

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 studies. 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.

Danksagung

Seit rund 10 Jahren führen wir in unseren Lehrveranstaltungen Computer-basierte Leistungskontrollen durch. Ich freue mich, dass ich mit dieser Arbeit die Möglichkeit erhalten habe, die Sicherheitsaspekte der dazu verwendeten CBA-Systeme zu beschreiben, ohne dabei die qualitativen Aspekte dieser prosperierenden Prüfungsform zu vernachlässigen.

Allen, die mich auf diesem Weg begleitet und unterstützt haben, möchte ich an dieser Stelle ein herzliches Dankeschön aussprechen: allen voran meinem langjährigen Vorgesetzten und Koreferenten Hans Hinterberger. Er war es, der aus Überzeugung früh damit begann, Computer-basierte Leistungskontrollen im Unterricht einzusetzen. Er tat dies aufgrund seiner Überzeugung, dass ein individualisierter Unterricht auch in Lehrveranstaltungen mit Hunderten von Studierenden möglich ist.

Ein grosser Dank gebührt Juraj Hromkovic. Ich achte ihn nicht nur als herausragenden Wissenschaftler und guten Pädagogen, sondern auch als Sportler, der es versteht, sich nicht auf Lorbeeren auszuruhen, sondern sich immer wieder neuen Herausforderungen zu stellen.

Ein herzlicher Dank geht nach Aachen zu Ulrik Schroeder. Er hat sich bereit erklärt, das Zweitgutachten für diese Arbeit zu erstellen. Mit seinem motivierenden Feedback in der Schlussphase hat er mir das gute Gefühl gegeben, dass der rote Faden in dieser Arbeit auch nach fünf Jahren Arbeit noch sichtbar war.

Mit seiner Doktorarbeit hat Lukas Fässler mir einen Steilpass für diese Arbeit gegeben. Er unterstützte mich in der Alltagsarbeit im Unterricht und gab mir in unzähligen Diskussionen immer wieder neue Impulse und Anregungen. Lukas, dafür danke ich dir.

Ein besonderer Dank möchte ich Peter Heinrich und Serkan Bozygit aussprechen. Mit ihren brillanten Programmierfähigkeiten ist es ihnen zu verdanken, dass Sioux sich im Prüfungsalltag als äusserst stabiles Prüfungssystem einen Namen gemacht hat.

Meinen Kollegen und Kolleginnen des Lehrstuhls Informationstechnologie und -Ausbildung möchte ich für das angenehme Arbeitsklima danken. Insbesondere gilt der Dank Hans-Joachim Böckenhauer, der wiederholt in gemeinsamen stundenlangen Korrekturmarathons seinem Ruf nach einer schnellen Auffassungsgabe numerischer Aufgabenstellungen gerecht wurde.

Fürs Korrekturlesen geht ein grosser Dank an meinen guten Freund Stefan Imper.

An dieser Stelle möchte ich auch die Mitarbeitenden des Instituts für Medizinische Lehre der Universität Bern erwähnen. Mit ihrem grossen Engagement im Bereich qualitativ hochwertiger Leistungskontrollen haben sie massgeblich dazu beigetragen, dass heute auch ausserhalb der Medizin in der Hochschullehre über die Qualität der Leistungsmessungen diskutiert wird.

Ein liebevoller Dank geht an meine Partnerin Lucia. Sie hat mir die für die Erstellung der Arbeit nötigen Freiheiten gegeben, mich aber auch immer wieder ermutigt, die Computertastatur an schönen Tagen gegen Tourenskies, Wanderschuhe oder das Mountainbike auszutauschen.

Meinen Eltern und meinen Geschwistern möchte ich für die emotionale Unterstützung danken. Es ist schön, euch an meiner Seite zu wissen.

Den Tausenden von Studierenden, welche bisher in unseren Kursen ihre Prüfungen am Computer abgelegt haben, gebührt ein grosses Dankeschön. Mit ihrer Offenheit gegenüber neuen Prüfungsformen und ihren freiwilligen, wertvollen Rückmeldungen sind sie massgeblich am Erfolg dieser Arbeit beteiligt.

Mein abschliessender Dank gilt dem Fonds Filep des Rektors der ETH Zürich und dem ETH Board via Switch/AAA für die grosszügige finanzielle Unterstützung unserer Projekte. Sie haben dadurch diese Arbeit erst möglich gemacht.

If you can not measure it,
you can not improve it.

Lord Kelvin

Inhaltsverzeichnis

Kapitel 1	1
Kontext und Ziele dieser Arbeit	1
1.1 Einleitung und Motivation	1
1.2 wissenschaftlicher Beitrag dieser Arbeit.....	2
1.2.1 Fragestellung 1: Erfüllen aktuelle CBA-Systeme die legalen Anforderungen?	2
1.2.2 Fragestellung 2: Welches sind geeignete technische Massnahmen, um in einem heterogenen Umfeld eine sichere Computer-basierte Leistungskontrolle durchzuführen?	2
1.2.3 Fragestellung 3: Können mittels automatisiert korrigierbaren Prüfungsfragen die in der Grundlagenausbildung vermittelten Informatik-Kompetenzen zuverlässig gemessen werden?	2
1.2.4 Fragestellung 4: Unter welchen Voraussetzungen bietet der Einbezug von Leistungsdaten einen Mehrwert in der Analyse lehrbezogener Daten?.....	3
1.3 Aufbau dieser Arbeit	3
1.4 Entwicklungskontext.....	7
Teil I: Grundlagen	9
Kapitel 2	11
Quantifizierung und Qualifizierung der Leistung von Studierenden	11
2.1 Kompetenzbegriff	11
2.1.1 Was ist eine "Kompetenz"?	12
2.1.2 Verschiedene Arten von Kompetenzen.....	12
2.1.3 Abgrenzung von "Kompetenzen" und "Qualifikation"	13
2.2 Wissensbegriff	14
2.2.1 Wissensarten.....	14

2.2.2	Fakten-Wissen	15
2.2.3	Konzept-Wissen	15
2.2.4	Prozedurales Wissen	15
2.2.5	Metakognitives Wissen.....	16
2.3	Taxonomie der kognitiven Leistungen	16
2.3.1	Relevanz der Taxonomie für Leistungskontrollen in der Hochschullehre	17
2.4	Taxonomie der kognitiven Leistungen nach Anderson	18
2.4.1	Die sechs kognitiven Prozessdimensionen	19
2.4.2	Prozessdimension K1: Erinnern (<i>Remember</i>)	19
2.4.3	Prozessdimension K2: Verstehen (<i>Understand</i>)	20
2.4.4	Prozessdimension K3: Anwenden (<i>Apply</i>)	20
2.4.5	Prozessdimension K4: Analysieren (<i>Analyze</i>)	21
2.4.6	Prozessdimension K5: Bewerten (<i>Evaluate</i>)	21
2.4.7	Prozessdimension K6: Erzeugen (<i>Create</i>)	21
2.4.8	Anmerkungen zur Anderson'schen Taxonomie.....	22
2.5	K3+ als Zielsetzung für die Grundlagenausbildung.....	23
2.6	Herausforderungen beim Messen der Leistung von Studierenden	24
2.6.1	Verbalisierbarkeit des Wissens	24
2.7	Zusammenfassung	26
2.8	Ausblick.....	27
	Kapitel 3	29
	Konstituierende Qualitätsmerkmale von Leistungskontrollen	29
3.1	Einleitung und Überblick.....	29
3.2	Klassierung der Leistungskontrollen	30
3.2.1	Klassierung aus verfahrenstechnischer Sicht	31
3.2.2	Klassierung aus didaktischen Sicht.....	32
3.3	Leistungskontrolle als Messinstrument für kognitive Leistungen	33
3.3.1	Diskrepanz von Messung und Realität.....	33
3.3.2	Qualitätsmetriken von Leistungskontrollen	34
3.3.3	Objektivität.....	35
3.3.4	Reliabilität (Zuverlässigkeit der Messung).....	36

3.3.5 Validität (Gültigkeit der Messung).....	41
3.4 Leistungskontrollen als Feedbackinstrument.....	43
3.4.1 Feedback aus studentischer Sicht: eigener Wissenstand beurteilen.....	44
3.4.2 Feedback zum Lerninhalt: LK zur Vermittlung des hidden Curriculums.....	45
3.4.3 Feedback aus Sicht der Dozierenden: Qualitätskontrolle.....	46
3.5 Möglichkeiten und Chancen von Computer gestützten Leistungskontrollen im Unterricht.....	48
3.5.1 Genese Computer gestützter Leistungskontrollen.....	48
3.5.2 Klassifizierung von Computer gestützten Leistungskontrollen (CAA).....	50
Kapitel 4	53
Sicherheitsaspekte Computer-basierter Leistungskontrollsysteme	53
4.1 Einleitung.....	53
4.2 Rechtliche Grundlagen von Leistungskontrollen an der ETH Zürich.....	54
4.2.1 Benotung studentischer Leistungen.....	55
4.2.2 Vorgehen bei der Anfechtung einer Notenverfügung	56
4.2.3 Gründe für die Anfechtung einer Notenverfügung.....	58
4.3 Paradigmenwechsel bei CBA.....	59
4.3.1 Verzicht auf ein physisches Prüfungsdokument	60
4.3.2 Rechtsunsicherheit aufgrund fehlender Rechtsprechungen	62
4.3.3 Auswirkungen der Rechtsunsicherheit im Prüfungsalltag.....	62
4.4 Risikomanagement-Modell für CBA.....	63
4.4.1 Statusanalyse	66
4.4.2 Risikoidentifikation.....	67
4.4.3 Risikotabelle.....	68
4.4.4 Risikoanalyse.....	70
4.4.5 Risikomanagement	75
4.4.6 <i>Go/no go</i> -Entscheidung.....	75
4.4.7 Durchführung CBA	76
4.4.8 Risikomanagement als Basis zur Entwicklung einer CBA-Strategie für Bildungsinstitutionen.....	76
4.5 Evaluationsresultate	77
4.6 Sicherheits-kritische Szenarien aktueller Online-Prüfungssysteme	82

4.6.1 Szenario 1: Teilausfall der Prüfungs-Infrastruktur	82
4.6.2 Szenario 2: "Das ist nicht meine Antwort!"	85
4.6.3 Szenario 3: "Wo ist meine Antwort der Prüfungsaufgabe 10?"	86
4.7 Diskussion der Fragestellung 1: Erfüllen aktuelle CBA-Systeme die legalen Anforderungen?	87
Teil II: Implementierung des CBA-Systems	89
Kapitel 5	91
Designprinzipien eines sicheren CBA-Systems	91
5.1 Ausgangslage.....	91
5.2 State-of-the-Art aktueller CBA-Systeme.....	92
5.2.1 Abgrenzung von CBA gegenüber ähnlichen sicherheitskritischen Client-Server-Applikationen	92
5.2.2 Übersicht über die häufigsten in der Schweiz eingesetzten CBA-Systeme.....	92
5.2.3 Manko 1: Ausfallsicherheit browserbasierter Systeme.....	94
5.2.4 Manko 2: Datensicherheit browserbasierter Systeme	95
5.3 Anforderungen an ein sicheres CBA-System	96
5.3.1 CBA-Designprinzip #1: Berücksichtigung der vorherrschenden Rahmenbedingungen.....	98
5.3.2 CBA-Designprinzip #2: Ausfallsicherheit dank Teilautonomie der Client-Applikation	99
5.3.3 CBA-Designprinzip #3: Belegbarkeit der Prüfungsergebnisse durch client-seitige, digitale Signaturen	100
5.3.4 CBA-Designprinzip #4: Nachweis der Vollständigkeit des Prüfungsdossiers dank durchgehender Nummerierung aller Antworten.....	104
5.3.5 CBA-Designprinzip #5: Reliable Prüfungsergebnisse werden erst durch eine hohe <i>Usability</i> der Client-Applikation erreicht.....	106
5.4 Auswirkungen der CBA-Designprinzipien auf die vier Sicherheitsphasen von CBA-Systemen	108
5.4.1 Sicherheitsphasen im CBA-Prozess	108
5.4.2 Verzicht auf spezialisierte Hardware für die CBA-Clients	110
5.4.3 Verzicht auf eine qualifizierte elektronische Signatur bei Prüfungsergebnissen	112

5.4.4 Verzicht auf den erweiterten Schutz der studentischen Antworten und des Fragenpools.....	117
Kapitel 6	119
<i>Proof-of-concept</i> der CBA-Designprinzipien: Sioux CBA-Suite	119
6.1 Einleitung.....	119
6.2 Sioux CBA-Suite: CBA-Designprinzipien im Prüfungsalltag	119
6.2.1 Die Sioux CBA-Suite	119
6.2.2 Systemübersicht.....	120
6.2.3 Client Applikationen	121
6.2.4 Server Applikationen	131
6.2.5 <i>Deployment</i>	133
6.2.6 Zertifikate zur Absicherung von Sioux.....	134
6.3 Sicherheitskomponenten von Sioux	135
6.3.1 Anmelden am System	137
6.3.2 Schlüsselpaar generieren.....	139
6.3.3 Leistungskontrolle starten	142
6.3.4 Antworten eingeben.....	143
6.3.5 Prüfungsdokument erstellen	145
6.3.6 Leistungskontrolle abgeben.....	147
6.4 Sioux im Praxiseinsatz.....	148
6.4.1 Organisation der Prüfungsdurchführung.....	148
6.4.2 Zeitaufwand von Leistungskontrolle mit Sioux.....	148
6.4.3 Organisation der Prüfungseinsicht.....	151
6.4.4 Dokumentation des Prüfprozesses	151
6.4.5 Performancemessungen.....	152
6.4.6 Bottleneck des Systems	154
6.4.7 Usability von Sioux	155
6.5 Weiterentwicklungen.....	157
6.5.1 MEASURED.....	158
6.5.2 Klärung der legalen Vorgaben bei CBA.....	159
6.5.3 Nachhaltige Dissemination von Sioux-basierten CBA-Systemen	160

6.6 Diskussion der Fragestellung 2: Welches sind geeignete technische Massnahmen, um in einem heterogenen Umfeld eine sichere Computer-basierte Leistungskontrolle durchzuführen?.....	161
--	-----

Teil III: Analyse der Leistungsdaten

Kapitel 7	167
------------------------	------------

Entwicklung und Evaluation valider Prüfungsfragen für CBA-Systeme.....	167
---	------------

7.1 Übergeordnete Entwicklungszyklen von CBA.....	167
---	-----

7.2 Entwicklungsmodell für die Fragenheft- und Fragenpool-Entwicklung	168
---	-----

7.2.1 Planung.....	169
--------------------	-----

7.2.2 Erstellung.....	170
-----------------------	-----

7.2.3 Durchführung	173
--------------------------	-----

7.2.4 Auswertung	174
------------------------	-----

7.3 Automatisch korrigierbare Prüfungsfragen für den Grundlagenunterricht Informatik	178
--	-----

7.3.1 Anforderungen an Fragen für CBA-Systeme.....	178
--	-----

7.3.2 Grundstruktur des Fragetyps.....	179
--	-----

7.3.3 Aufgabenstellung (Task).....	180
------------------------------------	-----

7.3.5 Fragetypen (Items)	181
--------------------------------	-----

7.4 Evaluation.....	184
---------------------	-----

7.4.1 Planung	186
---------------------	-----

7.4.2 Erstellung.....	186
-----------------------	-----

7.4.3 Durchführung.....	187
-------------------------	-----

7.4.4 Auswertung.....	188
-----------------------	-----

7.4.5 Qualitätszunahme des Fragenpools durch eine wiederholte Anwendung des Entwicklungsmodells.....	191
--	-----

7.5 Diskussion der Fragestellung 3: Können mittels automatisiert korrigierbaren Prüfungsfragen die in der Grundlagenausbildung vermittelten Informatik-Kompetenzen zuverlässig gemessen werden?.....	194
--	-----

Kapitel 8	197
------------------------	------------

Learning Analytics auf der Basis valider Leistungsdaten	197
--	------------

8.1 Analyse lehrbezogener Daten zur Unterrichtsentwicklung	197
--	-----

8.2 Möglichkeiten und Grenzen der isolierten Analyse lehrbezogener Daten	199
8.3 Aufbau des educational Data-Warehouses (eDWH).....	202
8.3.1 Eindeutige Kennung zur Aggregation der Daten	202
8.3.2 Speichern und Aggregation der Daten in einem eDWH.....	203
8.4 Kontext der Datenerhebung	204
8.5 Retrospektive Analyse des eDWH.....	205
8.5.1 Vergleich des Resultats am Probetest und der Semester-Endprüfung.....	205
8.5.2 Diskussion und Ausblick der retrospektiven Analyse.....	206
8.6 Prospektives Analysepotential des eDWH	208
8.6.1 Vergleich der Chronologie der Expertenurteilen mit dem Resultat an der Schlussprüfung.....	208
8.6.2 Vergleich der subjektiven mit den absoluten Leistungseinschätzungen in Abhängigkeit der Leistung an der Schlussprüfung.....	209
8.7 Diskussion der Fragestellung 4: Unter welchen Voraussetzungen bietet der Einbezug von Leistungsdaten einen Mehrwert in der Analyse lehrbezogener Daten?... 211	
Anhang	215
A.1 Auszug aus den Resultaten der Nachumfrage IBPW	215
A.2 Auszug aus den Resultaten der Nachumfrage EvIM.....	215
A.3 komplette Liste der Risikoanalyse	216
A.4 XML-Prüfungsdokument	218
A.5 Fehlerverteilungsmatrix.....	220
A.6 Vergleich des Lösungsverhaltens der Studierenden über die Zeit	222
A.7 Fragebogen Prozessanalyse.....	223
A.8 Prospektive Analyse der Selbst- und Fremdbeurteilung.....	224
Abbildungsverzeichnis	229
Tabellenverzeichnis	239
Literaturverzeichnis	243
Lebenslauf	255

Kapitel 1

Kontext und Ziele dieser Arbeit

1.1 Einleitung und Motivation

In der Schweiz haben sich seit 1980 die Studierendenzahlen nahezu verdreifacht (BFS, 2008). Zusammen mit der 1999 unterzeichneten Bologna-Erklärung und der damit verbundenen Verkürzung der voruniversitären Ausbildung lässt sich seither eine Diversifizierung der Vorbildung beobachten. Die dadurch resultierende Zunahme an Heterogenität der Lernvoraussetzungen und Forderung nach formalisierten, vergleichbaren Leistungsmessungen verlangt nach effizienten Prüfverfahren.

In Expertenkreisen¹ zählen Computer-basierte Leistungskontrollen (CBA, *Computer-based Assessment*) aufgrund der automatisierten Korrektur zu den effizientesten Prüfverfahren in der Hochschullehre. Zudem ermöglichen sie die Durchführung von formativen Lernerfolgskontrollen, wobei die Studierenden eine sofortige, automatisierte Leistungsrückmeldung erhalten (Black & William, 1998a). Obwohl das Potential von elektronisch gestützten Leistungskontrollsystemen seit längerem bekannt ist, sind brauchbare produktive Systeme noch rar (Bennett, 2002; Conole & Warburton, 2005). Gemäss Bennet liegt dies an den Kosten, der technischen Komplexität, der begrenzten Skalierbarkeit bei grossen Studierendenzahlen und der unzureichenden Qualität der Prüfungsfragen.

Mit dieser Arbeit werden die Grundlagen für zuverlässige, produktive CBA-Systeme erarbeitet. Insbesondere wird gezeigt, welche Systemvoraussetzungen zur Durchführung von sicheren CBA nötig sind und wie damit Informatik-Grundkompetenzen im Hochschulunterricht zuverlässig gemessen werden können. Abschliessend wird das Potential der Analyse lehrbezogener Daten auf der Basis valider Leistungsmessungen untersucht.

¹ Die im vorliegenden Dokument verwendeten männlichen Formen gelten auch für Personen weiblichen Geschlechts.

1.2 wissenschaftlicher Beitrag dieser Arbeit

Im Rahmen dieser informatik- und lehrbezogenen Arbeit werden folgende vier Fragestellungen untersucht und beantwortet:

1.2.1 Fragestellung 1: Erfüllen aktuelle CBA-Systeme die legalen Anforderungen?

Die Benotung studentischer Leistungen basiert auf der Prämisse, dass der rechtserhebliche Sachverhalt korrekt festgestellt wird. Dies ist unabdingbar, um faire und somit für alle Studierenden vergleichbare Prüfungsbedingungen zu schaffen.

Sicherheitslücken in den CBA-Systemen können dazu führen, dass die Vollständigkeit und die Integrität der Resultate nicht gewährleistet werden kann. In dieser Arbeit wird mittels einer Risikoanalyse untersucht, welches die aktuell gravierendsten Risiken bei CBA darstellen und wie allenfalls deren Auswirkungen reduziert werden können. Diese Fragestellung wird im Kapitel 4 beantwortet.

1.2.2 Fragestellung 2: Welches sind geeignete technische Massnahmen, um in einem heterogenen Umfeld eine sichere Computer-basierte Leistungskontrolle durchzuführen?

Ein heterogenes Umfeld stellt hohe Anforderungen an ein CBA-System. Einerseits lassen sich die einzelnen Faktoren nicht oder nur schwer kontrollieren und andererseits muss die Gültigkeit des Resultats trotzdem jederzeit überprüft werden können.

In dieser Arbeit sollen Designprinzipien zur Erhöhung der Sicherheit von CBA-Systemen erarbeitet und in der Praxis evaluiert werden. Ziel ist, dass ein auf diesen Prinzipien aufgebautes CBA-System ein im Vergleich zu aktuellen CBA-Systemen erhöhtes Sicherheitslevel aufweist. Diese Fragestellung wird im Kapitel 6 beantwortet.

1.2.3 Fragestellung 3: Können mittels automatisiert korrigierbaren Prüfungsfragen die in der Grundlagenausbildung vermittelten Informatik-Kompetenzen zuverlässig gemessen werden?

Automatisiert korrigierbare Prüfungsfragen sind nötig, damit CBA-Systeme die Leistungen der Studierenden unmittelbar und objektiv messen können. Diese Automatisierung erlaubt den flexiblen Einsatz von CBA-Systemen im Unterricht. Mittels automatisiert korrigierbaren Prüfungsfragen kann nicht nur der Korrekturaufwand reduziert werden, sondern sie ermöglichen auch das Versorgen der Studierenden mit individualisiertem Feedback bezüglich deren Lernzielerreichung.

In dieser Arbeit soll untersucht werden, inwieweit mit automatisiert korrigierbaren Prüfungsfragen die kognitiven Wissenslevels Informatik-Einführungsveranstaltungen reli-

abel und valide gemessen werden können. Diese Fragestellung wird im Kapitel 7 beantwortet.

1.2.4 Fragestellung 4: Unter welchen Voraussetzungen bietet der Einbezug von Leistungsdaten einen Mehrwert in der Analyse lehrbezogener Daten?

In Lernumgebungen fallen mit Daten aus Leistungskontrollen, Fragebögen, Tutoratsgruppen und Log-Daten eine Vielzahl lehrbezogener Daten an. Die Auswertung solcher Daten wurde vom *New Media Consortium* als Forschungsschwerpunkt im Bereich "*Learning Analytics*" deklariert.

Mit dieser Arbeit soll untersucht werden, unter welchen Voraussetzungen der Einbezug von qualitativ hochwertigen Leistungsdaten einen Mehrwert gegenüber der klassischen Unterrichtsevaluation liefern kann. Insbesondere soll untersucht werden, ob neben der retrospektiven Analyse auch prospektive Vorhersagen gemacht werden können. Dadurch liesse sich die Ursache und Wirkung von Lehrbemühungen unmittelbarer messen und gezieltes individualisiertes Feedback ausarbeiten. Durch zeitnahe individuelle Interventionen bei auffälligen Studierenden liesse sich womöglich auch bei grossen Studierendenzahlen der *Learning Outcome* generell erhöhen.

Diese Fragestellung wird im Kapitel 8 beantwortet und mit Beispielen aus der Praxis illustriert.

1.3 Aufbau dieser Arbeit

Im Zentrum dieser Arbeit stehen zum einen die Schaffung eines sicheren CBA-Systems und zum anderen den Entwurf von automatisiert korrigierbaren Items zur validen Messung von kognitiven Leistungen (Abbildung 1.1). In einem proof-of-concept wird die Gültigkeit dieser Designprinzipien in der Praxis validiert. Dazu wurde ein auf den Designprinzipien basiertes CBA-System entwickelt und mittels Fall-basierter Fragen in der Praxis evaluiert.

Abschliessend zeigt eine Analyse, inwiefern die mit dem CBA-System ermittelten Leistungsdaten einen Mehrwert bei der lehrbezogenen Analyse (*Learning Analytics*) von Lernumgebungen liefern können.

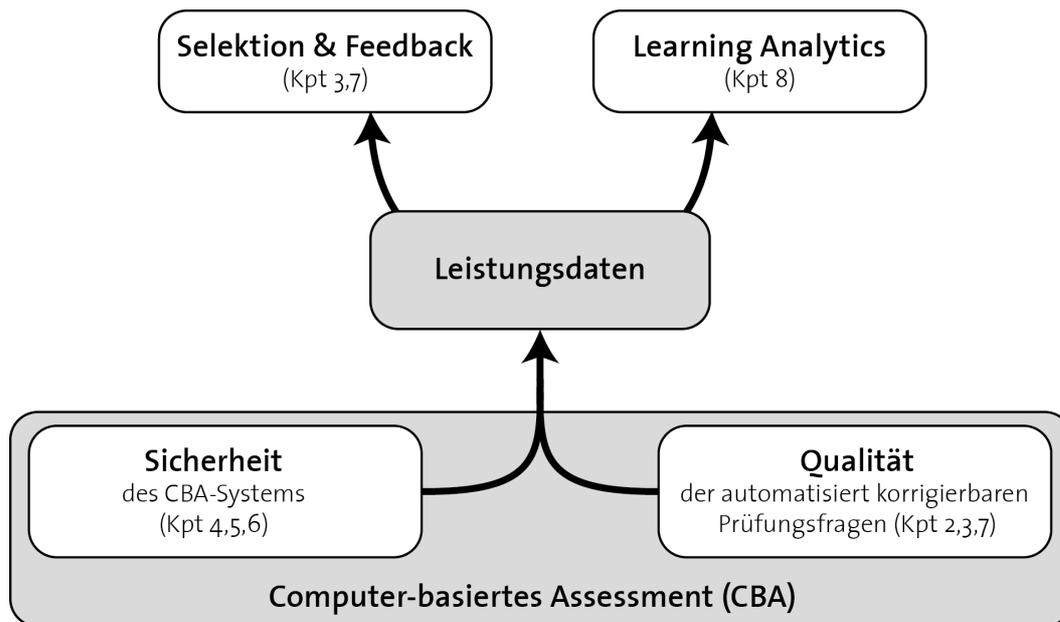


Abbildung 1.1: Illustration der vier Schwerpunkte dieser Arbeit. Das CBA-System bildet die Grundlage für die sichere Erfassung von qualitativ hochwertigen Leistungsdaten. Die Leistungsdaten dienen einerseits als Selektions- und Feedbackinstrument als auch als Grundlage für deren lehrbezogene Analyse (Learning Analytics). In Klammern findet sich ein Verweis auf das entsprechende Kapitel in dieser Arbeit.

