e-Assessment Behind the Scenes Common Perception of e-Assessment and How We See It Nowadays

Bastian Küppers^{1,2}, Thomas Eifert¹, Marius Politze¹ and Ulrik Schroeder² ¹IT Center, RWTH Aachen University, Germany

²Learning Technologies Research Group, RWTH Aachen University, Germany

Keywords: Electronic Examinations, Electronic Assessment, Computer Aided Assessment, Computer based Assessment.

Abstract: There exist common preconceptions about e-Assessment that, in the view of many, prohibit e-Assessment to come into operation. In this paper we examine the most commonly found preconceptions and derive how e-Assessment should be implemented from our point of view to overcome these obstacles, considering programming assessment as an example. While deriving an e-Assessment scenario, we concentrate on the integration of e-Assessment into the process workflow of institute of higher education, software for e-Assessment and suitable operational scenarios for e-Assessment.

1 INTRODUCTION

When talking to colleagues who do not deal with e-Assessment in their daily working life, it seems that some common preconceptions about e-Assessment exist, especially when it comes to summative as-These concern mainly how to implesessment. ment e-Assessment and in which situation to use e-Assessment. Altogether, these preconceptions and concerns often lead to a very reluctant position with respect to e-Assessment. The problem with this is, that even the deciders in institutes of higher education sometimes share these positions, which can potentially hinder the successful dissemination of e-Assessment. In this paper we will discuss the preconceptions most commonly expressed towards us in the past and state why we consider these as not justified any more (see section 2). Afterwards, we will discuss what in our point of view are strong and weak points of e-Assessment (see section 3). From the discussed points we will derive how e-Assessment should be used and implemented in our opinion (see section 4), using programming assessment as a use case (see section 5). The paper closes with a summary and an outlook (see section 6).

2 e-ASSESSMENT IS...

Either by colleagues from our own university, which is a public university in Germany, or by other researchers met at conferences and similar occasions, some beliefs about e-Assessment have been expressed towards us quite frequently - and apparently we are not the only ones experiencing this (Jeremias et al., 2015). In this section we examine the most frequently expressed preconceptions and state why we think these are not true nowadays.

2.1 ... Just About Multiple Choice

Admittedly, it is true that multiple choice questions (MCQ) are used intensively in e-Assessment nowadays (Oldfield et al., 2012) and if done right, there is nothing wrong with MCQ in particular (Nicol, 2007). One reasons for the prevalence of MCO in e-Assessment is certainly the ease of correcting these type of questions, as it can be done automatically. This spares the examiner of a manual correction and furthermore provides the students with their result immediately. MCQ are, however, by far not the end of the road for e-Assessment (Crisp, 2009). From our point of view, e-Assessment can only unfold its full potential, if for a particular lecture a fitting type of assignment is developed, if applicable. Otherwise, it could be better to relinquish e-Assessment in favour of another type of assessment (see also 2.2).

2.2 ... a Replacement for Traditional Exams

Some people tend to lose interest in e-Assessment, if they think that e-Assessment is only about MCQ, as discussed in the previous section. In contrast to that, other people are over-enthusiastic when it comes to e-Assessment and want to replace every Traditional Examination with e-Assessment (Curtis, 2009). From our point of view, that is also not the right way, since there are clearly cases where e-Assessment is not the right tool (Küppers and Schroeder, 2017). Consider for example examinations in a maths lecture, where students have to draw a function graph. This works a lot better in a paper-based examination than in e-Assessment, even if it could be made possible in e-Assessment if a graphics tablet is provided. Therefore, we think that e-Assessment is a valuable tool, which should be used complementary to already established types of examinations.

