HELSINKI UNIVERSITY OF TECHNOLOGY
Department of Engineering Physics and Mathematics

Henrik Ingo

Using Online Quizzes as a Study Aid in a Course on Probability and Statistics

Master's thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology

Espoo, January 17, 2005

Supervisor: professor Raimo P. Hämäläinen
Instructor: lecturer Ilkka Mellin

Teknillinen korkeakoulu
Teknillisen fysiikan ja matematiikan osasto

DIPLOMITYÖN TIIVISTELMÄ

Tekijä	Henrik Ingo
Osasto	Teknillisen fysiikan ja matematiikan osasto
Pääaine	Informaatiotekniikka
Sivuaine	Uuden median sisältötuotanto
Työn nimi	Todennäköisyyden ja tilastotieteen peruskurssin opiskelun tehostaminen webin kysymyssarjoilla
Title in English	Using Online Quizzes as a Study Aid in a Course on Probability and Statistics
Professuurin koodi ja nimi	Mat-2 Sovellettu matematiikka
Työn valvoja	Professori Raimo P. Hämäläinen
Työn ohjaaja	Erikoisopettaja Ilkka Mellin

Tiivistelmä

Tutkimus verkko-opetuksen ympärillä on ollut vilkasta ja kehittyy edelleen. On esitetty tuloksia sekä verkko-opetuksen puolesta että vastaan. Yleensä verkko-opetusta verrataan perinteisiin opetusmenetelmiin. Sen sijaan tutkimukset joissa keskityttäisiin yksittäisten verkko-opetusmenetelmien tutkimiseen ovat harvassa. Tässä työssä on tutkittu yhtä yleisimmistä verkko-opetuksen työkaluista, web-pohjaisten kysymyssarjojen käyttöä. Kysymyssarjoista ei ole julkaistu juurikaan tutkimuksia. Janicki & Steinberg (2003) löysivät tukea hypoteesilleen, että tehtävien ja "miniquizzien" käyttäminen edistää oppimista.

Osana työtä kehitettiin järjestelmä kysymyssarjojen tekoon ja julkaisemiseen. Järjestelmää käytettiin todennäköisyyslaskun ja tilastotieteen peruskurssilla. Järjestelmässä käytettiin XML-pohjaista kieltä kysymyssarjojen määrittämiseen. Tämä osoittautui käytännölliseksi ja sitä pidettiin parempana kuin selainkäyttöliittymään ja tietokantaan perustuvan järjestelmän käyttöä, mikä on nykyään vallitseva tapa luoda kysymyssarjoja.

35 % verkkokurssilaisista ja 6,8 % luentokurssilaisista käyttivät kysymyssarjoja. Harjoitukset olivat suositumpia, 88 % ja 77 % opiskelijoista palauttivat niitä. Merkittävää korrelaatiota kysymyssarjojen käytöllä ja koemenestyksen välillä ei ollut. Opiskelijat vastasivat usein uudelleen samaan kysymyssarjaan, jos eivät ensimmäisellä kerralla saaneet kaikkia vastauksia oikein. Verkkokurssilaisten ja luentokurssilaisten koemenestyksessä ei ollut eroja. Korrelaatio harjoitustehtäväpisteiden ja koepisteiden välillä oli 0,45.

Enimmäkseen kysymyssarjoja käytettiin kokeeseen valmistautumiseen. Opiskelijalle olisi hyödyllistä, jos tarjolla olisi kysymyssarjoja, jotka kattavat kerralla koko koealueen ja opiskelija voisi sen avulla selvittää mitkä asiat jo hallitsee ja mitä kannattaa vielä opiskella. Järjestelmään tulisi lisätä mahdollisuus, jossa opiskelija voi itse valita, haluaako nähdä oikeat vastaukset heti ensimmäisen vastauskerran jälkeen.

Sivumäärä	71	Avainsanat	verkko-opiskelu, monivalintakysymykset, kysymyssarjat
Täytetään osastolla			
Hyväksytty		Kirjasto	

Helsinki University of Technology

Department of Engineering Physics and Mathematics

ABSTRACT OF MASTER'S THESIS

Author	Henrik Ingo
Department	Department of Engineering Physics and Mathematics
Major subject	Information technology
Minor subjects	Content production
English title	Using Online Quizzes as a Study Aid in a Course on Probability and Statistics
Finnish title	Todennäköisyyden ja tilastotieteen peruskurssin opiskelun tehostaminen webin kysymyssarjoilla
Chair	Mat-2 Applied Mathematics
Supervisor	Professor Raimo P. Hämäläinen
Instructor(s)	Lecturer Ilkka Mellin

Abstract

Research on e-learning is still rapidly developing and results both in favor and against e-learning methods can be found. Most research focuses on comparing e-learning with traditional lecturing techniques. Almost completely missing are studies focusing on individual e-learning technologies. This thesis focuses on one of the most common tools used in e-learning, quizzes. Not much research on quizzes has been published. Janicki & Steinberg (2003) find support for their hypothesis that "including exercises and miniquizzes [...] will have a positive effect".

As part of this thesis a platform for authoring and publishing quizzes was developed. The system was used in a basic course on probability and statistics. The system used a markup language based approach. Authoring quizzes in XML turned out to be very convenient compared to a "web-based GUI & database" approach, which is the prevalent paradigm.

35 % of online students and 6.8 % of lecture-based students used the quizzes. The exercises were more popular with 88 % and 77 % of students returning exercises. There was no significant correlation between quiz usage and exam performance. Students will often retake a quiz if they don't get all questions right the first time. On the course online students performed equally to lecture-based students. There was a correlation of 0.45 between student exercise activity and exam performance.

Most use of quizzes was in preparation for an exam. It would benefit the student, if he could take a quiz covering the entire range of exam topics and then get feedback on which topics he did well and which should be given more study. The quiz system should be developed to support an option for the student to choose whether he wants to see the correct answers along with the results or not.

Number of pages	71	Keywords	e-learning, quizzes, online learning, ALN, CSCL
Study secretary fills			
Thesis approved		Library code	

Preface

This work was carried out at the Systems Analysis Laboratory at the Helsinki University of Technology. It was part of the laboratory's OtaStat project, which develops eLearning material for courses in probability and statistics.

I wish to thank my supervisor lecturer Ilkka Mellin for the opportunity to work in the project and for his guidance and cooperation. I also wish to thank other members of the OtaStat project, of whom I have learned new things both in statistics and XML, as well as the rest of the laboratory personnel for their friendship during my time here.

As this thesis will hopefully lead to my graduation I wish to thank my parents Sinikka and Bo Orvar for all their love and support and efforts during the many stages of my education, as well as my dear little brother, who also was my schoolmate during many years, for the courtesy of not graduating before me. I'm especially thankful for all the love and support my wife Sanna has provided me with as well as encouraging and inspiring me in my work.

I'm genuinely grateful for all the Finnish taxpayers, thanks to whom I have enjoyed altogether 19 years of free and top quality education.

Henrik Ingo

Table of Contents

Illustration Index

Index of Tables

Terms and acronyms

Term	*Explanation*
anti-aliasing	Computer screens have much smaller resolution than printed paper. Therefore text printed on the screen will not look as smooth as text printed on paper, round parts of the letters especially will become jagged. Anti-aliasing is used to make text (or any lines and graphics) smooth again by using different shades of gray at the border of lines. This fools the eye into not seeing the rough edges, even though the resolution has not really improved.
CSS	Cascading Stylesheets. CSS is the language used to define the visual aspects (the layout) of HTML documents (or any XML document). This is in contrast to the fact that HTML elements only define the structure of the document. For example: The term "heading" refers to structure while "large bold font" refers to layout.
GUI	Graphical User Interface. Most computer programs today (for instance, all Windows programs by definition) have a GUI that is typically used by pointing and clicking with a mouse. One alternative interface is the command-line interface, where the user will issue textual commands, one at a time, and the program also responds simply by printing text.
HTML	Hyper Text Markup Language. The language used to create web pages. HTML is not a programming language in that it is used for presentation, not functionality.
LaTeX	LaTeX is another markup system, based on the TeX typesetting program. It is mostly but not exclusively used to produce paper documents. (It is not XML-based.)
OSML	OtaStat Markup Language. An XML-based language developed within the OtaStat project that is used to publish course material, especially well suited for mathematical topics. Basic notation is identical with HTML.
WYSIWYG	What You See Is What You Get. An acronym used to refer to computer programs that present the data in it's final visual form also when it is being edited. Microsoft Word is a classic example, it tries to show the written text on the computer screen exactly as it will be printed on paper.
XML	Extensible Markup Language. A general specification, from which special markup languages can be created. XHTML (the newest version of HTML) is an example of one such special markup language.
XML element	Elements are the "commands" in XML, the parts of the document read by the computer but not part of the final document as seen by the reader.

1 Introduction

During the last ten years e-learning has been a focus of active development and research both in the academic and business world. We have found that most academic research has been focused on comparing e-learning with traditional lecturing techniques, often using student performance on a final exam as an important measure. What has instead been almost completely missing are studies focusing on the strength and weaknesses of individual e-learning technologies. This thesis therefore focuses on one of the most common tools used in e-learning, quizzes.

Quizzes are an integral part of e-learning, a tool for publishing quizzes can nowadays be found in all serious e-learning platforms. Quizzes have also been an important part of traditional teaching, but since computers now make it possible to completely automate the checking of answers, the interest in quizzes especially in an e-learning setting is understandable. Interestingly, quizzes are both interactive, yet still suitable for studying alone. They are one of the few e-learning technologies possessing both of these qualities that are seen as strengths of e-learning. Despite their importance, which is universally agreed upon, not much research on quizzes has been published.

As part of this thesis a platform for authoring and publishing quizzes was developed within the OtaStat project[1] at Helsinki University of Technology. The system was used in the spring 2004 during a basic course on probability and statistics "Mat-2.091 Sovellettu todennäköisyyslasku".

1 http://www.otastat.hut.fi/

Chapter 2 is a discussion and literature review on e-learning in general, while chapter 3 focuses on quizzes. In chapter 4 the OtaStat Quiz system is presented along with a discussion on decisions taken during it's development. Chapter 5 reports on the pilot courses and results. Chapter 6 concludes the thesis with a summary and discussion.

2 E-Learning

2.1 Terminology - too much of it

E-Learning as a word is clearly a child of the late 1990's, when the letter "e" was prepended to various words in order to transform traditional business into eBusiness. During that so called "dotcom-boom" e-learning was predicted to become a prosperous market for everybody involved and many technology companies and educational companies ventured into the field. Lots of software and many websites were born, although many of them later died out or were consolidated with each other as it became clear that like in most fields in IT at the time, market expectations had been set a bit too high.

Also universities and other educational institutions developed their own e-learning activities and research programs. It is possible that it's roots in the hype of the dotcom-era is one reason why academia has shun away from using the word "e-learning" and instead is trying to use terminology that has not been muddled by commercial marketing. But since it is so dominant in the popular press and marketing speak, the academic world has not been able to standardize on an alternative term. This has led to a situation, where reports on e-learning research can be found under an abundance of different search terms, except "e-learning", which academics refuse to include even as a keyword. Some terms appearing in the references of this thesis include "web-based" or "web-enhanced" education, teaching or courses (Brusilovsky & Miller, 1999; Pathak & Brusilovsky, 2002; McCreanor, 2000; Maki & Maki, 2002), "virtual university" (Brusilovsky & Miller, 2001), "computer-assisted learning" (Wegner, 2003), "computer-supported learning" (Janicki & Steinberg, 2003),"online learning" (McLaren, 2004), "learning networks" (Brooks, 1997) and the list could go on and on.

In addition to that list, two communities are especially worth mentioning here. Lot's of quality research is done under the term "Asynchronous Learning Networks" or ALN. The website [http://www.alnresearch.org/] features articles mostly focusing on measuring the impact of e-learning versus traditional education, for instance whether e-learning students will perform better than traditionally taught students on an exam. Another rather large community is the one gathered under the name Computer Supported Cooperative Learning or CSCL which owes it's name to the more general community of Computer Supported Cooperative Work. For a discussion of whether or not those two are the same thing, see (Schmidt 2001).

It is true that the word "e-learning" has probably seen some inflation due to it's use as a marketing term. After all, advertisers have the bad habit of calling anything they want to sell by a name that happens to be popular and "hot". As a consequence, it might be unclear to many people what e-learning actually is, or somebody might be afraid of using the term because you will never know what bizarre hype might be associated with it in the future. In fact this author has previously shared all those sentiments. It is this authors opinion, however, that the current situation is no better. For instance, when querying a database, it seriously hampers a researchers ability to find relevant articles and publications, when one has to use dozens of different search terms only to cover one subject. And even if the exact meaning of "e-learning" might not be well defined, how is "ALN" any better? Who knows what that means? Is it the same thing as "CSCL" or is it something different?

One very revealing example of the futility of this academic desire of trying to use a very specific and well defined name is found in (Spencer & Hiltz, 2001). They first define ALN: "The terminology of ALN was coined by the Sloan Foundation to distinguish anytime/anyplace computer-mediated communications (CMC), such as bulletin boards and email, from same time versions of distance learning." Bulletin boards and email definitely are good examples of *asynchronous* communication. But after this they expand their horizon such that they express as one of their articles three research questions: "Have various modes of delivering ALN (communications channels, such as *synchronous chat, audio conferencing*, asynchronous conferencing) been compared [...]?" (emphasis added). In all fairness it should also be noted, that consistently with their chosen name, they explicitly require studies included in their review to contain some asynchronous component, such as email or a bulletin board. But it is revealing to note the slip, that they consider synchronous communication to be a valid form of delivering "asynchronous learning networks". So it is somewhat disturbing to see that a paper explicitly purporting to discuss asynchronous learning networks actually contains studies of e-learning courses containing both synchronous and asynchronous components. Even more so when it becomes clear, that in none of the studies reviewed is it possible to distinguish whether some particular result was due to the asynchronous or synchronous components of the study, or both. In the end, the study turns out to be a study on what we define as e-learning in the next section, not a study on anything particularly asynchronous.

Another possible explanation to this unfortunate state of affairs is that academic research on e-learning seems to be rather scattered. There is no one community which we could invite to come together and agree on a term to be commonly used. In fact a lot of research is not reported at e-learning venues at all, but instead in journals and conferences related to the subject that was taught in the e-learning experiment. For instance the review by Spencer & Hiltz (2001) was reported in an "International Conference on System Sciences". Possible future reports from the OtaStat-project are likely to be reported in a journal on statistics or teaching statistics, not necessarily e-learning. Many fields might have dedicated e-learning activities and journals, but these are nevertheless only restricted to the field in question. One such example is the "Decision Sciences Journal of Innovative Education", which is a quality journal, but it's audience is naturally restricted to the community surrounding the Decision Sciences Institute.

That e-learning research is so scattered and conducted under so many different names is almost certain to lead to massive duplication of efforts on one hand, and difficulties in trying to synthesize results on the other. This is regrettable. Finally it must be pointed out that as sad as it is, the current situation might serve to the less noble desires of the academics involved. In a scattered field, every university can claim to be a leader and a central venue of research in it's field and neglect the fact that the same research is being done under other names all over the world. Similarly the individual academic is lured to think that he is the first in his field to publish something, only because earlier publications have happened under different terminology in a different community.

An important goal for the academic research on e-learning should therefore be to become more united, and at the very least try to find some commonly agreed upon terminology. Compared to other challenges in this young field, one would think that it would not be too hard to achieve, but the current situation does not offer much reason for optimism.

So currently in the academic world there is no one term that would have become universally agreed upon, but at least the popular press and business world, even the European Commission[2] is calling it "e-learning". For this reason, that will also be the term used in this thesis. The academic world might continue to be confused on this issue, but at least my mother will know what I'm writing about.

2.2 Benefits of e-learning

Before joining the OtaStat project to work on this thesis, the author developed and taught e-learning courses for a non-academic training business. In addition to the unwarranted commercial expectations towards e-learning, a widespread belief at the time was that e-learning was somehow magically going to revolutionize teaching and learning. More people would be able to learn more things in a shorter time and much better than before. Was this belief equally unfounded, or is e-learning a magic silver bullet?

In order to discuss this question, we should first briefly ask, what exactly we mean by e-learning? Is it a form of distance learning, accomplished through the use of Internet technologies? Or is distance learning e-learning also when traditional video-conferencing, or even video tapes are used to deliver a lecture? If a teacher uses a multimedia presentation in the classroom as part of a lecture, is that e-learning?

With the exclusion of trivial things like merely using a laptop to show simple PowerPoint-slides that could just as well have been printed, we'll define e-learning as being any type of teaching, where *electronic means* are used to deliver or enhance a course. After all, the "e" stems from the word electronic, so it's a fitting definition. But more importantly, this broad and inclusive definition is the one that will serve us best, as we'll conclude at end of this discussion.

In particular, e-learning thus encompasses both distance learning as well as in-classroom teaching enhanced with computers. This will now enable us to answer the original question of whether we can expect e-learning to deliver better learning results than traditional teaching.

