

DISS. ETH Nr. 22074

Designprinzipien und Evaluation eines reliablen CBA-Systems zur Erhebung valider Leistungsdaten

Abhandlung zur Erlangung des Titels

DOKTOR DER WISSENSCHAFTEN der ETH Zürich

(Dr. sc. ETH Zürich)

vorgelegt von

MARKUS DAHINDEN

MSc in Bioinformatik, Universität Lausanne und Genf

MSc in Biologie mit Schwerpunkt Mikrobiologie, Universität Bern

geboren am 14.05.1975

von Luzern

angenommen auf Antrag von

Prof. Dr. Juraj Hromkovic, Referent

Prof. em. Dr. Hans Hinterberger, Koreferent

Prof. Dr. Ulrik Schroeder, Koreferent

2014

Zusammenfassung

Das Messen von Kompetenzen stellt in der Hochschullehre eine Herausforderung dar: Einerseits ist die Formulierung von kompetenzorientierten Fragestellungen für Dozierende eine komplexe Aufgabe, andererseits verursachen die grossen Kohorten von Studierenden einen beträchtlichen Korrekturaufwand, welcher die beteiligten Experten oft tagelang beschäftigt. Durch den Einsatz eines Computer-basierten Systems kann die Entwicklung von Aufgabenstellungen durch Prozesssteuerung optimiert und der Korrekturaufwand erheblich reduziert werden.

Die vorliegende Dissertation hat zum Ziel, die Grundlagen für die nachhaltige Integration von Computer-basierten Leistungskontrollen (CBA) in der Hochschullehre zu schaffen. Dazu sind zwei Prämissen nötig: Erstens muss das System sicher sein, das heisst auch in heterogenen Systemumgebungen verlässliche und belegbare Resultate liefern. Zweitens müssen die entwickelten Fragestellungen die kognitiven Leistungen der Studierenden zuverlässig messen.

Um die Sicherheit der aktuell eingesetzten Systeme abschätzen zu können, wurden deren Funktionsweisen analysiert und potentielle Sicherheitsrisiken einer Risikoanalyse unterzogen. Dabei hat sich herausgestellt, dass die Systeme einerseits anfällig auf Netzwerkausfälle oder Überlastungen der Infrastruktur sind, andererseits nur rudimentäre technische Möglichkeiten bieten, um den rechtserheblichen Sachverhalt der Prüfungen zweifelsfrei feststellen zu können. Dies ist insofern kritisch, als dass der Wegfall eines physischen Prüfungsdokuments ein Paradigmenwechsel bei Prüfungsverfahren darstellt. Aufbauend auf den Erkenntnissen der Sicherheitsanalyse wurden deshalb fünf Designprinzipien für sichere CBA-Systeme ausgearbeitet. Diese verlangen unter anderem den Einsatz von digitalen Signaturen und Laufnummern für den Nachweis der Vollständigkeit der Resultate sowie eine Teilautonomie einzelner Komponenten des Prüfungssystems. Auf der Basis dieser Designprinzipien wurde mit der Sioux-Prüfungssuite ein *proof-of-concept* implementiert und in der Praxis mit unterschiedlichen Studierendenzahlen in sowohl unbenoteten Lernerfolgskontrollen als auch in summativen Schlussprüfungen evaluiert. Die Ergebnisse zeigen, dass ein auf den vorgeschlagenen Designprinzipien basierendes CBA-System die Anforderungen bezüglich eines sicheren und reliablen Systems erfüllen kann.

Zur zuverlässigen Messung der kognitiven Leistung der Studierenden wurden Fallbasierte Prüfungsfragen entwickelt. Diese münden in drei unterschiedlichen Fragetypen, welche sich automatisiert korrigieren lassen und somit für den Einsatz in CBA-Systemen geeignet sind. Um die Qualität dieser Fragen empirisch festzustellen, wurden

sie mittels eines in dieser Arbeit vorgestellten Entwicklungsmodells evaluiert. Dabei wurden sowohl die interne Konsistenz als auch die Gültigkeit der Resultate mittels Aussenkriterien ermittelt. Diese Analysen zeigen, dass durch den Einsatz von Fallbasierten Prüfungsfragen kompetenzorientierte, valide Leistungsmessungen möglich sind.