2.3 ... Less Secure than Traditional Exams

This is a point, which is regularly brought up by objectors of e-Assessment. In fact, it is true that making the scenario of the examination more complex also introduces more potential security threats and surely we can not postulate absolute security for e-Assessment, since this cannot be postulated for any complex system (Cook, 1998). However, that does not necessarily mean that e-Assessment is more vulnerable to cheating than paper-based examinations. First of all, it has to be distinguished between on-campus examinations and off-campus examinations. For the on-campus examinations is has to be further distinguished between e-Assessment using devices provided by the institute of higher education, e.g. in a lab, and e-Assessment using the students' own devices (BYOD). Even for off-campus examinations, solutions to retain integrity of the examination exist (Frank, 2010), however, in this paper we will concentrate on on-campus examinations. Hence, e-Assessment does not mean that the general setting of an examination has to change. Therefore, there are still invigilators in the examination room, who keep on eye on the students. Thus, uninterrupted cheating during the examination is still not possible, even if the e-Assessment is conducted on the students' devices and not on devices provided by the institute of higher education. Additionally, e-Assessment does not only introduce new ways of possible cheats, but also introduces new ways to prevent those (Küppers et al., 2017a). We think that, if implemented correctly, e-Assessment is even less likely

to be cheated in than paper-based examinations. The reason for our opinion is simply, that it is a lot easier to cheat in a traditional way, than to find and exploit security flaws in an e-Assessment system while being watched by invigilators. Therefore, even for e-Assessment we believe cribs and smartphones are the cheating tools that will be used rather than security flaws in the the e-Assessment system, which require a certain effort to be exploited.

2.4 ... a Simple Add-On to Assessment

Even if people are convinced that e-Assessment is worth implementing, they seem to underestimate the necessary effort to do so successfully. If it shall be done correctly, e-Assessment does not work as an Add-On. That means, that necessary processes need to be developed and integrated into the process landscape of an institute of higher education. At the least, students have to be identifiable in a digital workflow (Apampa et al., 2009) and results have to be stored reliably (Küppers et al., 2017b). This does not work, however, if these requirements are not met by the basic IT infrastructure of an institute of higher education. Therefore, it may be necessary to redesign existing processes and digital workflows in order to provide a valid basis for e-Assessment.

3 STRONG AND WEAK POINTS OF e-ASSESSMENT

In the last section, we discussed preconceptions about e-Assessment and why we think, these are not true nowadays. In this section we will discuss the strong and weak points of e-Assessment from our point of view to further illustrate the .

3.1 Strong Points

3.1.1 Life Cycle

During the life cycle of an assessment, which starts with creating the assignments and ends with archiving the sheets of paper that the students handed in, for a paper-based examination, several switches between analogue and digital occur: The assignments are created on a computer (digital), then are printed to sheets of paper and the students write their answers to that paper (analogue). Afterwards, the sheets of paper are corrected by pen (analogue) and the resulting marks are stored on a computer (digital), for example in an excel sheet. In the end the sheets of paper have be stored somewhere (analogue). For e-Assessment, these switches can be removed from the life cycle of the assessment, since every step can be accomplished digitally.

3.1.2 Domain-specific Tools

In most lectures, accompanying practicals and tutorials introduce domain-specific tools. For example, programming courses introduce integrated development environments, like Eclipse ¹ or NetBeans ². This way, students get already used to important tools during their studies, which may be important for later working life. In a paper-based assessment, these tools are obviously not available, which leads to a media disruption between the practised scenario and the assessment. This media disruption can be removed by switching to e-Assessment. Additionally, the students then have to show competency in using important domain-specific tools.

In addition, the niveau of the assessment can be raised. Considering Krathwohl's revised version of Bloom's Taxonomy of Educational Objectives (Krathwohl, 2002) (see Fig. 1), assessing the higher levels of the taxonomy, like Evaluate and Create, can be achieved in a more realistic fashion (Baumann et al., 2009). The tools available to the students can take care of the lower levels of the taxonomy, allowing the students to focus on the higher levels. Again, in programming courses, the integrated development environment for example provides auto completion, therefore the students do not need to remember every keyword of a programming language.

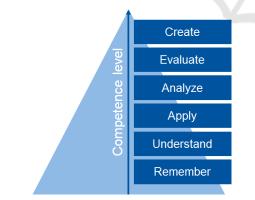


Figure 1: Revised Taxonomy of Educational Objectives by Krathwohl.

3.1.3 Simplified Correction

The same domain-specific tools that can be used by students during the assessment, can also be used by

the correctors. Sticking with the example of a programming course, this means the correctors can easily use a debugger to comprehend the internals of a piece of source code, instead of having to tediously work through handwritten code, which may in the worst case be crabbed.