2 http://www.elearningeuropa.info/

Because, had we chosen an alternative definition, where e-learning would only refer to teaching with no face-to-face activities at all (online web courses are a good example), we could have answered the question with the following reasoning. Unless we believe that traditional lecturing, a teaching method used for thousands of years, is somehow inherently harmful to students and their learning, we have to accept that online courses can at best be equally efficient to traditional lecturing. (We here use the term "online course" to refer to the alternative definition of e-learning as a form of distance learning, with no face-to-face activities at all.) This would follow directly from the fact that a teacher lecturing in the classroom obviously has the option to take advantage of any of the technologies used in the online setting, in addition to the methods available to him in the classroom setting. For instance, any video or multimedia material could be shown in front of the class, and students can take quizzes or participate in a discussion forum as a form of home exercises. Obviously online courses can have other benefits, like allowing students to participate in a course from a distant location or at a time that best suits the particular individual. (Or the possibility to teach a greater number of students with less resources - an oft mentioned dream we will only mention within parenthesis here.) If we would then find that such an online course is equally effective to face-to-face teaching, we could say that the benefit of the online course is that a broader audience now has access to the same learning possibilities that in-classroom students in any case will have.

If, despite of this reasoning, research would find that a group of students participating in an online course perform better than in-classroom students, it would have to be because of one or more of the following reasons:

1. Lecturing as a method of teaching is harmful to the students, and therefore those who were not subjected to this harmful treatment performed better. (This is a proposition that few would be willing to accept.)

2. The in-classroom students could have performed equally well, if successful methods from the online course had been used by the in-classroom lecturer. (The results are not valid, it was just a case of a bad lecturer.)

3. An online setting is indeed a more efficient way of learning, so students who don't waste time going to lectures use their time more efficiently and therefore perform better. For

instance, in class the pace of the lecture is determined by the lecturer, whereas an online student can divide his time in accordance to his individual needs, thus the time is used more efficiently.

4. There is some other element with the online course that makes the student perform better. For instance, a student may not be at his best at 8:00 in the morning, when the lecture takes place, but with the online course he can study in the evening, which suits him better. Or it could be, that in a lecture he is distracted by friends and therefore performs better when studying from his home.

5. The result is explained by some other contextual factor, and is therefore not a valid result. (For instance, say that online students in a computer programming course perform better than their in-classroom colleagues, and the reason some students are attending the lectures instead of the online course is that they don't have a computer at home. But this is then probably also a good explanation as to why they didn't perform as well in a course on programming.)

As said, few would be ready to accept proposition 1, that school as a place is something that should be categorically avoided. Should we conclude that the reason is number 2, it means that our broad and inclusive definition of e-learning was the right one, since the research would have shown that e-learning (as used both in distance learning and in-classroom) is a valuable addition to traditional teaching methods. 3 and 4 would be interesting conclusions, because we would then have concluded that e-learning methods are indeed significantly better and we should replace most of the traditional lecturing with online course activities. (Note that 3 and 4 differ from proposition 1, in that they don't purport traditional lecturing to be inherently harmful in itself, just that e-learning methods are better.) 3 and 4 would therefore come closest to validating the belief we began to question in this section, that e-learning is somehow magically better than traditional teaching methods. Proposition 5 is simply a catch-all for saying that the study is flawed.

In short, the question about benefits can be answered in two ways. We can hope for e-learning to be able to deliver a course to (distance) students who for some reason cannot attend the classroom lectures. We do not expect this course to be better than the one taught on campus, but if the distance students perform equally well, e-learning has benefited us in bringing the course to a wider audience. On the other hand e-learning material can possibly be used to enhance a traditional lecture.

As part of the discussion we have also come to choose the broad and inclusive definition for *e-learning*, meaning both distance and in-classroom teaching, whenever it is delivered or enhanced with some (non-trivial) electronic means. We use the term *"online course"* to refer to e-learning courses with no face to face activities at all.

2.3 E-learning research

Despite our reasoning in the previous section, most research seems focused on the assumption that e-learning is or should be better than traditional lecturing. Maki & Maki (2002) report that students in a web-based course on psychology learned more and performed better in examinations than their colleagues from a lecture-based version of the same course. In their report however, we find that the instructors in the lecture-based course only used traditional teaching methods like lecturing and discussion groups and did not make use of any of the possibilities that the web-based course could have offered their students. They also remark that part of their result could be explained by the fact that the web-based course turned out to have a higher work-load than the lecture-based course. Also the structure of the web-based course was different, in that it contained quizzes with a weekly deadline, which promoted a more regular rhythm of studying. It should not be surprising to find, that students that study harder, perform better in the examination. This study therefore supports both propositions 2 and 5 from the previous section, but none of the "e-learning is magic" propositions.

Maki & Maki (2002) also summarize the results of 11 earlier studies that compare web-based and lecture-based courses. Of these there are 7 with no significant difference, 2 report that the lecture-based courses delivered better results and only 2 earlier studies by R.H. Maki herself find the web-based courses to be better.

The study by Spencer & Hiltz (2001) summarizes the results of 15 different studies, which report on the performance of e-learning students[3] compared to students attending traditional face-to-face lectures. 10 out of 15, i.e. two thirds reported better results for e-learning, 5 reported no significant difference and none reported better results for the traditional

3 Which in their study obviously are called "ALN students". For the sake of consistency and not confusing the reader any more than necessary, whatever term was used in the original paper will be translated to "e-learning" in this thesis whenever appropriate.

face-to-face approach! These results again are in contrast to our earlier reasoning. The study explicitly calls the e-learning groups "distance students", so we are forced to believe that it is not a case of comparing a mix of the best of both worlds with pure traditional lecturing. The authors are cautious in drawing the positive conclusions supposedly apparent in their data, and instead speculate that it is possible that researchers are more motivated to report positive findings, whereas negative experiences don't get reported. In any case this study does seem to contradict our own reasoning, that distance students should not perform better than on-campus students in a well conducted experiment.

The approach by Benbunan-Fich & Hiltz (2002) is more along the lines of our own reasoning. They compare three groups of students: Pure e-learning, e-learning mixed with traditional face-to-face lecturing, and pure face-to-face methods. Their findings support our view, in that both of the e-learning groups achieve equally high grades, but the group with only face-to-face lecturing performs significantly lower.

It is important to note here, that when doing what scientists call *controlled experiments*, an important step in the experiment is to *randomly assign subjects* to treatment groups. In this context however, it is often not possible, or it is at least seen as ethically problematic to force students to enroll in some specific type of a course, possibly against their own wishes. In all of the studies presented here - with the only exception being (Janicki & Steinberg, 2003) that is presented next - students were not randomly assigned.

Although it is hard to see how it could be remedied, the lack of random assignment is a serious flaw we should always keep in mind when analyzing these results, including the results presented in this thesis! It is completely possible, that there are systematic errors in these comparisons. For instance, it is possible that students who choose to enroll in an online course are more zealous, perhaps more curious and interested in learning new things and better at individual studying. Or, perhaps the teacher giving the lecture class is better than the one teaching online. Or if it's the same teacher, perhaps he is bored of lecturing, having done it for 15 years, and instead very motivated in teaching the online groups. Dozens of such factors spring to mind that become open to speculation when random assignment is not used. In their review Spencer & Hiltz (2001) emphasize this problem by labeling studies as *quasi-experiments*, to caution the reader about the lack of this important feature.

Finally an interesting result is reported by Janicki & Steinberg (2003). They did not compare e-learning and traditional lecturing, but rather two different versions of the web-based course. Interestingly, they used as a control group 10% of the students, who were simply instructed to surf the web for the same topic. The surprising result was, that the control group performed better - although not significantly - than either of the groups that had used the prepared materials. This result suggests that - at least in their particular topics equity and expectancy - quality material is available on the Internet and the relevant materials were found by the students, presumably using a good search engine. They note that such a situation, if validated by future research, could open the door for their vision of a new role of the teacher, who has stepped down from the podium and become more of a coach pointing the students in the right direction.

In conclusion, even if there are already plenty of different e-learning research activities to be found, it looks like research on e-learning is still rapidly developing and results both in favor and against e-learning can be found. If the results are taken at face value, they could indicate that some of the e-learning systems and courses studied were not good, and therefore produced bad results, whereas others fared better. This is very probably one likely explanation, since there have been and currently exist many different e-learning platforms, and courses produced with them or also without any particular platform most probably are of a varying quality. Also the lack of random assignment leave the experiments open to possible external influences, which are difficult to control or account for.

It is appropriate to end this chapter with an enlightened quote from (Brooks, 1997: 27): "Face validity supports still another notion, namely, that certain kinds of teaching and interaction require face-to-face meetings and cannot be conducted electronically. Evidence in support of this notion is intuitive rather than based on the literature." Indeed, we'll have to agree with Mr Brooks, that our reasoning that online courses (e-learning in distance learning) can at best be equally effective with traditional lecturing, is in fact only based on intuitive arguments. As our reviewed experiments are divided on the issue, we will have to leave open the possibility that our intuitive argument might be misleading - like intuition so often can be.

3 Quizzes

3.1 Research on quizzes

Quizzes are one of the most important and popular components in developing e-learning material. Numerous are the occasions that this author has been told about the importance of including quizzes as part of a course and not simply be content on publishing static text and images (or worse, publish the lecture slides as is) and calling it e-learning. All e-learning platforms have tools to create and publish quizzes and many authors who do not use an off-the-shelf platform implement quizzes themselves, using JavaScript or some server-side technique.

With this background in mind it was surprising to find, that quizzes and their use is not widely reported on. Their importance is an accepted fact, like in (Brusilovsky & Miller 1999) which begins with the statement that "tests and quizzes are among the most widely used and well-developed tools in higher education". It is an opening statement that is not under discussion in the paper, and not supported by any statistics or references.

Similarly Wegner (2003) reports experiences from an e-learning course "Introduction to UNIX", and it's evolution during it's 20 years (!) in use. When enumerating the components of the course he mentions in passing that multiple-choice questions "were part of the course and later on turned out to be the major success factor for its survival." Frustratingly, he doesn't elaborate on the topic of this "major success factor" any more than this half-sentence.

It could be argued, that the positive results in (Maki & Maki, 2002) probably were in large part due to the quizzes that were part of the online version of the course, and the fact that the quizzes had a weekly deadline that imposed a more regular rhythm of studying onto students. This speculation is supported by evidence in (Baugher et al, 2003) where students who consistently accessed online course material performed better than those who didn't.

Only one study was found actually reporting on results regarding the use of quizzes. Janicki & Steinberg (2003) compared two randomly assigned groups, which both were observed studying e-learning material in a computerized classroom, at their own pace. One of the groups studied textual material produced with PowerPoint, Word or Notepad while the other group studied modules that had been produced with their WebTAS system, that incorporated methods based on pedagogical principles. The WebTAS system instructed the author to produce each course module such, that it would contain both examples as well as exercises and quizzes in addition to the narrative text. Another guiding principle in the study was self-control, that students should be free to decide for themselves, how much time to spend on each type of material.

They find support for their hypothesis that "including exercises and miniquizzes in the lesson content of Web-based tutorials will have a positive effect on posttest results", but it should be noted that the hypothesis is accepted with a p-value of 0.092 (they use an alpha level of 0.10). Other interesting findings include: no significant difference as to the number of examples included in the material; after the first few modules, different students will start taking different paths and spending different amounts of time in the different styles of learning materials (theory, example, exercise), thus students exercise self-selection as to which style of learning is best for them; and the observation that the control group that was instructed to freely surf the web for relevant information performed equally well with the treatment groups.

Other than this one example (and some case or other non-quantitative studies mentioning quizzes which we will mention later) we find that most research has until today been conducted under the "e-learning versus traditional methods" paradigm, comparing the two as alternatives. The individual components of e-learning are usually not investigated. This is also a result in Spencer & Hiltz (2001): "The various means of ALN delivery, alone or in combination, have not been compared" in any of the 32 studies reviewed! Benbunan-Fich & Hiltz (2002) end their study with a suggestion that "future research should 'open up the black box' to investigate what specific characteristics of the ALN medium and which pedagogical techniques help students to earn better grades and be more satisfied in this environment". This thesis tries to answer to these needs, with regards to one component: using quizzes in e-learning.

3.2 Anatomy and technology of quizzes

A quiz is a sequence of questions, that the student typically answers all at once. The questions are usually yes/no, multiple-choice/single-answer or multiple-choice/multiple-answer, but also open questions with a short and simple answer (fill-in questions) could be included.

More advanced types of questions could be matching-pairs, ordering, pointing on a figure and even graphing questions (drawing a simple figure or connecting lines). It should be noted, that even though most systems do not support the use of these more advanced types of questions, matching-pairs, ordering and some pointing questions can be emulated by breaking the question into one or more multiple-choice questions and possibly using static images as part of the question.

Especially if the quiz is not graded but used for self-assessment, the feedback is an important part of the students learning. In addition to indicating correct and false answers, a quiz should also contain some comments and possibly suggestions as to what topics the student might want to give further study.

3.2.1 JavaScript and other client-side technologies

When discussing web-based quizzes, which are the focus of this thesis and probably also by far the most popular form today, most techniques of implementing a quiz require the use of some server-side technology. Using JavaScript[4] is the only exception, but because of it's nature as a client-side technology it has some limitations. It cannot be used for graded quizzes, since there is no way to store the result outside of the browser. Even if there was a way, it's use for a graded quiz would be questionable because of the fact that the answers must be provided within the human readable HTML page. But for quizzes to be used

4 *JavaScript* is a programming language that can be embedded into HTML webpages in order to make them dynamic or interactive. Since the program, or *script*, is executed by the web-browser, it is a *client-side* technique.

only for self-assessment it is one possible technology. For an example of quizzes implemented with JavaScript, see SticiGui [http://www.stat.berkeley.edu/users/stark/SticiGui/index.htm].

One benefit of using a client-side technology like JavaScript is that there is no need for any kind of support from the webserver. In fact, there is no need for a webserver at all. Therefore JavaScript would be a good choice for implementing quizzes if one wants to make the course material available as a CD-ROM for instance.

There are other client-side technologies apart from JavaScript that could be used to implement a quiz. The most popular would be (in this order) Java, Macromedia Flash, Macromedia Shockwave and ActiveX. Out of these Java and ActiveX are tools for a programmer, while the Macromedia tools are more geared towards animation authoring. Unless there is a compelling reason to use some of these tools, they should not be considered for use in quizzes, because they are all non-standard plugin techniques and therefore a quiz author cannot trust that the students browser will be able to show material developed with these tools. This includes Java, which despite it's reputation as an omnipresent standard is lacking in a standard installation of Windows XP, as well as many Linux distributions. What's more, out of the four technologies ActiveX is not available for any other platforms than Microsoft Windows and Shockwave is only available for Windows and Mac OS. Normal quizzes, which have no need for animation or other advanced multimedia are simpler to produce and can be delivered more reliably without these plugin-based techniques.

3.2.2 Simple static quiz with server-side processing

When a server-side[5] technology is used, one has the option of programming the necessary components by hand or using tools provided in some e-learning platform[6]. Implementing a home-made solution is rather simple and many such implementations can easily be found on the web. In their simplest form, quizzes can be produced by standard HTML authoring software and the only server-side programming needed is for storing the students answers for later grading. In most cases however, this program will also be used to

5 With *server-side* techniques, such as PHP, Java Servlets and JSP, Perl, ASP, etc... the browser sends data entered by the user, such as the answer to the questions, back to the webserver, where the program processes the results and returns an answer to the server. This is a more common way of implementing quizzes than JavaScript. In the work presented in this thesis, a Java Servlet was used.

6 E-learning platforms are not a subject of this thesis, but we'll mention that WebCT [http://www.webct.com] and Blackboard [http://www.blackboard.com] are generally held to be the two most popular platforms. Popular Open Source alternatives include Moodle [http://moodle.org] and OpenUSS [http://www.openuss.org].

immediately check the answers, store the result, and usually also show some feedback to the student. Brusilovsky & Miller (1999) call this type of authoring "static quizzes", since the quiz is a normal web-page and will be the same for all students taking it, every time.

3.2.3 GUI and database

The common approach in almost all e-learning platforms is to provide the author with a (supposedly) easy browser-based graphical user interface (GUI) to create the individual questions. Each question, along with it's correct answer and other additional information is then stored separately in the systems database. Quizzes are then created by indicating which questions belong together and the system will generate the quiz by aggregating these questions. This setup has some advantages over creating static pages. Storing each question as an individual object in the database allows the same question to be re-used in many quizzes. A more advanced system can even generate personalized quizzes for each student, randomly drawing a number of questions from a larger pool. Compared to simpler solutions using a full-fledged e-learning platform has the overhead of installation and maintenance of the system and often also the maintenance of students usernames and passwords. But if an e-learning platform has already been deployed for use in other courses or it is planned to use it for many courses, this is not a serious issue.

3.2.4 QuizStar

There exists at least one such service that is freely available on the Internet, namely QuizStar [http://quizstar.4teachers.org/]. QuizStar could be used by a teacher who wishes to author quizzes in the GUI & database fashion, but doesn't have the option of maintaining his own e-learning platform and database. Even if the QuizStar service is very good and commendable, there are some obvious drawbacks and risks with using a gratis outside provider. There is obviously no guarantee as to the persistence of quizzes created in this free service. A minor annoyance is that the teacher and students need to create additional usernames and passwords to use the system, in addition to any user accounts they already have on the computing infrastructure of their own campus.

The QuizStar system was used in (Koskinen, 2003), but the quizzes never became popular with students taking the course. It is possible that this is because the course also contained other exercises deemed more important, but the registration and login needed to access the QuizStar could have been a factor. Also the risks with using a freely available but outside Internet service have materialized for Koskinen's course. As the QuizStar servers have been upgraded to a new version and even changed their address, the quizzes are no longer available as of this writing. Nevertheless, QuizStar, like many other services provided by 4teachers.org [http://www.4teachers.org/], have become popular with many teachers. Many quizzes are freely available even for the occasional surfer.

