Evaluating E-Assessment for Exercises That Require Higher-Order Cognitive Skills

Tim A. Majchrzak  
European Research Center for Information Systems (ERCIS), University of Münster  
Email: tina@ercis.de

Claus A. Usener  
European Research Center for Information Systems (ERCIS), University of Münster  
Email: usener@ercis.de

Abstract

Assessments are an integral part of teaching and learning. Since manual assessments are very time-consuming for teaching personnel, e-assessment systems have been created to support them. Today’s systems do not support exercises that require higher-order cognitive skills, though. Such exercises are prevalent in the Information Systems (IS) and Business & Information Systems Engineering (BISE) education. EASy is a system that provides e-assessments with a focus on computer science and mathematics exercises in courses taught in IS and BISE degree programs. While it provides a number of advanced modules, it has not been evaluated in a general manner, yet. Moreover, no empirical studies of e-assessment of complex exercises exist. We present a survey of the perception of e-assessment by students. Besides quantitative results, we describe qualitative findings. Results are discussed and generalized to propose future requirements for broadening the scope of advanced e-assessment systems.

1. Introduction

Assessments are an integral part of teaching and learning. They enable students to estimate their learning progress and teachers to check their teaching effectiveness. In most cases, assessments are also used for grading, i.e. they are used to measure the students’ ability to meet the learning goals. Additionally, they can be used to assess lecture improvements [1]. Assessments are time-consuming and bind resources, though. Therefore, e-assessment systems were introduced. E-assessment is a sub-discipline of e-learning, which describes the support of learning activities with information systems [2], [3]. By providing assessments on computer systems or mobile devices, and by integrating parts of the assessment workflow into an electronic system, teaching personnel can be disburdened of manual effort. Besides, it is aimed at carrying out examinations more precisely and with “greater neutrality and objectivity” [4].

While e-assessment receives increased interest [5], currently available systems are limited to simple exercise types (see Section 3). There is no question that automatic correction of multiple-choice questions relieves a teacher of a boring and error-prone task; multiple-choice is an option [6]. However, manual correction of multiple-choice sheets is much less time-consuming than manual correction of exercises that require complex cognitive abilities or creative-
rable scope and for the extension of the idea to other study subjects. Therefore, we conducted a survey with almost 200 undergraduate students of IS at a German university. We put emphasis on general questions with regard to e-assessment but also tried to get more detailed ideas about specific aspects of it. Besides, we gained qualitative data which asks for detailed analysis. Even though these findings lack the validation that the quantitative ones have, we consider them essential for learning how to build future systems.

Our work makes the following contributions. Firstly, we provide a detailed introduction of the status quo of e-assessment of exercises that require higher-cognitive skills in the CS, IS, and BISE education. It is based on a thorough study of available approaches. Secondly, we present significant results from a study on the perception of and the expectations towards e-assessment. Thirdly, we describe challenges of e-assessment in general and required improvements of EASy gained from the qualitative results. Moreover, we discuss our findings and provide a generalization of the insights. This helps to establish e-assessment with a broader scope and for other fields of studies.

This paper is structured as follows. Section 2 explains the background of our work. In Section 3 related work is discussed. Our research method is described in Section 4. In Section 5 we present detailed results from our study. Section 6 discusses our work and shows limitations. In Section 7 we draw a conclusion and highlight future work.

2. Background

Firstly, we describe the background of e-assessment in general. We then introduce our tool EASy.

2.1. E-Assessment

To understand the term e-assessment, we first put emphasis on assessment in general. In the educational context assessment can fulfill the following purposes [15]: The report function illustrates students the learning success in comparison to other students; from the feedback and diagnostic point of view assessment tells the teacher as well as the student whether learning goals were achieved and gives hints on improving teaching and learning processes. The chance to reach good marks is an extrinsic motivation, too. Some authors consider assessment to have a disciplinary effect by using marks as a disciplinary method.