Diese Arbeit gliedert sich in drei Teile:

Teil I "Grundlagen": Wodurch zeichnet sich die kognitive Leistung von Studierenden aus und wie kann sie gemessen werden? Wie kann die Qualität dieser Messung erkannt und abgeschätzt werden?

Teil II "Implementierung eines CBA-Systems": Entwurf und Evaluierung eines sicheren CBA-Systems unter Berücksichtigung der Heterogenität des Einsatzgebiets.

Teil III "Analyse der Leistungsdaten": Entwicklung von automatisiert korrigierbaren Prüfungsfragen und Qualitätserfassung der damit ermittelten Leistungsdaten für die Selektion und Feedbackerteilung. Untersuchung des Analysepotentials der auf validen Leistungsdaten basierten lehrbezogenen Analyse des Unterrichts.

Teil I: Grundlagen

Kapitel 2: Quantifizierung und Qualifizierung der Leistung von Studierenden

Leistungsmessungen sollen die Leistung von Studierenden messen. So verlangt es die allgemeine Verordnung der Leistungskontrolle der ETH Zürich (ETH Zürich, 2012). Doch wie lässt sich die "Leistung von Studierenden" formal beschreiben? Welche Möglichkeiten gibt es, verschiedene Leistungen zu bewerten und miteinander zu vergleichen?

In diesem Kapitel soll ausgehend vom Kompetenz- und Wissensbegriff die Taxonomie der kognitiven Leistungen erläutert werden. Mit diesem Werkzeug können die kognitiven Leistungen in eine hierarchische Form gebracht und so bewertbar gemacht werden. Durch die Projektion der Verbalisierbarkeit des Wissens auf die kognitive Taxonomie sollen zudem die Herausforderungen bei der Schaffung valider Leistungsmessungen illustriert werden.

Kapitel 3: Konstituierende Qualitätseigenschaften von Leistungskontrollen

Die Leistungskontrollen stellen quasi das Messinstrument zur Erfassung kognitiver Leistungen dar. Wie jedes Messinstrument haben auch Leistungskontrollen gewisse Messungenauigkeiten, welche zu falschen kognitiven Leistungsbeurteilungen führen können.

In diesem Kapitel wird die Grundlage zur qualitativen Beschreibung von Leistungskontrollen geschaffen. Diese setzen sich aus statistischen Verfahren der Item-Analyse und aus inhaltlichen und didaktischen Überlegungen zur Schaffung valider Leistungskontrollen zusammen.

Obwohl in diesem Kapitel der Fokus auf der qualitativen Beschreibung der Leistungskontrolle liegt, soll ein wichtiger Aspekt der Leistungskontrolle nicht ausser Acht gelassen werden: die qualitativen Auswirkungen der Leistungskontrollen auf den Unterricht. Die Grundlage dazu soll am Ende des Kapitels 3 mit dem Konzept des *hidden Curriculums* und der Genese von CBA-Systemen geschaffen werden.

Kapitel 4: Sicherheitsaspekte Computer-basierter Leistungskontrollsysteme

CBA-Systeme haben neben der schnellen automatischen Korrektur den Vorteil, dass sie den Lernenden in einer formativen Leistungskontrolle ein sofortiges Feedback zu ihrem individuellen Leistungsstand liefern können. Damit CBA-Systeme nachhaltig in eine heterogene Lernumgebung integriert werden können, müssen vorgängig die Sicherheitsaspekte solcher Systeme geklärt sein. Aufgrund fehlender Rechtsprechung zur Zeit der Drucklegung dieser Arbeit existieren noch keine verbindlichen Vorgaben für CBA. Es lässt sich feststellen, dass der Klärungsbedarf zwar akut ist, aber bei *Stakeholdern* von CBA oft das Bewusstsein für die spezifischen Herausforderungen von elektronischen Prüfungssystemen und das Verständnis für die involvierten Technologien fehlt. Es ist somit Ziel dieses Kapitels, das Bewusstsein in Bezug auf Risiken bei CBA-Entscheidungsträgern zu erhöhen, indem anhand einer Risikoanalyse die verfügbaren rechtlichen Vorgaben mit technischen, organisatorischen und formalen Überlegungen kombinieren werden.

Teil II: Implementierung des CBA-Systems

Kapitel 5: Designprinzipien eines sicheren CBA-Systems

Aus der Risikoanalyse lassen sich fünf Design-Prinzipien ableiten, welche in diesem Kapitel beschrieben werden. Diese basieren auf organisatorischen, technischen, finanziellen und didaktischen Faktoren und haben zum Ziel, ein CBA-System zu schaffen, welches auch in einer heterogenen Umgebung, wie sie beispielsweise in den öffentlichen Computerräumen an einer Hochschule anzutreffen ist, eingesetzt werden kann.

Kapitel 6: Proof-of-concept der CBA-Designprinzipien: Sioux CBA-Suite

In diesem Kapitel steht die Umsetzung der im Kapitel 5 beschriebenen CBA-Designprinzipien im Zentrum. Der so entstandene Prototyp eines sicheren CBA-Systems, "Sioux CBA-Suite" genannt, wird in der Praxis zur routinemässigen Durchführung von benoteten Prüfungen und unbenoteten Leistungskontrollen mit unmittelbaren Korrekturen der Resultate eingesetzt. Es werden die Erfahrungen besprochen, die mit diesem CBA-System im Praxisalltag an verschiedenen Bildungsinstitutionen im In- und Ausland gemacht werden. Dabei geht es hauptsächlich um die Frage, inwieweit mit der Sioux CBA-Suite das Ziel eines zuverlässigen, praxistauglichen Prüfungssystems auf der Basis der CBA-Designprinzipien erreicht wurde.

Teil III: Analyse der Leistungsdaten

Kapitel 7: Entwicklung und Evaluation valider Prüfungsfragen für CBA-Systeme

Damit mit einem CBA-System zuverlässige kognitive Leistungsmessungen gemacht werden können, ist neben einer sicheren Implementierung die Qualität der verwendeten Prüfungsfragen entscheidend. Im Kapitel 7 wird ein Entwicklungsmodell für CBA-Fragen vorgestellt. Dieses hat zum Ziel, die Prämissen für eine qualitativ hochwertige Leistungskontrolle zu beschreiben und deren Qualität empirisch zu bestimmen. Bei der wiederholten Anwendung sieht das Modell vor, die CBA-Fragen anhand ihrer individuellen, gemessenen Metriken zu optimieren. Dadurch kann der Fragen-Pool nachhaltig bewirtschaftet und die Qualität der Prüfung sukzessive gesteigert werden.

Im zweiten Teil dieses Kapitels wird das Entwicklungsmodell angewandt, um unsere CBA-Prüfungsfragen zu evaluieren und die Qualität unseres Fragen-Pools zu untersuchen. Unsere für CBA-Systeme entwickelten Fall-basierten Prüfungsfragen verwenden drei automatisiert korrigierbare Fragetypen und werden sowohl für unbenotete Lernerfolgskontrollen als auch für benotete Schlussprüfungen verwendet. Durch die im Entwicklungsmodell vorgegebene Qualitätsprüfung der Fall-basierten Fragen mittels Blueprint und Aussenkriterien, wird die Validität der Fragen ermittelt.

Kapitel 8: Learning Analytics auf der Basis valider Leistungsdaten

Leistungsdaten werden in der Praxis primär als Selektions- und als Feedbackinstrument verwendet.

Der Abschluss dieser Arbeit bildet die Untersuchung, inwiefern valide Leistungsdaten auch bei der Analyse lehrbezogener Daten einen Mehrwert bieten. Dazu werden die Leistungsdaten und die lehrbezogenen Daten integriert und basierend auf den Leistungsdaten gruppiert. Lehrbezogene Daten fallen in Rechner-gestützten Lernumgebungen durch die Aktivitäten der Studierenden an verschiedenen Orten an. Als Beispiele können Zugriffszahlen auf Webseiten, Nutzungsdaten von Diensten und Programmen sowie Rückmeldungen aus Fragebögen genannt werden.

Durch die Integration und das Clustering der lehrbezogenen Daten und der Leistungsdaten erhoffen wir uns prospektive Einschätzungen der individuellen Leistungsentwicklung von Studierenden. Mit Beispielen aus der Praxis soll das mögliche Potential der lehrbezogenen Analyse auf der Basis valider Leistungsdaten illustriert werden.

1.4 Entwicklungskontext

Diese Arbeit ist im Zusammenhang mit der Durchführung von zwei jährlichen Lehrveranstaltungen des Departements Informatik der Eidgenössischen Technischen Hochschule Zürich (ETH Zürich) entstanden. Die beiden Informatik-Einführungsveranstaltungen werden pro Jahr von rund 600-800 Studierenden der Naturwissenschaften im ersten Studienjahr besucht.

Diese beiden Lehrveranstaltungen wurden unter der Leitung von Prof. Dr. H. Hinterberger zwischen 2000 und 2004 komplett überarbeitet und seither im Praxisalltag laufend optimiert. Die beiden Lehrveranstaltungen sind für einen Arbeitsaufwand von rund 60-100 Stunden ausgelegt (2-4 ECTS Punkte) und basieren auf dem 4-Schritte-Modell für einen kompetenzorientierten Informatikunterricht (L. E. Fässler, 2007; Hinterberger, 2010). Nach der Emeritierung von Prof. Dr. H. Hinterberger im Herbst 2010 wurden die beiden Lehrveranstaltungen von Dr. H.-J. Böckenhauer übernommen und im bewährten Stil weitergeführt. Die Lehrveranstaltung "Grundlagen der Informatik" ersetzt "Informatik (für Biol./Pharm. Wiss/HST)" und setzt den Schwerpunkt stärker auf die Einführung in die Programmierung.

Nummer	Dozent	Titel	Zielgruppe	Typ	ECTS	Bemerkung
252-0852-00	Prof. Dr. J. Hromkovic & Dr. H.-J. Böckenhauer	Grundlagen der Informatik	BIOL, HST, CHAB, INF	O O O Z	4	Ab HS13
252-0839-00	Dr. H.-J. Böckenhauer Dr. L. E. Fässler	Einsatz von Informatikmitteln	AGRL, ERDW, UWIS, INFK	O O O Z	2	Ab HS10
252-0850-00	Dr. H.-J. Böckenhauer	Informatik (für Biol./Pharm.Wiss./ HST)	BIOL, HST, CHAB, INF	O O O Z	4	Ab FS11
251-0839-00	Prof. Dr. H. Hinterberger	Einsatz von Informatikmitteln	AGRL, ERDW, UWIS, INFK	O O O Z	2	Bis HS09
251-0850-00	Prof. Dr. H. Hinterberger	Informatik (für Biol./Pharm. Wiss.)	BIOL, BEWI, CHAB, INF	O O O Z	4	Bis FS10

Tabelle 1.1: Informatik-Einführungsveranstaltungen des Departement Informatik für Studierende der Naturwissenschaften. Typ O: Obligatorische Lehrveranstaltung, Typ Z: Zusatzangebot zum Vorlesungsverzeichnis, AGRL: Agrar- und Lebensmittelwissenschaften, ERDW: Erdwissenschaften, HST: Gesundheitswissenschaften und Technologie, INF: Informatik, CHAB: Chemie und angewandte Biowissenschaften, UWIS: Umweltnaturwissenschaften (Informationen zusammengestellt gemäss Vorlesungsverzeichnis der ETH Zürich, Dezember 2013).

Teil I

Grundlagen

Kapitel 2

Quantifizierung und Qualifizierung der Leistung von Studierenden

Die ETH Zürich definiert den Begriff der *Leistungskontrolle* als "Verfahren, mit dem die Leistung von Studierenden gemessen und bewertet wird, insbesondere Prüfungen, Prüfungsblöcke und schriftliche Arbeiten" (ETH Zürich, 2012, Art. 2d).

In diesem Kapitel soll der Begriff der "Leistung von Studierenden" umschrieben und definiert werden. Dies ist notwendig, um im weiteren Verlauf dieser Arbeit die Leistung von Studierenden zu adressieren und qualitativ fassbar zu machen. Diese qualitative Komponente ist in den Kapiteln 7 und 8 zentral, wo einerseits die Validität der im Rahmen dieser Arbeit erstellten Fragetypen bestimmt und andererseits deren Potential für die Evaluation und Weiterentwicklung einer Lernumgebung aufgezeigt werden sollen.

Die Leistung von Studierenden kann unter verschiedenen Gesichtspunkten beurteilt werden. So wird beispielsweise bei Studierenden der Sportwissenschaften die physische Leistungsfähigkeit ermittelt. Aber auch soziale Leistungen wie Einfühlvermögen oder Konfliktlösefähigkeit können Merkmale eines Individuums sein.

Da in der Hochschullehre vermehrt die Vermittlung von Kompetenzen gefordert wird, soll in den nachfolgenden Abschnitten die kognitive Leistung ausgehend vom Kompetenzbegriff definiert werden.

2.1 Kompetenzbegriff

Die Bologna Deklaration verlangt neben der strukturellen Harmonisierung der Europäischen Studiengänge auch ein Paradigmawechsel in der Lehre: so wird in der Hochschullehre der Wandel von der Inhalts- zur Kompetenzvermittlung vorangetrieben (Joint declaration of the European Ministers of Education, 1999). Dadurch soll das sogenannte *Lifelong Learning* (lebenslanges Lernen) unterstützt werden, welches das Bedürfnis unserer Wissensgesellschaft abdeckt, das Bildungsangebot auch auf Lebensphasen jenseits der klassischen Studienzeit auszudehnen.

Wie die Rektorenkonferenz der Schweizer Universitäten (CRUS) in ihrem aktuellsten Bericht zum Schluss kommt, befinden sich die Schweizer Universitäten jedoch noch am Anfang dieses Wandels (Rektorenkonferenz der Schweizer Universitäten, 2012).

Als Erklärung für diese langwierige Umsetzungsphase kann geltend gemacht werden, dass neben der Vermittlung von Kompetenzen auch deren Messung in der Praxis herausfordernd ist. Die Messung von Kompetenzen ist Inhalt der Kapitel 3, 7 und 8.

2.1.1 Was ist eine "Kompetenz"?

Als Kompetenzen werden die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten bezeichnet, welche es ihnen ermöglichen, bestimmte Probleme in variablen Situationen bewusst und selbständig zu lösen (Weinert, 1999). Kompetenzen sind nicht direkt messbar, sondern nur aus der beobachtbaren Handlung erschliess- und evaluierbar (Erpenbeck, 2007).

Abbildung 2.2 zeigt den Zusammenhang von Wissen, Kompetenz und Handlung: Während kompetente Studierende ihr Wissen in unbekanntem Problemstellung zur Lösungsfindung anwenden können, reicht bei Novizen das vorhandene Wissen entweder nicht aus oder sie können aufgrund geringer Erfahrung ihr Wissen nicht generalisieren, das heisst in unbekanntem Problemstellungen anwenden ("passives Wissen").

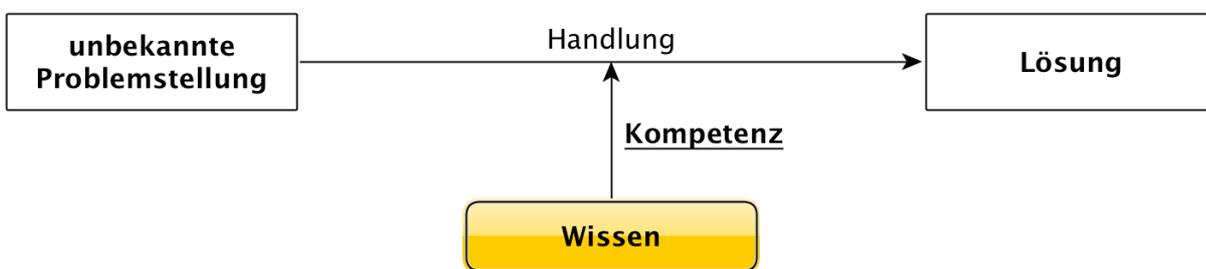


Abbildung 2.2: Als Kompetenz wird die bewusste, selbstorganisierte Anwendung von Wissen auf der Basis einer unbekanntem Problemstellung im Problemlösungsprozess verstanden. Eine Kompetenz kann nur beobachtet, nicht aber direkt gemessen werden (Erpenbeck, 2007).

2.1.2 Verschiedene Arten von Kompetenzen

Weinert unterscheidet vier Arten von Kompetenzen, welche ergänzend zu verstehen sind und sich nicht isoliert aneignen lassen:

- Personale Kompetenzen: Fähigkeit einer Person, sich selbst einzuschätzen und Motivationen zu entfalten, um reflexiv und selbstorganisiert zu handeln
- Fachlich-Methodische Kompetenzen: Disposition einer Person, ihr Wissen selbständig anzuwenden und weiterzuentwickeln

- Sozial-Kommunikative Kompetenzen: Fähigkeit einer Person, sich mit sich selber und anderen kreativ auseinander- und zusammzusetzen, um Pläne, Aufgaben und Ziele zu entwickeln
- Aktivitäts- und umsetzungsorientierte Kompetenzen: Fähigkeit einer Person, seine Kompetenzen dazu zu verwenden, um seine Pläne, Aufgabe und Ziele umzusetzen

In der Grundlagenausbildung der Schweizer Hochschulen steht primär die Vermittlung und Prüfung von Fachlich-Methodischen Kompetenzen im Zentrum. Diese Priorisierung wird gestützt und konserviert durch die auf nationaler und internationaler Ebene geforderte Vergleichbarkeit der Studienleistungen, welche sich primär auf die Fachlich-Methodischen Kompetenzen beschränken. Trotzdem ist es erfreulich zu sehen, dass viele neue Lernmethoden die anderen Kompetenzarten zum Teil bewusst adressieren und fördern. So wird beispielsweise mit Feedback-Elementen versucht, die Personalen Kompetenzen zur fördern und mittels Gruppenarbeiten und Problem-basiertem Lernen den Aufbau von Sozial-Kommunikative sowie Aktivitätsorientierte Kompetenzen zu unterstützen (Merkel, 2013).

Diese gesamtheitliche Förderung aller vier Kompetenzarten ist richtig und wichtig. So zeichnet sich beispielsweise ein guter Forscher dadurch aus, dass er nicht nur über herausragende Fachkompetenzen verfügt, sondern dass er Ziele und Visionen hat, seine Handlung selbstkritisch hinterfragt und seine Ideen auch mit anderen Personen teilen kann.

2.1.3 Abgrenzung von "Kompetenzen" und "Qualifikation"

Es ist wichtig zu betonen, dass aus einer beobachtbaren Handlung nicht zwingend auf eine Kompetenz geschlossen werden kann. Von einer Kompetenz kann erst dann die Rede sein, wenn die Handlung selbständig in einer für das Individuum neuen Problemstellung stattgefunden hat.

Darin besteht der entscheidende Unterschied zu einer *Qualifikation*, wobei das Wissen in normierten, identischen Problemstellungen angewandt wird. Die so beobachtbaren Ergebnisse widerspiegeln das aktuelle Wissen. Ob jemand davon ausgehend auch das Wissen auf neue Problemstellungen anwenden kann, darf aus einer guten Qualifikationsbewertung nicht abgeleitet werden (Erpenbeck, 2007). So ist es beispielsweise möglich, dass ein zertifizierter Informatiker ein Netzwerkproblem nicht zwingend schneller als sein unzertifizierter Kollege beheben kann.

Der in der Hochschullehre weit verbreitete Begriff von "transferfähigem Wissen" kann demnach auch als Abgrenzung von *Kompetenz* (transferfähig) und *Qualifikation* (nicht zwingend transferfähig) verstanden werden.

Die Frage stellt sich, ab wann eine kognitive Leistung als Kompetenz bezeichnet werden kann. Dazu ist es notwendig, die der Kompetenz zugrundeliegenden Wissensarten zu betrachten.

2.2 Wissensbegriff

Wissen dient als Grundlage für die kognitive Leistungsfähigkeit. Dessen Aneignung geht mit dem Aufbau von Kompetenzen einher. Wissen hat die Eigenschaft, eine persönliche, subjektive Eigenschaft einer Person zu sein, welche sich nicht direkt zwischen Personen übertragen lässt (Hey, 2004). Wissen kann nur transferiert werden, wenn dessen Träger sein Wissen in Information umwandelt und dem Empfänger kommuniziert (Abbildung 2.3). Über den wirksamen Wissenserwerb im Informatik-Unterricht sei an dieser Stelle auf andere Arbeiten verwiesen (Bauer-Messmer, Fässler, & Wyss, 2006; L. E. Fässler, 2007; Merkel, 2013; L. Snyder, 2008).

Auch bei Leistungskontrollen spielt die Transfereigenschaft des Wissens eine zentrale Rolle. Denn analog der Unmöglichkeit, Wissen direkt zu transferieren, kann auch bei Leistungskontrollen das Wissen nicht direkt abgefragt, sondern nur in schriftlicher, mündlicher oder handlungsgebundener Form externalisiert werden (Abbildung 2.3).

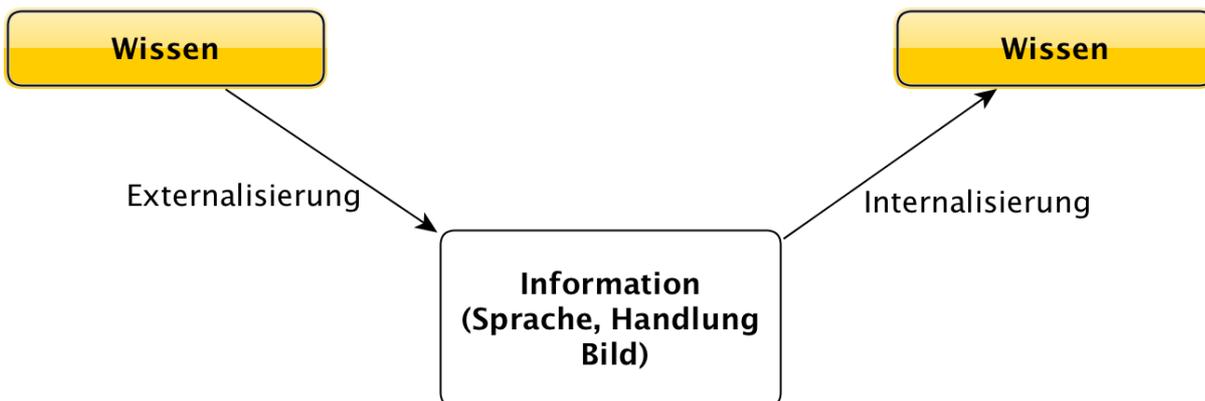


Abbildung 2.3: Wissen ist an eine Person gebunden und kann nur indirekt gemessen oder zu anderen Personen transferiert werden (Externalisierung von Wissen).

2.2.1 Wissensarten

Es gibt eine Vielzahl von Begriffen, um die Wissensarten zu definieren und vermutlich noch mehr, um diese zu beschreiben. "Begriffs-Wissen", "träges Wissen", "aktives Wissen" oder "mechanisches Wissen" sind nur einige, welche von Dozierenden im Praxisalltag häufig genannt werden.

Anderson und seine Kollegen führen 18 Wissensarten auf, die in der Literatur Erwähnung finden (L. Anderson, Krathwohl, & Bloom, 2001, p. 41). Aus dieser Liste haben sie vier generelle Wissenstypen abgeleitet (Tabelle 2.2):

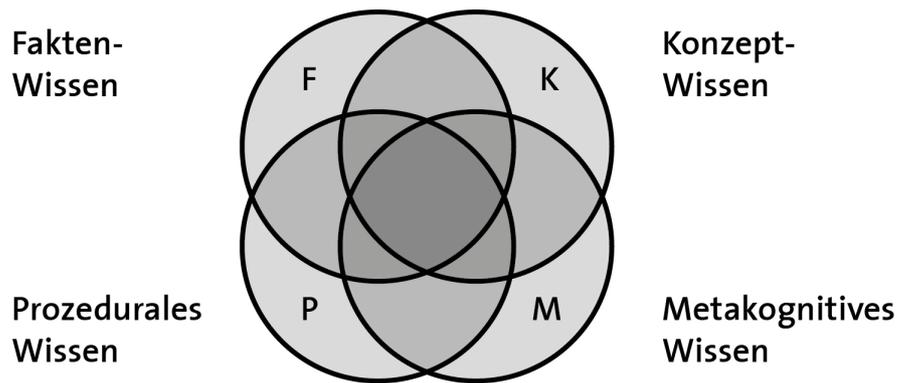


Tabelle 2.2: Die vier Wissensarten nach Anderson (L. Anderson et al., 2001). Diese vier Wissensarten sind nicht hierarchisch. Die Externalisierbarkeit hat Einfluss auf die Messbarkeit dieser Wissensart und wird im Abschnitt "2.6.1 Verbalisierbarkeit des Wissens" erläutert. Generell lässt sich festhalten, dass explizites Wissen einfacher messbar ist.