Mit dieser Arbeit ist die Grundlage zu einer zuverlässigen, automatisierten Erhebung valider Leistungsdaten geschaffen worden. Diese Daten werden primär für die Notengebung verwendet. Sie sind aber auch hilfreich, um mit individualisiertem Feedback die Lernaktivitäten zu fördern. Im Abschluss dieser Arbeit wurde untersucht, welches Potential zur Unterrichtsentwicklung in der Analyse der Leistungsdaten steckt. Dazu wurden die Leistungsdaten in die im Unterricht ermittelten Prozessdaten wie beispielsweise Umfrageresultate, Studierendenbeurteilungen durch Experten oder Selbstbeurteilungen integriert und auf der Basis der Leistungsdaten gruppiert. Die Ergebnisse der vorliegenden Arbeit zeigen, dass diese lehrbezogenen Daten neben der retrospektiven Analyse auch prospektive Aussagen über individuelle Lernfortschritte erlauben. Dadurch wird es möglich, die Ursache und Wirkung von Lehrbemühungen unmittelbarer zu ermitteln und so beispielsweise den Zusammenhang zwischen Feedback, Selbsteinschätzung und Motivation zu untersuchen.

Summary

Measuring competences in higher education is challenging: on the one hand, it can be difficult for lecturers to formulate competence-oriented exam questions; on the other hand, the evaluation of exams can require a lot of time and can often absorb lecturers for days. With the utilization of a computer-based system, the development of exam questions can be optimised through process management and the time spent on evaluating exams can be reduced significantly.

The aim of this dissertation is to create the basis for a sustainable integration of computer-based assessment (CBA) in higher education. This is based on two prerequisites: First, the system has to be secure. It has to deliver reliable and provable exam results even in heterogeneous system environments. Second, the designed exam questions need to validly measure the cognitive performance of the students.

To examine the security of existing systems, the operating modes of those systems were analysed and the potential risks were subjected to risk analysis. The results show that the systems are error prone due to data network failure or overload of the infrastructure. Additionally, the existing systems have only rudimentary technical features to determine the legal facts of the case of the exams. This is critical in the sense that eliminating a physical exam document from the exam process constitutes a shift of paradigm in this process. Based on the findings from the risk analysis, five design principles for secure CBA-systems were developed. These include the use of digital signatures and sequence numbers to verify the integrity of the results. Furthermore the partial autonomy of individual components of the exam system is required. Based on these design principles, the Sioux exam suite has been implemented as a proof of concept. It has been evaluated in practice with a varying amount of students in both formative evaluations of training success and through graded summative exams. The results show that a CBA-system built on these design principles complies with the requirements of a secure and reliable system.

Case-based exam questions were developed to validly measure the cognitive achievement of the students. They are based on three different types of questions, which can be corrected automatically and are therefore suitable for CBA-systems. To prove empirically the quality of those questions, a development model was created, which is presented in this thesis. Both the internal consistency and the validity of the exam results have been evaluated using external criteria. This analysis shows that competence-oriented and valid measurements are possible with case-based exam questions.

Through this research, the basis for reliable and automatic measurement of valid performance data has been laid. This data would mainly be used for grading students. However, it might also be useful to provide students with individualized feedback to develop future learning activities.

This study concludes with the analysis as to whether performance data might also be useful in developing educational instruction. In doing so, the performance data, *e.g.*, the survey results and the evaluation of the student performance through expert- and self-evaluation, were integrated with the teaching process data. The aggregated datasets have been clustered based on the performance data. The results show that educational data, along with retrospective and prospective analysis, can indeed support individual learning development. These analyses not only give the possibility to identify the cause and effect of teaching effort, but can also prove, for example, the causality between feedback, self-evaluation, and motivation.