level of focus they entail (see Appendix G, page 68). Both of these findings raise issues in line with the literature review of this thesis (see chapter 2) and converge towards careful handling of both the creation and the analysis of end-of-course evaluations.

Figure 5.2. Students vs. Instructors: Accuracy of Student Answers



When it comes to the use of data mining techniques for performing evaluation of service quality, results are mixed between the usage and opinion mining. The first technique appears to be in line with current student practices of interacting more often with online resources related to courses of higher interest (see Appendix E, page 57). Three out of four students favor a *usage mining* approach (see Appendix E, page 58), while instructors' responses are balanced (see Appendix G, page 69). This can be translated into a need for explaining the concept and the rationale behind it, and the strong relationship between usage and importance. While similar gross ratios apply for favoring an *opinion mining* approach (see Appendix E, page 59 and Appendix G, page 70), the practice of providing open-ended comments and simple ratings for educational services does not come as natural as for the interaction practice (see Appendix E, page 58). The student survey shows that 30 students strongly disagree and 20 disagree with the affirmation that they often post reactions on social networks, converging to 71% of the respondents in need of some sort of initiation and incentives for this practice to gain momentum. Reasons for this outcome need to be investigated as they might be relevant for similar feedback in academic information systems. One possible reason could be linked with issues around lack of anonymity—which is not always applicable to social networks—and around lack of incentives or benefits exposure (Donovan et al., 2006).

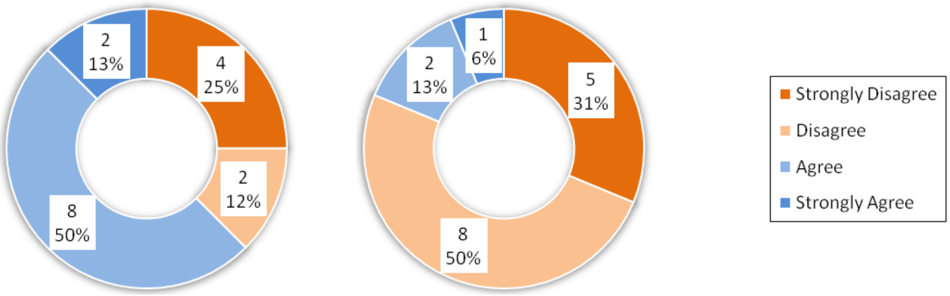
Figure 5.3. Students vs. Instructors: Course-unit Evaluation



These assessments also point out that despite a natural perception of reluctance towards the segmentation of the academic evaluation and towards balancing or even moving from a summative to a formative evaluation, it is actually a fair split between both students and instructors agreeing or disagreeing with such an approach. One can even argue that students show a tendency towards favoring per course-unit evaluation (see Figure 5.3).

The instructors were also surveyed regarding the concepts that are at the core of their end-of-course evaluations (see Appendix G, page 69). Nine have answered *Service Satisfaction*, while five chose *Teaching Effectiveness*. Two instructors faced problems in recognizing and distinguishing between the mentioned concepts. This poses as a reinforcement of some of the premises of this work, specifically of the acclaimed focus on service quality. Despite that, it does not come as a clear tendency.

Figure 5.4. Instructors: Ease of IPA Interpretation vs. Utility of IPA



Finally, a plotted chart using synthetic data showed an example of how a course unit can be evaluated using the Importance-Performance Analysis (IPA) technique and transposed into a visual resource. As Figure 5.4 shows, this was easily interpreted by 63% of the instructors, while 25% answered that it was a great obstacle. The IPA model is one of the core artifacts of service quality assessment. The result shows that even with a brief explanation of the model, the visual product can be translated into a meaningful resource. Despite that, 81% of the instructors did not find the graphic as being useful for assessing the specific course unit, nor did they see in it the possibility of shaping its content. Given the success of this artifact when used in the academia (see chapter 3), this implies that its usage does not come natural without a thorough explanation. This is possibly in close connection to the fairly balanced attitude towards data mining techniques used for academic evaluation.

To conclude, the survey results show that there are a number of issues, among others presented in chapter 2, that need to be addressed regarding the current practice of end-of-course evaluations. The method put forward in this thesis seems to have adequate support from the students, while it is to be carefully explained to the course responsibles in order to secure their trust in the model and its techniques. Last, but no least, a certain degree of awareness regarding the concept of service quality and student satisfaction is instated among the instructors of the DSV department, paving the way for developments in this area.

5.2 Difficulties Encountered

A series of issues arose during the data collection process. The most important one, previously mentioned, was the inability to gather a capacious amount of data during the short-term study. This is primarily connected to the low number of students that volunteered for the study—12 out of approximately 500 students. On top of that, there was a very low level of interaction with the Snowdrop information system. The student survey had to change its scope and to address this issue by asking the volunteers about the reasons for such a low turn out. Four respondents helped bringing some light to this matter and answered that the main issues that posed as an obstacle were that:

- Snowdrop is not an official DSV information system;
- students got accustomed to the other systems available and, more importantly, official at DSV;
- Snowdrop needs a more user-friendly interface.

While Snowdrop was comprehensive at interacting with all the official systems but one, it seems that a system that will perform data collection for the academic evaluation needs to be both labeled as an official system and equipped with a very user-friendly design and interaction functions. This is in spite of the result that Snowdrop did not lack any vital academic information that the students were after.

This low turn out has led to a change in structure of the instructor survey—which was supposed to include charts plotted with data collected during the study, instead of synthetic data—, but also to a change in structure—in the same manner—, scope and purpose of the student survey as well. Without usage and opinion data, it became impossible to carry out a more comprehensive survey on the study volunteers and also to discover differences and similarities between those who have and those who have not used an information system based on the extended IPA model.

On a secondary level, difficulties specific to the Snowdrop software have been encountered, yet ultimately resolved. Most of them do not make the scope of this thesis. However, it is worth mentioning here that the data extraction module was creating a load on the official information systems, leading to blacklisting of the Snowdrop server's Internet Protocol (IP) address and putting the system in an outdated state. While this was addressed by an official request to re-enable the access, it only adds to the importance of using a system officially supported and even promoted by the academic institution.

Chapter 6

Conclusion

This thesis had as a goal to study current views and methods of educational evaluation and to study opportunities to help fill in a gap both in the academic and in the research environment. This has been fulfilled by constructing, instantiating and evaluating an method for assessing *service quality* in the academia that taps into current technological and service-oriented trends. The instantiation study served as a proof-of-concept and it helped in shaping basic requirements for an academic information system, in general, and for an academic evaluation system, in particular.

The surveys conducted at the Computer and Systems Science (Swedish abbrev. DSV) (DSV) department of the joint campus of Royal Institute of Technology (Swedish abbrev. KTH) (KTH) and Stockholm University (SU) showed that current evaluation methods in the form of end-of-course student surveys have low response rates and that the student ratings are not trustworthy in the opinion of course instructors. Moreover, they recognize validity issues regarding the construction of their own survey questions. All of the above confirm key items of the literature review that acts as the foundation of this thesis.

Following the perspective of service quality, and applying this concept and its methods in the field of educational provision, reveal that students have a positive attitude towards data mining techniques being implemented in academic information systems for course evaluation purposes, while the attitude is more balanced and triggers careful handling for course instructors.