3.2.5 Specialized markup language

Yet another method of authoring quizzes is to use a specialized question markup language from which the e-learning platform can generate the quiz in HTML. Brusilovsky & Miller (1999) report that a browser interface approach is much more popular than using a markup language. Wegner (2003) used an XML[7]-based language to produce his course, including the quizzes in it. However, they also developed a graphical user interface to generate the markup language, so that the author would not typically write XML by hand, but rather the creation of quizzes much resembled the way "GUI and database" systems create theirs. Also Bridgeman et al. (2000) used a markup language based on LaTeX[8], although that was used for generating quizzes on paper. The system presented in this thesis also uses its own markup language, which is also XML-based.

A footnote in (Pathak & Brusilovsky, 2002) contains an interesting remark: "A number of our colleagues who were creating quizzes with the top-of-the-line Blackboard 5.1 system noted that the old 'paper and pencil' approach is better. Creating quizzes in a text processor is faster (especially taking into account slow Web-based authoring tools). Automatic grading saves time in large classes, however, in many cases, quizzes are graded by Teaching Assistants and thus this saving does not affect the professors themselves." Having used many different browser-based interfaces himself, this author emphatically agrees with that remark. Most browser-based interfaces are not very good to begin with, but even with a well designed interface the slow response times of the server are usually a problem. It can take seconds for the server to respond each time the user has clicked a button, and that will usually be more than once for each question being created. To the user it will feel that he is mostly waiting and nothing is getting done.

7 Extensible Markup Language. A general specification, from which special markup languages can be created. XHTML (the newest version of HTML) is an example of one such special markup language.
8 LaTeX is another markup system, based on the TeX typesetting program. It is mostly but not exclusively used to produce paper documents. (It is not XML-based.)

Using a solution based on a markup language therefore has an advantage over GUI-based solutions, especially browser interfaces, because one can typically write a whole quiz at once without clicking a single button and possibly even without having to touch the mouse at all. The system developed for this thesis requires the teacher to write fairly complex XML-code by hand, but even so as an authoring experience it was a vast improvement over any GUI-solution this author has used. (It would be possible to create a WYSIWYG[9]-editor for the creation of the quizzes, that would work much like WYSIWYG HTML-editors. This would hide some of the complexity of editing XML-code directly, while preserving the advantages of using a markup language system instead of a GUI-system. Development of such an interface would be easy when using tools such as Xmlspy [http://www.xmlspy.com/products_ide.html], Authentic [http://www.xmlspy.com/products_doc.html] or XMetaL Author [http://www.xmetal.com/], but has currently not been done, and might not be deemed necessary at all.)

A drawback of using a markup language is that one loses the ability to reuse the same question in multiple quizzes. It is possible to have such a feature in a markup language solution, but typically the questions will only live inside one document or quiz. This drawback is easily compensated with the ability to copy-paste questions from one document to another if necessary. A solution that is not completely acceptable by the most puritan standards of data management, where theory says that the same piece of data should never be stored in more than one place. In practice though, copy-pasting is usually a much simpler and more efficient way. In any case, this author's own experience is that questions pooled in a database are almost never reused anyway. In a system used by multiple teachers, one teacher will often not know what questions others have stored in the database and writing your own is usually faster than reading through all of the old ones or trying to query the database with lucky guesses. In one system this author has used, teachers had the habit of always making a copy when including material authored by someone else (at the rare occasions when that ever happened), because there was a valid fear that the original author, not knowing that the material was also in use in other courses than his own, might update the material in such a way that would not be compatible with it's other uses. This habit of copying everything obviously completely defeated the original reason to use such a pool of material in the first place.

9 "What You See Is What You Get". Used to refer to computer programs that present the data in it's final visual form also when it is being edited. Microsoft Word is a classic example, it tries to show the written text on the computer screen exactly as it will be printed on paper.

3.2.6 Variation and parametrization

Except for static quizzes, all of the approaches described above offer the possibility to variation through the use of randomization. Generating the quiz from a larger pool of questions provides each student with an individual quiz or gives one student the opportunity to take more than one quiz. A common approach is also to randomize the order of the questions within a quiz and in the case of multiple choice quizzes, to randomize the order of the choices. These simple techniques ensure that the student at the very least has to read through the question, even if he has seen it before or otherwise already knows the answer.

Yet one aspect of the technology of quizzes is the use of parameterized quizzes, a method that goes beyond merely shuffling the order of questions and choices. This is currently an area of much research, and many prototype technologies have been developed. (Pathak & Brusilovsky, 2002; Kashy et al, 2000; Brusilovsky & Miller, 1999; Bridgeman et al, 2000)

The concept of parametrization lends itself especially well to mathematical problems, where the answer is a number to be calculated. The idea is then to provide the student with a different set of (random) numbers each time the question is generated and to calculate the answer based on those numbers. While math is the most obvious field, Pathak & Brusilovsky (2002) have also successfully developed a system suitable for use in teaching some topics in programming language courses.

While simple in theory, the authoring of parameterized quizzes is usually not as easy as any of the other types we have described. Typically it means that together with each question a simple program has to be provided that knows how to calculate the answer. This means that the author of a quiz now must be familiar with a programming language. Alternatively, the system could contain some kind of general purpose solver (often called a "domain expert"), that would obviate the need for a specific program to go with each question. This is however not a small enterprise to undertake, wherefore such systems are not common. Likely a more successful approach in that case would be to make use of existing mathematics programs, like Mathematica or Matlab, and secretly run them on the webserver to help check the answers. In the case of programming languages it is similarly possible to use an interpreter or compiler of the programming language in question to run the program provided by the student and compare the result with a program known to be correct (Pathak & Brusilovsky, 2002).

See (Brusilovsky & Miller, 1999) for additional thoughts on the anatomy and technology of quizzes.

3.3 Self-assessment versus grading

As we have discussed, quizzes may or may not be graded. When quizzes are provided to the student as a tool for self-assessment, the aim is usually to enhance a narrative text with more interactive material. Student feedback both in (Pathak & Brusilovsky, 2002) and in our own experience suggests that (at least some) students would first not completely comprehend a topic presented in the theory section, but were then able to figure it out when taking a quiz on the same topic. Similarly Janicki & Steinberg (2003) suggest that students perform better when studying material that contains quizzes and exercises in addition to simple narrative text.

Using graded quizzes obviously contains these benefits too, but it is probably fair to characterize graded quizzes more as a tool for the teacher - it is a way of automating exams, at least partially. When quizzes are seen more as exam-like material, it creates additional burdens. First of all, there is now definitely a need to identify students, if nothing else, at least using usernames and passwords. Another issue is cheating. McCreanor (2000) and Kashy et al. (2000) both report interesting experiences on how students would cheat on graded quizzes. Most problems were however overcome with different techniques of variation.

Ultimately the solution to cheating is to use parameterized questions (Pathak & Brusilovsky, 2002; Bridgeman et al, 2000). In fact, when using parameterized questions, the nature of the quiz or exercise changes dramatically. When students traditionally are always told not to discuss or otherwise co-operate when doing graded exercises, they can now be encouraged to do so, because any discussion must necessarily now concern the principles of the solution, not the plain answer itself. Since discussion is also an important component of e-learning (and learning) this is actually an important feature of parameterized quizzes.

It's important to notice, that some important aspects of quizzes as learning material are lost, if the quiz is graded. With graded quizzes, there is usually a need to limit the amount of feedback given to the student immediately after completing the quiz. For instance, correct answers are usually not revealed before all students have completed the quiz. McCreanor (2000) finds out by experience the need to limit the amount of times a student is allowed to answer the same quiz. His students would simply start guessing the answers and retake the quiz until they got all correct.

3.4 Quizzes as a threshold

One idea that might come up when planning to use quizzes is that graded quizzes could be used as part of an e-learning system as thresholds. At the end of each module, the student would be required to take a quiz and would not be allowed to continue until a given percentage of the answers is correct. This would ensure that students don't take on more advanced topics before having mastered the basics. It could also be used as a guarantee that students get equally familiar with all topics of a course and not just specialize in a minimum number of easy topics only in order to pass the course.

We have taken a very cautious attitude towards this idea. As a general rule, a programmer should never underestimate the ways in which users will want to use his system. In this case it is easy to find examples of things that might go wrong. For instance, a student might want to enroll a course, because he is interested in some of the material in the last modules of the course even though he is already familiar with the topics in the first ten modules. This student is now forced to take ten quizzes that although to him are not impossible to overcome, are sure to be an annoyance. Similarly it could be, that the teacher makes a mistake when authoring some questions in a quiz, with the end result that all students get stuck at the point until the problem is fixed, because the computer will not allow them to continue, stubbornly believing that it is right and all others are wrong.

Brooks (1997: 140) is not enthusiastic about the idea either. "In a sense, it required that all of our students be above average."

We should also not forget that one important benefit of putting course material on the web, is that it is accessible to students as reference material even after the course. Students often sell their course books after the course, but online material will always be there. Even better, if access to the materials is not restricted by passwords, not only students, but anybody is able to find them with a search engine and can then benefit from them. But if the material is locked behind thresholds, this important benefit is lost.

However, in addition to this reasoning, a more compelling argument would be that such forced thresholds are unnecessary anyway. Kashy et al. (2000) have found that "allowing multiple tries on assigned problems is highly motivating; most students [voluntarily] strive to get all the work done correctly." We shall return to this question in our own analysis.

4 The OtaStat Quiz system

For this thesis a system to deliver online quizzes was developed as part of the OtaStat project. During the spring term 2004 we authored a number of quizzes covering an entire introductory course in probability and statistics and piloted them on a group attending an online version of the course as well as the students attending normal lectures.

4.1 Early choices affecting the systems design

In early discussion on design choices, the issue of graded quizzes came up. As is customary in many courses at Helsinki University of Technology, our course contains exercises from which students receive bonus points that affect the course grade. We thought that we should similarly motivate the students into using the quizzes. Quizzes would be a voluntary part of the course, but taking them would give bonus points positively affecting the students grade. But we also started to realize the constraints such a system implies. We would not be able to reveal correct answers directly after the students took the quiz, thus taking away the positive effect on learning that feedback has. Also, we might have to put deadlines on each quiz. Since we were interested to study how and when students would use our system, we decided not to grade the quizzes but to leave the situation as open as possible for the students themselves to decide the best way of using the quizzes.

For instance, we thought that one possible way of using quizzes could be to start the studying of a module with taking a quiz, even if the student would get most of the questions wrong. The student could then use the feedback to better understand what topics in the module he needs to focus on and possibly also find out which topics he already knows. A talented student might use such a quiz to decide whether he needs to study the module at all.

Also some students might want to take the same quiz both at the beginning and the end of a module, in order to prove their progress. None of these scenarios would be possible, or they would at least be discouraged, if the quizzes would affect students grades.

As said, we were very cautious of the idea of using quizzes as thresholds. In our case it would not be possible anyway, since we had decided that at least for now, the online course would follow the rhythm of the lectured course. Thresholds can obviously only be used in courses were students advance at their own pace.

Having made these choices, it became clear that there was no need for the system to identify students, thus we would not need a database to store usernames and passwords. This was one of the factors that eventually led us to choose a specialized markup language instead of a GUI and database approach. Not using a database at all would require less resources and maintenance on the server.

Another serious issue was that we would obviously make heavy use of mathematical formulas, a feature that is not supported by the HTML standard and therefore also lacking in most readily available systems. This issue alone made most ready made tools useless, although the most advanced e-learning platforms like Blackboard, WebCT or Moodle do contain support for mathematical notation, which they then convert to a presentation format web-browsers understand (typically an image).

4.2 The OtaStat Markup Language

The OtaStat project had previously developed a markup language to be used in delivering course material over the web. Since we eventually made a rather natural choice to base the OtaStat Quiz system on this earlier work, we will now briefly describe it here.

The OtaStat Markup Language (OSML) is an XML-based language primarily intended for publishing course material. It's basic notation is identical to HTML, but in addition to that OSML contains elements[10] for specifying metadata that is relevant to this kind of use. For instance, a document will contain a list of other documents, which are considered pre-requisite material and the system will then automatically add links to those documents. Similarly the language has support for keywords, from which an index can be generated.

10 *Elements* are the "commands" in XML, the parts of the document read by the computer but not part of the final document as seen by the reader.

The most important feature, and one of the reasons the development of a homegrown system was necessary, is the support for mathematical notation. OSML contains elements for the insertion of mathematical symbols or complete equations. Mathematical symbols are written within special elements with LaTeX[11] math-mode code. The system then extracts these codes and replaces them with an image representation of the symbol or equation.

We should mention, that there obviously exist other systems which also produce mathematical notation in similar ways. In fact, generating images from LaTeX-codes seems to be the most popular way of implementing this already. However, none of these systems included the other features needed for the OtaStat project, so they were not used. Some of the most popular such systems are in fact rather old and hence contain little or no support for techniques like CSS[12] stylesheets. Also some systems use LaTeX to write the complete document, so they are in fact more a way of converting LaTeX-documents into HTML, not a way of enhancing HTML with LaTeX-like functionality as in OSML. Since our aim is primarily to produce material for the web, basing the bulk of the markup language on HTML feels like a more natural choice and will possibly be familiar to more of our users.

In addition to advantages due to the heavy use of CSS stylesheets in documents published with the OtaStat system we have concluded that our OSML-based system is also better than older systems with respect to the representation of mathematical formulas. Some older systems lack anti-aliasing[13] or it is poorly implemented, making the mathematical symbols almost unpleasant to read. We also found that in most if not all implementations the baseline of a formula is not adjusted to match the rest of the text, thus making the formula stand out even more. The OtaStat system corrects these problems. One problem left in all of these systems, including ours, is that the size of an image in a web-page remains constant even if the font size is adjusted by the user. The use of CSS to explicitly define the font size of the document's text helps somewhat, but if a user intentionally wishes to override this setting - a feature available in

11 LaTeX is another markup system, based on the TeX typesetting program. It is mostly but not exclusively used to produce paper documents. (It is not XML-based.)

12 Cascading Stylesheets. CSS is the language used to define the visual aspects (the layout) of HTML documents (or any XML document). This is in contrast to the fact that HTML elements only define the structure of the document. For example: The term "heading" refers to structure while "large bold font" refers to layout.

13 Computer screens have much smaller resolution than printed paper. Therefore text printed on the screen will not look as smooth as text printed on paper, round parts of the letters especially will become jagged. Anti-aliasing is used to make text (or any lines and graphics) smooth again by using different shades of gray at the border of text. This fools the eye into not seeing the rough edges, even though the resolution has not really improved.

most browsers and of great importance for people with poor eyesight or healthy eyes but abnormal screen resolutions - the images will obviously not follow this setting. Out of the browsers commonly used, only Opera has a feature to zoom the whole page, both text and images.

MathML is an XML-based markup language that could be used to add mathematical notation to HTML. It is a standard promoted by the W3 Consortium[14], the same body that defines the HTML and XML standards. But at the time the project started browser support for MathML was still non-existent. Today Mozilla (also known as Netscape) supports MathML and Internet Explorer can be extended with a plugin, while Opera and Konqueror (also known as Safari on Mac OS) have no support. MathML is still not considered ready for widespread use, perhaps mainly because of the lack of out-of-the-box support from Internet Explorer.

It should be noted that the current implementation of OSML is forwards compatible with MathML. The part of the system that currently transforms LaTeX code to images could be replaced with a similar transformation into MathML. There are already such converters available as open source, so the transformation would be straightforward to implement. Also of note is, that MathML is not intended to be written by hand, but rather using WYSIWYG equation editors or some other input language, of which LaTeX is actually the most popular choice.