Assessment can also be distinguished by their type; there are summative, formative, and diagnostic assessments [16]. Most common are summative assessments—also called assessment of learning [17]. They record the overall learning success and therefore occur at the end of a learning process. An example is a final exam. Formative assessments in contrast are usually performed throughout the learning process and have a more conductive character. Examples are weekly exercises. These assessments are also called assessments for learning and should give the learner as well as the teacher feedback and report about the learning process [8]. The term diagnostic assessment describes so called self-assessments, which tell a learner’s learning progress. However, they are not reviewed and corrected by a third person and therefore these voluntary tests will not be marked. Alternatively, the term diagnostic assessment is also used for entrance exams, which have a selective character and are usually performed at the beginning of a learning process [8].

In this paper we focus on formative assessments in higher education to provide formative feedback. “Formative feedback is inseparable from teaching: […] the effectiveness of different teaching methods is directly related to their ability to provide formative feedback.” [18] Especially in universities, where mass lectures are common, usual formative assessment is almost impossible. E-assessment as an electronic support for the assessment process is a promising mean to face this problem. The term e-assessment was internationally published for the first time in 2000; nowadays it is about to replace terms that have similar meanings such as online assessment, computer-assisted assessment, and computer-based assessment [3]. COOKS and JENKINS [20] define the following main benefits: instant feedback, objective marking, scalability, time and place independence, and a wider range of tasks and activities. Particularly the ability to provide immediate feedback is considered very beneficial [21] and usually not possible with manual assessment. Nevertheless, e-assessment is by no means a trivial task and using it for the mere goal of saving money is prone to fail. There “is a constant danger that assessment systems are driven in undesirable ways, where things that are easy to measure are valued more highly than things that are more important to learn (but harder to assess)” [17]. Thus, systems have to be implemented with a clear didactical focus.

With the aim to research new assessment tasks and activities we developed the e-assessment system EASy, which we describe in the following section.

2.2. EASy

The E-Assessment System (EASy) has been designed at the University of Münster to provide a single research system for evaluating new e-assessment tasks and activities in computer science lectures. It focuses on assessing higher-order cognitive skills where students have to apply their analytic, creative, and constructive skills. Such skills cannot be assessed by simple e-assessment techniques such as fill in the gap or multiple choice exercises [14]. The system is designed as a modular Web application and currently provides four different assessment modules: a module for assessing software verification proofs, a mathematical proof

1. Automatic assessment techniques have been discussed earlier [19].
module, a module for programming exercises written in the Java programming language, and a multiple choice module.

The core component of EASY is the EASY platform. Conceptually, it provides services for the operational processes and thus the entire life cycle of examination from creation of exercises and their dissemination via solving them to correction, grading, and feedback [22]. Besides support for operational processes, the core component is responsible for administrative tasks such as user and course management. Technically the core component is a framework, which coordinates the assessment process by managing the data flow through well-defined interfaces and by triggering events. Single modules are responsible for exercise-specific processes. This architecture both allows changes to the core without changing the modules and rapid module development without touching the core.

EASY provides modules for different kinds of exercises. With the module for mathematical proofs, the ability to prove mathematical problems can be assessed. Therefore, EASY provides a special view for mathematical proof exercises, where students receive the mathematical problem that they should proof. During the proofing process, students can choose between different proof strategies and transform equations by successively applying arithmetic and Boolean rules on the problem. A detailed description can be found in [9]. The software verification module works similar to the module for mathematical proofs. It can be used for proving partial correctness of computer programs based on the rules of Hoare Logic [23], [24]. Similar to the mathematical module students prove the statement by applying transformation rules stepwise. USENER at al. [11] describe the module for assessing Hoare Logic in detail. The correctness of the proofing process is supervised in both the mathematical module and the Hoare module by giving feedback directly in case of wrong rule usage. This seems to be a limitation but solving an exercise by trial-and-error is almost impossible due to the multiplicity and variety of rules.