6.1 Discussion

While a quick conclusion would be to state that there is room for change and that change is desirable (see chapter 2), it is important not to neglect realities. Among them, current trends of educational provision play a major role, both in terms of its scale, goal and means. Decision factors need not fall under a common misconception that students' expectations and preferences should dictate the direction of change without any means of appeal. The goal is not to increase popularity by lowering the standards, but to increase satisfaction by understanding the expectations. The short-term aim is thus to evaluate the educational experience as accurately as possible, in order to provide educational means that students prefer, while the long-term aim must be kept to providing rigorous and effective teaching provision. (Sander et al., 2000)

It is obvious in that regard that student evaluation must not be the only criterion for evaluating education. This would result in an assessment as biased as one that is based only on views of course instructors. After all, it is the combination of multiple perspectives that leads to an objective outcome, and finally to a decisional outcome of planning future academic developments. The present work advocates for different views on academic quality and its assessment, as the traditional one cannot support valid and effective formative or summative judgments (Gilroy et al., 1999). It is in this context that instructors need to address both student expectations and satisfaction levels, and that students need to identify their expectations and express their satisfaction. If the ultimate goal is to provide both a higher quality education with a higher level of satisfaction, the two way communication is a must (Voss et al., 2007).

Another reality is that newly available knowledge and technology might trigger defensive mechanisms without a proper presentation strategy. Not only is it important to describe the necessary human resources needed for new evaluation methods, if any, but also to explain both the reasoning behind new methods of data collection for evaluation analysis and the presentation methods of the data. For this particular artifact part of this thesis, the above would translate into providing details on how data is gathered and analyzed, on how the analysis is then plotted on a Importance-Performance Analysis (IPA) chart, on how the chart can bring meaning to the evaluation, and last, but not least, on how this evaluation method does not add to the duties of course instructors.

Moreover, it is important to accept new technological and service-oriented realities, and to open up to new possibilities of assisting educational evaluation that are built on more sound grounds. Understanding that educational quality has more than one bevel, ambiguously defined as *teaching effectiveness* (see chapter 2), is yet another condition.

6.2 Future Work and Research

The outcome of the short-term study (see chapter 5) immediately defines one way to continue the hereby work: implementing the artifact in an officially supported system, and as an officially supported action. This is very much reinforced by an action coordinated by the Information Technology department of KTH, ongoing in parallel with the present work. The two universities mentioned in this thesis, SU and KTH, are building a social platform for cohesive learning and social networking (KTH, 2010). While the scope and the type of this future system is not clear, it does set the basis for an official platform to take the current research proposal further.

In terms of application improvements, Snowdrop does not associate message nor person resources with event or course resources, leaving out some usage mining data that could lead to higher importance levels for a course unit or for the overall course. The same applies for introducing course material directly in the Snowdrop information systems. Altogether, these aspects are closely knit to cloud-computing platforms (see section 1.5). Filters could also be added, in order to detect spamming techniques, both in usage and opinion mining.

Furthermore, due to the unique aspect of this research proposal, a pronounced variety of research continuations arise. The sentiment analysis process that is outlined in this thesis (see subsection 4.1.3) is far from a comprehensive method needed for yielding irrefutably valid results. Specific investigations are also called upon usage mining analysis and the rules that can shape the usage patterns into importance levels. As the surveys showed, instructors are aware of the *service satisfaction* concept, yet they do not see the benefits of the IPA process (see chapter 5), and this alone constitutes a possible path for future research at the DSV department: if satisfaction is assessed, what are the means, the reasoning and the understanding degree of this concept and the evaluation of this concept in academia?

Last, but not least, since only one technique of assessing service quality is used in this thesis, future research can address other techniques, like SERVQUAL, SERVERF and others (see section 3.1), that may use new technology to counteract present disadvantages, or may not have obvious means of using Educational Data Mining beyond conventional thought. For example, the 22-items SERVQUAL questionnaire could ultimately be answered through technological and observational means rather than by manually filling in with the recipient's opinion.

Bibliography

Appleton-Knapp, S. L. and Krentler, K. A. (2006). Measuring Student Expectations and Their Effects on Satisfaction: The Importance of Managing Student Expectations, *Journal of Marketing Education* **2**8(3): 254–264.

Argamon, S., Bloom, K., Esuli, A. and Sebastiani, F. (2009). Automatically Determining Attitude Type and Force for Sentiment Analysis, *Lecture Notes In Artificial Intelligence* pp. 218 – 231.

Borin, L., Dannélls, D., Forsberg, M., Gronostaj, M. T. and Kokkinakis, D. (2010). Swedish FrameNet [Online] [Accessed 15 May 2010].
URL: http://spraakbanken.gu.se/eng/swefn

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine, *Computer networks and ISDN systems* **3**0(1-7): 107–117.

Dawson, S. (2006). Online forum discussion interactions as an indicator of student community, *Australasian Journal of Educational Technology* **2**2(4).

Donovan, J., Mader, C. and Shinsky, J. (2006). Constructive student feedback: Online vs. traditional course evaluations, *Journal of Interactive Online Learning* **5**(3): 283–296.

Dougiamas, M. (2010). Moodle [Online] [Accessed 15 March 2010].
URL: http://www.moodle.com

Esuli, A. and Sebastiani, F. (2006). SENTIWORDNET: A Publicly Available Lexical Resource for Opinion Mining, *In Proceedings of the 5th Conference on Language Resources and Evaluation (LREC'06*, pp. 417–422.

Evolvera (2010). TimeEdit [Online] [Accessed 15 March 2010].
URL: http://www.evolvera.se/htmlinternational/eng/index.html

Feldman, K. a. (1977). Consistency and variability among college students in rating their teachers and courses: A review and analysis, *Research in Higher Education* **6**(3): 223–274.

FirstClass Division of OpenText (2010). FirstClass [Online] [Accessed 15 March 2010].
URL: http://www.firstclass.com/

Gilroy, P., Long, P., Rangecroft, M. and Tricker, T. (1999). The Evaluation of Course Quality Through a Service Template, *Evaluation* **5**(1): 80–91.

Google (2010). Corporate Information: Technology Overview [Online] [Accessed 15 March 2010].
URL: http://www.google.com/corporate/tech.html

Google Inc. (2010). Google Docs [Online] [Accessed 15 March 2010].
URL: http://docs.google.com

Harper, D. (2010a). Educate [Online] [Accessed 9 March 2010].
URL: http://www.etymonline.com/index.php?search=educate

Harper, D. (2010b). Education [Online] [Accessed 9 March 2010].
URL: http://www.etymonline.com/index.php?term=education

Harper, D. (2010c). Quality [Online] [Accessed 9 March 2010].
URL: http://www.etymonline.com/index.php?search=quality

Henderson, J. and Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations, *IBM systems Journal* **3**2(1): 4–16.

Hevner, A., March, S., Park, J. and Ram, S. (2004). Design science in information systems research, *Mis Quarterly* **2**8(1): 75–105.

Hill, F. M. (1995). Managing service quality in higher education: the role of the student as primary consumer, *Quality Assurance in Education* **3**(3): 10–21.

Hill, T. (2003). *Leveraging mobile technology for m-learning: 3rd generation threaded discussions*, IEEE.

Hislop, G. W. (2009). The Inevitability of Teaching Online, *Computer* **4**2(12): 94–96.