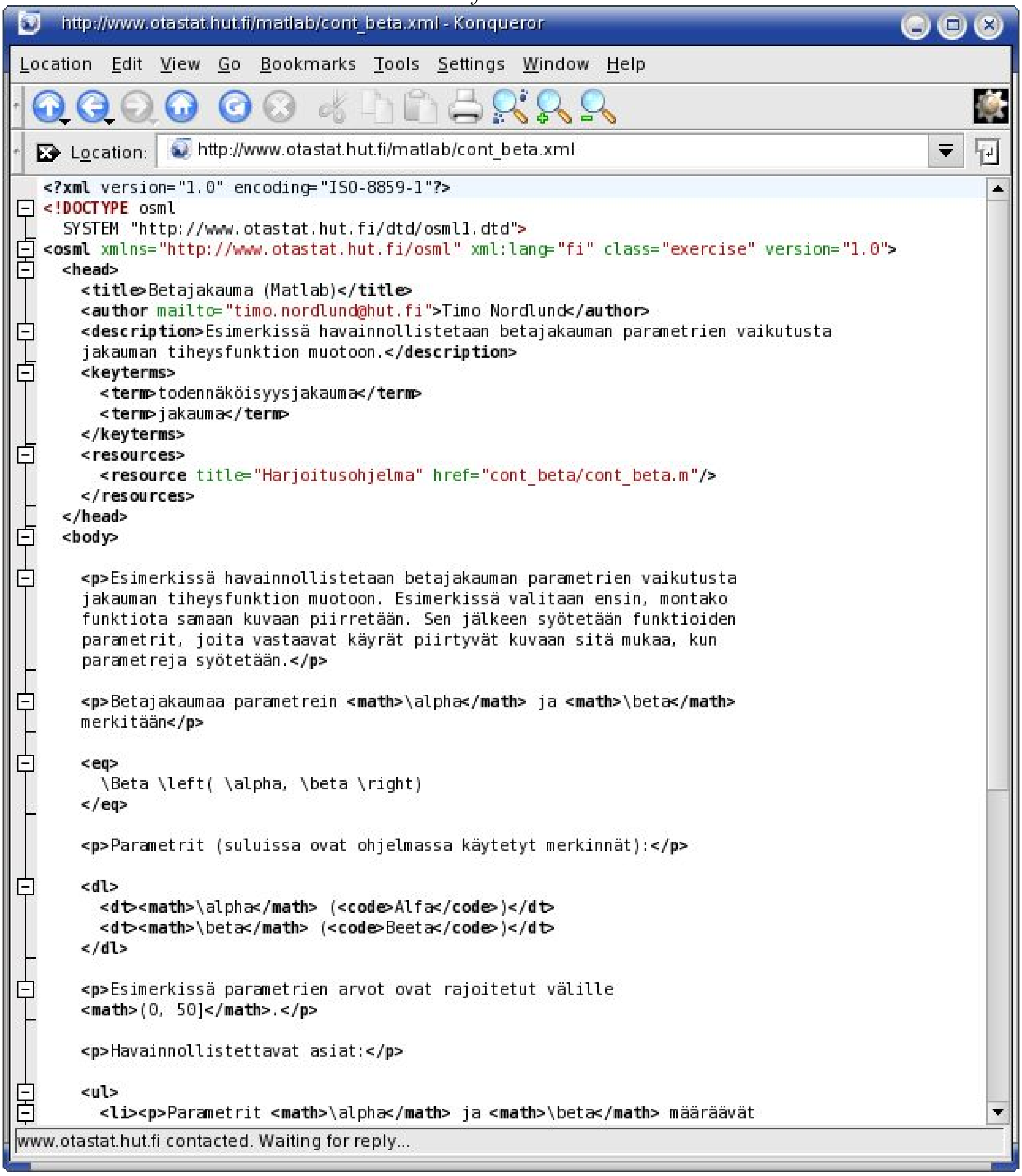
In the future the use of MahtML will therefore probably be reconsidered, since it could be done without having to update the actual documents. Advantages of MathML are that formulas will scale with the rest of the text, it is possible to copy-paste also the formulas and obviously they would be more accessible to seeing impaired than the current use of images. On that topic however, we claim that inserting the original LaTeX-code in the HTML alt-attribute of the images is not a bad workaround for now, especially since most screen readers or braille terminals currently do not support MathML anyway. For more background on the state of MathML, see (Rowlett, 2003).

The final OSML feature we'll present here is that the system is based on pre-processing. That is, all documents, as well as table of contents and index pages are generated

14 http://www.w3.org/

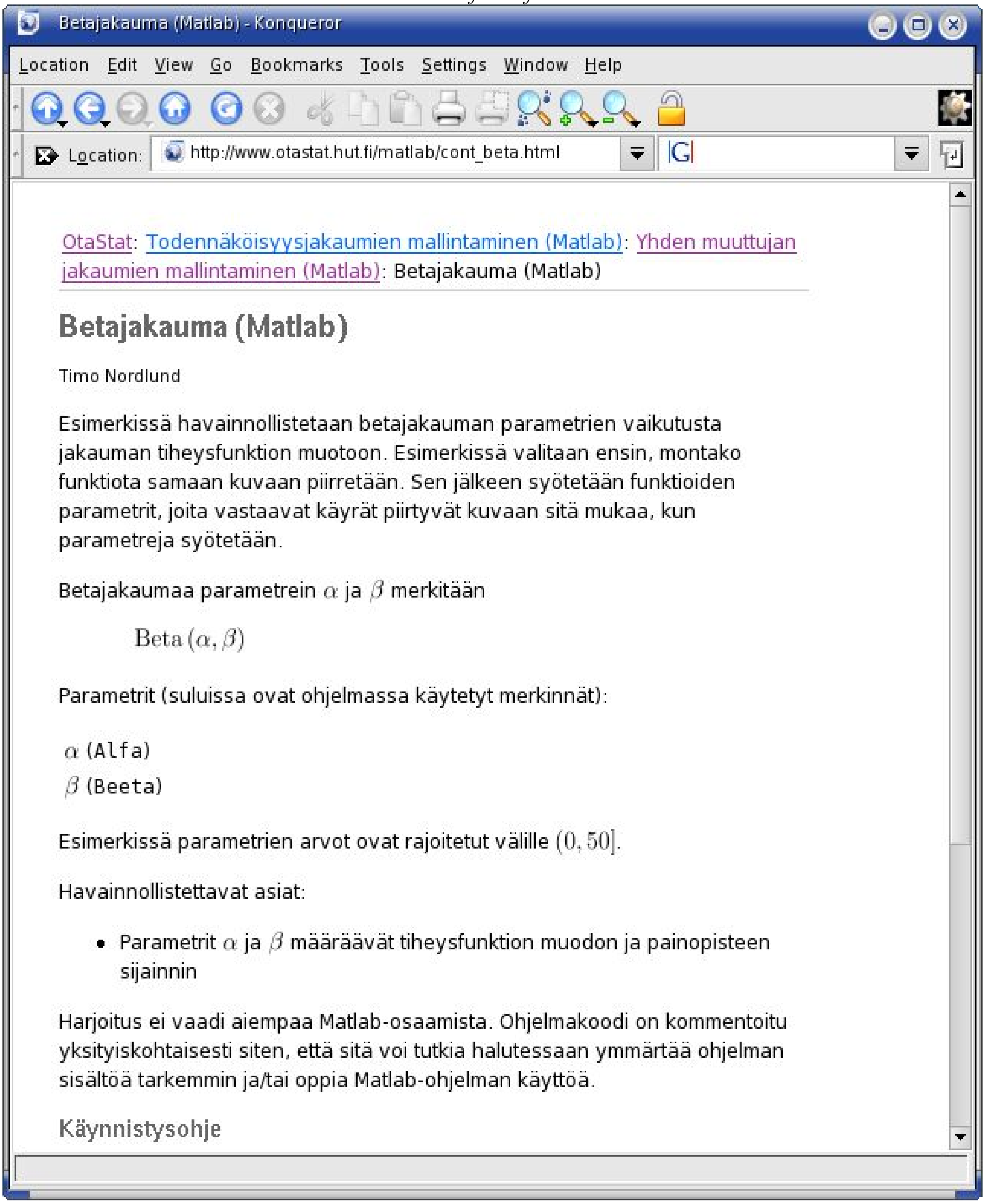
into HTML (and all mathematical notation into images) in advance, after which the files are copied to the webserver for publication. Once the preprocessing is done, the documents are then completely static, so there are no extra requirements on the server side beyond normal webserver capabilities.

Illustration 4.1: OSML document in source form.




```xml
<?xml version="1.0" encoding="ISO-8859-1"?>
<!DOCTYPE osml
  SYSTEM "http://www.otastat.hut.fi/dtd/osml1.dtd">
<osml xmlns="http://www.otastat.hut.fi/osml" xml:lang="fi" class="exercise" version="1.0">
  <head>
    <title>Betajakauma (Matlab)</title>
    <author mailto="timo.nordlund@hut.fi">Timo Nordlund</author>
    <description>Esimerkissä havainnollistetaan betajakauman parametrien vaikutusta
    jakauman tiheysfunktion muotoon.</description>
    <keyterms>
      <term>todennäköisyysjakauma</term>
      <term>jakauma</term>
    </keyterms>
    <resources>
      <resource title="Harjoitusohjelma" href="cont_beta/cont_beta.m"/>
    </resources>
  </head>
  <body>

    <p>Esimerkissä havainnollistetaan betajakauman parametrien vaikutusta
    jakauman tiheysfunktion muotoon. Esimerkissä valitaan ensin, montako
    funktiota samaan kuvaan piirretään. Sen jälkeen syötetään funktioiden
    parametrit, joita vastaavat käyrät piirtyvät kuvaan sitä mukaa, kun
    parametreja syötetään.</p>

    <p>Betajakaumaa parametrein <math>\alpha</math> ja <math>\beta</math>
    merkitään</p>

    <eq>
      \Beta \left( \alpha, \beta \right)
    </eq>

    <p>Parametrit (suluissa ovat ohjelmassa käytetyt merkinnät):</p>

    <dl>
      <dt><math>\alpha</math> (<code>Alfa</code>)</dt>
      <dt><math>\beta</math> (<code>Beeta</code>)</dt>
    </dl>

    <p>Esimerkissä parametrien arvot ovat rajoitetut välille
    <math>(0, 50]</math>.</p>

    <p>Havainnollistettavat asiat:</p>

    <ul>
      <li><p>Parametrit <math>\alpha</math> ja <math>\beta</math> määräävät
```

Illustration 4.2: OSML document seen in it's final form in a web browser.



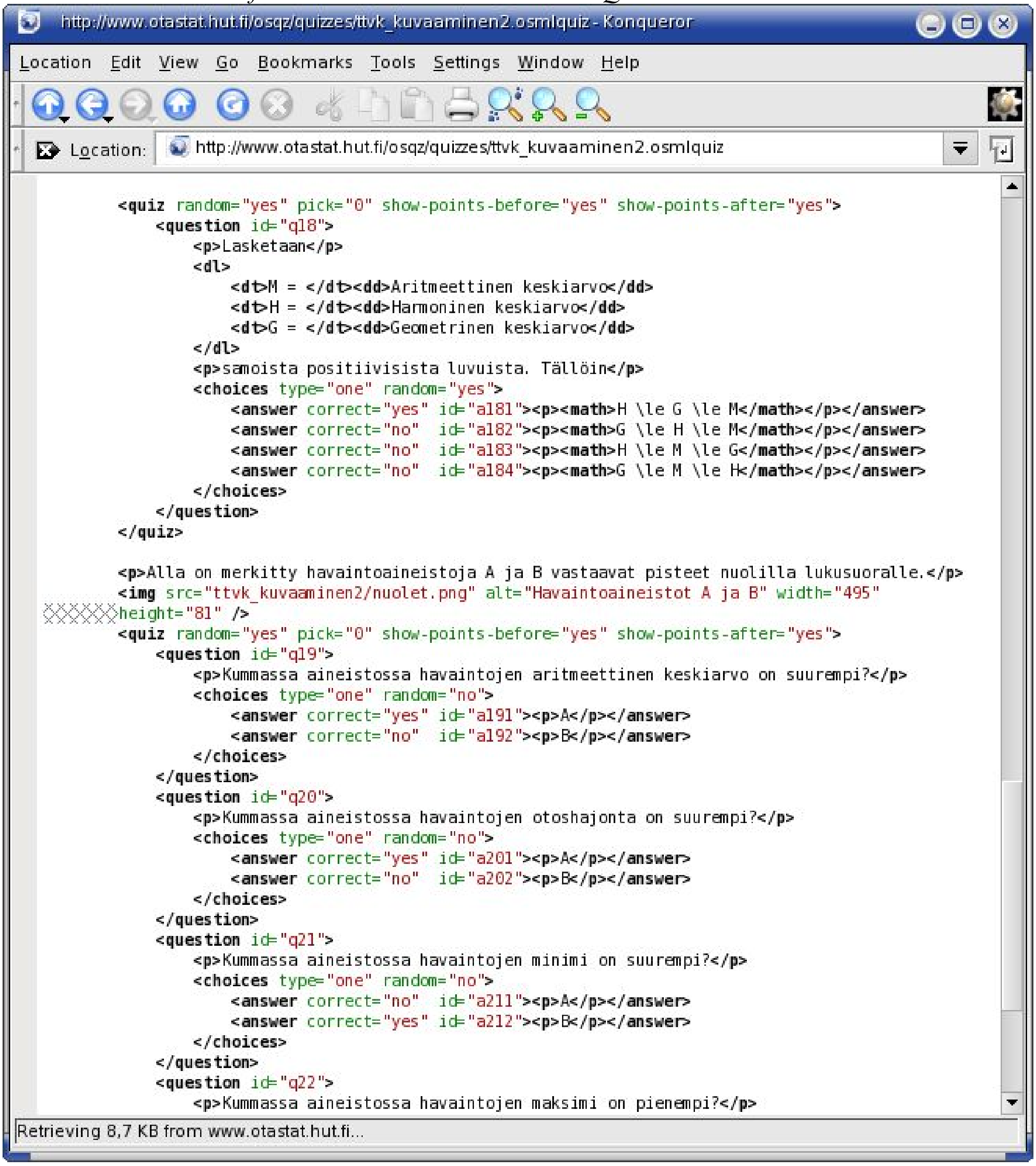
4.3 Developing a markup language for quizzes

We will now give a high level overview of the OtaStat Quiz system, that was developed as part of this thesis. It was decided to base the quiz system on the existing OSML architecture. It turned out to be a wise decision in that practically all OSML features were reused in the OtaStat Quiz system and in particular the handling of mathematical formulas. Another important benefit of this was that the quiz system now automatically has the same layout and "look and feel" as the static documents. Finally and slightly unexpectedly a positive consequence was that using a markup language turned out to be a very convenient way of authoring the quizzes.

The system has support for all kinds of multiple choice questions as well as open questions. It supports variation, both the order of questions and choices can be random. Questions within the same quiz can be grouped so that the variation only happens within the group, but the groups themselves always appear in the same order. This feature is needed if some questions are seen as introductory, whereas some others should be towards the end of the quiz but still in a randomized order. Grouping can also be used if the order of some questions should not be randomized while others should. In addition to varying the order, the system also supports picking a number of questions from the group. For instance one can include several variations of the same question and tell the system to randomly pick one of them. In multiple choice questions one can also include extra false choices. The correct choices have to always be present, but the false ones can be different from time to time.

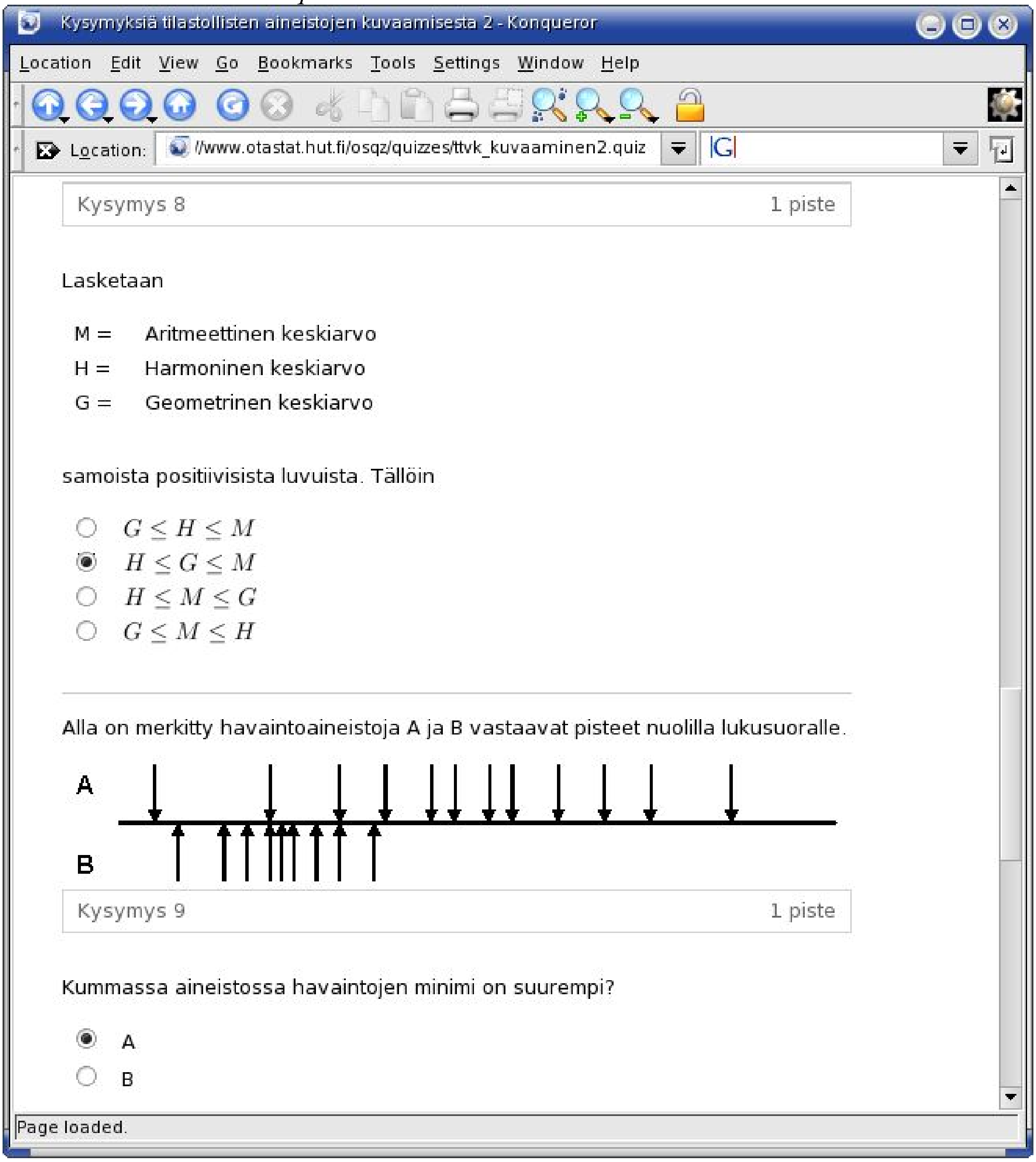
In order to provide these features the basic OSML had to be extended with new XML elements, that are used to describe these things. So in addition to writing the text of the question the author of a quiz will also use additional markup to show what is introductory text, what is a question, what are the choices and also which texts should only be shown after the student has answered the quiz. Some of the attributes to these new elements contain information regarding the variation of questions and choices. It is not our intention to describe the technical details of the system within this thesis, but readers familiar with XML or HTML will probably get a good sense of the markup language from the following examples.