Diese Klassifizierung mag zwar nicht zwingend den formalen psychologischen Modellen von der Wissensrepräsentation genügen, doch gerade in der Hochschullehre und in der Leistungsmessung wird die Unterscheidung zwischen Fakten und deren Verallgemeinerung in Form von Konzepten (und Modellen) als sinnvoll erachtet (L. Anderson et al., 2001; Mayer, 2002).

Ein häufig gehörtes Missverständnis in der Praxis der Leistungsprüfung ist, dass "Wissensfragen" bloss Fakten erfragen. Deshalb werden hier die vier Wissensarten noch etwas ausführlicher erläutert (L. Anderson et al., 2001, p. 46).

2.2.2 Fakten-Wissen

Fakten-Wissen besteht aus den Grundelementen, welche Studierende wissen müssen, um mit einer Disziplin vertraut zu sein und selbständig Probleme lösen zu können. Dieses Wissen beinhaltet Fakten, Terminologien und Begriffe. Zum Beispiel wissen die Studierenden, dass "XML" die Abkürzung für "Extensible Markup Language" ist oder sie können die acht primitiven Datentypen der Java Programmiersprache aufzählen.

2.2.3 Konzept-Wissen

Das Konzept-Wissen umfasst die Verknüpfung und Generalisierung von Fakten-Wissen zu Prinzipien, Modellen, Theorien und Konzepten. Das Konzept-Wissen ist nicht mehr wie das Fakten-Wissen an konkrete Situationen gebunden. So haben beispielsweise die Studierenden ein mentales Modell davon, wie ein Datenbankmanagementsystem funktioniert.

2.2.4 Prozedurales Wissen

Das Prozedurale Wissen umfasst Methoden und Entscheidungskriterien, um handlungsfähig zu sein. Dazu gehörten Prozeduren, Techniken und Methoden genauso wie

Kriterien, um diese zu nutzen. Beispielsweise können die Studierenden eine geeignete Visualisierungsmethode auswählen, um eine bestimmte Problemstellung zu lösen.

2.2.5 Metakognitives Wissen

Das Metakognitive Wissen beinhaltet die Kenntnis über das persönliche Wissen. Dazu gehört, dass man sich seinen eigenen Stärken und Schwächen bewusst ist. Der Aufbau von Metakognitivem Wissen ist eng mit der Fähigkeit zum selbstkritischen Handeln verbunden. Diese Eigenschaft ist die Grundlage für das zielgerichtete und effiziente Problemlösen. So ermöglicht beispielsweise ein solides Metakognitives Wissen einem Studierenden sich effizient auf eine Abschlussprüfung vorzubereiten, da er seinen Leistungsstand kennt und sich so gezielter vorbereiten kann.

2.3 Taxonomie der kognitiven Leistungen

Laien sind versucht zu sagen, dass Fakten-Wissen weniger kognitive Leistung verlangt als Meta-Kognitives Wissen. Aus wissenschaftlicher Sicht ist diese Aussage aber nicht haltbar, denn die vier im vorherigen Abschnitt beschriebenen Wissensarten sind ergänzend und nicht hierarchisch zu verstehen. Es ist also nicht so, dass beispielsweise Meta-Kognitives Wissen generell hochwertiger ist als Fakten-Wissen. Dieser Umstand soll anhand des folgenden plakativen Beispiels verdeutlicht werden:

Die berühmte Einstein'sche Formel der Speziellen Relativitätstheorie $E = mc^2$ besagt, dass Masse nichts anderes ist als Energie in einer anderen Erscheinungsform (Einstein, 1905). Es ist unbestritten, dass Albert Einstein zur Herleitung dieser Formel eine herausragende kognitive Leistung vollbracht hat. Durch diese Konzeptualisierung und Formalisierung seiner Forschungsarbeit hat er schlussendlich neues Fakten-Wissen geschaffen, welches seither Grundlage für viele moderne Wissenschaftsdisziplinen ist und deshalb vielfach gelehrt und geprüft wird. Aus didaktischer Sicht funktioniert jedoch der reversible Ansatz leider nicht: Durch die bloße Wiedergabe der Formel $E = mc^2$ kann keine Aussage über die kognitive Leistung seines Trägers abgeleitet werden, die er zur Internalisierung dieses Wissens aufgewendet hat. Die Frage bleibt also unbeantwortet, ob der Träger diese Formel bloss wiedergeben, erklären oder gar in neuen Situationen anwenden kann. Diese Unterscheidung ist jedoch in der Hochschullehre von zentraler Bedeutung.

Es kann also festgehalten werden, dass die kognitive Leistung erst durch die Unterscheidung der den Wissensarten zugrundeliegenden kognitiven Prozesse eine hierarchische, adressierbare Struktur erhält und so als Taxonomie der kognitiven Leistungen verstanden werden kann (Abbildung 2.4).



Abbildung 2.4: Von einer kognitiven Leistung kann erst dann gesprochen werden, wenn Wissen in einem kognitiven Prozess zur Anwendung kommt. Wissen alleine ist dimensionslos und kann nicht bewertet werden (F = Fakten-Wissen, K=Konzept-Wissen, P=Prozedurales-Wissen, M=Metakognitives-Wissen).

2.3.1 Relevanz der Taxonomie für Leistungskontrollen in der Hochschullehre

Eine Taxonomie [griech. táxis = Ordnung und nomos = Gesetz] ist ein Klassifikationssystem, mit dem Objekte oder Sachverhalte nach bestimmten Kriterien klassiert werden können (Koschnick, 1984). Durch die Klassierung werden Einzelfälle in einen einheitlichen Kontext gebracht, welcher summarische Aussagen ermöglicht und bis hin zur Erklärung von Zusammenhängen führen kann. Die wohl bekannteste Taxonomie ist jene von Linnaeus, welcher alle lebenden Organismen in einer baumartigen Hierarchie klassierte und so deren Beziehungen zueinander aufzeigte.

In der Bildungswissenschaft können Taxonomien der kognitiven Leistungen verwendet werden, um das Curriculum einer Lerneinheit detaillierter zu beschreiben, als man dies mit der blossen Auflistung der behandelten Themen tun würde. Der Mehrwert einer kognitiven Taxonomie rührt daher, dass man für Themen auch die Stufe der kognitiven Leistung mit definiert. So kann zum Beispiel das Lernziel ("Die Studierenden lernen die Rekursion kennen") so verstanden werden, dass die Studierenden die korrekte Lösung der Turm-von-Hanoi Aufgabe wiedergeben können. Oder aber, dass sie ihr Wissen über die Rekursion dazu verwenden können, um dieses in neuen Problemstellungen selbstständig anzuwenden. Diese qualitative Komponente der Lernzieldefinition ist schlussendlich auch die Voraussetzung für den kompetenzorientierten Unterricht. Der Zusammenhang von kognitiven Leistungen und Kompetenz wird Inhalt des Abschnitts 2.5 "K₃₊ als Zielsetzung für die Grundlagenausbildung" sein.

Es lassen sich sechs Vorteile einer kognitiven Taxonomie festhalten (abgeändert nach L. Anderson et al., 2001, p. 34ff):

Die Taxonomie der kognitiven Leistungen...

- 1) ...hilft den Lehrenden, den Standpunkt der Studierenden besser zu verstehen.
- 2) ...bildet die Grundlage für die kompetenzorientierte Curriculumentwicklung in der Hochschullehre.
- 3) ...dient als Werkzeug für das *Instructional Design* einer Lerneinheit, indem sie die integralen Beziehungen zwischen Wissen und kognitiven Prozessen darstellt.
- 4) ...bildet die Basis für die reliable und valide Messung der kognitiven Leistungen
- 5) ...dient als Grundlage zur Evaluation von Lerneinheiten (*Learning Analytics*), indem die Kategorisierung den Zusammenhang zwischen Lehre und Prüfung aufzeigen kann.
- 6) ...dient den Lehrenden und Bildungswissenschaftlern zur präzisen Kommunikation ihrer Erkenntnisse.

2.4 Taxonomie der kognitiven Leistungen nach Anderson

In der Bildungswissenschaft gilt Bloom als Pionier bei der Entwicklung einer Taxonomie der kognitiven Leistungen (Bloom, 1956). Seine eindimensionale Gliederung der Lehr- und Lernziele wurde später von Anderson und seinen Kollegen zu einer zweidimensionalen Taxonomie erweitert (L. Anderson et al., 2001). Bei dieser Erweiterung wurden die Wissensarten und die kognitive Prozessdimensionen getrennt, so dass sie explizit adressierbar wurden (Abbildung 2.5). Diese Taxonomie eignet sich deshalb gleichermaßen für die Planung des Unterrichts, des Curriculums sowie der Leistungsnachweise. Sie kann gar für die Überprüfung der internen Konsistenz dieser drei Planungsvorgänge verwendet werden (Baumgartner, 2011, p. 42ff). So können Fragen wie beispielsweise "Stimmen die vermittelten und die geprüften Inhalte überein?" beantwortet werden. Die Überprüfung der internen Konsistenz von studentischen Leistungen in der Praxis ist Inhalt des Kapitels 7.

Wissens- Dimensionen	6 kognitive Prozessdimensionen					
	K1 Erinnern	K2 Verstehen	K3 Anwenden	K4 Analysieren	K5 Bewerten	K6 Erzeugen
Fakten- Wissen						
Konzept- Wissen						
Prozedurales- Wissen						
Metakognitives Wissen						

Abbildung 2.5: Taxonomie der kognitiven Leistungen nach Anderson und Krathwol (L. Anderson et al., 2001). Sie bringt die Wissensarten und die kognitiven Prozessdimensionen in eine Beziehung und ermöglicht so die explizite Adressierung einer kognitiven Leistung.

2.4.1 Die sechs kognitiven Prozessdimensionen

Anderson und seine Kollegen haben insgesamt 19 verschiedene kognitive Prozesse identifiziert, die sie in sechs kognitive Prozessdimensionen gruppierten. Das Gruppierungskriterium ist die Komplexität der kognitiven Prozesse, die sich von einfachen ("Erinnern") bis zu komplexen Dimensionen ("Erzeugen") gliedern.

Nachfolgend werden die sechs kognitiven Prozessdimensionen kurz erläutert. Eine detaillierte Auflistung aller 19 kognitiven Prozessen findet sich im Buch "A taxonomy for Learning, Teaching and Assessing" von Anderson und seinen Kollegen (L. Anderson et al., 2001).

2.4.2 Prozessdimension K1: Erinnern (*Remember*)

Die erste Bloom'sche Prozessdimension (K1) umfasst das Abrufen von Informationen aus dem Langzeitgedächtnis, wobei diese in ähnlicher Form wiedergegeben werden, wie sie gelehrt wurden. Als Beispiel kann die Wiedergabe von Fakten, Definitionen oder Gesetzen genannt werden. Die erste kognitive Prozessdimension kann - wie die anderen Prozessdimensionen auch - auf alle vier Wissensarten einzeln oder in einer beliebigen Kombination angewandt werden.

Beispiele:

"Nennen Sie ein Clusteringverfahren, welches in der multivariaten Datenvisualisierung Anwendung findet" (Kombination aus Fakten-, Konzept- und Prozeduralem-Wissen)

"In Prüfungssituationen ist es hilfreich, ruhig zu bleiben" (Wissensart: Metakognitives-Wissen)

2.4.3 Prozessdimension K2: Verstehen (*Understand*)

Wenn es das Ziel ist, die Transferfähigkeit von vermitteltem Wissen zu fördern, müssen die Studierenden das Wissen nicht bloss abrufen können (KI), sondern sie müssen ein Verständnis für die vermittelten Inhalte aufbauen. Dieses wird aufgebaut, indem sie die vermittelten Inhalte mit bereits bestehendem Wissen verknüpfen. Das heisst, dass die Studierenden das neue Wissen in die bestehenden kognitiven Wissensstrukturen integrieren können.

Dem Verständnis zugrunde liegenden kognitiven Prozesse gehören unter anderem das Interpretieren, Klassifizieren, Zusammenfassen und Vergleichen von Konzepten. Einfache Ursache/Wirkung-Erklärungen werden möglich.

Beispiele:

"Erklären Sie, wieso folgendes SQL-Statement nicht effizient ist."

2.4.4 Prozessdimension K3: Anwenden (*Apply*)

Die in Prozessdimension zusammengefassten kognitiven Prozesse beinhalten den Gebrauch von Abläufen, um Übungen und Problemstellungen zu bearbeiten. In den Übungen kennen die Studierenden die für die Lösungsfindung notwendigen Abläufe. Bei einer Problemstellung hingegen ist für die Studierenden vorerst unklar, welche Abläufe für die Lösungsfindung beigezogen werden sollen. Das korrekte Vorgehen muss deshalb erst noch bestimmt werden. Entsprechend beinhaltet diese Prozessdimension zwei Prozesse: *ausführen* (execute) - wenn das Vorgehen bekannt ist und *umsetzen* (implement) - wenn das Vorgehen unbekannt ist. Diese dritte kognitive Prozessdimension stellt also faktisch den Übergang vom reproduzierbaren zum transferfähigem Wissen dar. Und Letzteres bildet, wie im Abschnitt 2.1 erläutert, die Grundlage einer Kompetenz.

Beispiele:

"Benutzen Sie die Zielwertsuche, um den Wendepunkt zu bestimmen, an dem sich die Dieselversion eines Autos gegenüber der Benzinversion zu lohnen beginnt." (bekanntes Vorgehen)

"Schlagen Sie eine Visualisierungsmethode vor, um den Zusammenhang zwischen der Ozonbelastung und der Temperatur aufzuzeigen". (unbekanntes Vorgehen)

2.4.5 Prozessdimension K4: Analysieren (*Analyze*)

Die vierte Prozessdimension vereint kognitive Prozesse, welche die Grundlagen bilden, damit eine Person eine neue Aufgabe analysieren kann. Dazu gehört, dass sie in der Lage sein muss, die Informationen in einer Aufgabe in ihre Einzelteile aufzuteilen, deren Zusammenhänge zu erkennen und mit dem Vorwissen abzugleichen. Diese drei Schritte sind die Grundlage dafür, dass eine Person in einer neuen Situation die wichtigen von den nebensächlichen Informationen unterscheiden kann.

Beispiele:

"Studierenden Sie folgenden Programmcode und beschreiben Sie die verwendeten Design-Pattern."

2.4.6 Prozessdimension K5: Bewerten (*Evaluate*)

Der kognitive Prozess des "Bewertens" verlangt, dass eine Person eine neue Situation aufgrund eigener oder gelernter Kriterien und Standards bewerten kann. Die dazu am häufigsten gebrauchten Kriterien betreffen die Qualität, Effektivität, Effizienz und Konsistenz.

Es muss an dieser Stelle angemerkt werden, dass nicht alle gefällten Entscheide eine Bewertung verlangen und somit eigentlich einer tieferen Prozessdimension zugeordnet werden müssen. Als Beispiel kann hier die Wahl der korrekten Schleife (*while* vs. *for* Schleife) in einer Programmierung verstanden werden. In diesem Fall geht es nur darum zu unterscheiden, ob das Resultat der beiden Schleifen gleich ist oder nicht. Diese Unterscheidung verlangt keine Bewertung der beiden Schleifentypen gemäss gelernter Kriterien oder Standards und könnte demnach eher der kognitiven Dimension K4 "Analysieren" zugeordnet werden.

Beispiele:

"Studieren Sie folgenden Programmcode und schreiben Sie ihn so um, dass er effizienter wird, die Ausgabe aber gleich bleibt."

"Studierenden Sie die folgenden Nährwertdaten und versuchen Sie daraus zwei Ernährungsempfehlungen abzuleiten."

2.4.7 Prozessdimension K6: Erzeugen (*Create*)

Die sechste und höchste kognitive Prozessdimension "Erzeugen" umfasst die Leistung, Wissen zu kombinieren, um etwas Neues zu schaffen. Dieser Prozess ist eng verknüpft mit der Erfahrung aus früheren Lernvorgängen.

Es ist wichtig zu betonen, dass der Begriff "Erzeugen" im Kontext der kognitiven Taxonomie nicht als manueller Vorgang verstanden werden darf. Vielmehr sind die kognitiven Vorgänge zu betrachten, die zum Erreichen des Produkts benötigt wurden.

Ein Student, der beispielsweise eine Darstellung eines komplexen wissenschaftlichen Sachverhalts *erzeugt*, leistet trotz der Schaffung eines kreativen Produkts möglicherweise nur kognitive Arbeit bis Stufe 3 ("Anwenden"). Das Erreichen der Prozessdimension 6 lässt sich also nicht anhand eines originellen oder kreativen Produkts erkennen. Vielmehr ist es entscheidend, dass die Studierenden mehrere Erfahrungselemente miteinander kombinieren, um in Abhängigkeit ihres Vorwissens neue Fähigkeiten zu *erzeugen*. Im Vergleich zu den Stufen K1-4 sind dabei die Elemente nicht mehr vorgegeben, sondern müssen von den Studierenden selber bestimmt werden.

Beispiele:

"Sie wollen den Zusammenhang der Treibhausgase auf die Klimaerwärmung untersuchen. Planen Sie die Datenerhebung im Hinblick auf die Forschungsfrage."

2.4.8 Anmerkungen zur Anderson'schen Taxonomie

Obwohl es intuitiv einsichtig wäre, dass die sechs kognitiven Prozessdimensionen nach Anderson einen streng hierarchischen Aufbau haben, so nehmen neuere empirische Studien im Gebiet der Informatik davon Abstand (Fuller et al., 2007). Diese Studien kommen zum Schluss, dass die kognitiven Prozesse der Dimensionen "Bewerten" und "Erzeugen" so komplex und vielschichtig sind, dass nicht mehr von einer klaren hierarchischen Abgrenzung gesprochen werden kann.

Fuller und ihre Kollegen haben deshalb eine für die Informatik adaptierte Form der Bloom'schen Taxonomie ausgearbeitet (Tabelle 2.3). Durch die Zweiteilung der sechs kognitiven Prozessdimensionen nach Anderson haben sie das Ziel erreicht, dass die beiden höchsten Dimensionen nicht mehr hierarchisch repräsentiert sind und somit die Stufe "Create" auch ohne vorgängige Beherrschung der Stufe *Evaluate* erreicht werden kann. Der Nachteil der Fuller'schen Taxonomie ist jedoch, dass die Stufe "Create" auch direkt ausgehend von der Stufe *Remember* erreicht werden kann, wobei die Vermutung nahe liegt, dass Fuller bei den Dimensionen *Apply* und *Create* die Wissensdimensionen mit den kognitiven Prozessdimensionen vermischt hat.

PRODUCING	Create				
	Apply				
	none				
		Remember	Understand	Analyse	Evaluate
		INTERPRETING			

Tabelle 2.3: Adaptation der Bloom'schen Taxonomie nach Fuller (Fuller et al., 2007). Im Hinblick auf die Leistungskontrolle ist vor allem die Unterscheidung der beiden Dimensionen "producing" und "interpreting" interessant. Sie sind ergänzend und nicht hierarchisch zu verstehen und zeigen, dass beide Verfahren zur Leistungsmessung gleichbedeutend verwendet werden können.

Ungeachtet dessen hat Fuller mit der Zweiteilung der Prozessdimensionen einen für die Informatik und für viele andere Ingenieurbereiche wichtige Feststellung gemacht: Das Erzeugen eines Produkts verlangt nicht zwingend eine höhere kognitive Leistung als dessen Interpretation.

Fuller begründet dies damit, dass zum Beispiel im Informatikunterricht das Schreiben von Programmen kognitiv nicht zwingend anspruchsvoller ist, als das Suchen nach Fehlern (*Bugfixing*) oder das Optimieren von vorgegebenen Programmen (*Refactoring*).

Diese Feststellung stellt eine zentrale Erkenntnis für Leistungskontrollen in Ingenieurwissenschaften dar. Es kann also festgehalten werden, dass das Erreichen von höheren kognitiven Stufen nicht alleine durch das Erzeugen eines Produkts, sondern auch durch das Interpretieren eines Sachverhalts gezeigt werden kann.

2.5 K3+ als Zielsetzung für die Grundlagenausbildung

Welche Bedeutung hat nun die eingangs des Kapitels erwähnte Kompetenz-Definition, wonach Individuen dann als kompetent bezeichnet werden können, wenn sie zur Problemlösung ihr Wissen in variablen Situationen einsetzen können?

Hierbei stellt sich die Frage, ab welcher kognitiven Prozessdimension von einer Kompetenz gesprochen werden kann. Werden die kognitiven Prozesse der Bloom'schen Taxonomie betrachtet, kann man erkennen, dass sich die ersten beiden Dimensionen "Erin-

uern" und "Verstehen" auf vorgegebene Inhalte beschränken. Eine Verknüpfung zu anderen Konzepten wird nicht vorausgesetzt. So kann beispielsweise die Syntax von html-Webseiten verstanden werden, aber dieses Wissen kann nicht auf andere strukturierte Textdokumente (wie beispielsweise LaTeX) übertragen werden. Diese Fähigkeit wird erst durch Erfahrungszuwachs in der Prozessdimension 3 "Anwenden" erreicht und bildet die Grundlage der Dimensionen 4-6.

Um die Kompetenzorientierung auf die in der Hochschullehre weit verbreitete Taxonomie nach Anderson zu übertragen, wurde der Begriff "K₃+" eingeführt. Darunter wird die Adressierung der kognitiven Prozessdimensionen K₃ und höher verstanden.

Wenn wir also den eingangs dieses Kapitels erwähnte Paradigmawechsel von der Inhalts- zur Kompetenzvermittlung wiedergeben wollen, dann sprechen wir fortan von "K₃+"-basiertem Unterricht. Hierbei ist es aber wichtig zu betonen, dass "K₃+" als Zieldimensionen einer Unterrichtseinheit verstanden werden soll. Aufgrund des hierarchischen Aufbaus der kognitiven Prozessdimensionen kann dieses Ziel nur erreicht werden, wenn im Unterricht auch die beiden ersten Dimensionen adressiert werden (L. E. Fässler, 2007, Kapitel 6.4).

2.6 Herausforderungen beim Messen der Leistung von Studierenden

Der Paradigmawechsel von der Inhalts- zur Kompetenzvermittlung stellt auch hohe Anforderungen an Leistungskontrollen. Es reicht nicht mehr, die vorgegebenen Inhalte abzufragen, sondern die Studierenden sollen zur Lösungsfindung diese Inhalte in neuen Situationen anwenden. Korrigiert werden dann die Lösungen, welche als Produkt der kognitiven Prozesse auf die Prüfungsblätter geschrieben werden. Es hängt von der Fragestellung und vom Unterricht ab, ob die Lösung aufgrund Auswendiglernens (K₁&2) oder aufgrund einer Transferleistung (K₃+) erreicht wurde.

Eine weitere Schwierigkeit von Leistungskontrollen beim Messen von K₃+-Leistungen ist die Tatsache, dass kognitive Leistungen vereinzelt überschätzt werden. So kann beispielsweise die Aufgabe "Beweisen Sie die Formel XY" bisweilen auch mit der blossen Wiedergabe von vorher vermitteltem Inhalts-Wissen (K₁&2) beantwortet werden, auch wenn die Herleitung der Formel für Aussenstehende als K₃+ erscheinen mag.

Wir können also festhalten, dass bei der Zuteilung einer Prüfungsfrage zu einer kognitiven Prozessdimension immer auch der Unterricht und die Verbalisierbarkeit des geprüften Wissens mitberücksichtigt werden müssen.

2.6.1 Verbalisierbarkeit des Wissens

Wie bereits erwähnt, kann Wissen nur in Form von individuellen Erkenntnisgewinnen angeeignet werden. Analog verhält es sich bei der Wiedergabe von Wissen: Der Träger muss sein Wissen erst verbalisieren oder in einer Handlung zum Ausdruck bringen, um

ihn als kompetente Person erscheinen zu lassen. Das direkte Wiedergeben (Externalisieren) des Wissens ist nur in wenigen Fällen direkt möglich. Nachfolgend werden die Hintergründe dazu erläutert.

In der Bildungswissenschaft teilt man das Wissen aufgrund dessen Externalisierbarkeit in zwei Gruppen: explizites und implizites Wissen. Während das explizite Wissen vom Träger einfach in Worte gefasst und beschrieben werden kann, kann das implizite Wissen nur in der Handlung beobachtet oder allenfalls vom Träger umschrieben werden. Oft ist die Umschreibung aber unzureichend und unangemessen (Nonaka & Takeuchi, 2012).

Abbildung 2.6 zeigt den Versuch einer Klassierung der Verbalisierbarkeit ausgehend von der kognitiven Taxonomie nach Anderson. Unsere Überlegungen basieren darauf, dass das Fakten-Wissen unabhängig von der Prozessdimension verbalisiert werden kann. Die restlichen drei Wissensdimensionen liegen nur in den beiden tiefsten Prozessdimensionen (K1 "erinnern" und K2 "verstehen") als explizites Wissen vor. Bei höheren kognitiven Leistungen ist dieses Wissen an Erfahrung geknüpft und kann vom Träger intuitiv richtig angewandt werden, ohne dass er es möglicherweise externalisieren kann. Polanyi formuliert dies folgerichtig mit den Worten: "Wir wissen mehr, als wir zu sagen wissen" (Polanyi, 1985).

Wissens-Dimensionen	6 kognitive Prozessdimensionen					
	K1 Erinnern	K2 Verstehen	K3 Anwenden	K4 Analysieren	K5 Bewerten	K6 Erzeugen
Fakten-Wissen						
Konzept-Wissen						
Prozedurales-Wissen						
Metakognitives Wissen						

Abbildung 2.6: Explizites (weiss) und implizites Wissen (grau) übertragen auf die Taxonomie der kognitiven Leistungen nach Anderson. Da sich explizites Wissen über alle kognitive Prozessdimensionen verteilt, können nur beim Messen von implizitem Wissen die Zieldimensionen von K3+ verlässlich erreicht werden.