Hunziker, H. (2006). Im Auge des Lesers-foveale und periphere Wahrnehmung: vom Buchstabieren zur Lesefreude, *the eye of the reader: foveal and peripheral perception-from letter recognition to the joy of reading), Transmedia Zurich, 2006* .

Kember, D. and Leung, D. (2008). Establishing the validity and reliability of course evaluation questionnaires, *Assessment & Evaluation in Higher Education* **3**3(4): 341–353.

Kent, W. (1912). Academic Efficiency, *Science (New York, N.Y.)* **3**6(938): 841–850.

KTH (2010). Sprintdemo 19 maj - KTHs nya sociala plattform [Online] [Accessed 20 May 2010].
URL: http://www.youtube.com/watch?v=i_kUpnC3bHw

KTH/SU (2010). Daisy [online] [accessed 15 march 2010].
URL: http://daisy.dsv.su.se

Lacan, J. (1975). Encore. Le Séminaire. Livre XX, *Paris, Seuil* .

Likert, R. (1932). A technique for the measurement of attitudes., *Archives of Psychology* **2**2(140): 1 – 55.

Liu, B. (2007). *Web data mining*, Springer.

Love, D. a. and J Kotchen, M. (2010). Grades, Course Evaluations, and Academic Incentives, *Eastern Economic Journal* **3**6(2): 151–163.

Manning, C. (2010). Statistical natural language processing and corpus-based computational linguistics: An annotated list of resources [Online] [Accessed 15 March 2010].
URL: http://www-nlp.stanford.edu/links/statnlp.html#Taggers

Marks, R. B. (2000). Determinants of Student Evaluations of Global Measures of Instructor and Course Value, *Journal of Marketing Education* **2**2(2): 108–119.

Miller, G., Beckwith, R., Fellbaum, C., Gross, D. and Miller, K. (1990). WordNet: An on-line lexical database, *International journal of lexicography* **3**(4): 235–312.

Möller, S. (2010). Delkurser och scheman VT2010 [Online] [Accessed 15 April 2010].
URL: http://dsv.su.se/utbildning/delkurser

Neculau, A. (2010a). Snowdrop "Study Trailer" [Online] [Accessed 15 May 2010].
URL: http://www.facebook.com/video/video.php?v=1392710970191

Neculau, A. (2010b). Snowdrop "Trailer" [Online] [Accessed 15 May 2010].
URL: http://www.facebook.com/video/video.php?v=1392682129470

Neculau, A., Henriksson, A., Azizyan, G. and Magarian Kenaraki, M. K. (2010). Clover: Mobile Student Community on Squace, *Journal of Information Systems & Operations Management* **3**(2).

O'Brien, K. (2010). What happened to studying? [Online] [Accessed 2010-07-17].
URL: http://www.boston.com/bostonglobe/ideas/articles/2010/07/04/what_happened_to_studying/?page=full

O'Neill, M. A. and Palmer, A. (2004). Importance-performance analysis: a useful tool for directing continuous quality improvement in higher education, *Quality Assurance in Education* **1**2(1): 39–52.

Ping Pong AB (2010). PING PONG [Online] [Accessed 15 March 2010].
URL: http://www.pingpong.se/

QS Quacquarelli Symonds Limited (2010). Top Universities [Online] [Accessed 15 March 2010].
URL: http://www.topuniversities.com/

Romero, C. and Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005, *Expert Systems with Applications* **3**3(1): 135–146.

Romero, C., Ventura, S. and Garcia, E. (2008). Data mining in course management systems: Moodle case study and tutorial, *Computers & Education* **5**1(1): 368–384.

Sander, P., Stevenson, K., King, M. and Coates, D. (2000). University students' expectations of teaching, *Studies in Higher Education* **2**5(3): 309–323.

Shanghai Ranking Consultancy (2009). Academic Ranking of World Universities [Online] [Accessed 15 March 2010].
URL: http://www.arwu.org/

Shevlin, M., Banyard, P., Davies, M. and Griffiths, M. (2000). The Validity of Student Evaluation of Teaching in Higher Education: love me, love my lectures?, *Assessment & Evaluation in Higher Education* **2**5(4): 397–405.

Simpson, K. and Tan, W. (2008). A Home Away From Home?: Chinese Student Evaluations of an Overseas Study Experience, *Journal of Studies in International Education* **1**3(1): 5–21.

Squace AB (2010). Squace [Online] [Accessed 15 March 2010].
URL: http://www.squace.com

Stieger, S. and Burger, C. (2009). Let's Go Formative: Continuous Student Ratings with Web 2.0 Application Twitter., *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society* p. 091019095317065.

Telford, R. and Masson, R. (2005). The congruence of quality values in higher education, *Quality Assurance in Education* **1**3(2): 107–119.

Toutanova, K., Klein, D., Manning, C., Morgan, W., Rafferty, A. and Galley, M. (2010). Stanford Log-linear Part-Of-Speech Tagger [Online] [Accessed 15 April 2010].
URL: http://nlp.stanford.edu/software/tagger.shtml

Tranæus, N. (2010). Fees make competition for students more equal [Online] [Accessed 15 March 2010].
URL: http://www.thelocal.se/25100/20100219/

Utbildningsdepartementet (2009). Studieavgifter för utländska studenter införs 2011 [Online] [Accessed 15 March 2010].
URL: http://www.regeringen.se/sb/d/11325/a/131230

Viberg, A., Lindmark, K., Lindvall, A. and Mellenius, I. (2002). The Swedish WordNet Project, *Euralex 2002*, Copenhagen University, Copenhagen, pp. 407–412.

Voss, R., Gruber, T. and Szmigin, I. (2007). Service quality in higher education: The role of student expectations, *Journal of Business Research* **6**0(9): 949–959.

Waara, A. (2010). Avgifter för utomeuropeiska studenter fastslagna [Online] [Accessed at 2010-07-06].
URL: http://www.uu.se/nyheter/nyhet.php?typ=artikel&id=1064

Appendices

Appendix A

Acronyms

AJAX Asynchronous Javascript and XML

DSV Computer and Systems Science (Swedish abbrev. DSV)

EEA European Economic Area

ELPO Expectation Led Planned Organization

EU European Union

GDP Gross Domestic Product

HTML HyperText Markup Language

IP Internet Protocol

IPA Importance-Performance Analysis

KTH Royal Institute of Technology (Swedish abbrev. KTH)

MVC Mobile Virtual Community

MVCs Mobile Virtual Communities

OECD Organization for Economic Co-operation and Development

SEQ Student Evaluation Questionnaire

SET Student Evaluation of Teaching

SML Squace Markup Language

SU Stockholm University

TEQ Teacher Evaluation Questionnaire

XML Extensible Markup Language

Appendix B

Student Study Registration

Please fill in this form in order to
- confirm your will to participate in this study
- allow your KTH,DSV,FC username to be used for Snowdrop authentication
- get full access to Snowdrop

Once you register, please allow 24h for your username to become valid for registration.

You can follow progress on
http://www.facebook.com/pages/Snowdrop-IS/285063654928 or
http://twitter.com/snowdrop_is

If you have any type of questions, don't hesitate to Email me at
neculau@kth.se .

Thank you!

Andrei Neculau 2009-2010
Master Thesis at KTH/SU, Stockholm, Sweden

* Required

Registration Data

In order to provide a higher level of security, and to allow for a high-performance environment, only registered participants will be allowed to use the system.