Programming exercises based on the programming language Java can be assessed with the EASY module for programming exercises. According to the exercise description students have to program a solution in their own development environment. After uploading the exercise, students get immediate feedback. First of all, source code is checked for compiling errors and wrong class structure such as missing methods and variables; the lecturer defines the essential class structure beforehand. Furthermore, predefined test cases are performed on uploaded solutions and commented results are directly returned to the student. To raise the students’ awareness of developing in a test-driven manner, we integrated a tool (Muggl [25], [26]) to enable automated test case generation. With this integration we can assess exercises in a back-to-back testing scenario, where students can generate test cases based on their solution and in a second step have to judge the quality of the uploaded code by comparing it with a given generated test case based on an example solution. Utilizing the immediate feedback, students can improve and upload own solutions until a given deadline is reached. Correction and marking is performed after the deadline. Several complex and extensive tests can check the code for errors concerning syntax and semantic automatically. In a second manual phase tutors finally grade the submissions with the help of the tests performed before.

In conjunction with the multiple choice module to assess multiple choice questions automatically, EASY can be used to assess exercises in mathematics as well as IS and CS lectures. For illustration purposes a screenshot of EASY is given in Figure 1 (see next page). It shows a student’s view of the module for mathematical proofs.

3. Related Work

The term e-assessment has been synonymously used with online assessment [27] for over a decade [28]. However, in this context it describes the software support of job assessments [29]. Using the term to describe electronic support for assessments in education [2] is a rather new emergence. Manual assessments have a long history [30].

Simple e-assessment functionality has been built into a number of E-learning platforms, for instance GPAM-WATA [31], ILIAS [32], and Moodle [33]. Commercial systems such as LPLUS [34] merely provide basic exercise types; their focus is legal certainty. Universities expect to use them in mass exams without the danger of students taking legal action against this kind of assessment. Despite the possibility to use exercise types such as multiple-choice questions in formative exams in CS courses [35], most systems are hardly comparable with our approach. In fact, most available systems are unsuitable for adoption in CS [11] unless the disadvantages of simple exercise types [36] are accepted. Nevertheless, a few systems provide complex exercise types for CS-related courses. We check them by exercise focus.

While automation of proofs is an active field of research, there are hardly any tools for assessing students’ ability to conduct proofs—in particular, if correction is meant to be done semi-automatically. For example, the system proposed by WANG is mainly used for didactical reasons [37]. The AIM system allows to specify assignments and to check students’ solutions using the computer-algebra tool Maple [38]. It does not support mathematical proofs, though.

Automated verification tools for non-educational contexts are widely used [39]. The educational tools we identified do not provide assessment capabilities. The Fridge Program Prover (FPP) is a Web-based tool thats main purpose is the visualization of proving strategies [40]. While FPP targets Ada programs, the New Palitz Program Verifier (NPPV) has been written for Pascal programs [41]. Finally, the Java Program Verifier (JPV) provides interactive step-by-step proving. However, none of them supports assessment.
Figure 1. Exemplary screenshot of EASY
With regard to programming exercises, four systems are notable that provide functionality for assessing Java programs. **Praktomat** offers students static and dynamic tests of their programs before they can upload them as exercise solutions [42]. **DUESIE** additionally checks the coding style and functionality implemented by students [43]. Both tools are based on style checking and unit testing tools. Correction resides a manual task of tutors. **ELP** addresses programming novices and provides self-assessment [44]. Uploaded solutions are compared with stored example solutions and immediate feedback is given based on the comparison of structural similarity. **ELP** is therefore limited to well-defined exercises. Finally, the system proposed by **STRIEWE** et al. is not limited to programming assessments but the latter has been described as a use case [45]. The tool seems to support most of the assessment life cycle. Solutions are automatically tested and marked if they are sufficiently correct; manual inspection follows.

Some papers discuss programming and testing in the context of e-assessment. **HANTOLA** describes a system that automatically generates test data and visualizes results by incorporating the **Java PathFinder** [46]. He also discusses assessment strategies. **Marmoset** motivates students to implement programs in a test-driven manner [47]. It however has been designed for didactical aims rather than as an assessment tool. Finally, **ASSYST** is a tool that supports tutors in assessing programming exercises [48]. Additionally, there are approaches that focus on specific details such as UML label matching as part of assessments [49].