Beispielsweise kann auf der Stufe K2 ein Student verstanden haben, dass er beim Aufbau einer Datenbank die referentielle Integrität sicherstellen muss. Trotz dieses Konzept-Wissens kann er die referentielle Integrität einer Datenbank im Alltag nicht überprüfen. Die Überprüfung ist erst ab K3 möglich ist, liegt doch auf Stufe K1 & K2 das Wissen vorerst nur als explizites Wissen vor. Mit steigender Erfahrung erweitern die Studierenden dann dieses Wissen und können es in neuen Aufgabenstellungen anwenden.

Die Vielschichtigkeit dieses Wissens und der direkte Erfahrungsbezug erschweren oder verunmöglichen es den Studierenden, das Wissen zu verbalisieren. Dieses nunmehr implizite Wissen kann jedoch in der Handlung zum Ausdruck gebracht werden. Diesen Umstand macht man sich bei der Messung des impliziten Wissens zu Nutze.

Aufgrund der besseren Verbalisierbarkeit lässt sich im Alltag beobachten, dass in der Grundlagenausbildung der Hochschullehre mehrheitlich explizites Wissen geprüft wird. Je nach angestrebtem *Learning Outcome* mag das Abfragen von explizitem Wissen ausreichend sein. Im Hinblick auf den Aufbau von Kompetenzen, als Grundlage für lebenslanges Lernen, sollte hingegen auch bewusst das implizite Wissen gemessen werden. Nur wenn mit einer Leistungskontrolle sowohl implizites als auch explizites Wissen gemessen werden kann, sind gültige (valide) Kompetenzmessungen möglich (vgl. Kapitel 3 "Konstituierende Qualitätsmerkmale von Leistungskontrollen").

Im Vergleich zum expliziten Wissen ist das Messen von implizitem Wissen anspruchsvoll, da sich dieses nur unzureichend verbalisieren lässt und mehrheitlich aus einer Handlung erschlossen werden muss. Bewältigt man diese Herausforderung, hat man die Gewissheit, dass die Messung auf die kognitiven Prozessdimensionen K3+ abzielen.

In Kapitel 7 werden wir empirisch untersuchen, ob wir mit unseren neu entwickelten Prüfungsfragen das Ziel der Kompetenzmessung im Informatikunterricht erreicht haben.

2.7 Zusammenfassung

Mit Leistungskontrollen soll die Leistung von Studierenden gemessen werden, so verlangt es unter anderem die "Allgemeine Verordnung über Leistungskontrollen" der ETH Zürich (ETH Zürich, 2012).

In diesem Kapitel haben wir deshalb den Leistungsbegriff ausgehend von dem ihm zugrunde liegenden Prozess definiert. Dieser reicht von der Wiedergabe von vordefinierten Inhalten (K1&K2), der Anwendung und dem Transfer von Wissen in neuen Situationen (K3&K4) bis zum Bewerten und Erzeugen von neuen Inhalten (K5&6).

In der Hochschullehre kommt der Prozessdimension K3 eine zentrale Bedeutung zu. Schliesslich stellt sie den Übergang zum transferfähigen Wissen dar und kann somit als minimale Zieldimension für einen kompetenzorientierten Unterricht verstanden werden.

Alle der sechs kognitiven Prozessdimensionen (K1-K6) nach Anderson basieren auf den vier Wissensarten (Fakten-, Konzept-, Prozedurales- und Metakognitives Wissen), welche sich unterschiedlich gut verbalisieren lassen.

Dies führt beim Prüfungen zu zwei Herausforderungen:

- 1) Es gibt auf jeder Stufe einfach zu prüfendes Fakten-Wissen. Es braucht jedoch den Einbezug des Unterrichts und der Aufgabenstellung, um die dafür notwendige kognitive Prozessdimension abschätzen zu können. So wäre beispielsweise die Frage nach der Formel der Speziellen Relativitätstheorie vor deren Veröffentlichung im Jahre 1905 der höchsten Prozessdimension K6 zuzuordnen gewesen. Die selbe Frage würde heute je nach Unterricht vermutlich noch maximal die kognitive Leistung der Stufe K2 erreichen.
- 2) Die Stufen K3-6 beinhalten handlungsbezogenes Wissen. Dieses kann bei schriftlichen Prüfungen nicht direkt geprüft werden. Festgehalten und korrigiert wird nur das Ergebnis einer Handlung; zum Beispiel in Form einer verbalisierten Handlungsbeschreibung oder eines Schlussresultats.

2.8 Ausblick

In der Hochschullehre haben Leistungskontrollen zum Ziel, die kognitive Leistung von Studierenden zu messen. Die so gewonnenen Daten dienen entweder den Studierenden als Rückmeldung zum aktuellen Wissensstand oder sie werden von den Dozierenden oder der Hochschule als promotionsrelevante Messwerte eingesetzt, um die Qualität des Ausbildungsstandortes zu gewährleisten.

Aufgrund dieser zentral wichtigen Stellung der Leistungskontrollen in der Grundlagenausbildung der Hochschullehre versucht diese Arbeit die Grundlagen für kompetenzorientierte Leistungsmessungen zu schaffen.

Basierend auf der theoretischen Grundlage des Wissens- und Kompetenzbegriffs und der kognitiven Prozessdimensionen, lassen sich folgende Erkenntnisse für die Entwicklung von Prüfungsfragen für den ICT- und Informatikunterricht festhalten:

A) Handlungsbezogene Prüfungen sind kompetenzorientierte Prüfungen, sofern es sich nicht um vordefinierte Handlungsabläufe mit vordefinierten Resultaten handelt. Die Prüfungen sollen handlungsbasiert sein und trotzdem teil-automatische Korrekturen ermöglichen.

B) Die Prüfungen sollen kein Fakten-Wissen erfragen, da die Gefahr besteht, dass dann primär auf Stufe K1/2 geprüft wird. In einem heterogenen Umfeld mit einem breiten Spektrum an Vorwissen lässt sich in Unterricht nur schwer feststellen, ob der Student das Fakten-Wissen höherer kognitiver Stufen tatsächlich selber erreicht hat (K3+) oder ob er sich an dieses Wissen bloss erinnert (K1/2).

C) Als Grundlage zur Entwicklung von kompetenzorientierten Prüfungsfragen soll die Erkenntnis von Fuller verwendet werden. Sie besagt, dass das Interpretieren und das Erzeugen von Sachverhalten vergleichbar hohe kognitive Anforderungen darstellen.

Die neu entwickelten, kompetenzorientierten Prüfungsfragen werden im Kapitel 7 beschrieben.

Da die Qualität dieser neu entwickelten Fragen empirisch ermittelt werden soll, wird im folgenden Kapitel die Qualität von Leistungskontrollen als Messinstrument für kognitive Leistungen umschrieben. Es soll aufgezeigt werden, mit welchen Methoden einerseits die Messzuverlässigkeit (Reliabilität) und andererseits die Messgenauigkeit (Validität) von Prüfungsfragen im Alltag ermittelt werden kann.

Kapitel 3

Konstituierende Qualitätsmerkmale von Leistungskontrollen

3.1 Einleitung und Überblick

Leistungskontrollen (LK) gehören neben der Vorlesung und den praktischen Übungen zu den zentralen Elementen der Hochschullehre. Obwohl LK den kleinsten effektiven Zeitanteil in einer Lehrveranstaltung einnehmen, lässt sich deren massiven Effekt auf den Unterricht in der Praxis immer wieder beobachten: Die Studierenden scheinen mehrheitlich an den prüfungsrelevanten Inhalten interessiert zu sein und fordern von den Dozierenden Transparenz bezüglich Prüfungsinhalten. Dieser Effekt wird noch verstärkt durch die Tatsache, dass die Schlussprüfung meist der einzige promotionsrelevante Faktor darstellt.

Dieses Kapitel wird sich deshalb den Qualitätsmerkmalen von Leistungskontrollen widmen.

In erster Linie stellen LK Messinstrumente dar, welche die kognitiven Leistungen der Studierenden messen. Wie jedes Messinstrument beinhalten auch die LK systematische und zufällige Messfehler, welche die Genauigkeit der Leistungsmessung reduzieren. Am Anfang dieses Kapitels sollen deshalb die Qualitätsmetriken von Leistungskontrollen beschrieben und die Grundlage für die Analyse der im Rahmen dieser Arbeit entwickelten kompetenzorientierten, automatisch korrigierbaren Prüfungsfragen geschaffen werden.

Leistungskontrollen können auch positive Effekte auf den Unterricht ausüben; nämlich dann, wenn der starke Einfluss der LK auf die Lernbereitschaft und den Lernfokus der Studierenden ausgenutzt werden kann. Dies ist in dem Moment besonders erfolgreich, wenn die LK auch als Feedbackinstrument verwendet wird, um den Studierenden realistische Rückmeldungen zu ihren aktuellen Wissensständen zu geben. Im Mittelteil dieser Arbeit soll deshalb der Schwerpunkt auf die qualitativen Aspekte von formativen LK als Feedbackinstrument gelegt werden.

Der Abschluss dieses Kapitels bildet ein kurzer Überblick über die Einsatzmöglichkeiten von CBA-Systemen. Durch das sofortige Feedback, die breite Verfügbarkeit von Daten und den reproduzierbaren Korrekturprozess stellen sie objektives und aus didaktischer

Sicht vielfältiges Messverfahren für kognitive Leistungen dar. Eine Klassifizierung dieses Systems zum Abschluss dieses Kapitels soll zeigen, welche Einsatzgebiete diese Systeme im Praxisalltag haben.

3.2 Klassierung der Leistungskontrollen

Die ETH Zürich definiert eine Leistungskontrolle als "Verfahren, mit dem die Leistung von Studierenden gemessen und bewertet wird" (ETH Zürich, 2012). Neben dieser Selektionsfunktion übernehmen Leistungskontrollen im Unterricht auch weitere Aufgaben. Flechsig unterscheidet insgesamt drei Hauptfunktionen von Leistungskontrollen (Flechsig, 1974). Tabelle 3.4 listet diese zusammen mit ihren Zielsetzungen auf.

1. Rekrutierungsfunktion (Selektionsfunktion)
<ul style="list-style-type: none"> • Platzierung in Kohorte • Selektion aus Kohorte • Qualifikations- und Kompetenznachweis
2. Didaktische Funktion
<ul style="list-style-type: none"> • Zeitliche und inhaltliche Gliederung des Studiengangs • Orientierung der Lehrenden und Lernenden über die Studienziele • Extrinsische Lernmotivation • Diagnoseinstrument • Rückmeldung des Lehr-Lernerfolgs an Lehrende und Lernende
3. Herrschafts- und Sozialisierungsfunktion (Systemreproduktion)
<ul style="list-style-type: none"> • Initiation • Statusverteilung • Legitimation

Tabelle 3.4: Funktionen und Zielsetzungen von Leistungskontrollen (abgeändert nach (Flechsig, 1974)). In der Hochschulpraxis sind primär die ersten beiden Funktionen wichtig.

Die Leistungskontrollen werden in der Praxis anhand ihrer Funktion und des verwendeten Verfahrens klassiert. Die Nachfolgende einfache Klassierung reicht, um die in dieser Arbeit erwähnten LK hinreichend zu beschreiben (Abbildung 3.7).

Die beiden Klassierungen ergänzen sich gegenseitig. Das heisst, eine LK kann immer aus verfahrenstechnischer und aus didaktischer Sicht klassiert werden.



Abbildung 3.7: Die Leistungskontrollen können aus verfahrenstechnischer und didaktischer Sicht klassiert werden. Diese beiden Klassierungen ergänzen sich.

3.2.1 Klassierung aus verfahrenstechnischer Sicht

Bei der Klassierung der LK aus verfahrenstechnischer Sicht werden diese aufgrund des eingesetzten Mediums, der Art der Durchführung oder des angestrebten kognitiven Levels klassiert.

Zu den beiden häufigsten Typen zählen die traditionellen schriftlichen und mündlichen LK, wobei diese in einer Vielzahl verschiedener Varianten durchgeführt werden. Angefangen bei Fliesstexten, psychologischen Fallbeschreibungen, mathematischen Beweisen bis hin zu *Multiple Choice* Fragen, Referaten und mündlichen Prüfungen in Form eines Expertengesprächs.

CBA-Systeme sind hinsichtlich der Fragetypen schriftlichen Prüfungen ähnlich. Durch den Einsatz digitaler Medien und der Möglichkeit sofortiger Korrekturen erfahren sie in den vergangenen Jahren zunehmender Beliebtheit². Verglichen mit den traditionellen Prüfungsformen sind sie jedoch an den meisten Hochschulen noch in der Einführungsphase, da der technische, administrative und personelle Aufwand vergleichsweise hoch ist.

Es wurden aber eine Vielzahl spezialisierter Prüfungsformen entwickelt, die in verschiedenen Wissenschaftsgebieten erfolgreich eingesetzt werden. Dazu zählen beispielsweise *clinical skill*-Leistungskontrollen (CS) in medizinischen Berufen oder Portfolios für praktische Arbeiten und Abschlussarbeiten (Chan & Yan, 2013; Reznick, Blackmore, Dauphinee, Rothman, & Smee, 1996).

² An der ETH Zürich verdoppelte sich die Anzahl der verfügbaren Noten bei CBA Prüfungen vom akademischen Jahr 2011/12 zum akademischen Jahr 2012/13 von ~2300 auf ~5000. Seit 2007 wurden insgesamt ca. 160 CBA-Prüfungen mit 14'00 verfügbaren Noten durchgeführt (Email vom 24.2.2014 von Tobias Halbherr, Zentrum für Lehrentwicklung und -Technologie, ETH Zürich).

3.2.2 Klassierung aus didaktischen Sicht

Bei der Klassierung aus didaktischer Sicht wird zwischen formativen und summativen LK unterschieden.

Eine formative LK ist Bestandteil des Lernprozesses und wird meist nicht benotet. Das Ergebnis einer formativen LK dient den Studierenden zur individuellen Leistungsbestimmung. Zentrales Element von formativen LK ist das Feedback. Die Qualität des Feedbacks beeinflusst dessen Effekt auf den Lernprozess (Peat & Franklin, 2002). Häufige Formen von formativen Leistungskontrollen sind Selbsttests, Übungsaufgaben und Besprechungen mit Lehr- und Assistenzpersonen.

Die summative LK bildet den Abschluss einer Unterrichtssequenz und misst abschliessend die kognitive Leistung von Studierenden. Dieses abschliessende Wissen wird in der Literatur als *Learning Outcome* bezeichnet (James & Brown, 2005; Knight & Yorke, 2003; Scott, 2011). Das Resultat wird in Form einer Note ausgedrückt. Das Feedback ist optional. Die häufigsten Formen von summativen Leistungskontrollen sind mündliche, schriftliche und CBA.

Abbildung 3.8 illustriert den Unterschied zwischen formativen und summativen Leistungskontrollen.

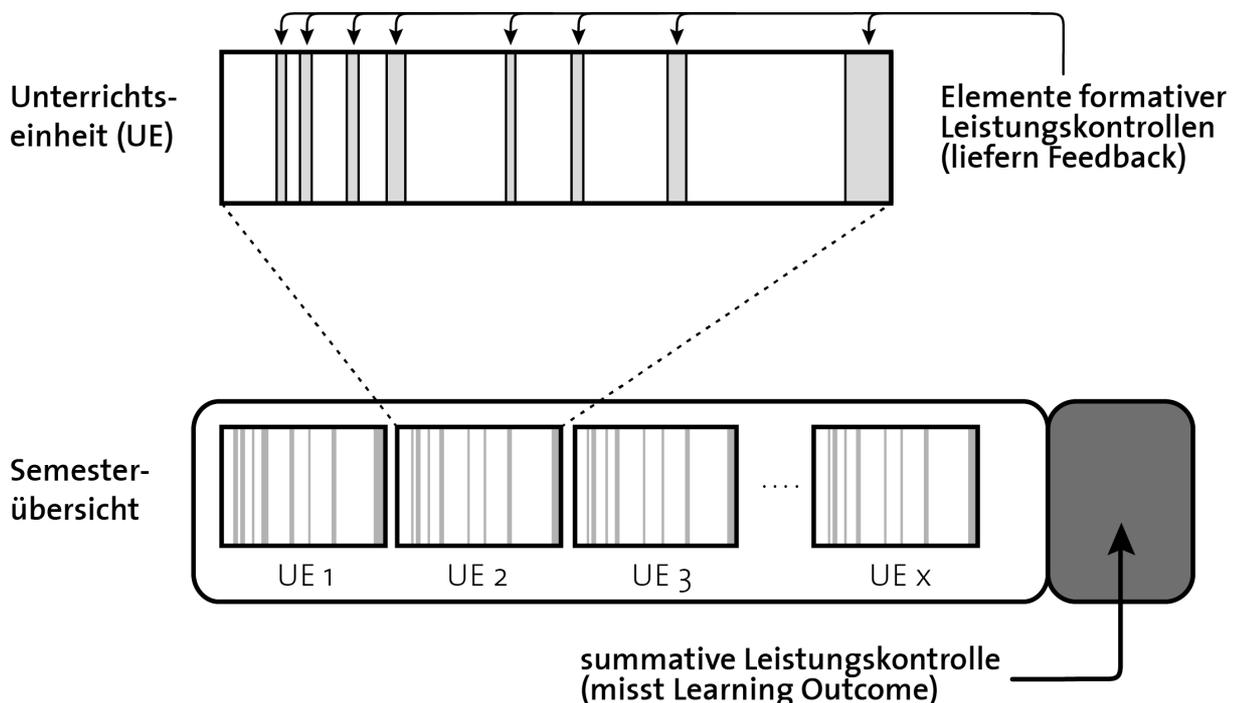


Abbildung 3.8: Unterscheidung zwischen formativen und summativen Leistungskontrollen. Die formativen LK sind Teil einer Unterrichtseinheit (UE) und liefern den Studierenden und Dozierenden Feedback zu den individuellen Leistungsständen. Das Feedback ist ein wichtiges Element im Lernprozess (vgl. dazu Abschnitt 3.4 "Leistungskontrollen als Feedbackinstrument"). Die summative LK bildet den Abschluss des Unterrichts und soll den *Learning Outcome* messen. Er wird benotet und das Feedback ist optional.

3.3 Leistungskontrolle als Messinstrument für kognitive Leistungen

Es gibt kein exaktes Verfahren, um die kognitive Leistung zweifelsfrei feststellen zu können. Alle Verfahren weisen eine systembedingte oder zufällige Messungenauigkeit auf. Diese zu beschreiben ist Inhalt dieses Abschnitts.

3.2.1 Diskrepanz von Messung und Realität

Die in der Praxis gemessenen kognitiven Leistungen sind Annäherungen an die tatsächlichen Leistungen. Abbildung 3.9 zeigt die schematische Darstellung des Messfehlers einer Leistungskontrolle. Je besser die Messung ist, desto näher befinden sich die Messpunkte bei der Regressionslinie. Die Abweichungen der Messpunkte von dieser Linie werden als Messfehler verstanden. Der Messfehler einer Prüfung führt dazu, dass die Leistung gewisser Studierenden überbewertet respektive unterschätzt wird.

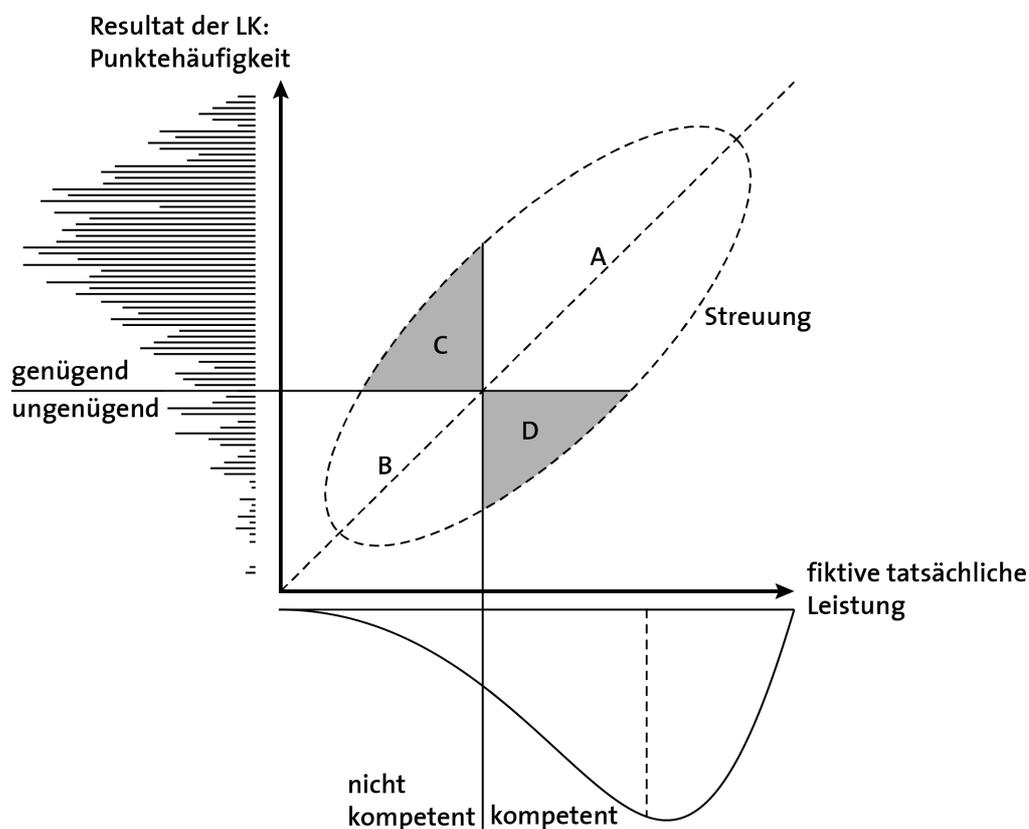


Abbildung 3.9: Schematische Darstellung des Messfehlers einer Leistungskontrolle. Die beiden Achsen zeigen die tatsächliche (x-Achse) und gemessene Leistung (y-Achse). Die Idealmessung ist als gestrichelte Regressionsgerade dargestellt. Je kleiner die Streuung ausfällt, desto besser ist die Messung. Grau eingezeichnet sind die Bereiche der heiklen Messfehler. Die Bereiche A-D werden im Text erläutert. In der Praxis ist die fiktive tatsächliche Leistung nicht bekannt und kann nur mittels Aussenkriterien annähernd bestimmt werden (vgl. Kapitel 3.3.5 "Validität (Gültigkeit der Messung)").

Die Studierenden im Bereich A sind kompetent und haben ein genügendes Resultat an der LK erreicht. Würde diese LK benotet, hätten sie die Prüfung verdientermassen bestanden.

Die Studierenden im Bereich B sind das Gegenteil der Studierenden im Bereich A. Sie sind nicht kompetent und haben entsprechend ein ungenügendes Prüfungsergebnis erzielt.

Bei den Studierenden im Bereich C macht sich der Messfehler unangenehm bemerkbar: Diese Studierenden sind eigentlich nicht kompetent, dennoch haben sie ein genügendes Ergebnis an der LK erzielt. Aufgrund dieses vermeintlich positiven Resultats wird ihre Leistung überschätzt.

Für die Studierenden im Bereich D hat der Messfehler unter Umständen massive Auswirkungen. Sie würden die kognitiven Lernziele dieses Kurses eigentlich erfüllen, dennoch erreichen sie nur ein ungenügendes Resultat an der LK. Ihre Leistung wird also unterschätzt. Das wird als unfair empfunden und kann zu Verunsicherung und Frustration führen.

In der Praxis steht das Mass der tatsächlichen Leistung nicht zur Verfügung. Entsprechend kann der Messfehler auch nicht direkt ermittelt werden. Die tatsächliche Leistung kann jedoch annähernd mit Aussenkriterien geschätzt und so die Qualität der eigentlichen Leistungskontrolle abgeschätzt werden. Auf diese Methode werden wir im Abschnitt 3.3.5 "Validität (Gültigkeit der Messung)" vertiefter eingehen. Angewandt wird sie dann im Kapitel 7, wo die im Rahmen dieser Arbeit entwickelten automatisch korrigierbaren Prüfungsfragen evaluiert werden.

3.3.2 Qualitätsmetriken von Leistungskontrollen

Es stellt sich nun die Frage, welche Faktoren eine gute Prüfung auszeichnen. In der Literatur findet man drei Gütekriterien einer Leistungskontrolle (Lienert & Raatz, 1998):

- Die Leistungskontrolle muss die Leistung der Studierenden objektiv, das heisst wert- und vorurteilsfrei messen.
- Die Resultate der Leistungskontrolle müssen verlässlich (reliabel) sein und auch bei wiederholten Messungen ein vergleichbares Resultat liefern.
- Die aus den Resultaten abgeleiteten Bewertungen müssen gültig (valide) sein, das heisst, sie müssen in einem direkten Zusammenhang mit dem stehen, was zu messen beabsichtigt war.

Neben den drei Gütekriterien erwähnen Lienert und Raatz mit den beiden Nebenkriterien "Ökonomie" und "Normierung" zwei weitere wichtige Prämissen für eine gute Leistungskontrolle. Der Ökonomie kommt in der Hochschullehre speziell bei grossen Studierendenzahlen eine zentrale Bedeutung zu, während die Normierung der Inhalte der Leistungskontrolle in Beschwerdefällen und in der Hochschulentwicklung Beachtung geschenkt werden sollte. In Abbildung 3.10 sind die Haupt- und zwei Nebenkriterien der Güte einer Leistungskontrolle illustriert.