3.1.4 Innovative Assessments

In a paper-based examination, every assignment has to be answerable only having pen and paper at hand. Additionally, the assignment can not include media, like audio files or video clips. Having digital tools available during the assessment, this is possible, therefore innovative methods of assessment are within reach (Baumann et al., 2009).

3.1.5 Simplified Archiving

In case of paper-based assessment, the archiving demands a lot of effort regarding personal and storage resources. Enough space has to be available where the paper sheets of the assessment can be kept. Additionally, staff members have to set up an archiving system which allows to retrieve a set of paper sheets later on. Last, but not least, the sheets of paper have to brought to the storage location and possibly have to be retrieved somewhen. In case of e-Assessment, archiving can be made a lot easier. Since the students results are already digitally available, archiving can even be automated. Additionally, a lot less space needs to be available and retrieving an archived exam can be done via a computer program.

3.2 Weak Points

3.2.1 Dependence on Infrastructure

To successfully conduct e-Assessment, a functioning infrastructure has to be available. Especially power supply and most often also network access have to be provided. In case mobile devices are used, a loss of power supply may be bridgeable for a short period of time due to the batteries in the devices. This is, however, not a reliable backup strategy. However, in a modern university almost everything relies on working IT-Infrastructure, from administrative tasks across all fields of research and education. From this perspective, an e-Assessment is "just another" aspect that relies on a 100% reliability, besides all other essential processes that keep a university running.

¹http://www.eclipse.org

²http://netbeans.org

3.2.2 Effort

Depending on the particular set up of the assessment, the switch from paper-based assessment to e-Assessment can demand a lot of effort. If, for example, individualized MCQ shall be utilized, a question pool has to be built up first. This question pool has to be peer reviewed to ensure that subsets of questions are of the same difficulty. This is important to be treat all students equally (see 4.1) and therefore to be able to give meaningful marks.

4 HOW TO DO e-ASSESSMENT

Based upon the previously discussed preconceptions, in this section we derive how e-Assessment should be implemented from our point of view in order to overcome the widely seen obstacles.

4.1 **Basic Requirements to e-Assessment**

There are some basic requirements to summative assessment, which do not change for summative e-Assessment. These requirements can be directly deduced from the purpose of summative assessment: Check whether a particular student reached the learning goals of a course. Therefore, the assessment has to be reliable in multiple manners, which has some implications. First, the writings produced by every student during the assessment have to be stored in a reliable way. That means especially, that the writings of the student cannot be altered once they were handed in, neither by the student nor by the examiner. Also, the method for storing the writings allows for correction and a later review of the correction. Second, the completion of the examination has to be reliable in the sense that it has to be ensured that the handed-in writings are indeed the work the particular student created by himself during the examination. This means in particular that cheating has to be prevented. The exact circumstances may differ, depending on the specific mode of the assessment, which may differ in several points, e.g. the location (on-campus or off-campus) or the allowed aids (closed book or open book).

Additionally, it is important that all students are treated equally during the exam. This is ethically important and, of course, required by law. In Germany, for example, Article 3 of the Basic Law of the Federal Republic of Germany demands an equality of treatment for all people (Bundestag, 2014). Therefore, it has to be ensured that all students have an equally difficult exam, though it has not to be exactly the same (Forgó et al., 2016). Additionally, also the circumstances for all students have to be similar enough to not handicap a particular student, since it is not possible to provide all students with the exact same circumstances. For e-Assessment, this applies also to the computing power of the used computer. Therefore, a student must not be handicapped by limited computing power of the used device.

4.2 Integration of e-Assessment Into Existing Processes

As previously discussed, authorship and integrity are important features of e-Assessment. In general, certificates are the current state-of-the art approach to ensure authorship and integrity of digital data by using these certificates in a digital signature (Kaur and Kaur, 2012). The problem that everyone is in principle able to generate a certificate - equivalent to a signature one can not compare with a known-good sample - is overcome by widely trusted Certification Authorities (CAs). For e-Assessment within an institute of higher education, there is no global communication, which could potentially include a wide variety of people and hence more than one CA, but only communication for processes within the institute of higher education itself. Therefore, exactly one CA has to be trusted, which ensures integrity for all certificates used in this setting. If such a CA is available, for example the DFN-PKI (DFG, 2016) for RWTH Aachen University, it should be used. If such a CA is not available, however, the institute of higher education can act as a CA itself. That would imply, of course, that the students have to trust the institute of higher education, but this seems to be the case since students already confide their personal data to the educating institutes and also trust them to handle examinations results and the like in an appropriate manner.