Obviously, there are many systems that offer some of the capabilities **EASy** has, target distinct tasks in education, or address didactical issues. Not a single approach holistically covers the whole assessment process, though.

Besides the approaches discussed above, related work can be identified for the distinctive modules of **EASy**. They consider the integration of work from other disciplines—e.g., automated test case generation [25]—into e-assessment systems. Such approaches are highlighted in the corresponding papers [9], [11], [13].

### 4. Methodology

To describe the underlying research design, research on **EASy** and the survey that this paper is based on have to be distinguished. The method chosen for building and evaluating **EASy** is design science [50]. Conducting a survey required the application of quantitative methods to analyze the data. A further particularity of the research is the inclusion of educational goals and didactical concepts [51].

Development of **EASy** is dominated by iterations of building and evaluating [52]. After the general platform had been implemented, we successively specified and designed modules. Requirements were derived in two ways. Firstly, we analyzed applicable lectures. This included studying their curriculum and checking paper-based exercises. Secondly, we talked to students about their perceptions of the exercises. Since a number of modules have already been implemented, the experiences we gained also influence the requirements analysis for new ones. We do not only have an idea of usability requirements demanded by students but we also are able to relate modules. For instance, developing a module for software verification proofs was fostered by the insights gained when developing the module for mathematical proofs (see Section 2.2). At the same time, completely novel modules could be designed—our approach allows proceeding without an existing theoretical base [53].

Each module was evaluated both quantitatively and qualitatively. We use **EASy** in actual courses from the IS bachelor and master degree programs of our institute. This helps us to understand the implications of its creations [54]. As a consequence, the base of knowledge on e-assessment grows, which enables the successful and rapid development of additional modules; we are “learning via making” [55]. This is particularly helpful for a field in which little progress has yet been reported (as discussed in Section 3). Moreover, existing modules are incrementally and iteratively improved. It has to be emphasized that modules are not used only once. Improved versions are employed in recurring courses. We thereby draw from our experience for future work. Continuously applying improvements to the artifacts built (the distinctive **EASy** modules) is typical for design-oriented research [54].

Conducting the survey for this paper and evaluating the modules beforehand required the inclusion of methods of qualitative and quantitative research [56], [57], [58]. However, the applied methods are rather simple. As our samples were relatively small (up to a few hundred students) and we did not compare various populations, no hypotheses testing was conducted. The aim was to understand the general tendencies and differences in perception. To give an example: when evaluating whether e-assessment would be accepted in formative assessments, it does not help to test a hypothesis such as **Students accept the usage of e-assessment in end-of-term exams**. Instead, we were interested in understanding to which degree students would accept such examinations and how shares looked like. Moreover, the qualitative parts of the surveys play an important role. For developing novel systems, it is vital to understand the users’ motivations of using them. In the above example, students that would “not” or “rather not” want to use e-assessment in exams could be asked to explain their skepticism. In combination with the design-orientation of our research, this straightforward approach is feasible. While results presented in this paper are significant, relevant and target-adequate, future work should also take a behavioristic point of view, which requires methods from the social sciences [59]. Broader studies including additional populations should be conducted to provide further empirical evidence.
5. Evaluation

To evaluate our approach, we used EASy within the tutorial process in the annual lecture Introduction to Programming. Each year around 300 freshmen (mainly CS and IS students) attend the course. To help students to understand the lecture content, weekly exercises are offered, which can be solved within groups of up to three students. Many exercises contain programming tasks. After student groups hand in their results, they are corrected by students of higher semesters (who are employed as tutors) and feedback is given in small groups of about 20 to 25 students. For motivational reasons students can earn bonuses for the final exam if exercise results are adequate. In the past, the hand-in process as well as the correction process was paper-based. This year students were free to use EASy; using EASy was not honored with extra bonus and completely voluntary.