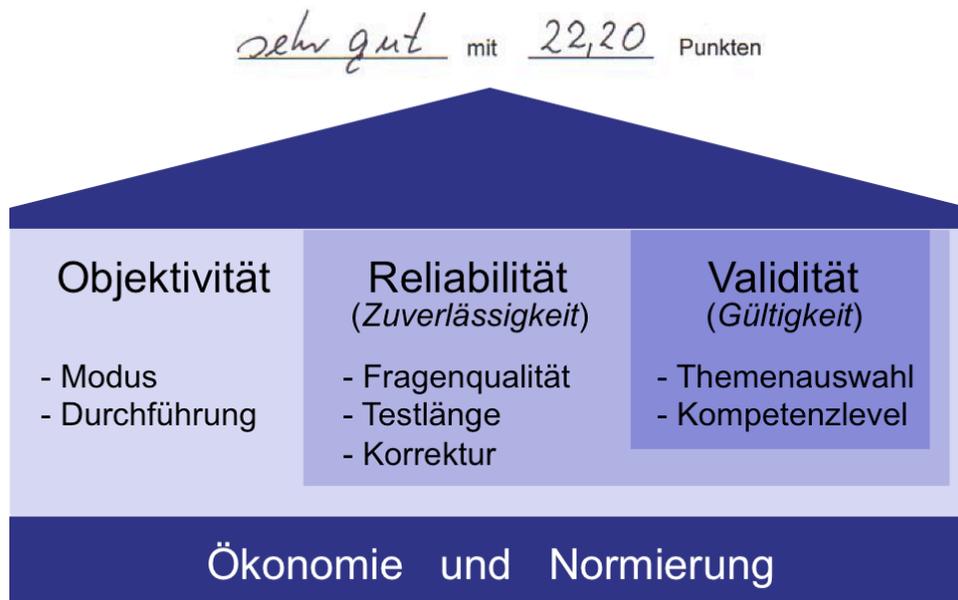


Abbildung 3.10: Die drei Hauptgütekriterien und zwei Nebenkriterien beeinflussen die Qualität des Resultats einer Leistungskontrolle. Diese ist abhängig von drei Hauptgütekriterien: Objektivität, Reliabilität und Validität. Wobei erstere meist durch den Modus und die anderen beiden durch den Inhalt der Leistungskontrolle beeinflusst werden.

Die drei Hauptgütekriterien sind einseitig voneinander abhängig: Objektivität beeinflusst Reliabilität beeinflusst Validität. Das heisst, je höher die Objektivität einer Leistungskontrolle, desto höher ist auch die Reliabilitäts- und Validitätserwartung. Daraus folgt, dass valide Leistungsmessungen auch reliabel und objektiv sind.

Nachfolgend soll aufgezeigt werden, wie die Qualität der drei Hauptgütekriterien in der Praxis festgestellt werden können.

3.3.3 Objektivität

Eine Leistungskontrolle ist dann objektiv, wenn die Korrektur auch bei unterschiedlichen Korrekturpersonen übereinstimmende Resultate liefert.

Die Prüfungsmethoden messen bezüglich der Hauptgütekriterien von LK unterschiedlich gut. Tabelle 3.5 zeigt eine pauschalisierte Einstufung der Erfassungsqualität. Daraus lässt sich ableiten, dass CBA allgemein die höchste Objektivität zugeschrieben wird. Das hängt primär damit zusammen, weil die Korrektur automatisiert abläuft und bei wiederholten Korrekturvorgängen identische Resultate liefert.

	Schriftliche Prüfung	Mündliche Prüfung	Computer-basierte Prüfung
Validität	* bis ***	* bis ***	* bis ***
Reliabilität	**	*	**
Objektivität	**	*	***

Tabelle 3.5: Die verschiedenen Prüfungsmethoden messen bezüglich der Gütekriterien unterschiedlich gut, was zu mehr oder weniger objektiven oder reliablen Prüfungen führt (abgeändert nach Krebs, 1999). * = geringe Güte, ** = mittlere Güte, *** = hohe Güte. Da die Validität primär vom Inhalt der LK abhängig ist, kann hier keine pauschale Aussage gemacht werden.

Die Objektivität einer LK wird im Hochschulalltag grösstenteils durch die Wahl der Methode festgelegt und kann durch die Prüfungsleitung nicht oder nur minimal beeinflusst werden. Ausgenommen sind mündliche Prüfungen, die allgemein eher eine tiefere Objektivität aufweisen und bei denen mittels geeigneter Kriterien- oder Checklisten die Objektivität noch günstig beeinflusst werden kann.

3.3.4 Reliabilität (Zuverlässigkeit der Messung)

Eine Leistungskontrolle ist dann reliabel, wenn sie bei wiederholter Messung dieselben zuverlässigen Resultate liefert. Die Reliabilität lässt sich anhand des Fiebermessers erläutern: dieser liefert auch bei wiederholter Messung immer sehr ähnliche, reliable Ergebnisse.

In der Praxis lässt sich die Reliabilität von LK nicht durch Wiederholung der LK feststellen, da mit jeder Durchführung der LK das Wissen der Studierenden verändert wird, was massgeblichen Einfluss auf die Beantwortung der einzelnen Fragen hat.

Zur Bestimmung der Reliabilität von LK werden in der Praxis häufig Metriken der Konsistenzanalyse beigezogen. Untersucht wird, wie gut die interne Konsistenz einer Prüfung ist, das heisst, wie viele Widersprüche die Prüfungsergebnisse beinhalten. Ein Widerspruch beispielsweise liegt dann vor, wenn eine Aufgabe von den schwachen, nicht aber von den guten Studierenden korrekt gelöst wird. Die von den Studierenden mit dieser Aufgabe erzielten Punkte widerspiegeln dann nämlich nicht das Schlussergebnis dieser LK.

Um die Reliabilität einer LK in der Praxis festzustellen, können in der Praxis die Schwierigkeit, die Trennschärfe und die interne Konsistenz mit dem Alpha-Koeffizienten bestimmt werden. Tabelle 3.6 zeigt eine Übersicht über die drei Metriken zur Bestimmung der Reliabilität einer LK.

Metrik	Wertebereich	Richtwerte	Quellenangabe
Schwierigkeitsindex (P)	0 ≤ P ≤ 100	Schwierige Frage: P < 20 Eisbrecherfrage: P ≥ 95	(Krebs, 2005)
		Minimalstreuung 20 < P < 80 (abhängig von Differenzierungsziel)	(Lienert & Raatz, 1998, p. 115)
Trennschärfe (r)	-1.00 ≤ r ≤ 1.00	Mindestanforderung r ≥ 0.2 (r ist abhängig von der Testgrösse)	(Hottinger, Krebs, Hofer, Feller, & Bloch, 2004) r in Abhängigkeit der Testgrösse: (Krebs, 2005)
		Gute Frage: r > 0.3 Akzeptabel: 0.2 > r > 0.3 Marginal: 0.1 > r > 0.2 Schlecht: 0.0 > r > 0.1 Schädlich: r < 0.0	(Möltner, Schellberg, & Jünger, 2006)
Alpha (α)	-1.00 ≤ α ≤ 1.00 (wobei nur α > 0 sinnvoll interpretiert werden kann)	Mindestanforderung α > 0.8	(Hottinger et al., 2004)
		Akzeptabler Bereich: 0.70-0.95 Idealerweise jedoch nicht >0.90 (da Redundanzen in den Fragen wahrscheinlich)	(Tavakol & Dennick, 2011)

Tabelle 3.6: Metriken zur Bestimmung der Reliabilität einer LK. Diese Werte können anhand der Prüfungsergebnisse berechnet werden.

Schwierigkeitsindex (P)

Der Schwierigkeitsindex ("P") gibt an, von wie vielen Probanden diese Aufgabe korrekt beantwortet wurde. Dieser prozentuale Wert variiert von 0 bis 100. Je höher der Wert ist, desto leichter ist diese Aufgabe.

In der Praxis wird der Schwierigkeitsindex oft mit der Formel

$$P = 100 \times \frac{N_R}{N_B}$$

berechnet.

Dabei gilt:

N_R = Anzahl der Studierenden, welche die Frage richtig beantwortet haben

N_B = Anzahl der Studierenden, welche diese Aufgabe bearbeitet haben

Diese Formel sollte jedoch nur verwendet werden, wenn zufällig korrekt abgegebene Antworten sehr unwahrscheinlich sind oder ausgeschlossen werden können (wie dies beispielsweise bei offenen Fragenformaten der Fall ist). In allen anderen Fällen, dazu gehören beispielsweise die klassischen *Multiple Choice* (MC) Fragenformate, sollte die von Lienert und Raatz vorgeschlagene *zufallskorrigierte* Formel zur Berechnung des Schwierigkeitsindex verwendet werden (Lienert & Raatz, 1998, p. 76):

$$P = 100 \frac{N_R - \frac{N_F}{m-1}}{N_B}$$

Dabei gilt:

N_F = Anzahl der Studierenden, die die Aufgabe falsch beantwortet haben

m = Anzahl der Antwortmöglichkeiten der Aufgabe (z.B. 4 beim Fragetyp Kprim)

Gelegentlich können sich durch die Zufallskorrektur bei MC-Fragen negative Schwierigkeitsindizes auftreten, die jedoch nicht interpretiert werden können. Gründe dafür sind entweder ein Fehler in den Musterlösung oder falsch formulierte Ablenker, welche von der korrekten Antwort ablenken.

Trennschärfe (r)

Die Trennschärfe (r) oder genauer der Trennschärfekoeffizient einer Aufgabe gibt an, wie gut die Frage zwischen "guten" und "schwachen" Studierenden zu trennen vermag.

Eine hohe Trennschärfe besagt, dass die guten Studierenden diese Frage richtig und die schwachen Studierenden falsch oder nicht beantwortet haben. Eine Frage mit Trennschärfe um null wurde von ungefähr gleich vielen guten wie schlechten Studierenden korrekt beantwortet. Derartige Fragen sind unbrauchbar. Bei Fragen mit negativen Trennschärfen handelt es sich um "Fangfragen" die mehrheitlich von schwachen Studierenden korrekt und von starken Studierenden falsch beantwortet wurden. Diese Fragen verfälschen das Schlussergebnis und sollten noch vor der Benotung eliminiert werden.

Es gibt verschiedene Methoden, um die Trennschärfe zu berechnen. Die am weitesten verbreitete Methode ist die Berechnung des Korrelationskoeffizienten zwischen der Aufgabenpunktzahl (meist biserial: 0/1, richtig/falsch) und der Summe der Gesamtpunktzahl respektive von Subgruppen (bei LK mit heterogenem Inhalt).

Der Trennschärfekorrelationskoeffizient wird in der Literatur üblicherweise der Korrelationskoeffizient r nach Pearson-Bravais vorgeschlagen (Möltner et al., 2006). Korreliert wird dabei die Punkteverteilung der Aufgabe mit Gesamtpunktzahl der LK.

In der Regel ist die Gesamtpunktzahl annähernd normalverteilt (meist mit einer leichten Verschiebung in Richtung der Maximalpunktzahl). Weicht diese jedoch stark von einer Normalverteilung ab oder verfügt sie über viele Ausreisser, dann empfehlen Möltner und seine Kollegen die Verwendung eines nicht-parametrischen Rangkorrelationskoeffizienten (z.B. Spearman).

Die Trennschärfe ist abhängig von der Schwierigkeit der Aufgabe. Zwischen den beiden Metriken besteht eine paraboloidische Abhängigkeit (Abbildung 3.11). Das heisst, bei mittlerer Schwierigkeit sind höhere Trennschärfen zu erwarten als bei schwereren oder leichteren Aufgaben.

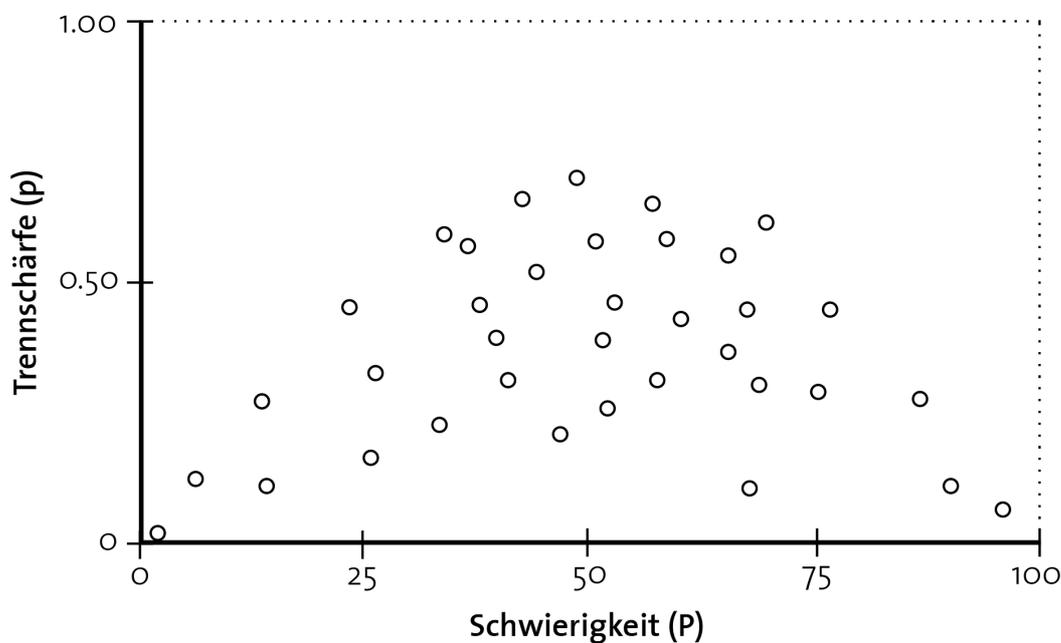


Abbildung 3.11: Abhängigkeit der Trennschärfe von der Aufgabenschwierigkeit (abgeändert nach Lienert & Raatz, 1998).

Lienert und seine Kollegen schlagen deshalb vor, bei der Berechnung der Trennschärfe eine "Schwierigkeitskorrektur" vorzunehmen (Lienert & Raatz, 1998, p. 91ff). Die paraboloidische Abhängigkeit verschwindet und die Trennschärfe kann dann auch bei schwierigen oder leichten Fragen hoch sein. Dies macht Sinn, denn eine schwierige Frage, die primär von guten Studierenden korrekt gelöst werden kann, weist tatsächlich eine hohe Trennschärfe auf (vgl. Linie F in Abbildung 3.12).

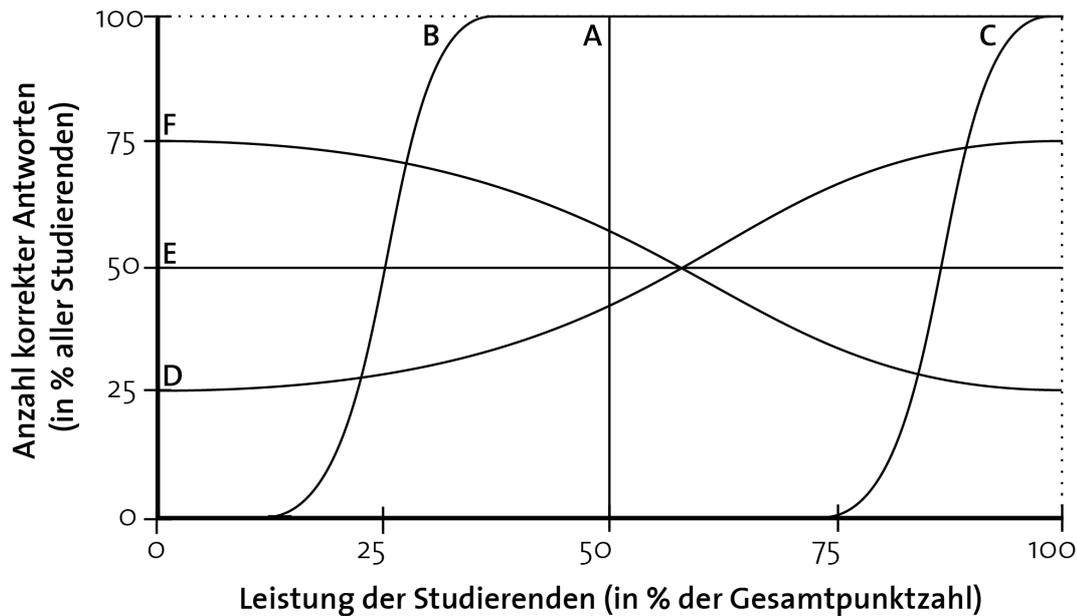


Abbildung 3.12: Schematische Darstellung der schwierigkeitskorrigierten Trennschärfe durch den Vergleich der korrekt abgegebenen Antwort mit der Leistung der Studierenden. Die Fläche unter der Linie stellt die Schwierigkeit (P) dar, die Steigung entspricht der schwierigkeitskorrigierten Trennschärfe (r_{corr}). Die mit A-G markierten Linien stellen unterschiedlich schwierige und trennscharfe Aufgaben dar. Weitere Informationen finden sich im Text (abgeändert nach Lienert & Raatz, 1998).

Aufgabe A ist ideal trennscharf. Die unterdurchschnittlichen Studierenden konnten die Aufgabe nicht lösen, während alle überdurchschnittlichen Studierenden diese Aufgabe lösen konnten. Eine derartig trennscharfe Frage gibt es in der Praxis leider nicht.

Aufgabe B hat eine gute Trennschärfe ($r_{corr} \approx 0.8$). Sie vermag zuverlässig die schwachen Studierenden vom Rest zu trennen. In der Praxis wäre das eine gute Aufgabe, um die Studierenden rund um die Bestehensgrenze zu trennen.

Aufgabe C hat eine gute Trennschärfe ($r_{corr} \approx 0.8$). Aufgrund der hohen Schwierigkeit dieser Aufgabe würde sie sich in der Praxis gut eignen, um die guten von den sehr guten Studierenden zu unterscheiden.

Aufgabe D weist eine geringe Trennschärfe auf ($r_{corr} \approx 0.1$). Sowohl schwache als auch starke Studierende haben diese Aufgabe korrekt beantwortet. Diese Art von Fragen lassen sich in der Praxis relativ häufig beobachten.

Aufgabe E ist nutzlos. Sie vermag nicht zwischen schwachen und starken Studierenden zu unterscheiden ($r_{corr} = 0.0$), obwohl sie mit dem Schwierigkeitsindex von 50 ideale Voraussetzungen dafür besäße.

Aufgabe F wird von den schwachen Studierenden besser gelöst als von den stärkeren. Sie besitzt eine negative Trennschärfe und verfälscht somit das Endergebnis der LK.

Alpha (α)

Zur Bestimmung der Reliabilität der gesamten LK kann anhand der Prüfungsergebnisse auch der α -Koeffizient nach Cronbach gemessen werden (Cronbach, 1951). Dieser liefert einen Wert über die Kohärenz aller Aufgaben in Bezug auf das Schlussresultat. Dieser Wert variiert von -1 bis 1, wobei negative Werte auf einen Fehler bei der Korrektur schliessen. Je höher α ist, desto höher ist die interne Konsistenz und somit die Reliabilität der LK.

Die Reliabilität einer LK hängt von der Anzahl und der Qualität der Fragen, der verwendeten Fragetypen sowie der Zuverlässigkeit der Korrektur ab.

Der Effekt der Testlänge, das heisst der Anzahl gestellter Fragen, wird unter anderem von Lienert und seinen Kollegen beschrieben (Lienert & Raatz, 1998, p. 209ff). Allerdings gilt es hierbei zu beachten, dass durch eine Testverlängerung auch die Anzahl der Störvariablen zunehmen (Ermüdung, Verspannungen etc) welche die Reliabilität negativ beeinflussen. Zudem steigt die Gefahr von Redundanzen in den abgefragten Themen, wodurch zwar der α -Koeffizient erhöht wird, die Validität aber nicht ansteigt.

Der gewählte Fragetyp hat einen grossen Einfluss auf die Korrektur oder den Ratefaktor. Auf diesen Aspekt wird im Kapitel 7 detaillierter eingegangen.

3.3.5 Validität (Gültigkeit der Messung)

Die Validität ist in der Praxis der bedeutendste Aspekt einer guten Leistungskontrolle, gleichzeitig ist sie aber auch am schwierigsten zu messen. Sie setzt eine objektive und reliable Leistungsmessung voraus. Weisen diese mangelhafte Metriken auf, so ist die LK auch nicht valide. Allerdings kann aufgrund guter Messmetriken bezüglich der Objektivität und der Reliabilität nicht auf eine valide LK geschlossen werden.

Die Validität einer LK kann nicht wie die Reliabilität alleine aufgrund der Resultate berechnet werden. Vielmehr werden Aussenkriterien benötigt, um bestimmen zu können, ob mit einer LK auch tatsächlich das gemessen wurde, was beabsichtigt war (Abbildung 3.13).

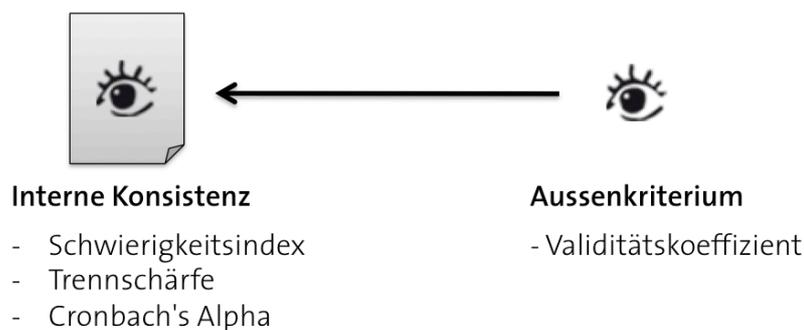


Abbildung 3.13: Qualitätsmetriken der Leistungskontrolle im Überblick. Für die Messung der Validität werden Aussenkriterien benötigt, die mit den Ergebnissen der LK verglichen werden.

Zwei Arten der Validität sind bei Leistungsmessungen in der Hochschullehre von zentraler Bedeutung: die inhaltliche und kriterienbezogene Validität.

Inhaltliche Validität

Bei der Überprüfung der inhaltlichen Validität wird der Inhalt einer LK mit den Lernzielen abgeglichen. Für jede Frage wird das Lernziel und die kognitive Prozessdimension bestimmt und in eine Lernzielmatrix, einen sogenannten "Blueprint", eingetragen (Newble, Hoare, & Elmslie, 1981). Mit Hilfe des Blueprints kann dann abgeschätzt werden, ob die im Unterricht vermittelten Inhalte auch in einem angemessenen Umfang und auf korrekter Kompetenzstufe geprüft werden (Tabelle 3.7).

<i>Fragenebene: Dimension Allgemeinmedizin</i>	<i>%</i>	<i>Dimension Fachwissen Allgemeinmedizin</i>	<i>%</i>
<i>Allgemeinmedizinisches Fachwissen</i>	<i>40</i>	<i>Chirurgie</i>	<i>10</i>
<i>Konsultation:</i>	<i>35</i>	<i>Innere Medizin</i>	<i>30</i>
<i>Anamnese</i>	<i>10</i>	<i>Arbeitsmedizin</i>	<i>2</i>
<i>Körperliche Untersuchungen</i>	<i>8</i>	<i>Dermatologie</i>	<i>5</i>
<i>Zusatzuntersuchungen</i>	<i>6</i>	<i>Gynäkologie/Geburtshilfe</i>	<i>6</i>
<i>Therapie, Management,</i>		<i>Neurologie</i>	<i>7</i>
<i>Entscheidungen</i>	<i>8</i>	<i>Ophthalmologie</i>	<i>4</i>
<i>Administration</i>	<i>3</i>	<i>Otorhinolaryngologie</i>	<i>7</i>
<i>Interaktion, Kommunikation,</i>		<i>Pädiatrie</i>	<i>8</i>
<i>Arzt-Patient-Beziehung</i>	<i>15</i>	<i>Psychiatrie</i>	<i>9</i>
<i>Ethik</i>	<i>5</i>	<i>Medizinische Radiologie (Radiodiagnostik)</i>	<i>5</i>
<i>Recht</i>	<i>5</i>	<i>Rheumatologie</i>	<i>7</i>
<i>TOTAL</i>	<i>100</i>	<i>TOTAL</i>	<i>100</i>
<i>Fallebene: Dimension Alter</i>	<i>%</i>	<i>Dimension Konsultationsart</i>	<i>%</i>
<i>Säuglinge (0-1)</i>	<i>5</i>	<i>Sprechstunde</i>	<i>70</i>
<i>Kinder (1-11)</i>	<i>10</i>	<i>Hausbesuch</i>	<i>10</i>
<i>Adoleszente (12-18)</i>	<i>10</i>	<i>Notfall</i>	<i>5</i>
<i>Erwachsene (19-64)</i>	<i>45</i>	<i>Telefon</i>	<i>10</i>
<i>ältere Menschen (>64)</i>	<i>30</i>	<i>Indirekte Konsultation (Akten, Gutachten)</i>	<i>5</i>
<i>TOTAL</i>	<i>100</i>	<i>TOTAL</i>	<i>100</i>

Tabelle 3.7: vierdimensionaler Blueprint zur Überprüfung der Validität der Schweizerischen Facharztprüfung "Allgemeinmedizin" (aus Krebs, 1999). Jede Prüfungsfrage wird genau einer Kategorie pro Dimension zugeteilt. Die prozentualen Angaben dienen als Zielvorgabe für die Gewichtung des Inhalts der LK. Die kognitive Kompetenzstufe ist in diesem Blueprint nicht berücksichtigt.

Da die Bestimmung der inhaltlichen Validität auf (subjektiven) Expertenurteilen basiert, besteht gemäss Lienert die Gefahr einer Übergeneralisierung (Lienert & Raatz, 1998, p. 225ff). So kann beispielsweise nicht von einer Programmierkompetenz gesprochen werden, nur weil ein Student die Datentypen aufzählen kann. Wenn bei der Zusammenstellung der LK im Blueprint die kognitive Prozessdimensionen berücksichtigt werden, ist die Gefahr der Übergeneralisierung um einiges geringer (Klauer, 1987).

Kriterienbezogene Validität

Die Bestimmung der kriterienbezogenen Validität geschieht vorwiegend empirisch mittels Korrelationen zu Aussenkriterien. Die Aussenkriterien werden dabei so gewählt, dass sie möglichst vergleichbare Eigenschaften messen sollen. So können beispielsweise Selbstbeurteilungen, Expertenbewertungen oder Prüfungsergebnisse eines ähnlichen Faches als Aussenkriterien verwendet werden. Korreliert das Resultat der LK mit einem oder idealerweise mehreren Aussenkriterien, kann dies Indiz für eine valide Messung sein.