In order to use certificates for e-Assessment in an institute of higher education, it is crucial that not only valid certificates are used, but these also have to be unambiguously relatable to a person. Therefore, information about the issued certificates has to be stored in the Identity Management (IdM) of the institute of higher education (Eifert, Th. and Bunsen, G., 2013), for example the corresponding public key. Only if this link between an identity and a certificate is established, authorship and integrity of the results of an e-Assessment can be ensured. This would be an improvement over the current situation in paper-based examinations. Nowadays the identification is normally done by looking at the students' identity cards and letting them sign an attendance list as well as the examination itself sometimes. The student card at this

point verifies the identity of the student, because it contains a picture of the particular student as well as the matriculation number, which links to an identity in the IdM. However, the signature on the examination may ensure authorship of the students' answers in an examination, but does not ensure integrity of the documents, as described before. For example it is not uncommon that additional sheets of paper get lost.

There is, however, a crucial point that has to be tackled: If the previously described certificates are only used for e-Assessment, this may tempt students to give their private key to other students, for example for letting someone else take the examination who can do it better than oneself. Therefore, the certificate which is linked to the student has to be of high value for the student by being used not only for e-Assessment, but for all or many processes of a student life cycle in order to prevent students from sharing their certificates. For example, viewing grades, enrolling for examinations and other high valued administrative processes can be secured on the basis of the students' certificates. Therefore, everyone possessing the certificate could carry out these processes in place of the certificate's true owner. Additionally, securing digital processes by the usage of certificates could improve the privacy of the communication between a student and the institute of higher education a lot in difference to unsigned and unencrypted emails.

4.3 (Security) Software for e-Assessment

When deciding about the software that is used for e-Assessment, it is important to take into account the hardware setup first. If devices can be provided by the institute of higher education, e.g. in a lab scenario, the use of so-called lockdown programs, for example the Safe Exam Browser developed at ETH Zürich as open source project ³, is from our point of view the best way to go. Since these devices can be preconfigured and the students have only access to these devices during the examinations without administrative privileges, lockdown programs can be considered as being secure for this scenario. However, not all institutes of higher education can afford a centrally managed IT-infrastructure that provides enough capacity to carry out e-Assessment properly, since such an infrastructure is costly in terms of building up and maintaining it (Biella et al., 2009)(Bücking, 2010). This issue can be resolved by letting the students use their own devices, since most of them already posses devices that are suitable for e-Assessment (Dahlstrom

et al., 2015)(Poll, 2015)(Willige, 2016). Additionally, suitable rooms are usually available, since these would also be needed for regular examinations. If a BYOD setting is targeted, a lockdown program is not the best way to go, since there are doubts about the security of this approach on untrusted platforms (Søgaard, 2016) and thus about the applicability of these programs in a BYOD setting. Therefore, we introduced an alternative approach (Küppers et al., 2017a), which employs monitoring and logging of the students' actions during the examination, instead of locking the operating system. For this approach, only the e-Assessment software itself has to be considered instead of the whole operating system. Therefore, this approach can be easily implemented platform independent, which is important in a BYOD setting. Additionally, this reduces the complexity of the security task, since only one software, which is well known, has to be taken into security considerations instead of a whole operating system. This results is a considerably sparer threat model, which is less likely to be error-prone and therefore better to handle.

If and which security measures are in fact necessary depends on how the assignments of the e-Assessment are designed. We discuss further details about suitable operational scenarios of e-Assessment in the next section.