At the end of the term, students were asked to answer a questionnaire concerning the use of EASy as well as their attitude towards e-assessment. It was made up of several free text and single-choice questions. Answering every question was not obligatory, skipping was possible. The questionnaire was answered by 187 students, most of them (88%) were male. Nearly everyone (95%) attended the course for the first time; the remaining students attended the course for the second time (and supposedly did not pass the exam in the preceding term). We asked whether EASy was the first e-assessment system they used. For more than 92% EASy was the first e-assessment tool they gained experiences with. If they had used a system before, it was either for an assessment of language competence or for an multiple choice exam; mainly systems were used that provide simple closed question types or systems that support open question types with low electronic assessment support.

Regarding the usage of EASy, we first asked for the frequency of usage; nearly 57% stated to use the system almost always and further 34% of the students said they used the system at least on occasion. Comparing to de facto usage, the denoted usage of EASy was a bit higher. Asking for the main benefit and, therefore, the reason of usage, results were unambiguous. The majority (89%) praised the system for the ability to check the functionality of their solutions beforehand and, if tests failed, to improve their solution. Saving time (3%) and other responses were only given sporadically, though students had to decide for one answer. According to the participants who did not use the system or used it only sporadically, the system’s availability was the main reason. Availability was indeed a problem of recurring tasks. The majority’s opinion concerning the relative amount of work while correcting with the help of EASy in comparison with the paper-based process was unambiguous: while assessment of low quality solutions is more laborious within EASy, a speedup is noticeable in case of mediocre and high quality solutions. Some tutors remarked that an in-place editor could soften this handicap. One tutor emphasized that in comparison with the last year the quality of correction rose due to the use of EASy.

Finally, we asked students to estimate the use of e-assessment in terms of summative, formative, and diagnostic e-assessment. Results are depicted in Figure 2. While most students (57%) judged voluntary diagnostic e-assessment as very useful, attitude towards formative or even summative e-assessment sinks. Formative e-assessment is judged by 47% of the students as very useful or useful, whereas 28% found it less useful or not useful at all. This is almost inverted for summative e-assessment where only 28% of the students have at least a positive attitude towards e-assessment and 43% refuse e-assessment.
Figure 2. Students’ attitude towards using e-assessment in different scenarios

After the final exam, we conducted an analysis and split the results into two groups: on the one hand students who used EASy for more than half of the exercises and on the other hand students who did not or only sporadically made use of EASy. Comparing the results, we noticed that students who used EASy intensively scored in average roughly 10% better than the other group. This is not a significant finding, particularly because other factors were ignored. Nevertheless, this figure underlines the positive overall perception.

6. Discussion

The findings allow for a discussion and are suited for generalization. However, some limitations have to be named.

6.1. Generalizing the Main Findings

First of all, our study supports the results we got in earlier evaluations [9], [11], [14]. E-Assessment is a feasible approach and EASy is an adequate tool for it. This has been confirmed by the high number of students that named the control of their learning progress as the main benefit of using EASy. Despite being a research tool, it has practical value and can be used to assess students in actual courses. At the same time, e-assessment has not been widely used (yet). Since most of the survey participants were first-term students and hardly any of them had used e-assessment before, it can be estimated that it is not used in primary or secondary education. It will also be subject of future research—but cannot be answered in this paper—whether e-assessment is suitable to be used in primary and secondary school education.

The fact that the majority of students did not read EASy’s documentation before using it can be generalized. Students obviously are not willing to put effort into understanding how a particular tool works. They expect it to be intuitive. This aligns with our experience from manual assessments: many errors in students’ solutions can be traced back to a superficial reading of the exercise text. It has to be demanded that students actually read assignments and put some effort into mastering the tools they need for assessments—there also are guidelines how to prepare paper-based exercise and ill-formatted solutions might be subject to deduction of points. Nevertheless, it is a finding that tools for e-assessment have to be as user-friendly as possible. This requires their design to be intuitive. Additionally, potential misusage should be prevented by sophisticated measurements. For example, students should only be provided with interface elements really needed at a given time. Entered data should not only be checked syntactically but also semantically. Think of an input mask that has two text fields: one field for the solutions in some notation and one for the comment. If the field for the solutions contains mostly text, a system should not simply accept it but ask the user whether he or she entered the desired text. Probably, the content of the two fields had been swapped. Developing e-assessment systems can draw from the vast body of knowledge in human-computer interaction (HCI) [60]—in our case, qualitative HCI research [61] is attractive. And we should take into account insights from learning psychology.