KTH,DSV,FC username *

e.g. neculau if your email is neculau@kth.se or neculau@dsv.su.se

Courses registered for Period 4, 2010

- ☐ ID2006 Software Evolution and Maintenance
- ☐ IV2016 Project and Power Games
- ☐ IV2031 Strategic Management of IT
- ☐ IV2035 Organisations, IT-Systems and Management
- ☐ IV2037 Business Process Design and Intelligence
- ☐ IV2039 System Integration

Other courses for Period 4, 2010

You can add other courses than the ones above. Fill in with the course's ID e.g. IV2015, IV2016

Signature

By signing with your name below, you commit yourself to be actively using the system during the period of the study and to allow collected usage & opinion data to be used for the benefit of the study. I also understand that you may need to contact me during the study by Email.

Full Name *

Surname FamilyName

Contact Email Address *

Submit

Appendix C

Student Survey

Course Analytics - Student Survey

Estimated time: 2-3 minutes
Availability: April 27 - June 3

Privacy
Your e-mail address has been extracted from Daisy/Bilda.
It was not nor will it be made available to any party.
You are required to input your e-mail address in order to filter spam.

If you have any type of questions, don't hesitate to Email me at neculau@kth.se .

Thank you!

Andrei Neculau 2009-2010
Master Thesis at KTH/SU, Stockholm, Sweden

http://www.facebook.com/pages/Snowdrop-IS/285063654928
http://twitter.com/snowdrop_is

* Required

E-mail Address *
Please enter the e-mail address where you received the invitation to this survey

I answer all end-of-course evaluations for the courses that I take. *

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I try to provide the best and most comprehensive answers possible. *

You don't rush, you contemplate, you think it through.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

When I'm interested in a course, I check course-specific resources more often than otherwise. *

Course-specific resources = course website, schedule, forums, coursebook, compendiums, etc.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I favor a less time-consuming approach to academic evaluation based on how much and how I use academic information systems. *

Academic information systems = Daisy, FirstClass, Bilda/PingPong, Moodle/VLE, etc

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I often post, like/dislike, bookmark or re-tweet reactions (status updates) regarding academic experiences (mine/others) on social networks. *

Social network = Facebook, Twitter, etc. e.g. "Exam tomorrow #fail"; "Exam + party = tomorrow"; " Finally got SciWri graded. Not bad."

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I favor a less time-consuming approach to academic evaluation based on simple ratings and comments per course-unit event, just like on current social-networks. *

Feedback would be entirely ANONYMOUS.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I favor a more direct approach to academic evaluation based on feedback at the end of each course-unit and not at the end of the course. *

Feedback would be entirely ANONYMOUS.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I used the Snowdrop application during April 5-25. *

- ○ Yes. I will answer the following questions as well
- ○ No. I will ignore the rest of the questions and press the SUBMIT button at the bottom

I would have used it more often if it was an official DSV system.

- Yes
- No

I would have used it more often if I hadn't already got used to the main information systems.

Main information systems = Daisy, FirstClass, Bilda/PingPong, etc.

- Yes
- No

I would have used it more often but some academic information was missing.

- Yes
- No

I would have used it more often if I didn't fear that my comments would be made public to the course leader(s).

- Yes
- No

I would have used it more often if the interface was more user-friendly.

- Yes
- No

Submit

Appendix D

Surveyed Students

Surveyed students were enrolled at the following master programmes:

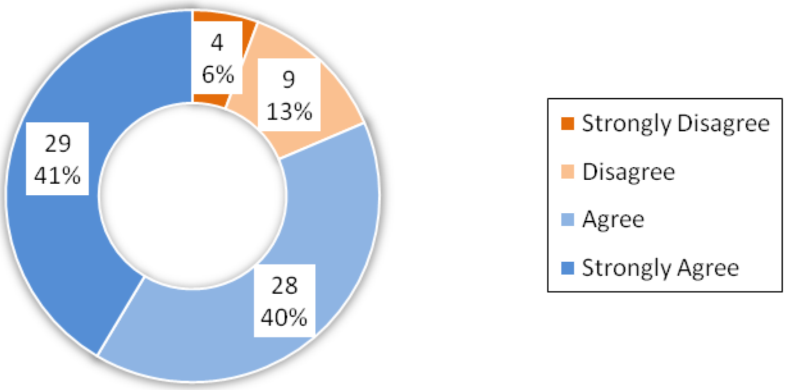
1. Engineering and Management of Information Systems (KTH)
2. Information and Communication Systems Theory (KTH)
3. Interactive Systems Engineering (KTH)
4. Information Security (SU)
5. Economics of Innovation and Growth (KTH)

and/or registered for the following courses:

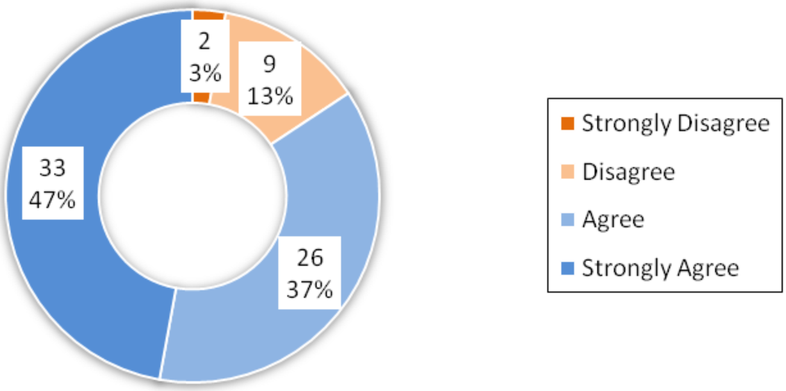
1. System Integration;
2. Strategic Management of IT;
3. Project and Power Games;
4. Business Process Design and Intelligence;
5. Organizations, IT-Systems and Management

Appendix E

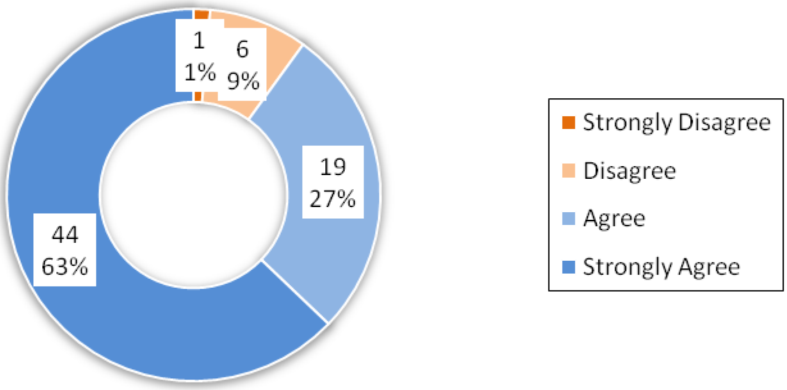
Student Survey Results



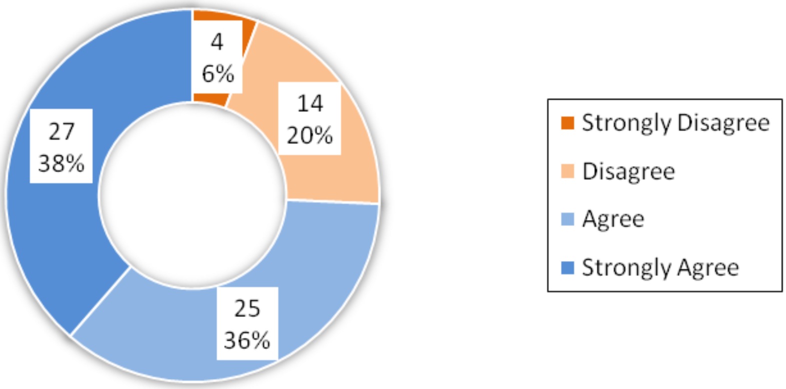
1. I answer all end-of-course evaluations for the courses that I take.