4.4 Operational Scenarios for e-Assessment

As already stated, we do not think that e-Assessment is a suitable replacement for every paper-based examination. We rather think that there are paper-based examinations for which the assessment can be drastically enhanced when being switched to e-Assessment. From our point of view, these are the examinations where e-Assessment can help to conduct the examination in a more realistic fashion. Basically speaking, these are the assessments where computers are nowadays part of the workflows related to the competences that shall be assessed. For example, programming assessment is obviously such a case, but there are also other examples like CAD design for architecture and mechanical engineering. How exactly the e-Assessment has to be set up in order to be considered properly set up, depends from our point of view heavily on the particular use case. In section 5 we discuss programming assessment as a representative case study on how to set up e-Assessment.

³https://safeexambrowser.org/

5 CASE STUDY: PROGRAMMING ASSESSMENT

Programming is a topic that is obviously suitable for e-Assessment, since the workflow of programming itself utilizes a computer. Therefore, e-Assessment can be used to design an assessment that is more realistic and thus better suited to assess important competences. For programming, it is not only important to know all the keywords and constructs of a programming language, but a successful programmer also needs to know how to combine those keywords and constructs into an algorithm that solves a particular problem. But that is still not sufficient, because on the way from a problem to a programmed solution, several tools are used, for example a debugger. Generally speaking, a successful programmer not only has to be ability to write working program code, but also needs to master the usage of all the tools that are necessarily part of a software development workflow. Especially this latter part of the software development process can not easily be assessed in a traditional examination, i.e. an examination on paper. When doing e-Assessment, however, it is very easy to assess these competences since the programmer's workflow and the e-Assessment-environment can be integrated with each other so that the complete workflow can get subject of the assessment. That means, that e-Assessment is not about using the computer for the same things that were carried out on paper before, but it is about providing a different workflow, which includes all necessary tools. For programming assessment, this means that an editor with syntax highlighting and probably auto-completion, a debugger and potentially a version control system not only have to be available but have to be part of the examination. This ensures that the outcome of the assessment tells something about a students' ability to write working software, and not only about her comprehension of a programming language.

Which editor is available and how this editor is integrated into a security concept does not matter that much, as long as the whole concept is consistent and fulfils all requirements regarding security and reliability. We think, however, that it is important that the setup in the examination does not differ from a setup that was used during practicals and tutorials, which potentially are held alongside the lecture. Otherwise, we observe that the students undergo separate trainings, for the exam and for the job they are trained for.

Additionally, tackling assessment like described above leads to a lot more meaningful marks. Considering a situation where an employer is looking for employees, a mark in a programming course that was obtained with the presented approach to e-Assessment tells a lot more about the skills of a potential employee and thus is way better suited to select potential appropriate applicants.

6 SUMMARY AND OUTLOOK

In this paper we discussed preconceptions regarding e-Assessment and why we think that these are nowadays not justified any more. From the discussion of the preconceptions, we derived how we think that e-Assessment should be applied to institutes of higher education, focusing on integration, security and suitable operational scenarios of e-Assessment. Based on the presented ideas, we discussed programming assessment as a case study.

Looking at the directions that e-Assessment already has taken in the last years, we are convinced that it will find its place as an important part of the examination system at institutes of higher education in the future. We do not know yet how e-Assessment will arrive there and what will probably change until then, but projecting some ten years into the future, we cannot imagine how e-Assessment could not be a part of the future of assessment - especially considering all the advantages that it offers.

REFERENCES

- Apampa, K. M., Wills, G., and Argles, D. (2009). Towards security goals in summative e-assessment security. In 2009 International Conference for Internet Technology and Secured Transactions, (ICITST), pages 1–5.
- Baumann, M., Steinmetzer, J., Karami, M., and Schäfer, G. (2009). Innovative electronic exams with voice inand output questions in medical terminology on a high taxonomic level. *Medical Teacher*, 31(10):e460–e463.
- Biella, D., Engert, S., and Huth, D. (2009). Design and Delivery of an E-assessment Solution at the University of Duisburg-Essen. In *Proceedings EUNIS 2009*, EUNIS Proceedings.
- Bücking, J. (2010). eKlausuren im Testcenter der Universität Bremen: Ein Praxisbericht. https://www.campussource.de/events/ e1010tudortmund/docs/Buecking.pdf. Accessed 2017-12-06.
- Bundestag, G. (2014). Basic law for the federal republic of germany. https://www.btgbestellservice.de/pdf/80201000.pdf.
- Cook, R. I. (1998). How complex systems fail. Cognitive Technologies Laboratory, University of Chicago. Chicago IL.
- Crisp, G. (2009). Interactive e-assessment: moving beyond multiple- choice questions.

https://www.researchgate.net/profile/Geoffrey_Crisp/ publication/228460929_Interactive_e-Assessment_moving_beyond_multiple-choice _questions/links/004635249f36897ffc000000.pdf. Accessed 2017-11-21.