Considering why students do not want to use e-assessment, no clear tendency could be found. With regard to EASy, a variety of reasons was named (see Section 5). Apparently, there is no single reason for students to be against using e-assessment. While this is a positive finding on the one hand, it on the other hand underlines that designing universally usable systems that are widely accepted will require solving many individual challenges. This finding is supported by the responses to our questions regarding the user interface of EASy and how fast students adapted to using it. Despite receiving mostly positive answers, it would be arrogant to claim that students that responded negatively were uncooperative. In fact, it shows that even with the experience from multiple evaluations that is built into EASy, it is a highly non-trivial task to create an e-assessment system that suits all students.

For the purpose of generalization, the students’ response to the questions on their attitude towards electronic assessment in general is specifically interesting. The general trend reveals a high level of skepticism: the more relevant e-assessment became for course grading, the less it would be accepted. Unsurprisingly, hardly any student is against offers to use e-assessment voluntarily. However, this response does not reveal whether such offers would be used. It is a common reaction for students to gather materials and opt-in for extra activities as long as there are no costs involved and they might step down from actually investing time.

It has to be asked why students are more in favor of using e-assessment in graded exercises throughout the term than
in end-of-term examinations. For both kinds of assessments, we got a high number of positive, neutral, and negative responses. However, there is a slightly positive overall attitude towards mandatory exercises whereas the attitude towards exams is slightly negative overall. In theory, both questions should be answered analogously since in both cases e-assessment has direct impact on the grade. There are several attempts to explain the skepticism towards exams:

- Students have made bad experience in other courses or at least heard of bad experiences with e-assessment in exams. While this is rather unlikely for first-term students, we often hear of reports of bad experiences with purely multiple choice exams that are conducted electronically. Thus, differences of e-assessment as conducted with tools like EASY to computerized multiple-choice exams should be communicated to the students.
- Work on exercises can be done collaboratively in groups—desired by teachers, or not. This greatly reduces the fear of not understanding an exercise, experiencing technical difficulties, or running into unexpected problems. There will be fellow students to back up.
- Technology might not be trusted. On any kind of technological problem during solving an exercise, there is plenty of time to check back with a teacher. This is not possible in an exam, which ends after a fixed amount of time.
- There might be an inherent fear of losing control. Even though students of IS and CS have a high affinity to technology and understand how computer systems work, they might feel that fixing ideas on paper is safer.

While above considerations are theories, our study cannot fully explain the differences in attitude between mandatory exercises and exams. Future research has to directly address the skepticism towards electronic exams. Particularly, the hidden aversion of students has to be fathomed. We deem this a major precondition for the successful application of e-assessment. To give an example: students frequently complain about imprecise exercise descriptions (whether they are—or not). For e-assessment, exercises have to be more precise. This could be seen as one reason for skeptical students to be less reluctant.

Furthermore, it also has to be checked why there is a negative attitude towards e-assessment in general—roughly 1/4 of the students already regarded e-assessment for mandatory exercises as bad practice. Qualitative analysis could give some insights. Interestingly, most feedback considered technical issues such as our tool’s stability, its usability, and its integration into the lecture. However, some of the students that had a negative attitude did not comment on their motives. Nevertheless, it will be very interesting to repeat our survey once we have reached a higher level of system stability and applied further usability improvements. Following the comments, the expectation would be an even more positive overall attitude.