Illustration 4.3: Part of the source XML to an OtaStat Quiz.



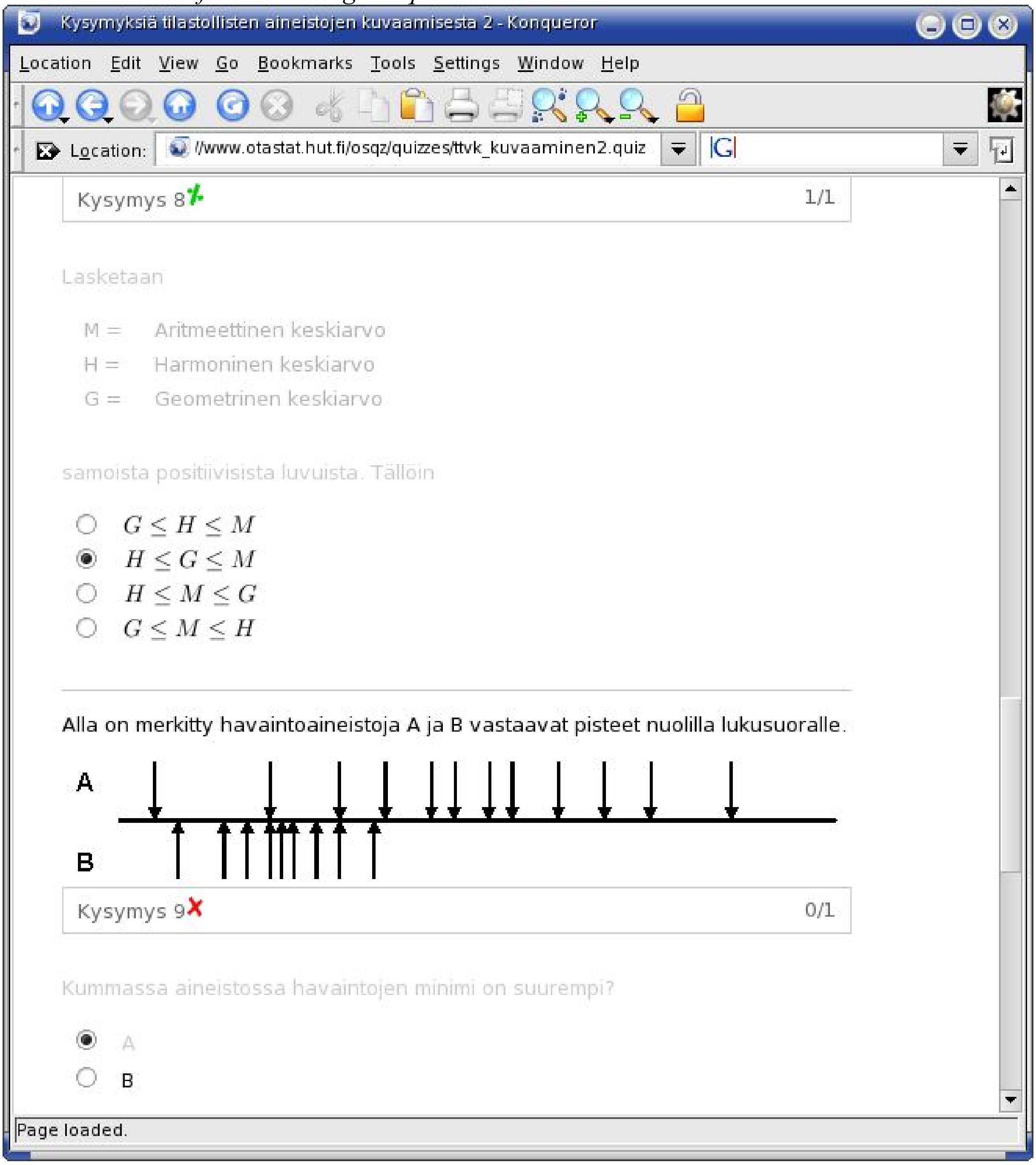

```
<quiz random="yes" pick="0" show-points-before="yes" show-points-after="yes">
    <question id="q18">
        <p>Lasketaan</p>
        <dl>
            <dt>M = </dt><dd>Aritmeettinen keskiarvo</dd>
            <dt>H = </dt><dd>Harmoninen keskiarvo</dd>
            <dt>G = </dt><dd>Geometrinen keskiarvo</dd>
        </dl>
        <p>samoista positiivisista luvuista. Tällöin</p>
        <choices type="one" random="yes">
            <answer correct="yes" id="a181"><p><math>H \le G \le M</math></p></answer>
            <answer correct="no"  id="a182"><p><math>G \le H \le M</math></p></answer>
            <answer correct="no"  id="a183"><p><math>H \le M \le G</math></p></answer>
            <answer correct="no"  id="a184"><p><math>G \le M \le H</math></p></answer>
        </choices>
    </question>
</quiz>

<p>Alla on merkitty havaintoaineistoja A ja B vastaavat pisteet nuolilla lukusuoralle.</p>
<img src="ttvk_kuvaaminen2/nuolet.png" alt="Havaintoaineistot A ja B" width="495"
height="81" />
<quiz random="yes" pick="0" show-points-before="yes" show-points-after="yes">
    <question id="q19">
        <p>Kummassa aineistossa havaintojen aritmeettinen keskiarvo on suurempi?</p>
        <choices type="one" random="no">
            <answer correct="yes" id="a191"><p>A</p></answer>
            <answer correct="no"  id="a192"><p>B</p></answer>
        </choices>
    </question>
    <question id="q20">
        <p>Kummassa aineistossa havaintojen otoshajonta on suurempi?</p>
        <choices type="one" random="no">
            <answer correct="yes" id="a201"><p>A</p></answer>
            <answer correct="no"  id="a202"><p>B</p></answer>
        </choices>
    </question>
    <question id="q21">
        <p>Kummassa aineistossa havaintojen minimi on suurempi?</p>
        <choices type="one" random="no">
            <answer correct="no"  id="a211"><p>A</p></answer>
            <answer correct="yes" id="a212"><p>B</p></answer>
        </choices>
    </question>
    <question id="q22">
        <p>Kummassa aineistossa havaintojen maksimi on pienempi?</p>
```

Illustration 4.4: The same quiz as seen in a browser.



Note that the question after the image (Kysymys 9) is not the same question that is defined after the image in the source code, because the order of questions is defined to be randomized. Similarly the choices in the question above the image (Kysymys 8) are in a random order, whereas the choices to the latter question (9) are not randomized.

Illustration 4.5: After answering the quiz the results are shown.



For each question there is a "correct" or "false" indicator and optionally the points for each answer are shown. Note that all questions and wrong choices are grayed out, in order to highlight the correct option. This feature doesn't work for the mathematical formulas however, which are simply images and cannot be grayed out as easily as text. It is also possible to add an "explanation" element (not seen in this image) to a question or one particular choice, which will only be shown after the student has answered the quiz. The content of an explanation can be conditional on how the student answered.

We did not implement support for matching-pairs or other types of advanced questions, but as we mentioned in chapter 3.2 (page 19), many of these can be emulated with one or more multiple-choice questions. At first we planned to have parameterized questions as a feature, but once we realized that it would not be a trivial thing to implement, that was left out. Support for parameterized questions is however now being developed in the OtaStat project as an independent effort. (It should be noted that using a markup language approach to authoring the quizzes does make it a possibility to generate hundreds of similar questions with a simple Perl script or Microsoft Office macro. The current system could then draw one question from this very large pool, in effect emulating the parameterized questions feature. We have not done anything like this, but the possibility exists and this once again shows the power of a markup language system over a GUI and database approach.)

While the static OSML documents are just normal files on a web-server, an interactive component like a quiz needs some kind of server-side programming. This was a departure from the simplistic design of the OSML system, which after the pre-processing stage doesn't have any non-standard requirements on the web-server. The OtaStat Quiz system makes use of the pre-processing tools from OSML to produce an intermediary "almost HTML" version of the quiz. In the intermediary file all formulas are already converted to images and CSS styles and other general layout features have been dealt with. However, the XML elements concerning the variation features are still present.

The final HTML page is then produced with a Java servlet from this intermediary file each time a student opens a quiz in his browser. This servlet applies the variations as defined in the XML elements and the result is a normal HTML document, with no special-purpose elements left. The same servlet is also used to check the students answer and give feedback. All in all this solution is still the simplest possible - as much as possible is done in the pre-processing stage and one servlet is all that is needed on the web-server.

Checking the answers turned out to be more complicated than one would imagine. In theory, checking multiple-choice questions is easy, but the task is complicated by the fact that there are different kinds of multiple-choice questions. Multiple-choice / single-answer is the simplest, the student is to pick one choice and it is either correct or false. But multiple-choice / multiple-answer questions are more complicated. Is the answer correct if the student has checked only some correct choices or do we require the student to find all correct choices?

Naturally we wanted the system to allow all different scenarios and this resulted in extra branches in our Java code.

Checking open questions has its own challenges. While it is simple for a computer to match an answer with pre-defined choices in a multiple-choice question (once the issues we just mentioned are tackled), an open answer is a vague creature for a computer to deal with. The main problem is with knowing the data type. As an example "05054321" is equal to "5054321" if we are talking about numbers (integers) but if the answer is a phone number, leaving out the leading zero is clearly a mistake. So in order to make open questions work, one also has to define the data type the answer is to be checked in. This is only an issue for the author and not a burden for the student, which usually will do the right thing naturally. If needed it is possible for the author to specify multiple correct answers that might also be of different types, for instance "1" and "1.0" and even "one". Supported data types are integer, floating point, text and regexp (short for "regular expression"). The text type requires the answer to match exactly, while the regexp type gives the possibility to provide a pattern to which the answer will be compared. The answer can be defined as "equal to", "not equal to" or "less than" and "greater than".

Before the checked answers and the result of the quiz can be shown to the student there is one final obstacle that is a side-effect of the variation feature. Since the questions and choices are randomized, each student has received an individual version of the quiz. Therefore the servlet has to store all individual variations of a quiz that it creates, and when it receives an answer it has to find that correct set of questions to check against.

Both the OSML and the OtaStat Quiz systems are currently going through some updates, based on experiences and feedback we received from the pilots. The intention is to release all source codes on the project website http://www.otastat.hut.fi/ once all updates are finished. The quizzes that were developed for the first pilot can be found at http://www.otastat.hut.fi/osqz/quizzes/.

5 Results and experiences

5.1 The pilot course

We ran a pilot with our OtaStat Quiz system during the spring 2004. The course in question was a basic course on probability and statistics, "Mat-2.091 Sovellettu todennäköisyyslasku". Almost all students at Helsinki University of Technology attend one of two such courses. Out of the two, this course is slightly easier and about 2/3 therefore attend it. The course was organized both as a traditional lecture-based course and an online course which both ran simultaneously side by side. The online course was also offered as a summer course.

5.1.1 The lecture-based course

The lecture-based course contained 4 hours of lectures and 2-3 hours of exercise labs each week, altogether 12 weeks. There were two alternative course books: Milton & Arnold: "Introduction to Probability and Statistics" (Milton & Arnold, 2002) and Pertti Laininen: "Todennäköisyys ja sen tilastollinen soveltaminen" (Laininen, 1998). Also of note is that the lecture slides, which were available to students from the course website, were very comprehensive and of high quality, often consisting of more than 100 slides per lecture, in effect being a third alternative to the books. Finally there was a compendium of formulas and tables, which the students were also allowed to use in exams.

The students in the lecture-based course were encouraged to also use the e-learning materials provided at the course website. In addition to the quizzes there were Java-applets that interactively demonstrate many topics of the course. Both lecture slides and exercises were also available from the course website.

The course contained two exams, one midterm and one final. There were two opportunities for taking both exams, the second opportunity being two weeks after the first. One reason for this arrangement was because the course is so big and contains students from so many different university departments that there would almost always be some students with conflicting exams, if only one date was given. For students who wished to do so, it was however possible to attend both exams and it that case the better score would be used to calculate the final course grade. About 15 % of the students used the option to take the same exam twice and 85 % of those performed better the second time.

In order to motivate students to do the weekly exercises, every week there were some exercises to be returned to the lab assistant. By returning enough exercises the student could collect extra points that positively affected his grade. These bonus points would at a maximum correspond to one exam problem. The weekly exercises were however completely voluntary and it was possible to both pass the course or get the best grade without returning a single exercise. The students could also receive two additional bonus points by filling in two online feedback forms. As we previously discussed, it was decided not to give any bonus points for doing the quizzes.

It is also possible to pass the course by taking only one exam at given dates during the year. Most students however at least try the two exams given when the course is lectured, especially since they cannot benefit from the bonus points on a standalone exam. This is an intentional effect of the bonus points system, to get the students to study the topics more regularly than just one night before one exam. This study only includes results from students that took the midterm exams, those that have later passed the course by taking the stand-alone exam are considered to have dropped out.

During the spring 2004 there were 469 that attended the course and 403 that made it all the way to the second exam.

5.1.2 The online course

The online course ran for the same 12 weeks than the lecture-based course. The online students took the same exams with the lecture-based course. Other than the exams there were no activities that would have required them to be on campus at any time. The central meeting point for the online course was a web page that announced the program for each week.

The online students were supposed to read either one of the books or the lecture slides. Each weeks topic was announced on the web page. Links to the relevant quizzes and Java-applets for each topic were also part of each weeks program.

They also had weekly exercises both to be returned and to be calculated individually and they too received bonus points for the returned exercises. The exercises could be returned either by email, fax, traditional mail and even by hand to the teachers desk. When using email any file format was accepted. This flexible policy was intended to make it as easy as possible for students to return their exercises and turned out to work well.

An important part of the online course was a discussion forum on the school's internal Usenet server. In addition to the 8 bonus points for exercises and feedback forms, the online students were awarded 0 to 3 bonus points for active participation in the online discussion.

As the online course was a pilot in many respects, only 20 students were allowed to enroll. The selection was on a first come, first serve basis. Out of the 20 that enrolled, only 17 actually participated in the course (took at least one exam or returned at least one exercise), and out of those only 11 made it all the way to the final exam. While it is not a topic of this thesis, we will just briefly note here that higher drop-out rates seem to be a common problem of online courses. A typical scenario is a student with children and/or a full-time job, who opts for the online course as an alternative because he is too busy to attend the lectures. This seems to be an Internet-era version of - and just as effective as - putting the course book under your pillow while you sleep in the hope that you might learn something even though you didn't have time to read the book during the day. Emails from students confirm at least a couple such cases in this course too. (Too busy at work, children got ill.) See (McClaren, 2004) for more on this topic.

5.1.3 The summer course

Since the pilot was deemed successful, the online course was also given as a summer course in June-July 2004. The course was basically identical to the one in the spring, but the program had been squeezed into only seven weeks, with two exercise rounds each week.

The summer course did not have a lecture-based group alongside and also some of the statistics we gathered during the spring were not collected during the summer. For this reason the summer course is not analyzed in as much detail as the spring courses, but we will refer to it and compare it to the spring courses where appropriate.

This offering turned out to be hugely popular, with 186 students enrolling but only 75 making it all the way to the second exam.

5.2 Data collected

5.2.1 Quiz usage

During the pilot, we logged usage of the quizzes. We logged both how each question was answered as well as the overall score attained. We also added an extra field to the quizzes, where students could voluntarily give their student number to identify themselves. This way we would be able to track how each person used the quizzes and also compare it to other personal variables, like the students performance in the exams. During the pilot there were 675 times a a quiz was answered and 370 of those hits contained a student number while 305 are anonymous.

The quizzes were authored during the course. For a couple of weeks the teaching staff fell behind schedule, such that quizzes for a given module only became available after the deadline for that modules exercises had passed. Even if this could have impacted our results, according to student feedback it did not. Those students who wanted to use the quizzes used them when they had become available and those who didn't use the quizzes didn't miss them.

5.2.2 Exam performance

We will use results from the exams to measure student performance, i.e. learning. Obviously, all bonus points will be excluded and only the plain exam results are used.

5.2.3 Questionnaires

During the pilots the students got to fill in a number of questionnaires. After both exams there was a standard feedback questionnaire that is used every year. The online students got to fill in their own version of this questionnaire. These questionnaires were anonymous.

Additionally we gathered background information on students, with a questionnaire where they would identify themselves. Data collected in this questionnaire falls into three categories:

1. **Basic demographic data:** student gender, department, class.

2. **Interest in the course:** how interested is the student in the topics, how important is the topic to the students future studies or work career. These aspects are not analyzed in this thesis.

3. **Grades from earlier courses:** Grades from introductory math courses and math courses in high school as well as average grade from high school and past university courses.

Since the students were free to use or not use the quizzes, this is not a controlled experiment with random assignment. We have tried to use data from this questionnaire to compensate for this. Now we can at least compare earlier course grades or student motivation with quiz usage as explaining factors for student performance.

5.2.4 Other data

Data on students participation in the weekly exercise rounds is available for all courses. We also have available data on how actively online students participated in the discussion forum, but this data is too scarce to be used.