Im Kapitel 7 werden wir die im Rahmen dieser Arbeit entwickelten automatisch korrigierbaren Prüfungsfragen bezüglich der Reliabilität und Validität mit den hier vorgestellten Verfahren empirisch untersuchen.

3.4 Leistungskontrollen als Feedbackinstrument

Feedback spielt in der Lehre eine zentrale Rolle (Schulmeister, 2007). Es beeinflusst die Leistung, die Motivation und steigert das Bewusstsein über die Selbstwirksamkeit der lehrbezogenen Aktivitäten von Studierenden. Einige Studien zeigen gar, dass Feedbacks zu den effektivsten didaktischen Mitteln gehören, um die Leistung von Studierenden zu steigern (Marzano, 2007). Auch schwächere Studierende mit weniger Vorwissen können vom Feedback profitieren (Fuchs & Fuchs, 1986). Dieser Aspekt ist speziell in der Grundlagenausbildung der Hochschullehre wichtig, da das Vorwissen der Studierenden oft sehr heterogen ist und die Gefahr besteht, die Chancengleichheit der Studierenden nicht zu wahren. Hier kann gutes Feedback helfen, die Streuung des Wissensspektrums zu reduzieren und den *Learning Outcome* generell zu erhöhen.

Gutes Feedback zeichnet sich dadurch aus, dass es nachvollziehbar, das heisst valide ist und unmittelbar, individualisiert, konstruktiv (Bacon, 2011) und spezifisch (C. Miller, Doering, & Scharber, 2010) ausfällt. Zudem sollen durch das Feedback die Lernziele und die Erwartungen an die Studierenden klar werden (Marzano, 2007). Letzteres ist ein sehr effektives Instrument zur Steuerung der Lernaktivitäten. Welche Rolle dabei die Vermittlung des *hidden Curriculums* - der prüfungsrelevanten Lernziele - spielt, wird im Abschnitt 3.4.2 "Feedback zum Lerninhalt: LK zur Vermittlung des hidden Curriculums" erläutert.

Feedback kann auf verschiedene Arten gegeben werden. Zu den in der Literatur am häufigsten genannten Feedbackarten zählen: selber oder von Tutoren korrigierte Übungen, mündliches Expertenfeedback, Peer-Feedback von Mitstudierenden, automatisiertes Feedback und Leistungskontrollen (Herding, 2013).

Das Feedback der formativen Leistungskontrollen nimmt eine Sonderrolle ein, da sie den summativen LK meist ziemlich ähnlich sind. Entsprechend hoch ist die extrinsische Motivation der Studierenden, diese gewissenhaft zu bearbeiten und das Feedback zu

studieren, um Ihren Leistungsstand mit den Zielen und Erwartungen der Dozierenden abzugleichen.

Neben dem Feedback für Studierende liefert die Leistungskontrolle aber auch wichtige Informationen für Dozierende. In Abbildung 3.14 sind die drei Komponenten des Feedbacks von formativen LK dargestellt.

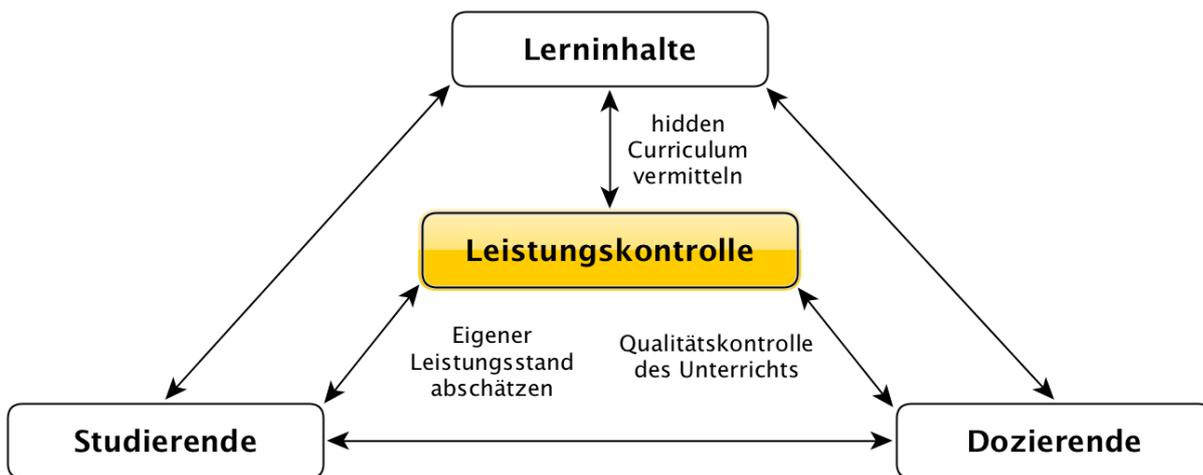


Abbildung 3.14: Das erweiterte didaktische Dreieck zeigt die formative Leistungskontrolle als Feedbackelement bezüglich der drei Komponenten: Studierende, Lerninhalte und Dozierende.

Nachfolgend werden die qualitativen Aspekte des Feedbacks näher beschrieben.

3.4.1 Feedback aus studentischer Sicht: eigener Wissenstand beurteilen

Das Feedback der formativen LK hilft den Studierenden ihren persönlichen Leistungsstand abzuschätzen. Dieses meta-kognitive Wissen führt dazu, dass sie ihre Fähigkeiten und Fertigkeiten besser beurteilen können, anspruchsvollere Aufgaben wählen und so den *Learning Outcome* erhöhen (Ames & Archer, 1988; Dweck, 1986; Gibbs & Simpson, 2004). So hielt beispielsweise Dweck in Ihrer Arbeit folgendes fest: "*confidence in their current ability must be high and must remain high if they are to choose appropriately challenging tasks and pursue them in effective ways*".

Beim Feedback spielen quantitative Faktoren eine wichtige Rolle. So wurde beispielsweise festgestellt, dass die Häufigkeit von Feedbacks einen positiven Einfluss auf den *Learning Outcome* hat (Bangert-Drowns, Kulik, Kulik, & Morgan, 1991; Fuchs & Fuchs, 1986). Zudem ist der Effekt grösser, wenn die Studierenden das Feedback unmittelbar nach Abschluss der formativen LK erhalten (Butler & Roediger, 2008).

Den positiven Effekt des Feedbacks von formativen Leistungskontrollen konnte im eigenen Unterricht festgestellt werden. Wir führen seit mehreren Jahren formative, unbenotete, aber korrigierte Semester-Endtests durch. Wie sich gezeigt hat, erachten nur 4% der Studierenden die formative LK als mehrheitlich nutzlos für die Prüfungsvorbereitung (Diagramm 3.1).

Ich erachte den Semesterend-Test als nutzlos für die Prüfungsvorbereitung und er könnte somit ersatzlos gestrichen werden.

trifft voll zu	2 (1%)	
trifft zu	8 (3%)	
trifft teilweise zu	20 (8%)	■
trifft weniger zu	43 (17%)	■
trifft nicht zu	167 (67%)	■
keine Antwort	11 (4%)	

Diagramm 3.1: Frage nach der Sinnhaftigkeit des formativen, unbenoteten Semester-Endtests. Befragt wurden die Studierenden direkt nach dem Semester-Endtest (Nachumfrage "Informatik (Biologie, Pharm. Wiss & HEST)", Frühlingsemester 2012, n = 251).

89% der Studierenden konnten vom automatischen, individuellen Korrekturhinweis profitieren und 74% der Studierenden gaben an, dass Sie aufgrund der formativen LK ihren Wissensstand besser abschätzen konnten³. Diese Evaluationsergebnisse decken sich mit jenen aus früheren Jahren oder den Resultaten anderer, ebenfalls eigener Lehrveranstaltungen.

Unsere Zahlen stützen demnach auch die eingangs erwähnte These, dass Studierende vom Feedback der formativen Leistungskontrollen profitieren und ihren individuellen Leistungsstand besser abschätzen können.

3.4.2 Feedback zum Lerninhalt: LK zur Vermittlung des hidden Curriculums

Ein mächtiges didaktisches Werkzeug zur Steuerung der studentischen Lernaktivitäten ist das Kommunizieren der Lernziele. Dabei gilt es, zwischen den scheinbaren (*formal curriculum*) und den echten Lernzielen (*hidden curriculum*) zu unterscheiden (C. M. L. Miller & Parlett, 1974; B. R. Snyder, 1971).

Das *formale Curriculum* wird den Studierenden transparent kommuniziert und ist oft Bestandteil der Beschreibung einer Lehrveranstaltung. Es listet alle behandelten Themengebiete auf und gibt idealerweise auch die kognitive Prozessdimension mit an (Bsp: "Die Studierenden können die fünf primitiven Datentypen der Java Programmiersprache aufzählen"). Die Information aus dem formalen Curriculum gibt den Studierenden eine Idee, was sie in dieser Lehrveranstaltung erwartet. Im Praxisalltag lässt sich allerdings immer wieder beobachten, dass Dozierende dieses formale Curriculum bewusst flexibel gestalten und die Prioritäten meist erst im Unterricht unter Berücksichtigung der Heterogenität und der Wünsche der Studierenden festlegen.

Deshalb ist für Studierende das formale Curriculum für die Vorbereitung der summativen LK unbedeutend. Für sie ist es jedoch entscheidend, das *hidden Curriculum* zu ken-

³ Beiden Werte stammten aus der Nachumfrage "Informatik (Biologie, Pharm. Wiss & HEST)", Frühlingsemester 2012, n = 251. Die Rohdaten finden sich im Anhang A.1 dieser Arbeit.

nen. Das *hidden Curriculum* umfasst die prüfungsrelevanten Themen und deren kognitive Prozessdimensionen (B. R. Snyder, 1971). Es existiert nicht in formaler, schriftlicher Form, sondern das Wissen über das *hidden Curriculum* wird von den Studierenden ganz selbstverständlich aufgebaut. Sie kombinieren dabei die Informationen, die sie bezüglich der summativen LK in Erfahrung bringen können. Dazu gehören unter anderem die besprochenen Übungen, Vorlesungen, Kommentare von Dozierenden und Tutoratspersonen, Beispielprüfungen und das Feedback aus formativen LK. Letzteres spielt deshalb eine wichtige Rolle, weil es, wie bereits erwähnt, für die Studierenden relativ transparent die Lernziele und kognitiven Anforderungen aufzeigt.

Wie wir in unserem Unterricht messen konnten, führten die formativen LK zur Repetition gewisser Unterrichtsinhalte. So gaben 76% der Studierenden an, dass sie aufgrund der freiwilligen Probeproofung gewisse Inhalte noch einmal gezielt repetiert haben⁴.

Es kann festgehalten werden, dass die Dozierenden mit den formativen LK ein mächtiges didaktisches Werkzeug zur Verfügung haben, um das *hidden Curriculum* den Studierenden zu vermitteln und damit ihre Lernaktivitäten zu beeinflussen.

3.4.3 Feedback aus Sicht der Dozierenden: Qualitätskontrolle

Die Entwicklung einer Lernumgebung ist anspruchsvoll. Meist sind mehrere Iterationen und das Feedback vieler Studierenden nötig, um aus dem ersten Entwurf eine funktionierende Lernumgebung zu gestalten. In den siebziger Jahren, als die Professionalisierung der Hochschullehre ihren Anfang nahm, war man der Überzeugung, dass die Optimierung der Lernumgebung geprägt sei von einer guten Intuition von Dozierenden (R. C. Anderson, 1961). Zweifellos hat dies bis heute die Gültigkeit behalten und zeichnet gute Dozierende aus. Allerdings sind die Anforderungen an eine moderne Lernumgebung seither massiv gestiegen, so dass Intuition meist nicht mehr ausreicht. Einige Lehrveranstaltungen, meist solche im Grundlagenunterricht mit hohen Studierendenzahlen und heterogenem Vorwissen, entwickeln sich vom klassischen Frontal- und Gruppenunterricht zu *blended learning* Umgebungen. Diese zeichnen sich dadurch aus, dass sie neben den Präsenzzeiten vermehrt auch Selbststudiumsequenzen unter Verwendung von elektronischen und analogen Lernmedien beinhalten. Das Zeitverhältnis wird dabei oft zugunsten der selbständigen Arbeit verschoben (L. E. Fässler, 2007, p. 192).

Zudem stehen den Dozierenden eine breite Palette von elektronischen Tools zur Verfügung, mit denen der Unterricht angereichert werden kann. So können sie beispielsweise Abstimmungen in Echtzeit in der Vorlesung machen⁵ und sie aufzeichnen lassen, so

⁴ Dieser Wert stammt aus der Nachumfrage "Einsatz von Informatikmitteln", Herbstsemester 2011, n=67. Die Rohdaten finden sich im Anhang A.2 dieser Arbeit.

⁵ Einsatz von "Clicker" im Unterricht: <http://www.let.ethz.ch/projekte/closed/eduApp> (abgerufen am: 23.05.2014)

dass Studierende diese orts- und zeitunabhängig auf ihren mobilen Geräten anschauen können. Weiter kann die Kommunikation mit den Studierenden ausserhalb des Präsenzunterrichts auf einem Learning Management System (LMS) oder einem Forum weitergeführt und traditionelle Prüfungen durch CBA ersetzt werden.

Die Optimierung einer Lernumgebung ist mit Aufwand verbunden. Hier gilt es den Implementationsaufwand abzuschätzen und mögliche Wechsel- oder Nebenwirkungen mit der bestehenden Lerneinheit zu erkennen. Black und William vergleichen eine derartig komplexe Lernumgebung als "Black Box" (Black & William, 1998b). In ihrem systemischen Modell erfährt die Black Box einen Initialinput in Form von Lernmaterialien, Rahmenbedingungen, Anzahl der beteiligten Personen etc. Während des Unterrichts wirken verschiedenste Einflüsse auf diese Black Box ein und am Ende resultiert ein Output in Form eines *Learning Outcomes* (Abbildung 3.15). Black und William gehen davon aus, dass Umstellungen im Unterricht in Form eines veränderten Outputs messbar gemacht werden können.

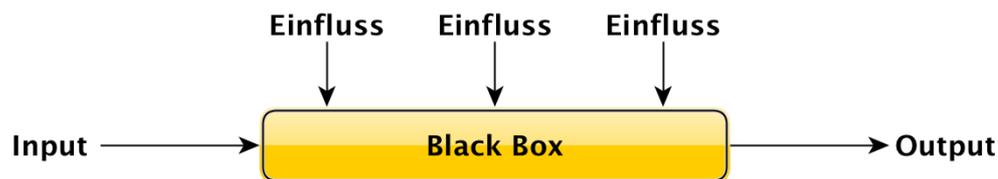


Abbildung 3.15: Die Lernumgebung dargestellt als Black Box. Je nach Input, sowie der Art und dem Zeitpunkt der Einflüsse verändert sich der Output.

Aus eigener Erfahrung wissen wir, dass unsere Stakeholder (vor allem Studierende, Assistierende und Vertreter der Departemente) die oben beschriebenen technischen Möglichkeiten begrüßen und einfordern. Um den eigenen Prinzipien treu zu bleiben und sich dennoch vor einer Weiterentwicklung des Unterrichts nicht zu verschliessen, setzen wir seit mehreren Jahren unter anderem das Feedback von formativen und summativen LK zur Überprüfung unserer Lehrqualität ein. Dabei vergleichen wir die Erfahrungswerte mit den neuen Messungen und versuchen so die Auswirkungen unserer Verbesserungen messbar zu machen. Diese empirischen Resultate geben uns einerseits die Gewissheit, unsere Lernumgebung stetig zu optimieren. Andererseits lassen sich damit unserer Lehrbemühungen für unsere Stakeholder sichtbar machen.

Getrieben vom jungen Wissenschaftsgebiet der *Learning Analytics* (Johnson, Smith, Willis, Levine, & Haywood, 2011) ergänzen wir unser Feedback aus den formativen und summativen LK mit den lehrbezogenen Daten, welche im Verlauf eines Semesters anfallen. Dazu gehören Selbstbeurteilungen, Resultate aus Peer-Evaluationen, Prozessdaten (L. E. Fässler, 2007, p. 212ff) und Log-Daten (Scheuner & Faessler, 2010). Im Kapitel 8 beschreiben wir, wie diese Daten zu einem educational Datawarehouse kombiniert wurden, um die Fortschritte in der Lehrqualität zu überprüfen und zu dokumentieren.

3.5 Möglichkeiten und Chancen von Computer gestützten Leistungskontrollen im Unterricht

Der Abschluss dieses Kapitels bildet eine wichtige Grundlage für diese Arbeit: wir zeigen auf, dass CBA in der Hochschullehre bei grossen Studierendenzahlen aus mehreren Gründen wichtig sein können: sie messen objektiv und reliabel und können flexibel sowohl im Rahmen summativer als auch formativer Leistungskontrollen eingesetzt werden. Der letzte Abschnitt dieses Kapitels gibt einen Überblick über die geschichtliche Entwicklung von Computer-basierten respektive Computer-gestützten Leistungskontrollsystemen und zeigt, dass derartige Systeme bereits seit mehr als 40 Jahren im Unterricht eingesetzt werden. Der Durchbruch von CBA scheint jedoch hauptsächlich mit der in den vergangenen Jahren eingesetzten Professionalisierung der Hochschullehre und dem Ausbau der technischen Infrastruktur möglich zu werden.

3.5.1 Genese Computer gestützter Leistungskontrollen

Bereits in den frühen 60er Jahren wurden erste Systeme für Computer gestützte Prüfungskorrekturen entwickelt. Sie entstanden aus der Notwendigkeit, die begrenzten finanziellen und personellen Mittel effizienter einzusetzen, um die Übungsbetreuung auch bei den rasant wachsenden Studierendenzahlen sicherstellen zu können. So berichtet Jack Hollingsworth vom Computer Labor des Polytechnischen Instituts in New York, dass die Studierendenzahlen seiner Programmierlehrveranstaltung innerhalb von zwei Jahren von 20 auf 120 Studierende zugenommen habe (Hollingsworth, 1960). Er entwickelte deshalb den "Automatic Grader", ein Programm zur Korrektur von Numerik-Übungsaufgaben. Die Studierenden mussten ihre Lösungen in Maschinencode auf Lochkarten stanzen. Mittels lediglich 108 Instruktionen entschied das Programm, ob die Lösung korrekt oder falsch war oder ob ein Assistent die Lösung zum Beispiel aufgrund einer Zeitüberschreitung manuell überprüfen musste. Der Korrekturvorgang einer einzelnen Übung dauerte pro Student zwischen 30 bis 60 Sekunden. Georges Forsythe entwickelte kurze Zeit später einen auf Algol basierten Grader, welcher 30 Programme von Studierenden in 41 Sekunden korrigieren und deren Effizienz in Form der individuellen Programmlaufzeit bestimmen konnte (Forsythe & Wirth, 1965). Wie bereits Hollingsworth musste auch Forsythe für jede Übungsaufgabe einen eigenen Grader schreiben. Forsythe kam in seiner Arbeit zum Schluss, dass der Grader nicht nur eine grosse Zeitersparnis bedeutet, sondern dass *"the machine may be more objective in grading than the human, because of its notable lack of prejudice and its inability to become bored."*

In der Folge wurden die Programme parallel mit den sich schnell entwickelnden Technologien weiterentwickelt und zu echten Prüfungs-Korrekturprogrammen ausgebaut.

1971 präsentierten Duder und White von der Universität in Auckland ein Programm zur Korrektur und Auswertung von *Multiple Choice* Fragen (Duder & White, 1971). Allerdings

mussten die Antworten der Studierenden von den Prüfungsverantwortlichen noch mittels eines Lochers auf eine Lochkarte übertragen und digitalisiert werden. Der Aufwand fürs Digitalisieren entsprach in etwa dem Aufwand einer manuellen Korrektur. Gemäss Duder und White liegt der grosse Vorteil dieses Systems in einer hohen Korrekturqualität und einer Prüfungsanalyse, welche durch das Programm quasi kostenlos erstellt wird.

1986 wurde an der *Texas Tech University School of Medicine* das Programm "Techscore" präsentiert, mit welchem *Multiple Choice* Fragen korrigiert und die prozentualen Anteile der Antworten berechnet werden konnten (Posteraro, Blackwell, & Huddleston, 1986). Allerdings mussten damals die Antworten der Studierenden noch von Hand in den Rechner übertragen werden. Ein optisches Lesegerät zum automatisierten Einlesen der Antworten war zur Zeit der Drucklegung der Publikation noch in der Testphase.

1988 wurde die Firma *Questionmark* gegründet, welche als eine der ersten Softwarefirmen ein Framework zur Erstellung und Durchführung von Leistungskontrollen auf Personalcomputern entwickelte. Die Prüfungs-Software *Questionmark™ Perception™* wird heute noch als kommerzielles Softwareprodukt angeboten⁶.

1994 präsentierte Urs von Matt das automatische Korrektursystem "Kassandra", welches an der ETH Zürich in der Einführungsvorlesung "Wissenschaftliches Rechnen" eingesetzt wurde. Pro Jahr konnten damit rund 1000 Übungsabgaben korrigiert werden. Kassandra wurde direkt von den Studierenden über einen Terminalservice bedient. Sie hatten somit grundsätzlich Zugriff auf die Musterlösungen und auf die Datei, in welcher die Bewertungen der Studierenden gespeichert wurden. Um dies zu verhindern, kümmerte man sich bei "Kassandra" erstmals auch um Sicherheitsaspekte von automatischen Korrektursystemen.

Mit der Verbreitung des Internets an den Hochschulen wurden die hochspezialisierten automatischen Korrektursysteme durch Lernplattformen, sogenannten Learning Managementsystemen (LMS), abgelöst. Dazu zählen zum Beispiel OLAT (*Online Learning And Training*) welches im Jahr 2000 erstmalig erschien und 2004 komplett neu programmiert wurde oder Moodle, welches im Jahr 2002 erschien und seither das mit weltweit 50'000 Installationen⁷ am weitesten verbreitete LMS ist. Ziel dieser komplexen Online-Softwaresysteme ist das Bereitstellen von Lerninhalten und die Organisation von Lernvorgängen. Diese Systeme enthielten anfänglich nur einfache Quiz-Komponenten, welche zur Durchführung von formativen Lernerfolgskontrollen konzipiert waren. Aufgrund der grossen Nachfrage wurden diese nach und nach erweitert. Es wurden damit auch benotete Leistungskontrollen durchgeführt.

⁶ Stand 23.05.2014, gemäss Webseite <https://www.questionmark.com/us/perception/Pages/default.aspx>

⁷ Stand 23.05.2014, gemäss Webseite <http://www.moodle.de/>

Begleitet von einer Reihe sicherheitskritischer Publikationen bezüglich Web-basierten Prüfungssystemen (Heinrich, 2008; Wolf, 2007) wurden für sicherheits-relevante Leistungskontrollen wieder vermehrt dedizierte Prüfungssysteme oder Prüfungs-Umgebungen entwickelt. Dazu zählen unter anderem EvaExam (Gesellschaft für Softwarelösungen mbH, 2007), Sioux (Secure Interactive Online eXam) (Hinterberger & Dahinden, 2008), der Safe Exam Browser (RaHS, 2009) und MEASURED (Guttormsen, Rieder, Hinterberger, & Dahinden, 2010).

Die Sicherheitsaspekte von Computer gestützten Leistungskontrollen wird in Kapitel 4 "Sicherheitsaspekte Computer-basierter Leistungskontrollsysteme" detaillierter besprochen.

3.5.2 Klassifizierung von Computer gestützten Leistungskontrollen (CAA)

Wie im vorherigen Abschnitt beispielhaft dargelegt, wurden in den vergangenen Jahren verschiedene Computer gestützte Leistungskontrollsysteme entwickelt. Diese Systeme zeichnen sich dadurch aus, dass sie eine automatische oder zumindest teilautomatische Korrektur der Leistungskontrolle ermöglichen. Bull und McKenna definierten CAA als "*the use of computers for assessing student learning*" (Bull & McKenna, 2003). CAA Systeme bilden somit die Obergruppe aller Leistungskontrollen, welche den Computer punktuell oder konsequent im Prüfprozess einsetzen. Ein typischer Prüfprozess beinhaltet die Erstellung, die Durchführung, die Korrektur, die Analyse und die Archivierung der Leistungskontrolle. Abbildung 3.16 gruppiert die drei Arten von Leistungskontrollsystemen hierarchisch anhand des Penetrationsgrads des Computers im Prüfprozess.

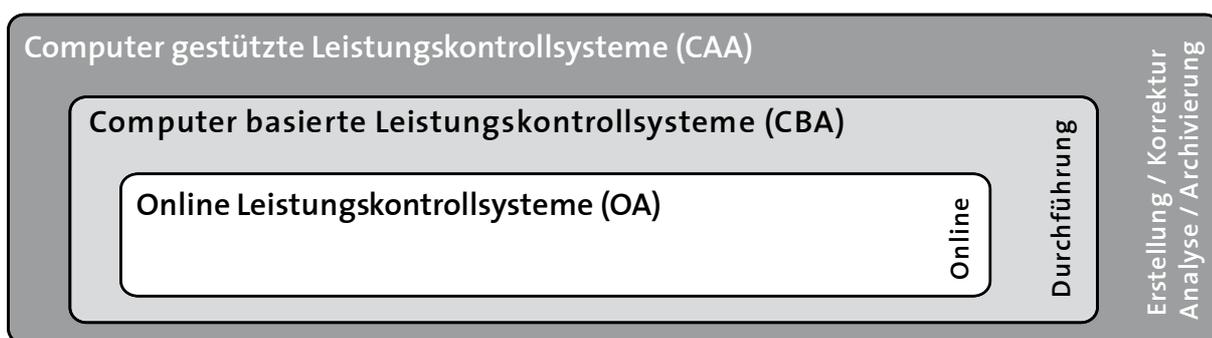


Abbildung 3.16: Hierarchie der Computer gestützten Leistungskontrollen in der Übersicht. CAA bildet dabei die Obergruppe und beinhaltet die CBA und OA Systeme. Prozesstechnisch unterscheiden sich die Systeme hauptsächlich in der Art, wie die Prüfung durchgeführt wird. Abbildung abgeändert nach (Conole & Warburton, 2005).