- Curtis, P. (2009). Computerised testing likely to replace traditional exams, says head of board. https://www.theguardian.com/education/2009/jul/12/ written-exams-computerised-testing. Accessed 2017-11-21.
- Dahlstrom, E., Brooks, C., Grajek, S., and Reeves, J. (2015). Ecar study of students and information technology, 2015. https://library.educause.edu//media/files/library/2015/8/ers1510ss.pdf?la=en. Accessed 2017-12-06.
- DFG (2016). Certification policy of the DFN-PKI. https://www.pki.dfn.de/fileadmin/PKI/Policy-Archiv/DFN-PKI_CP-EN.pdf. Accessed 2017-11-22.
- Eifert, Th. and Bunsen, G. (2013). Grundlagen und Entwicklung von Identity Management an der RWTH Aachen. *PIK - Praxis der Informationsverarbeitung und Kommunikation*.
- Forgó, N., Graupe, S., and Pfeiffenbring, J. (2016). Rechtliche Aspekte von E-Assessments an Hochschulen.
- Frank, A. J. (2010). Dependable distributed testing: Can the online proctor be reliably computerized? In 2010 International Conference on e-Business (ICE-B), pages 1–10.
- Jeremias, X. V., Sellmer, B., and Tippe, U. (2015). Eassessments in der hochschullehre: Chancen und herausforderungen. https://opus4.kobv.de/opus4th-wildau/files/361/11_E-Assessments-in-der-Hochschullehre.pdf. Accessed 2017-12-06.
- Kaur, R. and Kaur, A. (2012). Digital signature. In 2012 International Conference on Computing Sciences, pages 295–301.
- Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*, 41(4):212– 218.
- Küppers, B., Kerber, F., Meyer, U., and Schroeder, U. (2017a). Beyond lockdown: Towards reliable eassessment. In *GI-Edition - Lecture Notes in Informatics*, volume 273, pages 191–196.
- Küppers, B., Politze, M., and Schroeder, U. (2017b). Reliable e-assessment with git - practical considerations and implementation. In *EUNIS 2017 Shaping the Digital Future of Universities*, pages 253–262.
- Küppers, B. and Schroeder, U. (2017). Vergleich von papierklausuren und elektronischen pr
 üfungen. In Eibl, M. and Gaedke, M., editors, *INFORMATIK 2017*, pages 307–318. Gesellschaft f
 ür Informatik, Bonn.
- Nicol, D. (2007). E-assessment by design: using multiplechoice tests to good effect. *Journal of Further and Higher Education*, 31(1):53–64.
- Oldfield, A., Broadfoot, P., Sutherland, R., and Timmis, S. (2012). Assessment in a digital age: A research review. http://www.bristol.ac.uk/medialibrary/sites/education/documents/researchreview.pdf. Accessed 2017-11-21.

- Poll, H. (2015). Student Mobile Device Survey 2015: National Report: College Students. https://www.pearsoned.com/wp-content/uploads/2015-Pearson-Student-Mobile-Device-Survey-College.pdf. Accessed 2017-12-06.
- Søgaard, T. M. (2016). Mitigation of Cheating Threats in Digital BYOD exams. Master's thesis, Norwegian University of Science and Technology, Trondheim, Norway.
- Willige, J. (2016). Auslandsmobilität und digitale Medien: Arbeitspapier Nr. 23. https://hochschulforumdigitalisierung.de/sites/ default/files/dateien/HFD_AP_Nr23_Digitale _Medien_und_Mobilitaet.pdf. Accessed 2017-12-06.