5.3 Research hypotheses

5.3.1 Questions regarding quizzes

The first question we were interested in was, how would the students want to use the quizzes? In order to explore this, we tried to influence the students as little as possible with regard to the quizzes. No deadlines or bonuses were imposed, but the quizzes were completely voluntary and available at all times. (Except for the fact that the quizzes were not available before having been authored.)

Being a mathematical subject, with exercises as an important, perhaps even the most important part of the course, one possible outcome we expected was that the quizzes would not be used much at all, but students would focus on doing the exercises. Especially so, since doing exercises was rewarded with bonus points but the quizzes were not. It should however be noted, that the content of the quizzes covered a slightly different angle than the exercises. The questions in the quizzes were about more basic issues, like what is the formula or definition for something, or very elementary exercises like what is the probability to get a spade when drawing a card. (See Appendix 1 for some example questions.)

We had undocumented anecdotal evidence from the Chemistry department at HUT, which suggested that quizzes are a popular tool when students study before an exam. This led us to expect a surge in quiz usage for the days before exams.

Even though not all of our research questions will be answered with a statistical test, for clarity we have listed them all using the H# convention:

H1: The students will not use quizzes very much (not as much as the exercises).

H2: Students will use the quizzes when studying for an exam.

The obvious question we are interested in is whether using the quizzes have had a positive impact on the students' learning. We measure this with students' performance in the exams.

H3: Students using quizzes will perform better on the exam than those who don't.

The obvious flaw in our setting is that when quizzes are voluntary, they will be used by the most diligent students, who will get a good grade, as they always do. We will try to remedy this problem by also taking into account students' grades from earlier math courses and high school. We can then use *partial correlation* to try to control for this behavior and try to find out whether a student who used quizzes actually performed better than his own average level.

Partial correlation is used when there is reason to believe that the correlation between random variables X_1 and X_2 is influenced by one or more other random variables X_3 ... X_n. The partial correlation

$$r_{1,2\cdot 3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{(1-r_{13})(1-r_{23})}}$$

is the correlation between X_1 and X_2 with the effect from X_3 eliminated. (r_{ij} is the correlation between X_i and X_j.)

The more important question therefore is:

H4: Students using quizzes will ***not*** *perform better than others, once other explaining factors, such as earlier grades in math courses, are taken into account.*

And finally we are interested in addressing the issue raised in chapter 3.4 by Kashy et al. (2000) that "allowing multiple tries on assigned problems is highly motivating; most students strive to get all the work done correctly." Can this kind of behavior be found in our data? We hypothesize:

H5: Students will take a quiz multiple times, if they don't succeed on their first attempt.

5.3.2 Other hypotheses

Since we have the data available, we will also make the following hypotheses that are of interest in the more general topic of e-learning, even if they don't pertain to our main issue of quizzes.

As is customary, we might just as well compare the online group to the lecture-based group, formulating the following null hypothesis.

H6: Students in the online course will perform equally well to students in the lecture group.

Finally we will assess the impact of exercises in the same way we do for quizzes:

H7: Students doing exercises will perform better on the exam than those who don't.

H8: Students doing exercises will ***not*** *perform better than others, once other explaining factors, such as earlier grades in math courses, are taken into account.*

5.4 Results

5.4.1 Quiz usage (H1-H2)

Table 5.1 and Illustration 5.1 give a nice overview of how the quizzes were used. The table has all the quizzes of the course in the left column and dates on the first row. Each cell contains the number of times somebody took a specific quiz on a given day. The bottom row and rightmost column contain totals for each column and row.

Since the quizzes are listed in the same order they would appear in the course, the progress of the course is clearly seen. Students start by taking the first quizzes and as time goes by they continue downward in the table. The 4 exam dates have been marked with vertical black lines and the horizontal line divides the quizzes belonging to topics covered in the first exam (above) from quizzes belonging to the second exam (below). Both in the table and the illustration it is apparent how activity grows in the days before an exam. Therefore we can already establish that H2 is clearly supported by our data. Especially of note is that the exam dates themselves have a high hit rate. This is both due to the fact that students have studied past midnight the evening before, and that the exams were in the evening, such that some students have used the morning for studying.

Table 5.1: Quiz usage / day (2 pages).

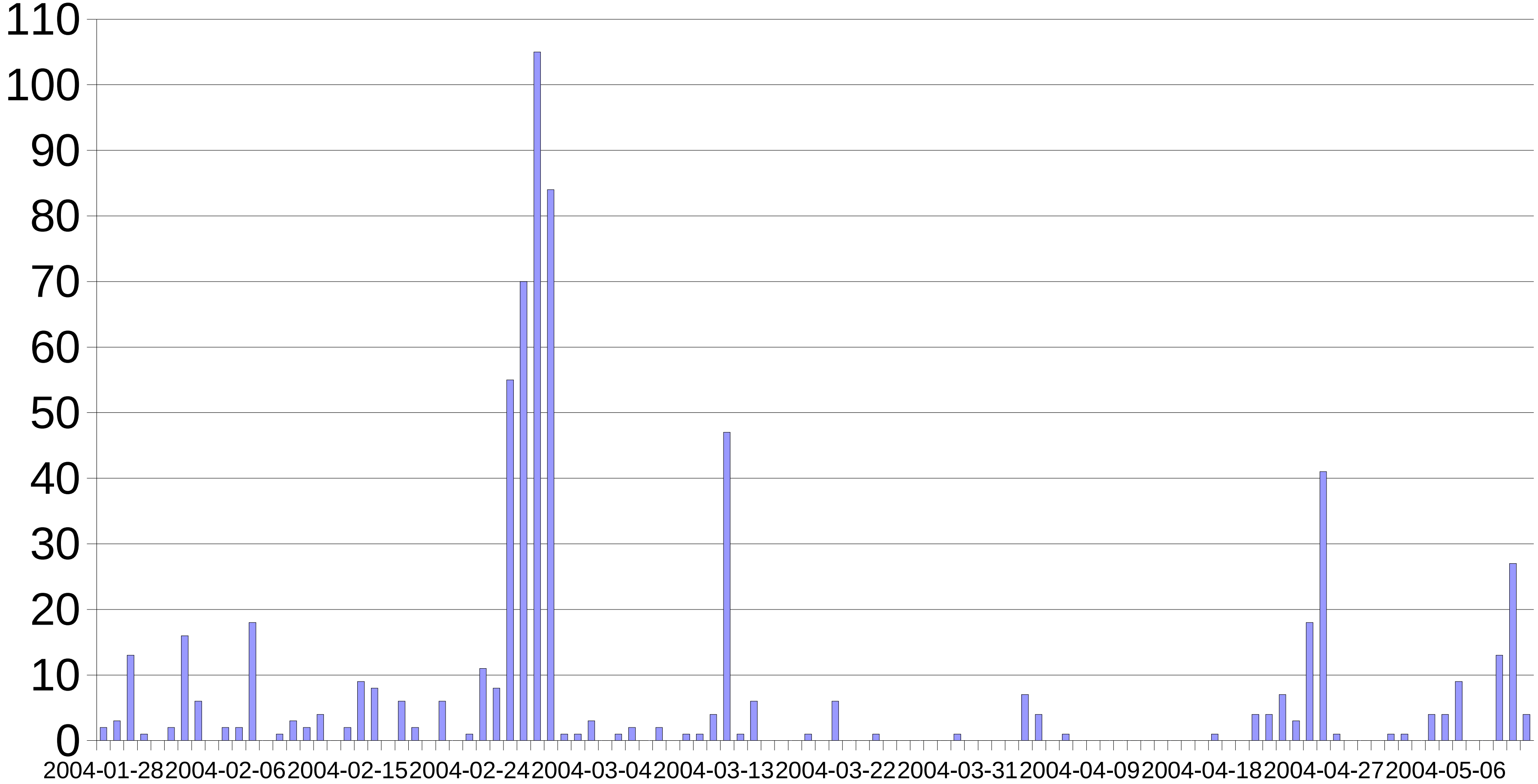
	2004-01-28	2004-01-29	2004-01-30	2004-01-31	2004-02-01	2004-02-02	2004-02-03	2004-02-04	2004-02-05	2004-02-06	2004-02-07	2004-02-08	2004-02-09	2004-02-10	2004-02-11	2004-02-12	2004-02-13	2004-02-14	2004-02-15	2004-02-16	2004-02-17	2004-02-18	2004-02-19	2004-02-20	2004-02-21	2004-02-22	2004-02-23	2004-02-24	2004-02-25	2004-02-26	2004-02-27	2004-02-28	2004-02-29	2004-03-01	2004-03-02	2004-03-03	2004-03-04	2004-03-05	2004-03-06	2004-03-07	2004-03-08	2004-03-09	2004-03-10	2004-03-11	2004-03-12	2004-03-13	2004-03-14	2004-03-15
/osqz/quizzes/ttvk_joukko-oppi1.quiz	2		3				1	1			1	1				1				5	2		1						2	1	5	7	5	5										1	1	1		
/osqz/quizzes/ttvk_joukko-oppi2.quiz			1					2			1	1								1									2		3	4	3															
/osqz/quizzes/ttvk_joukko-oppi3.quiz			1									1					1						1								2	1	1															
/osqz/quizzes/ttvk_tn_peruskasitteet.quiz			3				1	1				1					1			1	2		2	1					4		4	6	11	5													4	
/osqz/quizzes/ttvk_tn_peruslaskusaannot1.quiz			2				1	1				1					1			1	2		1	1					1		3	6	10	5						1							4	
/osqz/quizzes/ttvk_tn_peruslaskusaannot2.quiz			1									1					1			1	1								1		3	4	5	3													4	
/osqz/quizzes/ttvk_klassinen_tn.quiz		3	1				11	1		2		1									1					1				2	20	3	11	10								1					10	
/osqz/quizzes/ttvk_kombinatoriikka.quiz				1		2	1					1																		1	5	3	4	4		1											1	
/osqz/quizzes/ttvk_kombinatoriikka2.quiz			1				1					1																		1	6	4	5	2						1							1	
/osqz/quizzes/ttvk_kombinatoriikka3.quiz												1																		1	2	4	2	2														
/osqz/quizzes/ttvk_bayes.quiz												3																		1		2	6	2								1					1	
/osqz/quizzes/ttvk_verkot.quiz												2		1	3											4					1	10	10	5												3	1	
/osqz/quizzes/ttvk_satunnaismuuttujista.quiz												3				1										1			1			4	8	8													3	
/osqz/quizzes/ttvk_tunnusluvuista.quiz																							1								1	2	6	2													1	
/osqz/quizzes/ttvk_tunnusluvuista2.quiz																																1		1														
/osqz/quizzes/ttvk_tunnusluvuista3.quiz																																1		1														
/osqz/quizzes/ttvk_diskreetit.quiz																			2									1				2	4	7			3										1	
/osqz/quizzes/ttvk_jatkuvat.quiz																														1			5	18	1												13	
/osqz/quizzes/ttvk_jatkuvat2.quiz																																5	5	1													2	
/osqz/quizzes/ttvk_normaalijakauma.quiz																																1	4	3			1										1	1
/osqz/quizzes/ttvk_moniulotteiset_diskreetit.quiz																																																
/osqz/quizzes/ttvk_moniulotteiset_diskreetit2.quiz																																																
/osqz/quizzes/ttvk_moniulotteiset_diskreetit3.quiz																																																
/osqz/quizzes/ttvk_kaksiulotteinen_normaalijakauma.quiz																																																
/osqz/quizzes/ttvk_kaksiulotteinen_normaalijakauma2.quiz																																																
/osqz/quizzes/ttvk_kuvaaminen.quiz																																																
/osqz/quizzes/ttvk_kuvaaminen2.quiz																																																
/osqz/quizzes/ttvk_otantajakaumat.quiz																																																
/osqz/quizzes/ttvk_otantajakaumat2.quiz																																																
/osqz/quizzes/ttvk_luottamusvalit.quiz																																																
/osqz/quizzes/ttvk_luottamusvalit2.quiz																																																
/osqz/quizzes/ttvk_luottamusvalit3.quiz																																																
/osqz/quizzes/ttvk_luottamusvalit4.quiz																																																
/osqz/quizzes/ttvk_asteikot.quiz																																																
/osqz/quizzes/ttvk_asteikot2.quiz																																																
/osqz/quizzes/ttvk_asteikot3.quiz																																																
/osqz/quizzes/ttvk_asteikot4.quiz																																																
/osqz/quizzes/ttvk_asteikot5.quiz																																																
/osqz/quizzes/ttvk_asteikot6.quiz																																																
/osqz/quizzes/ttvk_asteikot7.quiz																																																
/osqz/quizzes/ttvk_asteikot8.quiz																																																
/osqz/quizzes/ttvk_asteikot9.quiz																																																
/osqz/quizzes/ttvk_asteikot10.quiz																																																
/osqz/quizzes/ttvk_asteikot11.quiz																																																
/osqz/quizzes/ttvk_asteikot12.quiz																																																
/osqz/quizzes/ttvk_asteikot13.quiz																																																
/osqz/quizzes/ttvk_asteikot14.quiz																																																
/osqz/quizzes/ttvk_asteikot15.quiz																																																
/osqz/quizzes/ttvk_asteikot16.quiz																																																
/osqz/quizzes/ttvk_asteikot17.quiz																																																
/osqz/quizzes/ttvk_testaus.quiz																																																
/osqz/quizzes/ttvk_testaus2.quiz																																																
/osqz/quizzes/ttvk_testaus3.quiz																																																
/osqz/quizzes/ttvk_testaus4.quiz																																																
/osqz/quizzes/ttvk_testaus5.quiz																																																
/osqz/quizzes/ttvk_testaus6.quiz																																																
/osqz/quizzes/ttvk_testaus7.quiz																																																
/osqz/quizzes/ttvk_testaus8.quiz																																																
/osqz/quizzes/ttvk_testaus9.quiz																																																
/osqz/quizzes/ttvk_lineaarinen_regressio.quiz																																																
/osqz/quizzes/ttvk_lineaarinen_regressio2.quiz																																																
/osqz/quizzes/ttvk_lineaarinen_regressio3.quiz																																																

Quiz	2004-03-16	2004-03-17	2004-03-18	2004-03-19	2004-03-20	2004-03-21	2004-03-22	2004-03-23	2004-03-24	2004-03-25	2004-03-26	2004-03-27	2004-03-28	2004-03-29	2004-03-30	2004-03-31	2004-04-01	2004-04-02	2004-04-03	2004-04-04	2004-04-05	2004-04-06	2004-04-07	2004-04-08	2004-04-09	2004-04-10	2004-04-11	2004-04-12	2004-04-13	2004-04-14	2004-04-15	2004-04-16	2004-04-17	2004-04-18	2004-04-19	2004-04-20	2004-04-21	2004-04-22	2004-04-23	2004-04-24	2004-04-25	2004-04-26	2004-04-27	2004-04-28	2004-04-29	2004-04-30	2004-05-01	2004-05-02	2004-05-03	2004-05-04	2004-05-05	2004-05-06	2004-05-07	2004-05-08	2004-05-09	2004-05-10	2004-05-11	2004-05-12	Total
/osqz/quizzes/ttvk_joukko-oppi1.quiz																							1																	1											1	3			3	1		56
/osqz/quizzes/ttvk_joukko-oppi2.quiz																																																					1			2			21
/osqz/quizzes/ttvk_joukko-oppi3.quiz																																											1													1			10
/osqz/quizzes/ttvk_tn_peruskasitteet.quiz																																																									5	1	53
/osqz/quizzes/ttvk_tn_peruslaskusaannot1.quiz																																																									2	1	44
/osqz/quizzes/ttvk_tn_peruslaskusaannot2.quiz																																																									1		26
/osqz/quizzes/ttvk_klassinen_tn.quiz																																										2															8		88
/osqz/quizzes/ttvk_kombinatoriikka.quiz																																										1															2		27
/osqz/quizzes/ttvk_kombinatoriikka2.quiz																																										1															1		25
/osqz/quizzes/ttvk_kombinatoriikka3.quiz																																										1															1		14
/osqz/quizzes/ttvk_bayes.quiz																																										1										1					5		23
/osqz/quizzes/ttvk_verkot.quiz																																										4									4					7			55
/osqz/quizzes/ttvk_satunnaismuuttujista.quiz																																										1	1										1					2	34
/osqz/quizzes/ttvk_tunnusluvuista.quiz																																			1									1															15
/osqz/quizzes/ttvk_tunnusluvuista2.quiz																																																	1										3
/osqz/quizzes/ttvk_tunnusluvuista3.quiz																																																												2
/osqz/quizzes/ttvk_diskreetit.quiz					1																	1																					2										1						25
/osqz/quizzes/ttvk_jatkuvat.quiz																																																												38
/osqz/quizzes/ttvk_jatkuvat2.quiz																																																												13
/osqz/quizzes/ttvk_normaalijakauma.quiz																																																					1						12
/osqz/quizzes/ttvk_moniulotteiset_diskreetit.quiz	1																																							5	1	1																	8
/osqz/quizzes/ttvk_moniulotteiset_diskreetit2.quiz	1																																							2	1	1																	5
/osqz/quizzes/ttvk_moniulotteiset_diskreetit3.quiz	1																																																										1
/osqz/quizzes/ttvk_kaksiulotteinen_normaalijakauma.quiz	2																																									1																	3
/osqz/quizzes/ttvk_kaksiulotteinen_normaalijakauma2.quiz	1																																																										1
/osqz/quizzes/ttvk_kuvaaminen.quiz							4			1						1																						1				2	1									2	2						14
/osqz/quizzes/ttvk_kuvaaminen2.quiz							2																															1					1																4
/osqz/quizzes/ttvk_otantajakaumat.quiz																					1																		1				1	1																4
/osqz/quizzes/ttvk_otantajakaumat2.quiz																					4																		1					1																6
/osqz/quizzes/ttvk_luottamusvalit.quiz																																							1				2																3
/osqz/quizzes/ttvk_luottamusvalit2.quiz																																							1				1																2
/osqz/quizzes/ttvk_luottamusvalit3.quiz																																							1																				1
/osqz/quizzes/ttvk_luottamusvalit4.quiz																																							1									1												2
/osqz/quizzes/ttvk_asteikot.quiz																					1	1																				1	1														1		5
/osqz/quizzes/ttvk_asteikot2.quiz																					1	1																																						2
/osqz/quizzes/ttvk_asteikot3.quiz																						1																					1																2
/osqz/quizzes/ttvk_asteikot4.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot5.quiz																																											2																2
/osqz/quizzes/ttvk_asteikot6.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot7.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot8.quiz																																																												0
/osqz/quizzes/ttvk_asteikot9.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot10.quiz																																											2																2
/osqz/quizzes/ttvk_asteikot11.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot12.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot13.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot14.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot15.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot16.quiz																																											1																1
/osqz/quizzes/ttvk_asteikot17.quiz																																											2																2
/osqz/quizzes/ttvk_testaus.quiz																																											1																1
/osqz/quizzes/ttvk_testaus2.quiz																																											2																2
/osqz/quizzes/ttvk_testaus3.quiz																																											1																1
/osqz/quizzes/ttvk_testaus4.quiz																																											1																1
/osqz/quizzes/ttvk_testaus5.quiz																																											1																1
/osqz/quizzes/ttvk_testaus6.quiz																																											1																1
/osqz/quizzes/ttvk_testaus7.quiz																																											1																1
/osqz/quizzes/ttvk_testaus8.quiz																																											1																1
/osqz/quizzes/ttvk_testaus9.quiz																																											2																2
/osqz/quizzes/ttvk_lineaarinen_regressio.quiz																																																												0
/osqz/quizzes/ttvk_lineaarinen_regressio2.quiz																																											1																1
/osqz/quizzes/ttvk_lineaarinen_regressio3.quiz																																																												0

Quizzes (left column) are in chronological order top down. Vertical lines mark the exam days and the horizontal line the divide between topics belonging to exam 1 or exam 2.