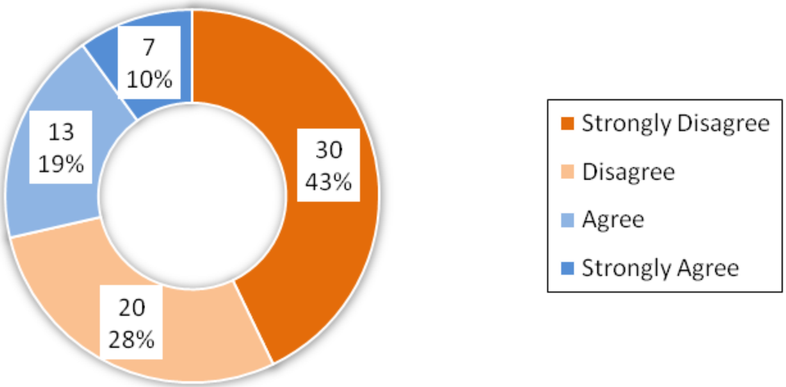
2. I try to provide the best and most comprehensive answers possible.



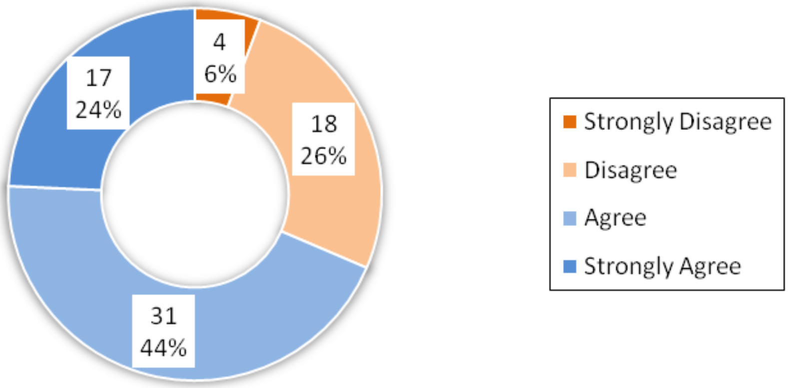
3. When I'm interested in a course, I check course-specific resources more often than otherwise.



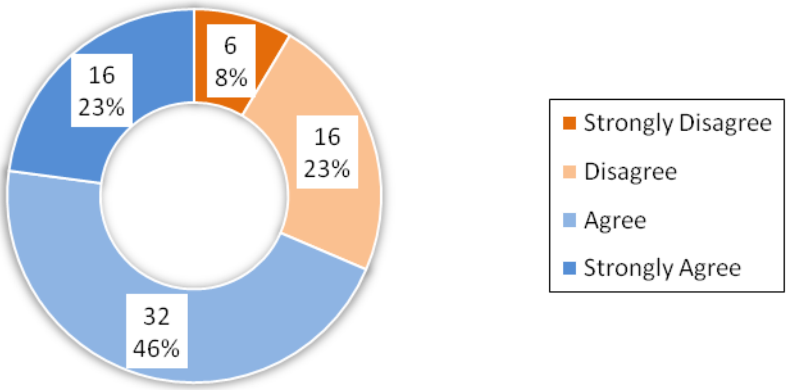
4. I favor a less time-consuming approach to academic evaluation based on how much and how I use academic information systems.



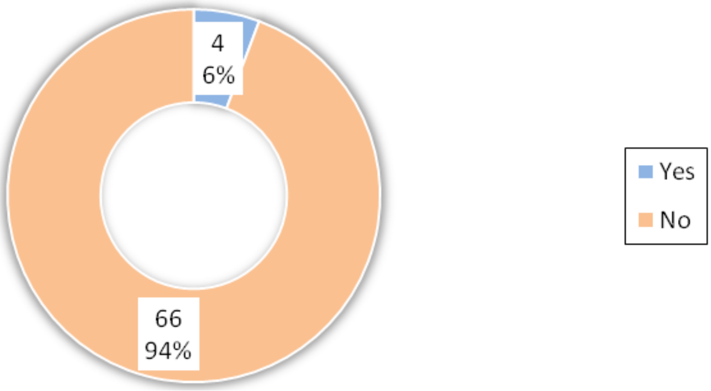
5. I often post, like/dislike, bookmark or re-tweet reactions (status updates) regarding academic experiences (mine/others) on social networks.



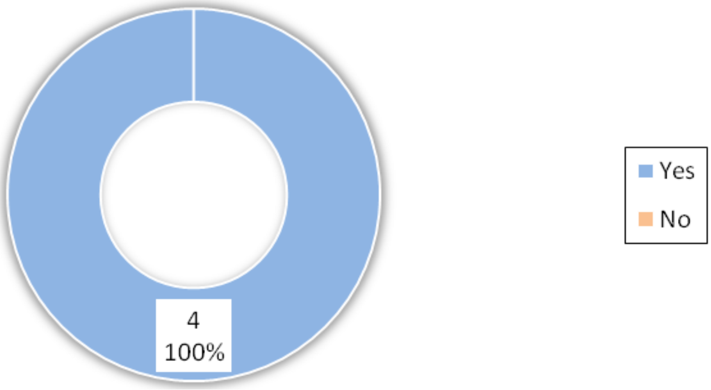
6. I favor a less time-consuming approach to academic evaluation based on simple ratings and comments per course-unit event, just like on current social-networks.



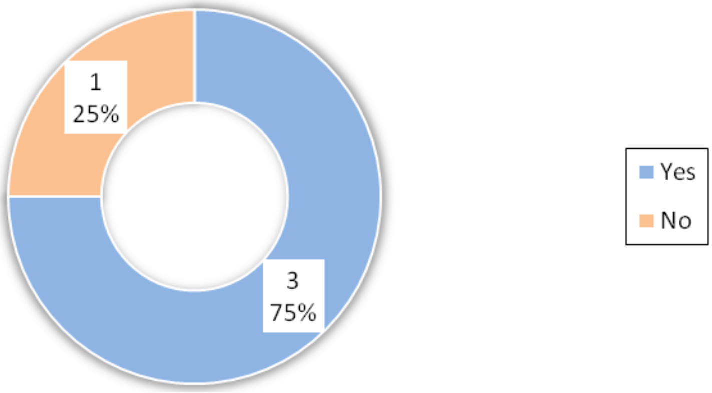
7. I favor a more direct approach to academic evaluation based on feedback at the end of each course-unit and not at the end of the course.