There are two more notable findings. Firstly, we got a very high number of replies for the questions of further subjects that EASY could be used for. To generalize these findings: students that are in favor of using e-assessment obviously are willing to extend it to other subjects that require higher-order cognitive skills. The named courses commonly make usage of methods from mathematics and computer science. Secondly, several students stressed that they were glad to be part of a research project and that they could contribute to research. Even though we did not announce the work with EASY to be an experiment—after all, the system is almost mature and the exercises were graded—, students felt that they helped us in our research and were proud of it. This is notable, specifically in the light of observations that students usually behave rather opportunistically. We deem this finding extremely encouraging and hope that other researchers in our field can reproduce it in their own faculties.

After all, it was surprising that students used the system very frequently and with no reservation. However, half of the tutors who had to correct the results reported at least some reservations with using the system. While test results helped tutors and were widely used to get a general overview, the code review process was criticized. This was not traced back to the usage of EASY but to the general commenting and viewing process, which is, comparing to paper and pencil work, less comfortable due to too small screens and worse abilities. Comparing screen work to handling simple paper sheets, it is not as easy to swap between files or to jump between top and tail of files. This can be aligned with the continuing but rather unsuccessful effort to establish paperless offices, which was in 1975 pretending to be the office of the future. Keep in mind that use of cellulose (in the USA) almost doubled between 1980 and 2000 [62]. Much research is required to adequately transfer paper-based reviewing processes to electronic reviews. These incapabilities seem to be the biggest problem for tutors. With regard to the tutors’ motivation, it seems to be possible to distinguish them into early-adopters and skeptics.

6.2. Limitations

Our work has some limitations. First and foremost, the survey was conducted with a relatively small group from one course at one university. While we had enough participants to get significant results, our survey has to be validated by additional studies on a broader level. It has to be both checked whether findings can be replicated at other universities and whether they apply to other courses as well. Nevertheless, there is no doubt that general conclusions can be drawn since we already did surveys along with the development of each EASY module.

The additional insights of the qualitative analysis of individual feedback are not significant—they lack the rigor of
the findings gained from the quantitative study. Nevertheless, they can be used as tendencies and to prepare future work particularly regarding improvements of EASy. Despite their limited applicability they help by enabling us to understand phenomena “from the point of view of the participants” [63]—namely, how e-assessment is perceived by students.

Finally, work on a prototypic system such as EASy is constrained by limited resources, which force us to implement new modules step-by-step. While EASy is intended to relieve teachers of manual work, developing new modules and testing them for the first time causes much overhead. While we are satisfied with our—in comparison with progress in the field fast—progress and while we believe in the feasibility of e-assessment in general, it would be naive to assume that e-assessment of arbitrary creative and complex tasks from a variety of subjects will soon be possible.

7. Conclusion and Future Work

We presented a survey of e-assessment usage in the computer science education. After introducing e-assessment in general, we introduced EASy, our modular Web-based tool for exercises that require higher-order cognitive skills. EASy comprises several modules and is a rather mature tool considering the state of related work. Our survey provided us with both quantitative results and qualitative findings. Quantitative results were diverse and addressed the usage of EASy as well as e-assessment in general. We got feedback from students who could (if desired collaboratively) solve exercises using EASy and tutors who used it for correction. This gave us many ideas for improvements and helps to understand the status of e-assessment.

Discussing our findings revealed that further research is required while e-assessment as conducted with EASy is a feasible approach. Assessment systems have to be intuitive to use and support students as good as possible. At the moment, there is a negative attitude towards graded e-assessment. Understanding this attitude and changing it by providing a new generation of tools remains the greatest challenge of the future.

Our work on EASy and on e-assessment is not finished. The presented study is a prerequisite for intensified future work. Open questions and future tasks can be identified with regard to new modules for EASy, improvements of EASy, and the progress of e-assessment in general.

We will extend EASy with new modules and refine existing ones. Our goal is not only to frequently use it in the courses our faculty offers but also to provide EASy as a tool other lecturers can use. Moreover, we aim at promoting e-assessment as a solution for better assessments. Thus, we will investigate in students’ needs for e-assessment tools. Eventually, we hope that we can promote EASy to be used by other institutions, which would allow for a longitudinal study about its effectiveness.

Acknowledgments

We thank the tutors and students for taking part in the survey, and in particular for providing much feedback.

References