Illustration 5.1: Quiz usage (total hits) / day. The spikes before exam days are obvious.

.



On the first page of the table the bottom half is completely empty, because the latter quizzes were not available on those dates. The second page covers the dates of the second half of the course and most activity there is on the quizzes belonging to that part of the course, but four times a student has chosen to check out a quiz from the first part of the course. Whether this is purely a random instance of curiosity, or because the student wanted to remind himself about something in those quizzes is unknown. We should mention that on May 12th along with the second instance of the final midterm exam there was also a possibility to take a standalone exam covering the complete course. This explains why in the last two weeks there is renewed interest in the first quizzes. In fact, it seems that students studying for the standalone exam have been more interested in the quizzes than those studying for the final midterm that only covers the latter half of the course topics. Especially since there were only 40 students taking the standalone exam and 115 taking the latter final midterm.

It is also apparent in both the table and illustration that students were more active in the beginning of the course. Also the students studying for the standalone exam were more interested in the first quizzes. This is a known feature of this course and can also be observed in how students do the exercise rounds. The first half of the course covers probability and the second half statistics. Since the first half is more or less familiar to students from high school, it is generally held to be easier and many students focus on those topics to pass the course. Another factor is the obvious fact that students always become more busy towards the end of a semester.

In the following 3 tables we have broken down quiz usage data between students from the online course and students from the lecture-based course. For comparison we have only included the number of students who did exercises, which is relevant to hypothesis H1. Slightly less than half of the hits were anonymous, that is, the user did not give his student number when submitting the quiz answers. Only six hits belong to students who identified themselves but are not students of either of the two groups. Four out of these six hits seem to be from one student who has been studying for the first midterm exam, but then for whatever reason have not participated in the exam.

Because there are so few hits in the "Others" category, and also based on analysis of a table similar to Table 5.1 but with only anonymous hits included, we have strong reason to believe that almost all anonymous hits are also from students of these courses, not some random surfers of the web. The only exception are the hits from the last two weeks, that are mostly from unidentified students studying for the standalone exam. It is possible that those were not students in the course, but rather flunked students from earlier semesters or students who are trying to pass the exam by going straight to an exam.[15] We believe that most of the anonymous hits should belong to the lecture-based group. This is based on reasoning that the online students knew that quiz usage would be studied and reported on in this thesis, and the author of the thesis was their teacher with whom they had daily contact, therefore they should have been more motivated to identify themselves. Of course, simply using pure a priori probabilities we could already assign 96 % of the anonymous hits to the lecture based group.

Table 5.2: Quiz usage in the online group versus the lecture based group.

	Online course	*Lecture-based*	*Anonymous*	*Others*	*Total*
Total hits[1]	73	291	305	6	675
Students (% of course)	(41 %) 7	(9.2 %) 42	-	3	52
Students with >1 hits[1]	(35 %) 6[2]	(6.8 %) 32	-	1	39
Students who did exercises[3]	(88 %) 15	(77 %) 362	-	-	377

1) With hits we mean the number of times a quiz was answered. (Not the number of times it was viewed and possibly left unanswered).
2) From student feedback we know that only 2 of the online students used the quizzes regularly as a routine part of their studying. So based on these statistics we see that an additional 4 students used the quizzes more than once, even if only occasionally.
3) 1 online student and about 22 lecture-based students returned only one exercise, others returned more.

From this table we can already see support for hypothesis H1, in the sense that exercises are clearly more popular than the quizzes. It seems that the lecture-based students used the quizzes much less than the online students. But we should remember that most of the anonymous hits probably belong in the lecture-based column and therefore the actual percentages might be as high as 20 % and 14 % (corresponding to 9.2 % and 6.8 % respectively). But those are still less than the percentages in the online group. Part of this is explained by the fact that the quizzes were all directly linked to from the home page of the online group, nicely grouped under each topic they belonged to. In contrast, the home page for the lecture based course only contained a link to "additional material" where all of the quizzes were listed (in addition to the link on the homepage the quizzes were also promoted by the

15 Also note that the identification string that was used by the teacher when testing quizzes has been removed from all data that is used in this thesis. Those hits are therefore not a factor in any of the categories.

lecturer and lab assistants). This difference in how the quizzes were promoted is possibly a good explanation to the difference. Finally, if we believe that the percentages for the online course truly are higher than the lecture-based course, we could deduce that students who don't attend lectures or exercise labs are more interested in interactive material like quizzes. This conclusion is also supported by our earlier observation from Table 5.1 that anonymous students preparing for the standalone exam used the quizzes more than the students that had attended the course during the semester.

In the second table we have removed data from the most active days before any of the midterm exams. Which days to remove was decided by looking at the highest bars in Illustration 5.1.The last two weeks were completely removed, because the course was by then finished in the sense that no lectures or exercise rounds took place.

Table 5.3: Quiz usage data with days before exams removed.

	Online course	*Lecture-based*	*Anonymous*	*Others*	*Total*
Total hits[1]	30	22	99	2	153
Students (% of course)	(35 %) 6	(2.1 %) 10	-	2	18
Students with >1 hits[1]	(29 %) 5	(1.5 %) 6	-	0	11

1) With hits we mean the number of times a quiz was answered. (Not the number of times it was viewed and possibly left unanswered).

The third table contains the percentages that are derived when dividing a number in Table 5.3 with the corresponding total in Table 5.2. Those percentages then represent the amount of activity during the course that was not in direct preparation for an exam.

Table 5.4: Percentage of quiz usage ***not occurring*** *before exam days.*

	Online course	*Lecture-based*	*Anonymous*	*Others*	*Total*
Total hits	41 %	7.6 %	32 %	33 %	27 %
Students	86 %	24 %	-	67 %	35 %
Students with >1 hits	83 %	19 %	-	0 %	28 %

These tables confirm what was evident in Table 5.1 and Illustration 5.1 that a majority of quiz usage is in preparation for one of 5 exams (4 midterms and the one standalone exam that clearly has to be included as a factor in our analysis). A new observation is that this difference is much bigger in the lecture-based course than the online course. This is another piece of evidence that supports our belief that online students are more interested than lecture-based students in using quizzes as part of their studying.

Based on this analysis, we can now answer the first two hypotheses.

H1: The students will not use quizzes very much (not as much as the exercises).

True. But especially in the online group one third of the students did use the quizzes (more than once), which is not a bad result since the quizzes were completely voluntary.

H2: Students will use the quizzes when studying for an exam.

True. This result was very clear.

An important additional observation was, that the online students and students studying for the standalone exam seemed to be more interested in using the quizzes than the students in the lecture-based course.

5.4.2 Correlation with student performance (H3 - H4)

For measuring the correlation, we included all students that had at least one point from the second exam. This in effect excludes drop outs. From the quiz usage data, obviously only those hits could be included where the student had identified himself with a student number. Out of 401 students thus selected, there were 41 who had used the quizzes one or more times.

The Pearson correlation between amount of quizzes answered and total exam points was

$$r_{quizzes,\,exam} = 0.14$$

No significant correlation therefore between quiz usage and exam performance, but we'll note as a positive aspect that at least the correlation is not negative.

Since the number of students who had not used the quizzes at all was very high, we also calculated the same correlation for only those students who had answered at least one quiz. This correlation was even smaller:

$$r_{quizzes\,,\,exam}=0.057$$

Therefore the answer to hypothesis H3 is

H3: Students using quizzes will perform better on the exam than those who don't.

Not supported. But they don't perform any worse either.

203 students answered the questionnaire in which data on their earlier high school and university grades was collected. We asked for students average and math grade from their high school diploma, as well as their current average grade from university courses and their grades for the basic university calculus courses 1 and 2. Out of these 5 variables, the average on all university courses had the clearly highest correlation with exam performance in this course.

Table 5.5: Correlations between exam and some earlier school performances.

	high school average	*high school math*	*university average*	*university calculus 1*	*university calculus 2*
Prob&statistics exam performance	0.18	0.22	**0.51**	0.30	0.27

Correlations between student performance on the exam and some of their earlier performances. Average grade from all university courses is clearly the best predictor.

The average on all university courses was therefore chosen to be the control variable in the equation for partial correlation. There were 27 students for which all three of these statistics (quiz usage >0, exam performance, university average) were available. Among these the correlation is

$$r_{quizzes\,,\,exam}=0.088$$

And the partial correlation, with university average as the controlling variable, is 24 % lower.

$r_{quizzes, exam \cdot average} = 0.067$

H4: Students using quizzes will ***not*** *perform better than others, once other explaining factors, such as earlier grades in math courses, are taken into account.*

True. But this was already true in H3 before introducing a control variable and the partial correlation is only slightly lower.

5.4.3 Taking a quiz multiple times (H5)

We collected all hits were a student had answered the same quiz more than once. This table was small enough to be analyzed by hand. There were 20 such students and the table contains all in all 46 such student-quiz pairs.

Table 5.6: Quizzes taken more than once by the same student.

	Student id	Quiz	Date								Date 2				Date 3			
1	xxxxx1	tn_peruslaskusaannot2.quiz	2004-02-17	0,69							2004-02-28	0,96					+	
2	xxxxx1	tn_peruslaskusaannot1.quiz	2004-02-17	0,91							2004-02-28	0,91					+	
3	xxxxx1	verkot.quiz	2004-02-10	0,88							2004-02-11	**0,13**			2004-02-28	1	+	
4	xxxxx1	jatkuvat.quiz	2004-02-29	0,71	**0,14**												+	
5	xxxxx1	jatkuvat2.quiz	2004-02-29	0,67	1												+	
6	xxxxx1	tn_peruskasitteet.quiz	2004-02-17	1							2004-02-28	1					+	
7	xxxxx1	klassinen_tn.quiz	2004-02-28	0,6							2004-02-29	**0**	0	0,2			o	
8	xxxxx2	joukko-oppi3.quiz	2004-02-27	0,94	1												+	
9	xxxxx2	verkot.quiz	2004-04-26	0,75	0,75	1	1										+	
10	xxxxx2	kombinatoriikka.quiz	2004-02-27	0,8							2004-04-26	0,93					+	
11	xxxxx2	kombinatoriikka2.quiz	2004-02-27	0,93							2004-04-26	0,93					+	
12	xxxxx2	klassinen_tn.quiz	2004-02-27	0,4	**0,2**						2004-04-26	0,4	0,6				o	
13	xxxxx3	tn_peruslaskusaannot2.quiz	2004-02-29	0,92							2004-03-01	0,96					+	
14	xxxxx3	tn_peruslaskusaannot1.quiz	2004-02-29	0,82							2004-03-01	1					+	
15	xxxxx3	satunnaismuuttujista.quiz	2004-03-01	0,03	**0**												-	
16	xxxxx3	bayes.quiz	2004-02-29	0,32							2004-03-01	0,89					+	
17	xxxxx3	klassinen_tn.quiz	2004-02-29	0,6							2004-03-01	1					+	
18	xxxxx4	tn_peruslaskusaannot2.quiz	2004-03-14	0,19	0,77												o	
19	xxxxx4	tn_peruslaskusaannot1.quiz	2004-03-14	0,64	0,73												o	
20	xxxxx4	tn_peruskasitteet.quiz	2004-03-14	0,5	0,5												-	
21	xxxxx4	klassinen_tn.quiz	2004-03-14	0,2	0,2	0,2	0,6	**0,4**	0,6	0,8							o	(1
22	xxxxx5	diskreetit.quiz	2004-03-01	0,88	0,88	0,94											+	
23	xxxxx5	verkot.quiz	2004-03-01	0,75	1												+	
24	xxxxx5	jatkuvat.quiz	2004-03-01	0,71	**0**												-	
25	xxxxx5	normaalijakauma.quiz	2004-03-01	1	1												-	
26	xxxxx6	diskreetit.quiz	2004-02-29	0,81	0,88	1											+	
27	xxxxx6	jatkuvat2.quiz	2004-02-29	0,67	1												+	
28	xxxxx6	tn_peruskasitteet.quiz	2004-02-28	0,83	1												+	
29	xxxxx7	tunnusluvuista.quiz	2004-02-29	0,13	0,5												o	
30	xxxxx7	satunnaismuuttujista.quiz	2004-02-29	0,74							2004-05-12	0,11	0,92				+	
31	xxxxx7	klassinen_tn.quiz	2004-02-29	0,8							2004-05-11	0,8					-	
32	xxxxx8	tn_peruskasitteet.quiz	2004-05-11	0,5	0,5	1											+	
33	xxxxx8	klassinen_tn.quiz	2004-05-11	0,2	0,2	0,2	0,2	0,2	0,6	1							+	
34	xxxxx9	jatkuvat.quiz	2004-03-14	0,43	0,43												-	
35	xxxxx9	moniulotteiset_diskreetit.quiz	2004-04-24	0,94	1												+	
36	xxxxxa	klassinen_tn.quiz	2004-03-01	0,6	0,6	0,6	0,8	0,8	1								+	
37	xxxxxb	jatkuvat2.quiz	2004-02-28	0	0,33	0,67	0,67	1									+	
38	xxxxxc	klassinen_tn.quiz	2004-02-03	0	0,8	**0**											o	
39	xxxxxd	klassinen_tn.quiz	2004-01-29	0	0,6	1											+	
40	xxxxxe	tn_peruskasitteet.quiz	2004-02-19	0,5	1												+	
41	xxxxxf	verkot.quiz	2004-03-13	0,63	**0,38**												+	
42	xxxxxg	verkot.quiz	2004-02-11	0,88	1												+	
43	xxxxxh	kaksiulotteinen_normaalijakauma.quiz	2004-03-16	0,9	1												+	
44	xxxxxi	tn_peruskasitteet.quiz	2004-03-01	0,5	1												+	
45	xxxxxj	diskreetit.quiz	2004-02-15	0,56	0,88												o	
46	xxxxxk	tn_peruskasitteet.quiz	2004-02-25	0,5	0,83												o	

46 rows (14 online, 32 lecture)
20 students (6 online, 14 lecture)

Finishes above 90 % (+)	31	67%
Improves but finishes below 90% (o)	9	20%
Identical or the error is a technicality (-)	6	13%

1) The 0,4 on this row is the only occasion where the student actually makes an error on a calculation he previously got right.

Notes:
The longest series are from quizzes with open questions (simple calculations)

From the table we clearly see, that the reason students retake a quiz is to get all or some questions right that they didn't get the first time. The times when the result from a quiz is lower on a consecutive try than the previous are marked with a gray background. For most of these cases we have found good explanations. Sometimes the student has on the second try only answered the questions he got wrong and left the other questions empty. Sometimes the lower number is based on a simple technicality, for instance students were often confused whether to use a period or a comma as a decimal sign. Only on row 21 there is a case were the student seems to have made an error, but he then corrects that and retakes the quiz again.

Therefore the table clearly shows, that these students have strived to get all questions correct. Especially encouraging is that 67 % of them don't stop before they have got at least 90 % of maximum points.