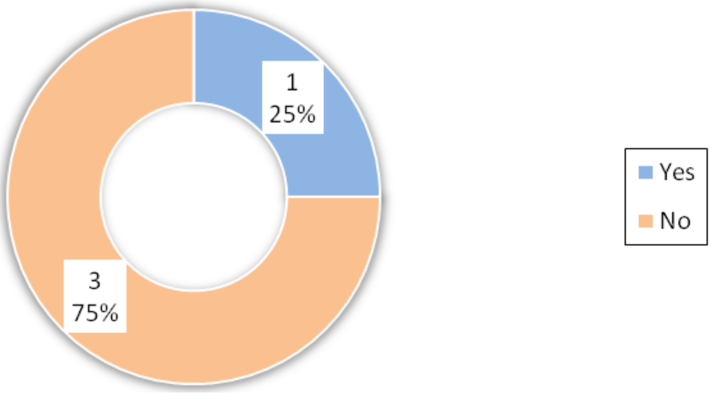
8. I used the Snowdrop application during April 5-25.



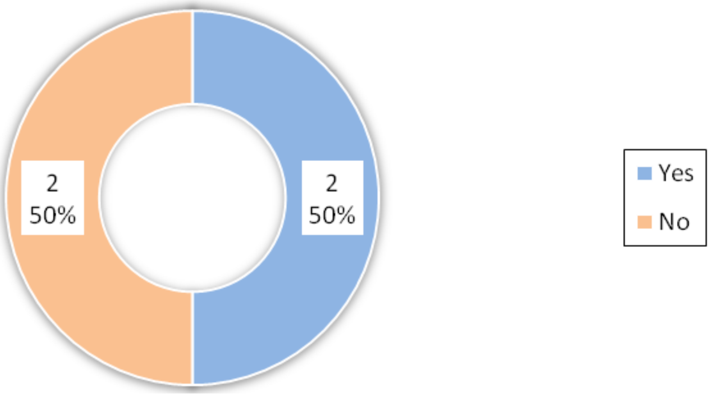
9. I would have used it more often if it was an official DSV system.



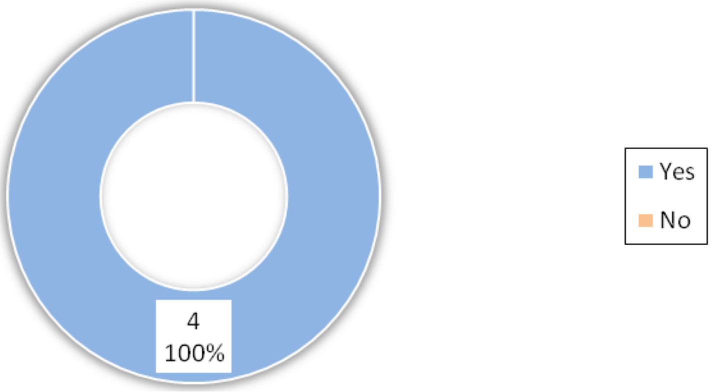
10. I would have used it more often if I hadn't already got used to the main information systems.



11. I would have used it more often but some academic information was missing.



12. I would have used it more often if I didn't fear that my comments would be made public to the course leader(s).



13. I would have used it more often if the interface was more user-friendly.

Appendix F

Instructor Survey

Course Analytics - Instructor Survey

Estimated time: 2-3 minutes
Availability: April 27 - June 3

If you have any type of questions, don't hesitate to Email me at neculau@kth.se .

Thank you!

Andrei Neculau 2009-2010
Master Thesis at KTH/SU, Stockholm, Sweden

http://www.facebook.com/pages/Snowdrop-IS/285063654928
http://twitter.com/snowdrop_is

*Obligatorisk

All students give course feedback through end-of-course evaluations. *

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I trust that the evaluation questions are highly accurate and focused. *
A question may be contaminated by leaving space for interpretation.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I trust that the evaluation answers are highly accurate and focused. *
An answer may be contaminated by allowing little time for a thorough answer (i.e. due to lack of interest/motivation in giving correct feedback).

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I evaluate my course(s) by asking questions related to *

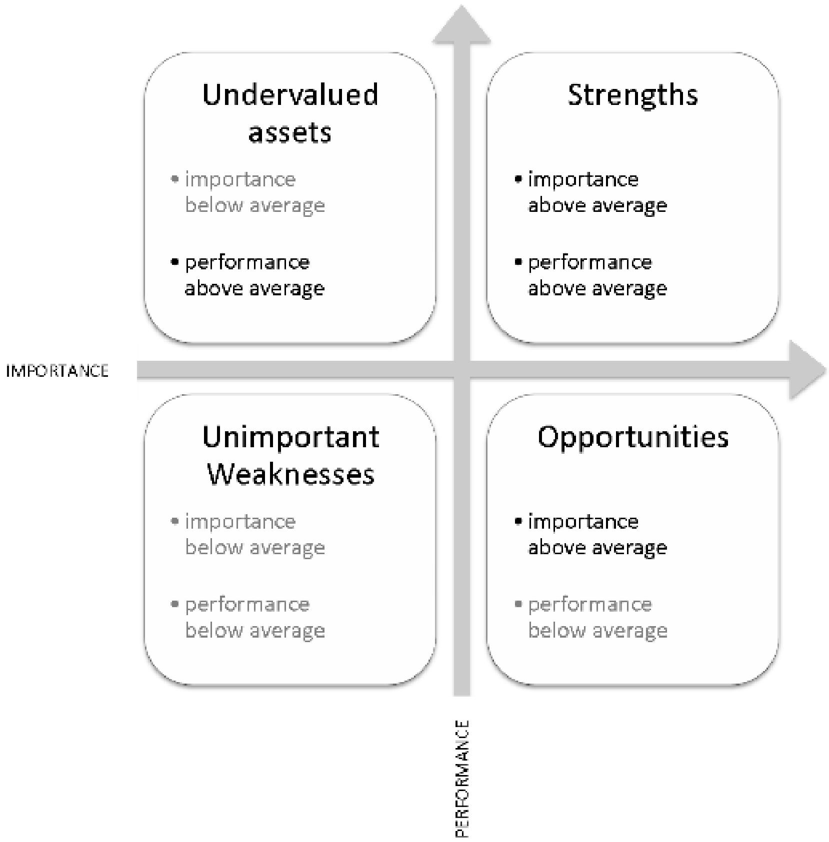
If both options apply, pleace go with the one that covers MOST of your questions - which is the focus of your evaluations?