Computer gestützte Leistungskontrollsysteme (CAA)

Computer-gestützte Leistungskontrollsysteme (CAA) verzichten bei der Durchführung der Leistungskontrolle auf den Einsatz des Computers. Stattdessen kommen physische Prüfungsdokumente zum Einsatz, die anschliessend manuell (Techscore) oder automa-

tisch, zum Beispiel per Scanner, digitalisiert werden (EvaExam). Die restlichen Arbeitsschritte des Prüfungsprozesses findet Computer gestützt statt: Erstellung, Korrektur, Analyse und Archivierung der Leistungskontrolle.

Die Gründe, warum der Computer bei CAA Systemen nicht zur eigentlichen Prüfungsdurchführung verwendet wird, sind vielseitig. Bei älteren Systemen ist dies meist technisch bedingt, da damals die Infrastruktur noch nicht oder nur unzuverlässig zur Verfügung stand. Bei neueren Systemen wird oft aus Sicherheitsgründen auf die Computer gestützte Durchführung verzichtet. Dies meist deshalb, um bei juristischen Problemen auf ein physisches und unterschriebenes Prüfungsdokument zurückgreifen zu können. Mehr zur Problematik des fehlenden physischen Prüfungsdokuments finden Sie im Kapitel 4.3 "Paradigmenwechsel bei CBA".

Die restlichen beiden Systeme, CBA und OA, setzen den Computer auch bei der Durchführung der Leistungskontrolle ein. Sie unterscheiden sich lediglich darin, dass bei den Online Leistungskontrollsystemen (OA) die Prüfungsclients eine permanente Verbindung zum Server benötigen.

Computer-basierte Leistungskontrollsysteme (CBA)

Die CBA-Systeme zeichnen sich dadurch aus, dass die Leistungskontrollen nicht nur am Computer erstellt und korrigiert, sondern dass die Prüfungen durch die Studierenden direkt am Computer abgelegt werden. In früheren Publikationen wurden die CBA-Systeme als *stand-alone* Prüfungsapplikationen bezeichnet, (Conole & Warburton, 2005) welche nicht durch ein Netzwerk verbunden waren. Somit war es damals noch nicht möglich, die Resultate der Studierenden zentral zu erfassen und zu speichern. Neue CBA-Systeme (Dahinden & Hinterberger, 2008; Guttormsen et al., 2010) verfügen allerdings über technische Möglichkeiten, wie die Resultate der Studierenden auch ohne permanente Netzwerkverbindung zentral gesammelt und ausgewertet werden können.

Online Leistungskontrollsysteme (OA)

Die Online Leistungskontrollsysteme bilden eine Untergruppe der CBA-Systeme. Sie sind meist als Webapplikationen konzipiert. Die OA-Clients benötigen deshalb während der Durchführung einer Leistungskontrolle eine permanente Verbindung zum Server (Abbildung 3.17). Dies hat drei Vorteile:

- Die Resultate der Studierenden werden zentral auf dem Webserver gespeichert und können somit einfach korrigiert und analysiert werden.
- Kurze Entwicklungszeit aufgrund effizienter und weitverbreiteter Webapplikations-Technologie

- Moderne Webbrowser ermöglicht den Einsatz verschiedenartiger Medien (hochauflösende Röntgenbilder, Töne, Filme, etc.)

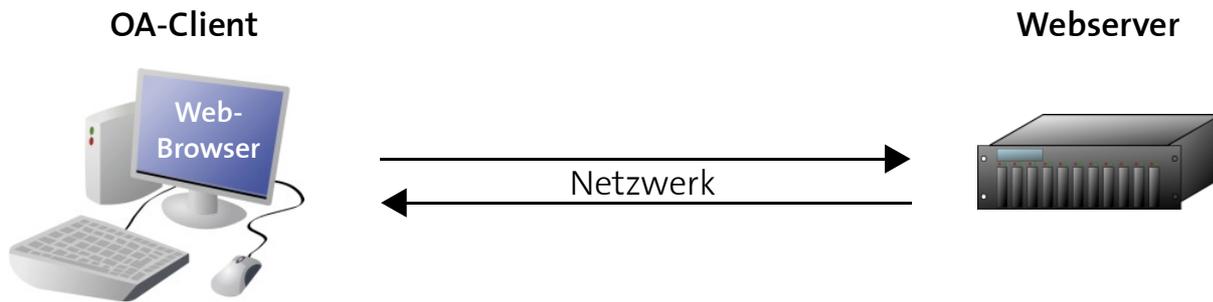


Abbildung 3.17: Vereinfachte Funktionsweise einer OA Architektur: Der Client ist über ein Netzwerk permanent mit dem Server verbunden. Der Client speichert keine Prüfungsinhalte, sondern lädt bei Bedarf alle Inhalte vom Server. Die Inhalte werden auf dem Client mittels eines Web-Browsers gerendert und angezeigt.

Durch die permanente Netzwerkverbindung und der damit verbundenen Möglichkeit zur Kommunikation zwischen den Prüfungsclients, müssen im Vergleich zu reinen CBA Leistungskontrollen zusätzliche Sicherheitsvorkehrungen getroffen werden. Dazu zählen:

- Kommunikation einschränken: die Kommunikation der Clients untereinander und mit dem Internet muss effektiv eingeschränkt werden können (z.B. via *Firewall*-Einstellungen oder durch den Einsatz von geschlossenen Netzwerken).
- Sichere Netzwerkarchitektur: die Netzwerkarchitektur muss vor Ausfällen geschützt werden. Idealerweise werden dazu redundante Netzwerke und mindestens zwei Prüfungsserver eingesetzt.
- Testen von Drittapplikationen: da OA Systeme Drittapplikationen wie zum Beispiel Webbrowser einsetzen, muss vor jeder Leistungskontrolle deren korrektes Funktionieren überprüft werden. So muss unter anderem sichergestellt werden, dass die Leistungskontrolle erwartungsgemäss am Bildschirm dargestellt wird. Der Zeitaufwand für diese Tests können durch den Einsatz von kontrollierten Boot-Medien reduziert werden (Aegler, Kraft, Utiger, & Hassenstein, 2010), da dort das Einspielen von Updates besser kontrolliert werden kann.

Im nachfolgenden Kapitel werden die Sicherheitsaspekte von CBA-Systeme beschreiben im Detail beschrieben.

Kapitel 4

Sicherheitsaspekte Computer-basierter Leistungskontrollsysteme

4.1 Einleitung

An der ETH Zürich zählen Computer-basierte Leistungskontrollen neben den schriftlichen und mündlichen Prüfungen zu den anerkannten Messmethoden kognitiver Leistungen. Die Schulleitung der ETH Zürich hat im Jahr 2006 eine Strategie zur einheitlichen Organisation und Nutzung von Informations- und Kommunikationstechnologien verabschiedet. Eine Massnahme davon umfasste den Aufbau einer Infrastruktur zur Durchführung von CBA (Schmucki, 2010). Dass dieser Entscheid wichtig war, zeigt sich 8 Jahre später daran, dass seither an der ETH Zürich schon zahlreiche benotete Prüfungen am Computer abgelegt wurden. Gemessen an der Tatsache, dass es bisher bei CBA noch zu keinem Rekurs gekommen ist, zeigt, dass die Startphase von CBA an der ETH Zürich aus juristischer Sicht als erfolgreich bezeichnet werden kann.

Aufgrund zunehmender Studierendenzahlen und der wachsenden Nachfrage an summarischen sowie formativen Leistungskontrollen wird die Anzahl der CBA in Zukunft weiter steigen. Durch deren Zunahme erhöht sich auch das Risiko eines Rekurses. Da die Bildungsinstitution in einem Rekurs- und Beschwerdefall in der Beweispflicht ist und belegen muss, wie ein Prüfungsergebnis zustande kam, wird sich zeigen, inwiefern aktuelle CBA-Systeme die legalen Anforderungen erfüllen. Es ist schon jetzt offensichtlich, dass ein allfälliger Rechtsspruch wegweisend für den zukünftigen Einsatz von CBA sein wird und dass es dabei zu einer Selektion der bestehenden CBA-Systeme kommen wird.

Im Sinne einer nachhaltigen Etablierung von CBA befasst sich dieses Kapitel mit den denkbaren Sicherheitsproblemen aktueller CBA-Systemen. Im ersten Teil wird die aktuelle rechtliche Grundlage von Leistungskontrollen an der ETH Zürich beschrieben. Diese wird im zweiten Teil mit den im Vergleich zu traditionellen Prüfungen veränderten Eigenschaften von CBA-Systemen gegenübergestellt. Diese Analyse dient dann als Grundlage für die Erstellung einer Risikoeinschätzung aktueller CBA-Systeme am Beispiel der ETH Zürich. Eine sorgfältige Risikoabschätzung ist für die ressourceneffektive Etablierung von CBA essentiell. Diese für die ETH Zürich spezifische Risikoeinschätzung wird im Abschnitt 3.5.8 in einem generischen Modell erweitert, welches beschreibt, wie das Risikomodell an weiteren Bildungsinstitutionen zur Ausarbeitung einer CBA-

Strategie verwendet werden kann. Abschliessend werden die Resultate der quantitativen Evaluation unserer Prüfungsdurchführung gezeigt und drei konkrete sicherheitskritische Szenarien beschrieben, wie sie bei aktuellen CBA-Systemen auftreten können. Diese Szenarien werden die Grundlage für die Neuentwicklung eines auf aktuellen Erkenntnissen basierenden CBA-Systems sein. Dieses CBA-System wird Inhalt des nachfolgenden Kapitels 5 sein.

4.2 Rechtliche Grundlagen von Leistungskontrollen an der ETH Zürich

An der ETH Zürich bildet die allgemeine Verordnung über Leistungskontrollen die Grundlage für die Messung studentischer Leistungen in den Bachelor- und Masterstudiengängen (ETH Zürich, 2012). Diese Verordnung wird ergänzt durch verschiedene Merkblätter und Weisungen. Tabelle 2.1 fasst die für Leistungskontrollen relevanten Dokumente zusammen:

Verordnung der ETH Zürich über Lerneinheiten und Leistungskontrollen an der ETH Zürich (Leistungskontrollverordnung ETH Zürich)⁸ (ETH Zürich, 2012)	
Richtet sich an:	alle an Leistungskontrollen beteiligten Personen
Inhalt:	Diese Verordnung legt die Grundsätze für sämtliche Lerneinheiten und Leistungskontrollen in den Bachelor- und Master-Studiengängen an der ETH Zürich fest.
Ausführungsbestimmungen des Rektors zur Leistungskontrollverordnung der ETH Zürich (ETH Zürich, 2013a)	
Richtet sich an:	alle an Leistungskontrollen beteiligten Personen
Inhalt:	Beschreibt die konkrete Umsetzung der Leistungskontrollverordnung ETH Zürich im Praxisalltag.
Merkblatt: "Rechtsmittel und Rechtsbehelfe bei Noten und weiteren Leistungsbewertungen" (ETH Zürich, 2014)	
Richtet sich an:	Studierende, Dozierende
Inhalt:	Beschreibt die rechtlichen Möglichkeiten der Studierenden zur Anfechtung einer Notenverfügung.
Weisungen zum Prüfungsplan (ETH Zürich, 2013c)	
Richtet sich an:	Studierende
Inhalt:	Allgemeine Hinweise zur Prüfung (Rechtliches, Fristen, Abmeldung) verfasst von der Rektorin der ETH Zürich.
Schreiben des Prorektors an die Examinatorinnen und Examinatoren einer Prüfungssession	
Richtet sich an:	Dozierende

⁸ SR 414.135.1

Inhalt:	Checkliste für die in besonderen Situationen zu treffenden Massnahmen sowie einige reglementarische und administrative Punkte
Merkblatt: Akteneinsicht und Aktenweitergabe im Rahmen von Leistungskontrollen (ETH Zürich, 2013b)	
Richtet sich an:	Dozierende und Studierende
Inhalt:	Weisung der Rektorin der ETH Zürich, wie eine Prüfungseinsicht in der Praxis ablaufen werden soll.

Tabelle 4.8: Übersicht über die im Rahmen von Leistungskontrollen relevanten Dokumente.

Diese Dokumente bilden die rechtliche Grundlage aller an der ETH Zürich durchgeführten Leistungskontrollen der gestuften Studiengänge. Somit sind bei der Durchführung von Leistungskontrollen in den Bachelor- und Masterstudiengängen zwingend diese rechtlichen Vorgaben zu beachten. In den nachfolgenden Abschnitten wird erläutert, welche rechtlichen Anforderungen im Detail an die Leistungskontrollen gestellt werden und welche Möglichkeiten die Studierenden haben, eine Leistungsbewertung anzufechten.

4.2.1 Benotung studentischer Leistungen

An der ETH Zürich sind als Leistungskontrolle alle Verfahren zugelassen, welche die Leistungen von Studierenden messen und bewerten, insbesondere Prüfungen, Prüfungsblöcke und schriftliche Arbeiten (ETH Zürich, 2012, Art. 2d). Falls diese Note Teil eines als Ganzes zu bestehenden Prüfungsteils ist (Prüfungsblock), beantragt die Notenkonferenz beim Rektor oder der Rektorin das Bestehen oder Nichtbestehen einer Prüfungsstufe zu verfügen (ETH Zürich, 2012, Art. 17). Handelt es sich bei der Note um eine eigenständige Benotung (z.B. Semesterend-Prüfung), verlangt der Examinator oder die Examinatorin deren Verfügung direkt beim Rektorat der ETH Zürich. Die gemessenen Leistungen werden den Studierenden in Form einer Notenmitteilung kommuniziert. Die Studierenden haben dann 30 Tage Zeit, eine beschwerdefähige Notenverfügung zu verlangen. In Abbildung 4.18 sind die beiden Verfahren zur Verfügung einer Note illustriert.

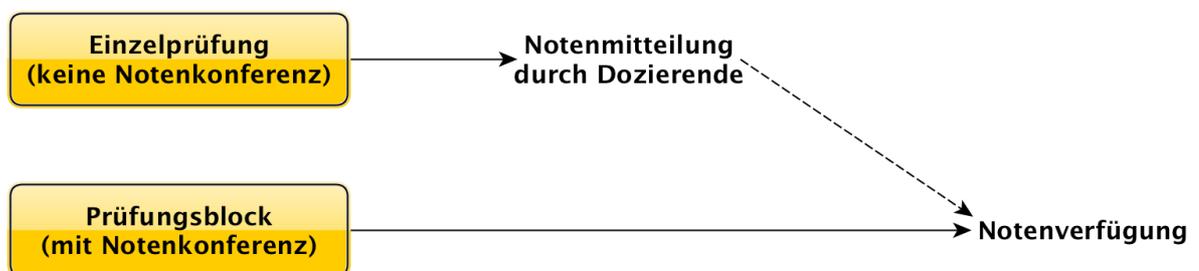


Abbildung 4.18: Verfahren zur Verfügung einer Note. Bei einem Prüfungsblock (bestehend aus mehreren Einzelprüfungen, welche es als Ganzes zu bestehen gilt) beantragt die Notenkonferenz, die Verfügung

der Noten beim Rektorat. Bei Einzelprüfungen teilt in der Regel der Examinator die Noten den Studierenden direkt mit. Beschwerdefähige Notenverfügungen werden nur auf Verlangen ausgestellt.

4.2.2 Vorgehen bei der Anfechtung einer Notenverfügung

Falls Studierende eine Ihrer Ansicht nach ungerechtfertigte Leistungsbewertung erhalten haben, stehen ihnen drei Wege offen, diese anzufechten:

- 1) Persönliches Gespräch mit der Prüfungsleitung oder dem/der Studientelegierten
- 2) Rechtsbehelfs-Weg: Einsprache bei der verfügenden Stelle
- 3) Rechtsmittel-Weg: Beschwerde bei einer höheren Instanz

Die Rechtsberatung der ETH Zürich empfiehlt auf ihrer Webseite, diese drei Möglichkeiten in der hier aufgeführten Reihenfolge zu benutzen.

1) Persönliches Gespräch mit der Prüfungsleitung oder dem/der Studientelegierten

In einem ersten Schritt sollte das direkte Gespräch mit der Prüfungsleitung oder der Studientelegation gesucht werden. Ein solches Treffen findet meist nach der ordentlichen Prüfungseinsicht statt (ETH Zürich, 2013b). Ist ein solches Gespräch nicht möglich oder findet keine Einigung statt, so empfiehlt sich eine Beratung bei der Studienadministration⁹. Dort werden die betroffenen Studierenden ausführlich und objektiv über das Vorgehen und die Erfolgchancen des weiteren Vorgehens beraten. Dazu gehört ein Gesuch zur Neubeurteilung oder Annullierung der Note (Rechtsbehelfs-Weg) oder eine Beschwerde bei der ETH-Beschwerdekommision (Rechtsmittel-Weg). Abbildung 4.19 illustriert diese beiden juristischen Wege.

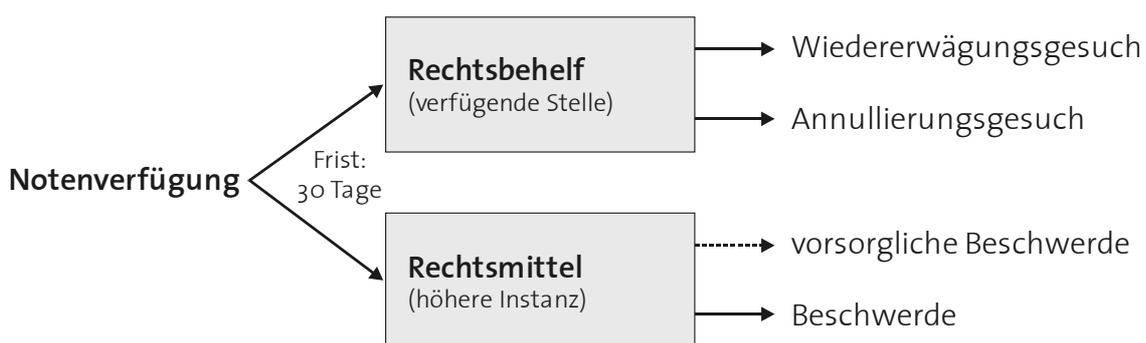


Abbildung 4.19: Juristische Wege zur Anfechtung einer Notenverfügung.

⁹ <https://www.ethz.ch/studierende/de/beratung/rechtsberatung/kritische-studiensituationen.html> (abgerufen am: 23.05.2014)

2) Rechtsbehelfe: Wiedererwägungsgesuch und Annullierungsgesuch

Rechtsbehelfe richten sich an die verfügende Stelle selber (also an die ETH Zürich). Diese wird ersucht, die Verfügung zu überprüfen, abzuändern oder aufzuheben. Dabei unterscheidet man zwischen zwei Rechtsbehelfen: das Wiedererwägungsgesuch und das Annullierungsgesuch. Das Gesuch muss spätestens 30 Tage nach Empfang der Notenverfügung dem Prorektor Lehre eingereicht werden. Dieser entscheidet abschliessend über das Gesuch, nach Einholung von Stellungnahmen aller Beteiligten. Gemäss Auskunft des Rektorats ist das Einreichen des Gesuchs an der ETH Zürich kostenlos (im Gegensatz zur ETH Lausanne).

Pro Jahr werden an der ETH Zürich rund 30 Wiedererwägungs-/Annullationsgesuche im Zusammenhang mit Notenverfügungen eingereicht. Die Chancen, dass einem Gesuch entsprochen wird, belaufen sich gemäss Auskunft des Rektorats auf "deutlich unter 50 Prozent"¹⁰.

Nach Ablehnung des Gesuchs bleibt noch die Möglichkeit einer Beschwerdeeinreichung bei der ETH-Beschwerdekommision.

3) Rechtsmittel: Beschwerde bei der ETH-Beschwerdekommision

Mit dem Rechtsmittel-Weg wird der Fall einer höheren Instanz zum Entscheid vorgelegt. Die Beschwerdeinstanz ist die ETH-Beschwerdekommision in Bern. Die Beschwerde muss innerhalb von 30 Tagen nach Empfang der Notenverfügung eingereicht werden. Da diese Frist nicht verlängerbar ist, kann durch die Einreichung einer vorsorglichen Beschwerde verhindert werden, dass die eigentliche Beschwerde - während des Annullierungsgesuchs - aus formalen Gründen abgelehnt wird (ETH Zürich, 2014). Der Entscheid der ETH-Beschwerdekommision kann ans Bundesverwaltungsgericht weitergezogen werden.

Mit einer Beschwerde können Rechtsverletzungen gerügt werden. Dazu zählen zum Beispiel ein nicht reglements- oder prüfungsplankonformer Prüfungsablauf oder eine Ungleichbehandlung der Studierenden. Eine Ungleichbehandlung liegt zum Beispiel dann vor, wenn der Notendurchschnitt fehlerhaft berechnet wurde oder wenn nicht alle gelösten Aufgaben zur Notenberechnung verwendet wurden.

Bei Beschwerden beträgt der Kostenvorschuss CHF 500. Wenn der Beschwerdeführer verliert, wird in der Regel dieser Vorschuss als Verfahrenskosten vereinnahmt. Im Erfolgsfall wird er zurückerstattet.

¹⁰ Datenbasis: 2009 und 2010. Email von Dieter Wüest, Leiter Rektorat ETH Zürich, 15.03.2011

Pro Jahr werden an der ETH Zürich rund 15 Beschwerden im Zusammenhang mit Notenverfügungen bei der ETH-Beschwerdekommision eingereicht. Die Chancen, dass einem Rekurs stattgegeben wird, sind ähnlich tief wie beim Rechtsbehelfs-Verfahren.

4.2.3 Gründe für die Anfechtung einer Notenverfügung

Pro Jahr finden gemäss Auskunft des Rektorats an der ETH Zürich rund 100'000 Prüfungen statt. In den Jahren 2009 und 2010 kam es pro Jahr durchschnittlich in 30 Fällen zu einer Anfechtung der Notenverfügung. Im Merkblatt "Rechtsmittel und Rechtsbehelfe bei Noten und weiteren Leistungsbewertungen" werden die beiden wichtigsten Anfechtungskategorien erwähnt. Sie werden nachfolgend kurz erläutert.

Willkürliche, objektiv nicht nachvollziehbare Leistungsbewertung

In den meisten Fällen wird dabei die Leistungsbewertung gerügt und die Erteilung einer höheren Note beantragt. Dabei müssen die Beschwerdeführenden nachweisen können, dass "klarerweise eine willkürliche, objektiv nicht nachvollziehbare Punkteerteilung bzw. Benotung"¹¹ stattgefunden hat. Dies gelingt nur in seltenen Fällen, denn die Examinatoren und Examinatorinnen verfügen über einen Beurteilungsspielraum, in den die ETH-Beschwerdekommision von Gesetzes wegen nicht eingreifen darf. So ist es zum Beispiel allein den Examinatoren überlassen, wie viele Punkte oder Teilpunkte sie für eine gegebene Antwort vergeben.

Im Rahmen der Entscheidungsfindung überprüft die ETH-Beschwerdekommision unter anderem, ob der rechtserhebliche Sachverhalt, das Prüfungsdokument, vollständig und korrekt festgestellt wurde. Damit soll überprüft werden, ob im Grundsatz bereits eine willkürliche Bewertung vorliegt. Während bei Papier-basierten Prüfungen dieser Nachweis einfach zu erbringen ist, stellt es CBA vor eine grosse Herausforderung. Mehr dazu im Abschnitt 4.3 "Paradigmenwechsel bei CBA".

Physische oder psychische Beeinträchtigung

An der ETH Zürich gilt der Grundsatz, dass eine Prüfung nur einmal wiederholt werden kann (SR 414.135.1, Art. 14). Gelegentlich versuchen die Studierenden die Annullierung einer Prüfung zu erwirken. Sie begründen dieses Vorgehen mit einer physischen oder psychischen Beeinträchtigung während der Prüfung. Eine Annullierung öffnet den Weg, eine Leistungskontrolle zu wiederholen und diese somit öfter als zweimal abzulegen. Damit die Gleichbehandlung aller Studierenden gewährleistet bleibt, werden an die Voraussetzung für eine Annullierung strenge Anforderungen gestellt. So wird zum Beispiel eine Annullierung grundsätzlich ausgeschlossen, wenn jemand im Bewusst-

¹¹ Merkblatt "Rechtsmittel und Rechtsbehelfe bei Noten und weiteren Leistungsbewertungen", Ziffer 2, Seite 2

sein gesundheitlicher Probleme eine Leistungskontrolle ablegt¹². Begründet wird diese Tatsache dadurch, dass durch dieses Verhalten bewusst der Prüfungsmisserfolg in Kauf genommen wird.

Für CBA kommt den Annullierungsgesuchen aufgrund physischer Beeinträchtigungen keine entscheidende Bedeutung zu, da sie unabhängig vom Prüfungsmodus auch bei Papier-basierten oder mündlichen Prüfungen gleichermaßen auftreten können. Stattdessen gibt es entscheidende Unterschiede bei der psychischen Beeinträchtigung, da CBA aufgrund der limitierten Raumkapazität oft in mehreren Serien und in verschiedenen Räumen stattfinden. Die bei CBA im Vergleich zu traditionellen Prüfungen veränderten Voraussetzungen und deren Auswirkungen auf den rechtskonformen Ablauf der Leistungsmessung ist Inhalt der nachfolgenden Abschnitte.

4.3 Paradigmenwechsel bei CBA

Prüfungstechnisch weisen CBA und Papier-basierte Prüfungen zwei Hauptunterschiede auf: 1) CBA finden meist in verschiedenen Räumen zu verschiedenen Zeiten statt und 2) gibt es bei CBA kein physisches Prüfungsdokument.