An interesting observation is, that most of the quizzes in this table contain open questions, i.e. simple calculations that are to be answered with a number, and not multiple choice.

There are altogether 122 hits in this table, which is 33 % of all non-anonymous hits. Of all 370 non-anonymous hits, 95 got full points, 25 of which are in this table. So instead of 370 hits there are really only 275 that can be improved. Therefore a more correct percentage is 122 / (275 + 25) = 41 %. Especially when taking into account the following reasoning, this is a very large percentage.

One problem with this experiment was, that all quizzes were designed such that once the student has answered the quiz, the correct answer is shown along with the points. This was perhaps a bad design decision and in a future upgrade the system will be changed to give the student the choice whether he wants to see the correct answers, or only have his own answers checked for correctness. The issue is of importance to our question at hand, because when the student has seen the correct answers after his first try, there isn't much point in trying again. It was therefore surprising to find that this many students did take a quiz multiple times.

H5: Students will take a quiz multiple times, if they don't succeed on their first attempt.

True. 41 % of the times a student did not get full points, he tried to improve.

5.4.4 Other hypotheses (H6 - H8)

Statistics for exam performance for lecture-based versus online students is presented in the table below.

Table 5.7: Exam performance for lecture-based versus online students.

	lecture-based course	*online course*
number of students (n)	387	11
average	33.2	34.09
standard deviation	8.1	8.03

The T-test yields a P-value of 0.72.

H6: Students in the online course will perform equally well to students in the lecture group.

True.

The calculations for the last two hypotheses are analogous to those of H3 and H4. There were 377 students who had returned at least one of the weekly exercises. The correlation between exercise and exam performance was

$$r_{exercises, exam} = 0.45$$

While not a strong correlation, it is clearly significant.

This data is also available for the summer course. There the correlation was 0.31.

H7: Students doing exercises will perform better on the exam than those who don't.

True. With correlation 0.45.

There were 182 students for which all three variables (exercise performance, exam performance, university average) were available. Among those students the previous correlation is again lower

$$r_{exercises, exam} = 0.34$$

This calls for an interesting observation. Since leaving out students that didn't fill in the questionnaire seems to consistently lower the correlation, it seems that whether a student answered our questionnaire might be a good predictor for the students exam performance.

The partial correlation, with university average as the controlling variable, is 41 % lower

$$r_{exercises, exam \cdot average} = 0.20$$

H8: Students doing exercises will ***not*** *perform better than others, once other explaining factors, such as earlier grades in math courses, are taken into account.*

Undecided. Taking into account students average on earlier university courses partially explains an already weak correlation, but not completely.

6 Summary and discussion

6.1 State of e-learning research

Academic research on e-learning is currently very fragmented. Even so much so, that there is disagreement and confusion as to the terminology to be used. An important goal for the academic research on e-learning should therefore be to become more united, and at the very least try to find some commonly agreed upon terminology. For this thesis we chose to use the term "e-learning" because of it's widespread acceptance outside the academia. As part of the introductory discussion we also came to choose a broad and inclusive definition for e-learning, meaning both distance and in-classroom teaching, whenever it is delivered or enhanced with some (non-trivial) electronic means.

It looks like research on e-learning is still rapidly developing and results both in favor and against e-learning can be found. Most probably some of the e-learning systems and courses studied were not good, and therefore produced bad results, whereas others fared better. Many different e-learning platforms exist, and courses produced with them or without any particular platform most probably are of a varying quality. The lack of random assignment is a common problem in most studies in this field.

An interesting direction for future research arises from (Janicki & Steinberg, 2003). They found that a group instructed to independently surf the web for information on a lecture topic performed better - although not significantly - than either of the groups that had used the prepared materials. This result suggests that quality material is available on the Internet and the relevant materials were found by the students, presumably using a good search engine. Such a situation, if validated by future research, could open the door for their vision of a new role of the teacher, who has stepped down from the podium and become more of a coach pointing the students in the right direction.

Most e-learning research to date has focused on the question whether e-learning methods in general are more effective than traditional lecturing methods. The usual approach is to compare two groups of students, both subjected to one of these methods. What is almost completely lacking, and sometimes suggested as a topic for further study (Spencer & Hiltz, 2001; Benbunan-Fich & Hiltz, 2002), are studies who would focus on individual e-learning technologies. This was one of the reasons behind the decision to completely focus on quizzes in this thesis.

There have been some articles written about quizzes from a technological or otherwise general viewpoint (see f.i. Brusilovsky & Miller, 1999). But we were only able to find one study were the impact on quizzes on student performance had been experimentally studied. Janicki & Steinberg (2003) compared two randomly assigned groups, which both were observed studying e-learning material in a computerized classroom, at their own pace. They find support for their hypothesis that "including exercises and miniquizzes in the lesson content of web-based tutorials will have a positive effect on posttest results".

6.2 System design

For various reasons we decided to use a markup language based approach for our system. Authoring quizzes in XML turned out to be very convenient compared to a "web-based GUI & database" approach, which is the prevalent paradigm.

The system supports randomization of the order of questions and also of the choices in a multiple choice question. It is also possible to define a set of similar questions, out of which only one or a few is randomly shown at a time. In a similar fashion it is possible to variate the false choices in a multiple choice question.

The quizzes were a voluntary part of the course and not graded. This allowed for students to use the quizzes most flexibly. It also made it possible to reveal correct answers directly after the student submits the quiz, which is good feedback and an important part of the learning process.

6.3 Results

In our pilot we found that 35 % of online students and 6.8 % of lecture-based students used the quizzes more than once. Only 2 of the online students used the quizzes routinely as a regular part of their study, but there are indications that online students are more interested in quizzes than lecture-based students. Being a mathematical subject the course also contained weekly exercises. The exercises were clearly more popular with 88 % and 77 % of students returning at least one exercise (H1). For the online group more than half and for the lecture-based group almost all quiz usage was in days preceding an exam (H2). There was no significant correlation between quiz usage and exam performance (H3). The partial correlation, with students average from earlier university course grades as a controlling variable, was not significantly different from the plain correlation (H4).

It seems that in a mathematical subject the exercises of a course are clearly the most important ingredient - in fact, both students and lecturers often see them as even more important than the lectures. Another reason for our result is that students were awarded with bonus points for returned exercises, while the quizzes were completely voluntary with no attached bonus system.

We know from student feedback that the quizzes were appreciated by those that used them and they sometimes helped a student to better comprehend a topic at hand. Even though the exercises are clearly the most important part of the course, we argue that the quizzes were a valuable addition to the mix of teaching material. Some students chose to use them, while others did not. Those that did use them performed equally well to those who didn't. Therefore the quizzes did not magically make some students smarter, but they helped cater to some students individual preferences of learning style. We also expect that in a non-mathematical course that lacks actual exercises such as the ones in this course, the use of quizzes can play a greater role.

Another important result was, that about half of the identified hits to the quiz system were related to instances were a student took the same quiz more than once, mostly in order to correct mistakes made on the first attempt. Students voluntarily strive to get either full points or at least perform satisfactorily on the quizzes. This result confirms similar observations made by Kashy et al. (2000) (H5).

Some research questions did not address the use of quizzes. Comparing the online group with the lecture-based group we found no difference in exam performance (H6). There was a correlation of 0.45 between student exercise activity and exam performance (H7). Using partial correlation we found that this is partially explained by students performance in earlier university courses, i.e. the students that do most of the exercises are the "good students" who will always get good grades. But also the partial correlation was still clearly on the positive side and we conclude that to some extent doing exercises will positively affect a students performance on the exam (H8).

6.4 Future work

The finding that quizzes are mostly used in preparation for an exam leads to one obvious conclusion. For this pilot we authored the quizzes to go along with the lectures and exercises, so that there was one or more quizzes available for each topic. In preparing for an exam it would probably benefit the student more, if he could take a quiz covering the entire range of the exam topics and then get feedback on which topics he did well and which should be given more study. In the future the project should focus on authoring a set of such quizzes.

In this implementation of the OtaStat Quiz system, the author could decide whether correct answers should be shown to the student after he has answered the quiz and we used this option in all quizzes. It would however be beneficial if the author could also have a third option: let the student decide. In that case the student would be provided with two buttons: one to submit answers and see which questions are right and wrong, another to both show the results and correct answers. Then the student would not need to see correct answers directly after his first attempt, but could try to correct his mistakes first and only eventually opt to see the correct answers.

An interesting area related to quizzes is the field of parameterized quizzes. There is currently work ongoing to implement that feature as well.

Bibliography

Baugher, D., Varanelli, A. & Weisbord, E. (2003): Student Hits in an Internet-Supported Course: How Can Instructors Use Them and What Do They Mean? Decision Sciences Journal of Innovative Education, vol 1 issue 2, 2003: 159-179.

Benbunan-Fich, R. & Hiltz, S.R. (2002): Correlates of Effectiveness of Learning Networks: The Effects of Course Level, Course Type, and Gender on Outcomes. 35th Annual Hawaii International Conference on System Sciences (HICSS'02)-Volume 1, 2002.

Bridgeman, S., Goodrich, M.T., Kobourov, S.G. & Tamassia, R. (2000): SAIL: a system for generating, archiving, and retrieving specialized assignments using LATEX. Proceedings of the thirty-first SIGCSE technical symposium on Computer science education, 2000: 300-304.

Brooks, D.W. (1997): Web-Teaching, A Guide to Designing Interactive Teaching for the World Wide Web. Plenum Press, New York, 1997.

Brusilovsky, P. & Miller P.L. (1999): Web-based Testing for Distance Education. WebNet 1999: 149-155 [http://www.contrib.andrew.cmu.edu/~plb/papers/WebNet99.html]

Brusilovsky, P. & Miller P.L. (2001): Course Delivery Systems for the Virtual University. In: Tschang F.T. & Senta T.D. (eds): Access to Knowledge: New Information Technologies and the Emergence of the Virtual University. Elsevier Science and International Association of Universities, Amsterdam 2001: 167-206.

Janicki, T. & Steinberg, G. (2003): Evaluation of a Computer-Supported Learning System. Decision Sciences 1(2), 2003: 203-223.

Koskinen, V. (2003): e-Learning and Teaching Mathematical Models of Negotiation Analysis. Master's thesis. Helsinki University of Technology, Department of Engineering Physics and Mathematics, 2003. [http://www.sal.hut.fi/Publications/pdf-files/tkos03.pdf]

Laininen, P. (1998): Todennäköisyys ja sen tilastollinen soveltaminen, Otatieto 1998, numero 586.

Maki, W.S. & Maki, R. H. (2002): Multimedia comprehension skill predicts differential outcomes of web-based and lecture courses. Journal of Experimental Psychology: Applied, 8 (2), 2003: 85-98.

McClaren C.H. (2004): A Comparison of Student Persistence and Performance in Online and Classroom Business Statistics Experiences. Decision Sciences Journal of Innovative Education, vol 2, nr 1, Spring 2004: 1-10.

McCreanor, P.T. (2000): Developing a Web-enhanced Course: a Case Study. 30th ASEE/IEEE Frontiers in Education Conference, 2000. [http://fie.engrng.pitt.edu/fie2000/papers/1514.pdf]

Milton, J.S. & Arnold, J.C. (2002): Introduction to Probability and Statistics : Principles and Applications for Engineering and the Computing Sciences. McGraw-Hill Science/Engineering/Math; 4 edition, 2002.

Pathak, S. & Brusilovsky, P. (2002). Assessing Student Programming Knowledge with Web-based Dynamic Parameterized Quizzes. World Conference on Educational Multimedia, Hypermedia and Telecommunications 2002(1), 1548-1553. [http://www2.sis.pitt.edu/~peterb/papers/EDMED01QUIZ.pdf]

Schmidt, K. (2001): Computer-supported cooperative work and learning. FREREF ICT workshop, July 2001.

Spencer, D. & Hiltz, S. (2001): Studies of ALN: An Empirical Assessment. 34th Annual Hawaii International Conference on System Sciences (HICSS-34)-Volume 1, 2001: 1007.

Wegner, L (2003): Teaching an Old Course New Tricks: A Portable Courseware Language Based on XML. in: "Computer Science in a Perspective - Essays dedicated to Thomas Ottmann" (Klein, Six, Wegner Eds.), Springer LNCS 2598, 2003: 343-355 .

Websites

4Teachers.org: **http://www.4teachers.org/**

ALN WebCenter: **http://www.alnresearch.org**

Blackboard (e-learning platform): **http://www.blackboard.com**

eLearning programme of the European Commission: **http://www.elearningeuropa.info/**

Moodle (e-learning platform): **http://moodle.org**

OpenUSS (e-learning platform): **http://www.openuss.org**

OtaStat: **http://www.otastat.hut.fi/**

QuizStar: **http://quizstar.4teachers.org/**

Quizzes authored with OtaStat Quiz: **http://www.otastat.hut.fi/osqz/quizzes/**

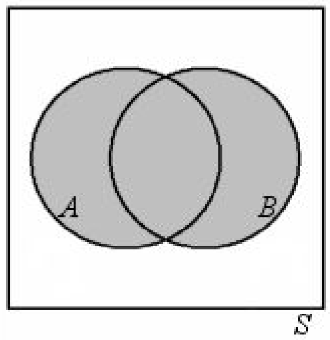
SticiGui: Statistics Tools for Internet and Classroom Instruction with a Graphical User Interface: **http://www.stat.berkeley.edu/users/stark/SticiGui/index.htm**

W3 Consortium: **http://www.w3.org/**

WebCT (e-learning platform): **http://www.webct.com**

Appendix 1: Some example questions from the quizzes

Kumpi seuraavista vaihtoehdoista on tosi *kaikille* joukoille *A* ja *B*?



Valitse tosi vaihtoehto:

Kysymys 1	1 piste

- ○ $x \in A \setminus B \Rightarrow x \in A$ ja $x \notin B$
- ○ $x \in A \setminus B \Rightarrow x \in B$ja $x \notin A$

Kysymys 2	1 piste

- ○ $x \in A \cap B \Rightarrow x \in A$ ja $x \in B$
- ○ $x \in A \cap B \Rightarrow x \in A$tai $x \in B$

○ $Pr(S) = 0$
○ $Pr(S) = 1$

Kysymys 2 1 piste

○ $0 \leq Pr(A) \leq 1$
○ $0 < Pr(A) < 1$

Kysymys 3 1 piste

○ $Pr(\varnothing) = 0$
○ $Pr(\varnothing) = 1$

Oletetaan, että otosavaruuden $S = \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}\}$ alkeistapahtumat ovat symmetrisiä.

Kysymys 4 1 piste

Määrää $Pr(x_3)$

Määrää todennäköisyys, että tapahtuma *A* sattuu 1. kerran 10:ssä toistossa.

- ○ Binomijakauma
- ○ Geometrinen jakauma
- ○ Negatiivinen binomijakauma

Riippumaton satunnaisotos voidaan poimia äärellisestä perusjoukosta S kahdella eri tavalla: takaisinpanolla ja ilman takaisinpanoa.

Haluamme määrätä todennäköisyyden saada otannan tulokseksi osajoukon $A \subset S$.

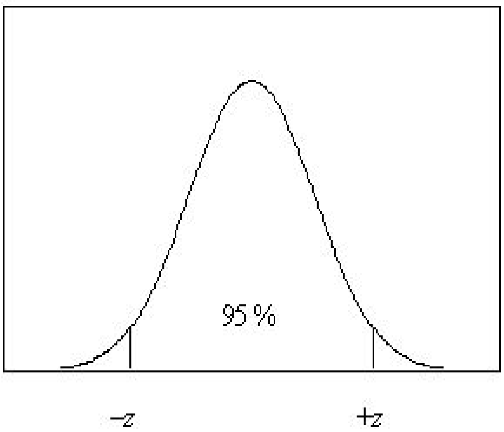
Joukon *A* todennäköisyys määrätään otantamenetelmästä riippuen binomijakaumasta tai hypergeometrisesta jakaumasta.

Kysymys 8	1 piste

Otanta ilman takaisinpanoa.

- ○ Binomijakauma
- ○ Hypergeometrinen jakauma

Allaoleva kuva esittää standardoitua normaalijakaumaa $N(0,1)$.



Pisteet $-z$ ja $+z$ on valittu niin, että niiden väliin jää 95% jakauman todennäköisyysmassasta.

Kysymys 1	1 piste

Mikä on *z*:n arvo?

$$p_i = \frac{E_i}{n}$$

Odotetut frekvenssit E_i on määrätty nollahypoteesin H_0 mukaan.

χ^2-yhteensopivuustestin testisuureen oikea muoto on:

Kysymys 1 1 piste

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$$

- Tosi
- Epätosi

Kysymys 2 1 piste

$$\chi^2 = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{O_i}$$

- Tosi
- Epätosi

Kysymys 4 1 piste

Väli $(\hat{\theta} - a, \hat{\theta} + a)$ peittää estimaattorin $\hat{\theta}$ arvon todennäköisyydellä α.

- Tosi
- Epätosi

Kysymys 5 1 piste

Jos otantaa toistetaan, parametrille θ konstruoiduista luottamusväleistä $(\hat{\theta} - a, \hat{\theta} + a)$ suunnilleen $100 \times (1 - \alpha)\%$ sisältää parametrin θ todellisen arvon.

- Tosi
- Epätosi

Kysymys 6 1 piste

Väli $(\hat{\theta} - a, \hat{\theta} + a)$ peittää estimaattorin $\hat{\theta}$ arvon todennäköisyydellä $1 - \alpha$.

- Tosi
- Epätosi