- Teaching Effectiveness (e.g. if the stated course goals were fulfilled, if the course material was appropriate for the course, etc.)
- Satisfaction achieved through the Educational Service (e.g. if students are satisfied with the student-instructor interaction, method of teaching, etc)
- I do not understand the options above

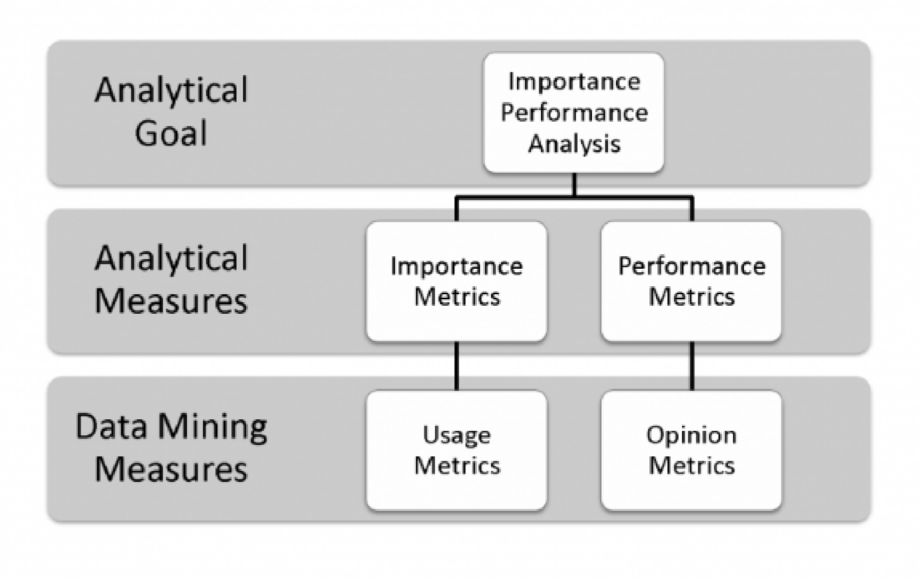
Importance-Performance Analysis

The Important-Performance Analysis is a model based on the core of service quality research - the confirmation-disconfirmation paradigm (are the expectations being met and result in satisfaction or not).

This type of analysis can evaluate different criteria and categorize them as undervalued assets, strengths, opportunities or, respectively, as unimportant weaknesses.



While applied successfully in the academia to evaluate the educational service, new technological means can assist in collecting data to process using this type of analysis.



I favor a less time-consuming approach to academic evaluation based on how much and how students use academic information systems. *

Academic information systems = Daisy, FirstClass, Moodle/VLE, etc

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

I favor a less time-consuming approach to academic evaluation based on simple ratings and comments per course-unit event, just like on current social-networks. *

Feedback would be entirely ANONYMOUS.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

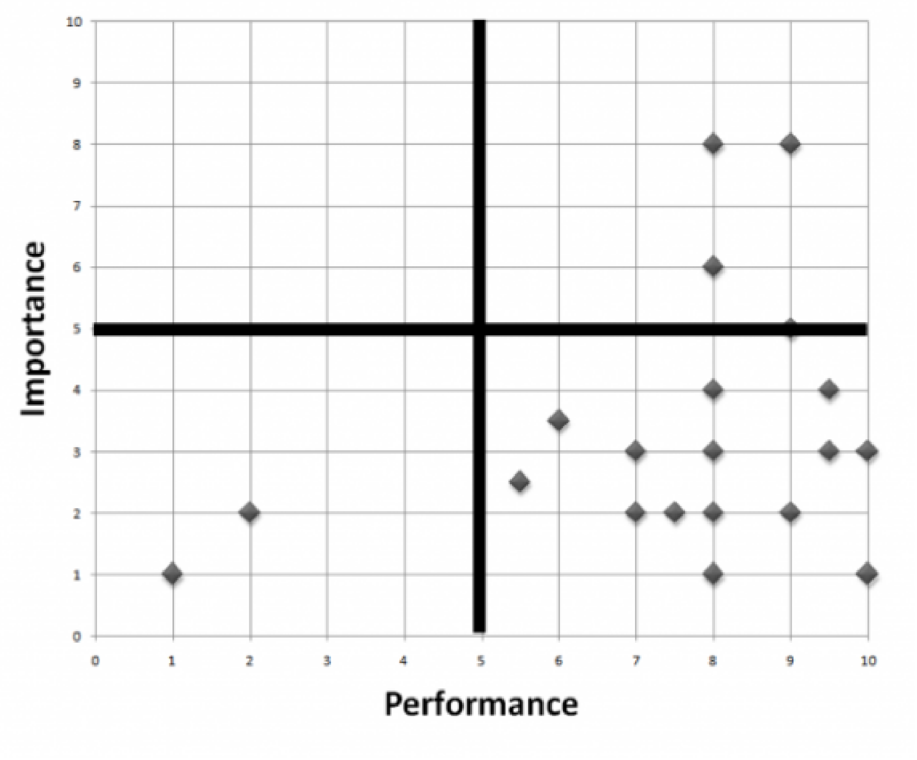
I favor a more direct approach to academic evaluation based on feedback at the end of each course-unit and not at the end of the course. *

Feedback would be entirely ANONYMOUS.

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

Lecture/Seminar/Exam Evaluation

Evaluation data recorded through the two above approaches can result in a graphic for each course unit (lecture, seminar, exam, etc) similar to the following one:



I can very well interpret such a graphic. *

	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

Such a graphic would be highly useful in assessing a course unit, and some times even shape the basis for later improvements. *

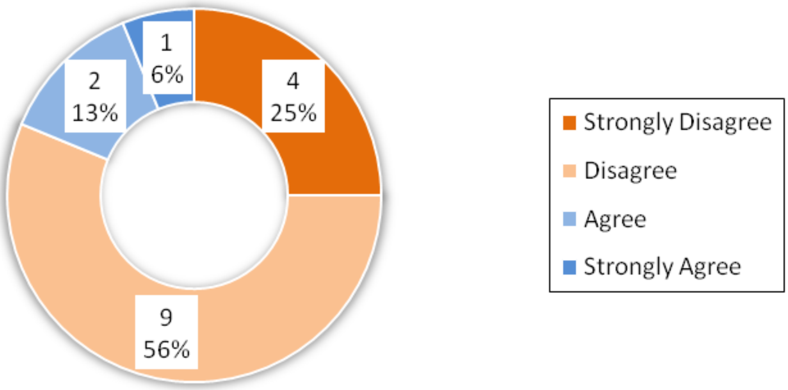
	1	2	3	4	
Strongly Disagree	○	○	○	○	Strongly Agree

Skicka

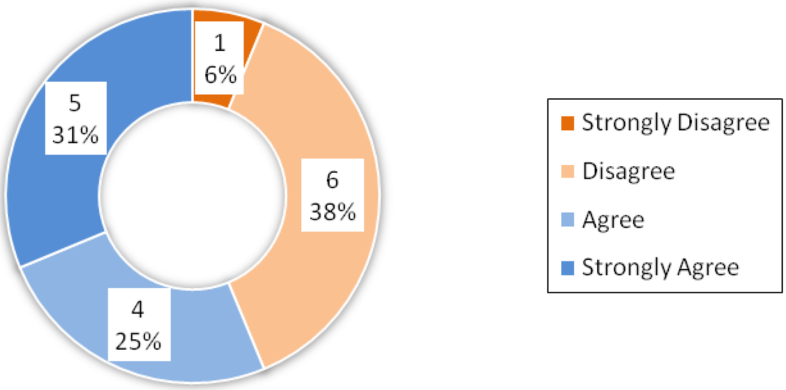
Från Google Dokument

Appendix G

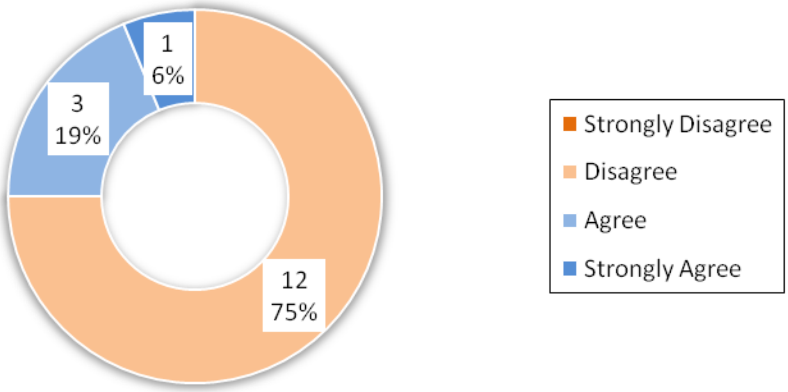
Instructor Survey Results



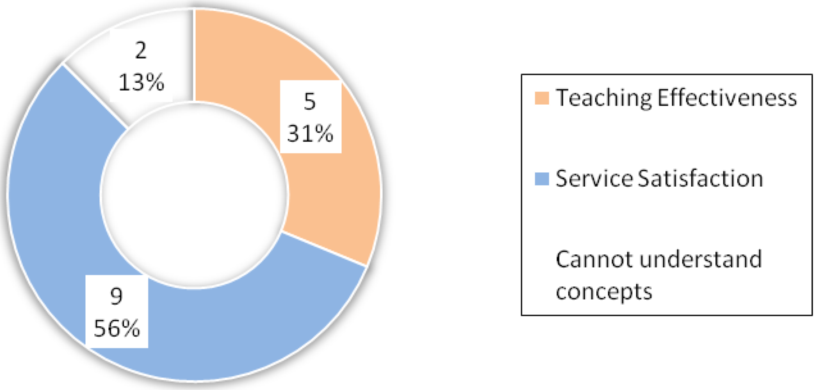
1. All students give course feedback through end-of-course evaluations.



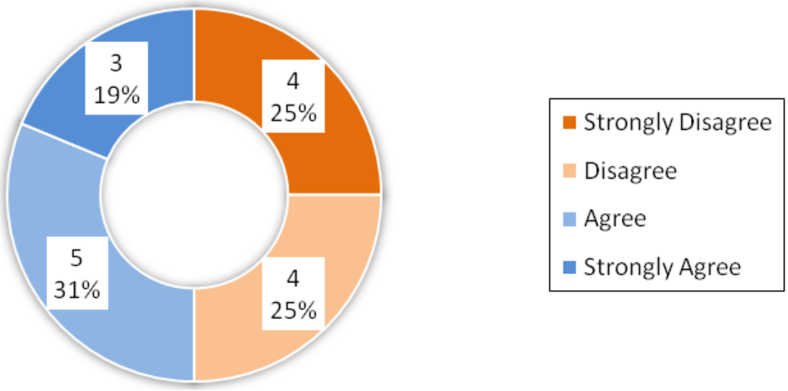
2. I trust that the evaluation questions are highly accurate and focused.



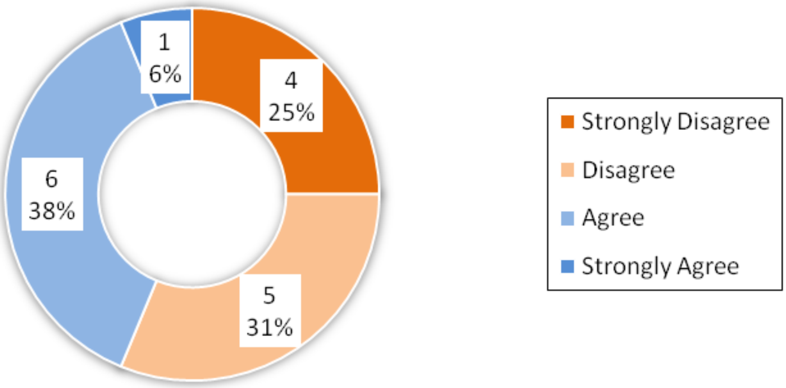
3. I trust that the evaluation answers are highly accurate and focused.



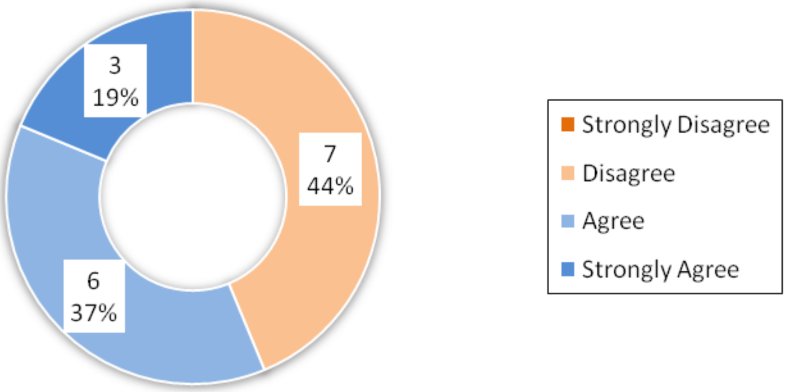
4. I evaluate my course(s) by asking questions related to



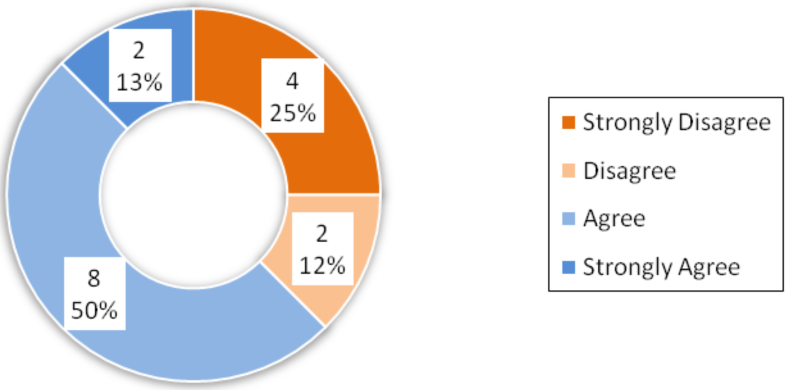
5. I favor a less time-consuming approach to academic evaluation based on how much and how students use academic information systems.



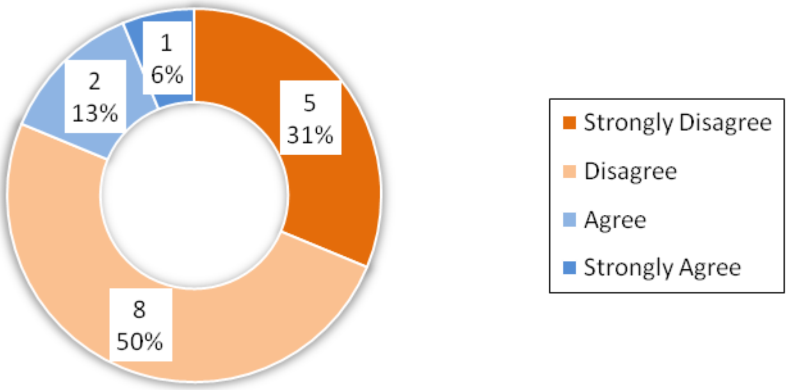
6. I favor a less time-consuming approach to academic evaluation based on simple ratings and comments per course-unit event, just like on current social-networks.



7. I favor a more direct approach to academic evaluation based on feedback at the end of each course-unit and not at the end of the course.



8. I can very well interpret such a graphic.



9. Such a graphic would be highly useful in assessing a course unit, and some times even shape the basis for later improvements.