Diese Umstände machen es schwierig, für alle Studierenden vergleichbare Prüfungsbedingungen zu schaffen und den rechtserheblichen Sachverhalt zweifelsfrei nachweisen zu können (Abbildung 3.20).



Abbildung 3.20: Zwei juristische Herausforderungen bei CBA: Nachvollziehbarkeit des Prüfungsergebnisses und Gleichbehandlung aller Studierenden.

Erschwerend kommt hinzu, dass es bis zur Drucklegung dieser Arbeit im Jahr 2014 in der Schweiz noch kein juristisches Verfahren gab, welches sich mit den Sicherheitsanforderungen von CBA-Systemen befasste. Es ist somit auch Ziel dieser Arbeit, eine Diskussionsgrundlage zu schaffen und nach einer praxisfähigen Lösung zu suchen, wie man trotz der Sicherheitsbedenken CBA unter Einhaltung der juristischen Rahmbedingungen durchführen kann. Die folgenden Abschnitte befassen sich mit dieser Thematik.

¹² Merkblatt "Rechtsmittel und Rechtsbehelfe bei Noten und weiteren Leistungsbewertungen", Ziffer 3b, Seite 3

4.3.1 Verzicht auf ein physisches Prüfungsdokument

Seit der Einführung von CBA Anfang der 60er Jahre hat sich einiges geändert: Die damals zum Erfassen der Resultate verwendeten Lochkarten und mit optischen Geräten digitalisierbaren Papierbögen wurden mit der fortschreitenden Vernetzung der Computer durch Prüfungsclients abgelöst, welche die Antworten der Studierenden direkt in digitaler Form dem Prüfungsserver zur Speicherung und Korrektur weiterleiteten. Somit fehlt bei CBA das physische Prüfungsdokument, welches in einem juristischen Verfahren als zentrales Dokument zur Überprüfung der Korrektheit des rechtserheblichen Sachverhalts angesehen wird (vgl. Kapitel 4.2 "Rechtliche Grundlagen von Leistungskontrollen an der ETH Zürich"). Aufgrund der fehlenden Rechtsprechungen ist es aus legaler Sicht noch unklar, ob und unter welchen Bedingungen digitale Prüfungsergebnisse ein physisches Prüfungsdokument ersetzen können.

Als Diskussionsgrundlage ist es nützlich, die Gemeinsamkeiten von CBA-Systemen zu beschreiben:

- Sie bestehen aus einem oder mehreren Prüfungsclients und einem zentralen Server zur Speicherung der Zwischenresultate.
- Die Clients kommunizieren über eine verschlüsselte Netzwerkverbindung mit dem Server unter Verwendung des Secure Socket Layer (SSL) Protokolls.
- Die Zwischenresultate der Studierenden werden auf dem Server in einer Textdatei oder einer Datenbank gespeichert (Log-Daten).
- Nach Abschluss der Prüfung werden die Zwischenresultate zu einem individuellen Resultatdossier zusammengestellt.
- Dieses Dossier dient als Grundlage für eine automatische oder halb-automatische Leistungsbeurteilung.

Die Resultatdossiers sind funktionell identisch mit den Prüfungsdokumenten bei Papierprüfungen. Im Vergleich mit einem elektronischen weist das physische Prüfungsdossier jedoch einen bedeutenden Vorteil auf: Die Probanden erstellen die individuellen Dossiers direkt im Prüfungsraum und bezeugen deren Vollständigkeit und Herkunft handschriftlich.

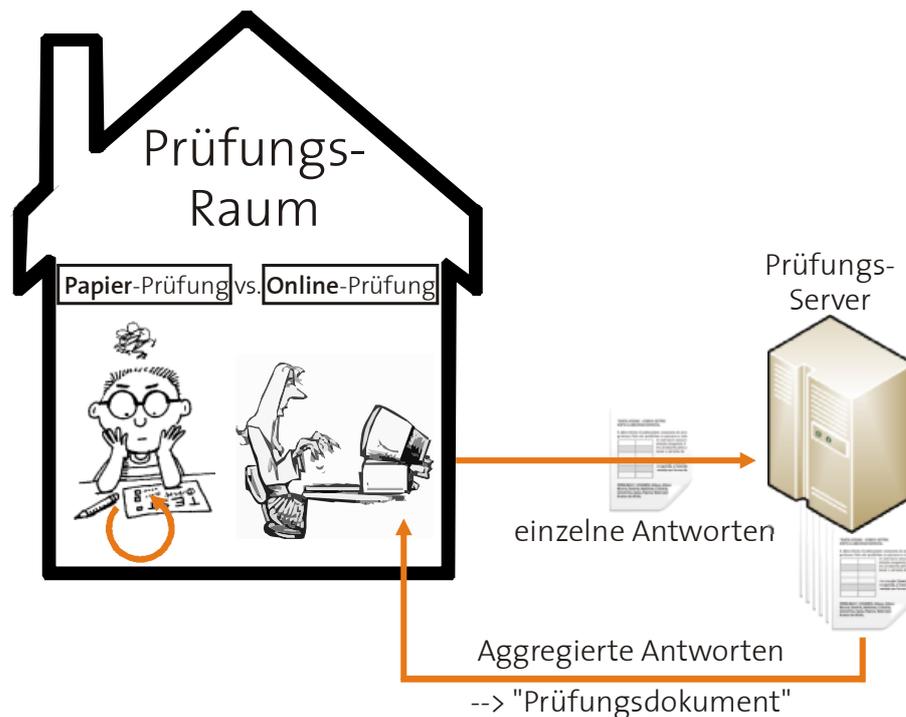


Abbildung 4.21: Bei einer Papier-basierten Prüfung bleibt das "Prüfungsdokument" ganzheitlich bei den Studierenden und ist für diese immer sichtbar. Bei CBA verlassen hingegen die Zwischenresultate den Prüfungsraum und sind für die Studierenden nicht mehr als Ganzes einsehbar.

Anders ist die Situation bei CBA, wenn digitale Zwischenresultate eines virtuellen Dossiers bereits während der Prüfung den Prüfungsraum verlassen, um auf dem Server gespeichert zu werden. Da stellt sich die Frage nach der Vollständigkeit und Authentizität der auf dem Server gespeicherten Antworten (Abbildung 4.21). Für die Studierenden ist dadurch der Online-Prüfungsvorgang weniger transparent, weil sie ihre Resultate nicht mehr physisch unter ihrer Kontrolle auf dem Schreibtisch haben. Es ist somit für die Studierenden nicht mehr nachvollziehbar, ob ihre Resultate vollständig und unverändert zu Korrektur verwendet werden.

Um ein mögliches subjektives Ungerechtigkeitsempfinden unter den Studierenden zu verhindern und die Risiken abzuschätzen, welche durch das Fehlen des physischen Prüfungsdokuments entstehen, steht folgende Frage im Zentrum:

Wie kann sichergestellt werden, dass die Daten, welche die Studierenden auf dem Bildschirm sehen, unverändert und vollständig zur Korrektur verwendet werden?

Diese Frage wurde bis heute noch nicht schlüssig beantwortet. Einerseits liegt dies daran, dass es zurzeit aufgrund fehlender Rechtsprechungen noch keine verbindlichen technischen Anforderungen an CBA-Systeme gibt. Andererseits ist die Nachfrage nach CBA derart gross, dass diese ungeachtet der technischen Risiken durchgeführt werden.

4.3.2 Rechtsunsicherheit aufgrund fehlender Rechtsprechungen

Die Rechtsunsicherheit aufgrund fehlender Rechtsprechungen in Bezug auf die Sicherheitsanforderungen von CBA zwingt Bildungsinstitutionen eigene Sicherheitsvorgaben zu definieren. Karsten Wolf brachte 2007 dieses Problem mit folgender Aussage auf den Punkt: "In der jetzigen Experimentier- und Einführungsphase herrscht keineswegs Rechtssicherheit. An den bundesdeutschen Hochschulen verfährt man mehr nach dem Motto, offensichtliche Punkte nach den besten Kenntnissen zu klären und etwaige Klagen abzuwarten." (Wolf, 2007).

4.3.3 Auswirkungen der Rechtsunsicherheit im Prüfungsalltag

Die Rechtsunsicherheit führt in der aktuellen Hochschullandschaft zu einer Vielzahl verschiedener CBA-Szenarien:

- Szenario "Verzicht auf CBA": Das Institut für Medizinische Lehre (IML) in Bern verzichtet derzeit beim Medizin-Staatsexamen aus rechtlichen Gründen auf die Durchführung von Online-Prüfungen. Und dies, obwohl das Institut aufgrund der grossen Anzahl Prüfungen und der qualitativ hochwertigen, multimedialen Prüfungsfragen eigentlich für CBA prädestiniert wäre.
- Szenario "Outsourcing": Die Hochschule Hannover hat aus Personal- und Raum-mangel die Durchführung der Online-Prüfungen einem externen Dienstleister übertragen. Dieser stellt die Hard- und Software zur Verfügung und sorgt für die Wartung sowie die technische Betreuung während der Prüfungen (Benning, 2008). Dabei profitiert die Hochschule vom Vorteil, dass durch die externe Zusammenarbeit die schulinternen Arbeitsabläufe transparent gemacht werden müssen und sich der Dienstleister vollumfänglich für die Sicherheit der Prüfungsdurchführung verantwortlich zeigt. Verschiedene Sabotageversuche und die bei Medizinstudierenden erhöhte Anzahl angestrebter Rekurse bedeutet für die Bildungsinstitution ein ernst zu nehmendes finanzielles Risiko, welches aber durch vertragliche Abmachungen durch den Dienstleister getragen wird.
- Szenario "Erhöhter Sicherheitsstandard durch Neuentwicklungen": Die ETH Zürich hat 2008 erstmals eine benotete Online-Prüfung durchgeführt (Schmucki, 2008). Die Prüfung fand in öffentlichen Computerräumen statt. Dabei kam ein neu entwickelter Lockdown-Browser zum Einsatz, um das Starten von Drittapplikationen zu verhindern. Zudem wurde 2008 das Projekt Sioux gestartet, welches sich zum Ziel gesetzt hat, ein CBA-System zu entwickeln, mit welchem der für high-stake Prüfungen benötigte rechtserhebliche Sachverhalt korrekt festgestellt werden kann.
- Szenario "Grossflächiger Einsatz": Die Freie Universität Berlin führt schon seit mehreren Jahren CBA durch. Die speziell umgerüsteten Räume bieten Platz für bis zu

240 Studierende, welche die Prüfungen teilweise am eigenen Laptop ablegen (Apostolopoulos, 2007).

- Szenario "Erstellen eines physisches Prüfungsdokuments": Die Universität Karlsruhe hat sich aus rechtlichen Gründen dafür entschieden, die Antworten der Studierenden nach Ende der Leistungskontrolle auszudrucken und von den Studierenden unterschreiben zu lassen. So verfügte der Prüfungsleiter trotz CBA ein schriftliches Dokument, dessen Authentizität durch die Unterschriften der Studierenden garantiert wurde. Die elektronischen Daten wurden anschliessend gelöscht (Küstermann, Melcher, Toussaint, & Seese, 2005).

So unterschiedlich diese individuellen Lösungen sind, eine Eigenschaft haben sie gemeinsam: Im Falle eines Rekurses sind die Bildungsinstitutionen in der Beweispflicht und müssen rechtsgültige Informationen über die Prozess- und Datensicherheit des eingesetzten CBA-Systems vorlegen können (vgl. Kapitel 4.2 "Rechtliche Grundlagen von Leistungskontrollen an der ETH Zürich"). Dass sich die Institutionen dieses Umstands bewusst sind, zeigen die vor Kurzem initiierten Projekte zur Sicherheitssteigerung von CBA: Safe Exam Browser (Rahs, 2009), Sioux (Dahinden & Hinterberger, 2008) und QualiAss (C. Meier, 2008) sind drei davon.

Welchen Effekt hat nun der Verzicht auf ein physisches Prüfungsdokument im Prüfungsalltag und welche Risiken geht eine Bildungsinstitution bei der Durchführung von CBA ein? Diese Fragen werden im nachfolgenden Abschnitt 4.4 erläutert.

4.4 Risikomanagement-Modell für CBA

Der Wechsel von Papier-basierten Leistungskontrollen zu CBA ist in vielerlei Hinsicht eine Herausforderung. Nicht zuletzt deshalb, weil viele *Stakeholder* involviert sind, welche unterschiedliche Bedenken gegenüber CBA haben. Abbildung 4.22 listet die *Stakeholder* auf, welche bei der Planung, Durchführung oder Analyse von CBA involviert sind.

- Aus Sicht der Rechtsvertreter birgt das Fehlen eines physischen Prüfungsdokuments ein Risiko.
- Für Dozierenden ist die Entwicklung von automatisch korrigierbaren Prüfungsfragen zur Messung von höheren kognitiven Fähigkeiten aufwändig.
- Für das technische Supportteam und die Prüfungsleitung bedeutet CBA mit unzuverlässigen und wartungsintensiven Prüfungssystemen einen bedeutenden Arbeitsaufwand.
- Für die Prüfungsleitung und die Prüfungsaufsicht bedeutet CBA ohne institutionalisierte Standards einen beträchtlichen Zusatzaufwand gegenüber traditionellen Leistungskontrollen.

- Die Schulleitung muss personelle und monetäre Ressourcen bereitstellen, um die für CBA nötige Infrastruktur zu Verfügung zu stellen und eine nachhaltige Etablierung von CBA an einer Bildungsinstitution zu ermöglichen.
- Für die Studierenden kann eine ungenügende Leistung an einer CBA zum Ausschluss vom weiteren Studium führen. Sie müssen sich deshalb auf das CBA-System verlassen können, wobei sie davon auszugehen haben, dass dieses die Antworten ähnlich zuverlässig speichert wie Papier und zudem die Vorteile einer automatischen Korrektur und eines individuellen Feedbacks bietet.

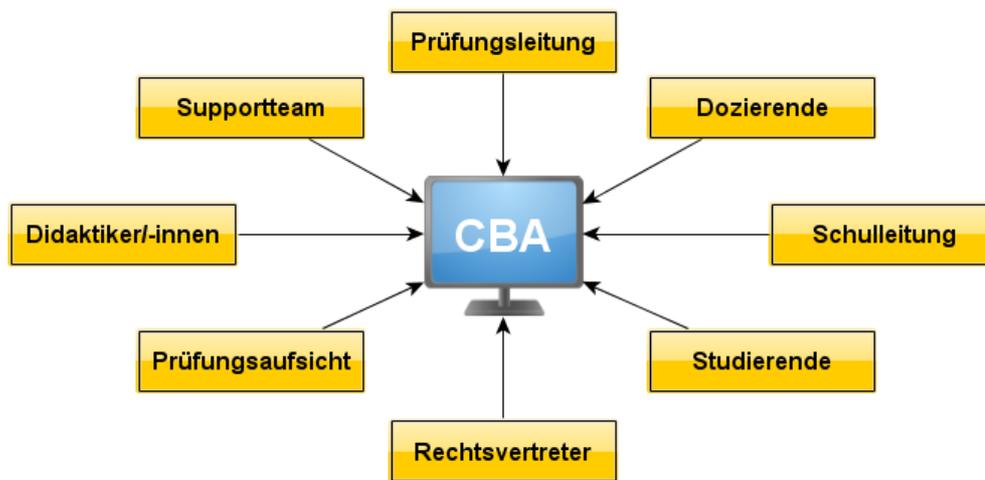


Abbildung 4.22: Stakeholder bei CBA.

Damit CBA an einer Bildungsinstitution nachhaltig etabliert werden kann, müssen diese vielfältigen Bedenken bei der Planung von CBA berücksichtigt werden. Unberücksichtigte Bedenken zählen zu den zentralen Risiken von CBA und können dazu führen, dass eine Gruppe von *Stakeholdern* sich aktiv gegen den Einsatz von CBA ausspricht. Eine Risikomanagement-Strategie kann helfen, die abschätzbaren Risiken zu identifizieren. Ebenso können Massnahmen zu deren Verhinderung oder Reduktion getroffen werden.

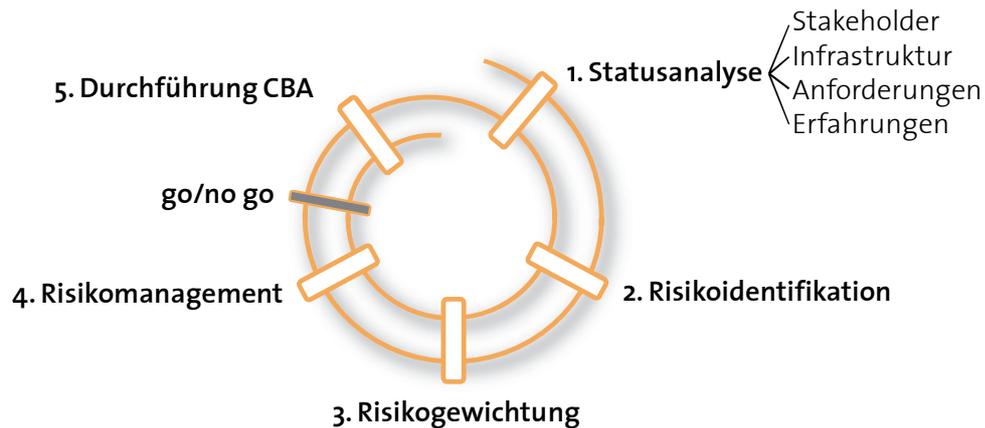


Abbildung 4.23: Risikomanagement Modell bei CBA. Die Risiken bei CBA variieren je nach Bildungsinstitution. Im Schritt 1 "Statusanalyse" wird den unterschiedlichen Bedingungen Rechnung getragen. Diese fließen in die Schritte 2 bis 5 ein.

Abbildung 4.23 illustriert die 5-Schritte bei der Ausarbeitung einer Risikomanagement-Strategie. Die Schritte 2-4 basieren im Grundsatz auf bekannten Verfahren, wie sie in Software Engineering Prozessen angewandt werden (Pressman, 2010):

1. Die **Statusanalyse** beinhaltet die Identifikation der Bildungsinstitutions-spezifischen Voraussetzungen zur Durchführung von CBA.
2. Die **Risikoidentifikation** führt zu einer Auflistung aller erdenklichen Risiken bei CBA.
3. Die **Risikogewichtung** priorisiert die Risiken gemäss ihren Eintrittswahrscheinlichkeiten und Auswirkungen beim Risikoeintritt.
4. Das **Risikomanagement** beinhaltet die Ausarbeitung von Strategien zur Reduktion der Eintrittswahrscheinlichkeiten von Risiken oder zur Abschwächung von deren Auswirkungen auf den CBA-Prozess. Dieser Schritt bietet auch die Möglichkeit, jene Risiken von der weiteren Bearbeitung auszuschliessen, deren Auswirkungen auf den Prüfprozess als vernachlässigbar eingestuft werden können. Dies hilft Kosten zu sparen. Basierend auf dem verbleibenden Restrisiko wird die **go/no go-Entscheidung** gefällt. Sie bietet die letzte Möglichkeit, den CBA-Prozess abubrechen, sollten die Risiken unkontrollierbar sein.
5. Die **Durchführung der CBA** geschieht unter Berücksichtigung der Risikomanagement-Strategien zur Reduktion der Risikoeintrittswahrscheinlichkeiten oder zur Abschwächung der Risikoauswirkungen auf den Prüfprozess. Der Prüfungsablauf sollte detailliert protokolliert werden und Informationen für eine erneute Statusanalyse liefern.

In den folgenden Abschnitten haben wir anhand konkreter Daten aus unseren CBA-Praxiseinsätzen eine Risikomanagement-Strategie für CBA am Beispiel der ETH Zürich

durchgeführt. Bei anderen Bildungsinstitutionen können die Risikoliste und die Risikogewichtung aufgrund der unterschiedlichen Rahmenbedingungen variieren.

Im Abschnitt 4.4.8 Risikomanagement als Basis zur Entwicklung einer CBA-Strategie für Bildungsinstitutionen wird das hier erläuterte Vorgehen verallgemeinert. Danach wird beschrieben, wie das Modell für das Risikomanagement von CBA zur Entwicklung einer CBA-Strategie für Bildungsinstitutionen verwendet werden kann.

4.4.1 Statusanalyse

Die Statusanalyse setzt sich zum Ziel, die bildungsinstitutionsspezifischen Rahmenbedingungen für CBA festzustellen. Dazu zählt die Identifikation der *Stakeholder* sowie die Analyse der Infrastruktur, der Anforderungen an CBA und die allfällig bereits gemachten Erfahrungen.

Rahmenbedingung	Zentrale Fragen
Stakeholder	<ul style="list-style-type: none"> • Welche Personengruppen können als <i>Stakeholder</i> im CBA-Prozess identifiziert werden? • Was sind die Interessen der jeweiligen <i>Stakeholder</i> in Bezug auf CBA? • Inwiefern beeinflussen diese Interessen die Gewichtung der einzelnen Risiken? (vgl. Tabelle 4.10)
Infrastruktur	<ul style="list-style-type: none"> • Welche Infrastruktur steht für die Durchführung von CBA zur Verfügung? • Wie zuverlässig ist sie? • Welches CBA-System wird verwendet? • Welches sind Schwachpunkte der Infrastruktur? • Inwiefern beeinflusst die Infrastruktur die Eintrittswahrscheinlichkeit der einzelnen Risiken? (vgl. Tabelle 4.9)
Anforderungen	<ul style="list-style-type: none"> • Welche legalen Anforderungen gelten für CBA? • Inwiefern beeinflussen diese Anforderungen die Gewichtung der einzelnen Risiken?
Erfahrungen	<ul style="list-style-type: none"> • Welche Erfahrungen wurden an der Bildungsinstitution bereits gemacht?

	<ul style="list-style-type: none">• Existieren Operationalisierungslisten oder Erfahrungsberichte, welche allfällige Probleme erwähnen?• Inwiefern beeinflussen diese Erfahrungen die Gewichtung der einzelnen Risiken?
--	--

Eine sorgfältige Statusanalyse dient als Basis für die Identifikation und Gewichtung der Risiken und bildet somit die Grundlage für die Risikoabschätzung bei CBA.

4.4.2 Risikoidentifikation

Unter einer Risikoidentifikation versteht man das Bestreben, alle bekannten und absehbaren Gefahren in einem CBA-System zu beschreiben. Durch die systematische Identifikation der Gefahren wird die Grundlage geschaffen, um diese wenn möglich zu vermeiden oder deren Auswirkungen auf den Prüfprozess zu kontrollieren.

Wir definieren vier Risikobereiche bei der Durchführung von CBA-Systemen:

Didaktische Risiken

Die didaktischen Risiken umfassen die Entwicklung von automatisch korrigierbaren Prüfungsfragen und deren Abstimmung auf die zu prüfenden curricularen Unterrichtsinhalten. Die didaktischen Risiken bergen die Gefahr, die Validität und die Reliabilität einer Prüfung massiv negativ zu beeinflussen und so die Aussagekraft eines Prüfungsergebnisses zu minimieren. Die meisten didaktischen Risiken sind unabhängig vom Prüfungsmodus und treten bei CBA genauso auf wie bei traditionellen schriftlichen oder mündlichen Prüfungen.

Operationale Risiken

Die operationalen Risiken beinhalten verfahrenstechnische Gefahren, welche bei der Planung, Durchführung und Analyse von CBA auftreten können. Erfahrungsgemäss sind die operationalen Risiken einfacher abschätzbar als die didaktischen und technischen Risiken. Zudem können deren Auswirkungen durch den Erfahrungsgewinn bei wiederholter Prüfungsdurchführung bedeutend reduziert werden.

Technische Risiken

Die technischen Risiken umfassen alle absehbaren Gefahren, welche im Zusammenhang mit der technischen Durchführung von CBA zu erwarten sind. Dazu zählen primär Sicherheitsrisiken wie: Ausfall des Prüfungssystems, nicht eindeutige Nachweiseinbringung der Integrität und Vollständigkeit des Prüfungsergebnisses und eine unvollständige Protokollierung des Prüfungsprozesses. Die technischen Risiken treten primär bei CBA auf und variieren je nach technischem Konzept des Prüfungssystems (Web-basiert, Ser-

ver-Client Applikation). Die in dieser Arbeit definierten Risiken beziehen sich primär auf Web-basierte Applikationen, da sie die grosse Mehrheit der existierenden Systeme umfassen.

Finanzielle Risiken

Finanzielle Risiken umfassen Aspekte des nachhaltigen Umgangs mit den personellen und materiellen Ressourcen. Dazu zählen unter anderem die Kosteneffizienz des Prüfungssystems und die ideelle Unterstützung der Schulleitung gegenüber CBA. Hier gilt zu beachten, dass erste CB-Pilotprüfungen meist durch spezielle Projektfördergelder abgedeckt sind, während zur nachhaltigen Etablierung von CBA längerfristige Finanzierungen sichergestellt werden müssen (z.B. für den technischen Unterhalt der Infrastruktur).

4.4.3 Risikotabelle

Die Risikotabelle stellt das zentrale Dokument einer Risikoidentifikation dar. Sie listet alle vorhersehbaren Risiken auf, die im Zusammenhang mit CBA auftreten können. Tabelle 4.9 listet die Risiken für CBA an der ETH Zürich auf. Sie basiert auf einer Risikotabelle der Universität Plymouth und der London Metropolitan Universität (Ricketts & Zakrzewski, 2005; Zakrzewski, Steven, & Ricketts, 2009). Die Tabelle wurde von uns angepasst und ergänzt, so dass die legalen Anforderungen und unsere Erfahrungen berücksichtigt werden. Die insgesamt 49 beschriebenen Risiken werden in vier Risikokategorien eingeteilt. Jedes Risiko wurde mit einer Identifikation (R-ID) versehen, welche die Risikogruppe (**D**idaktisch, **O**perational, **T**echnisch, **F**inanziell) und eine Ziffer beinhaltet. Diese Risiko-ID wird in späteren Kapiteln referenziert und dient der eindeutigen Kennzeichnung der Risiken.

Z-ID	R-ID	Beschreibung
P1	D1	CBA nicht im Curriculum (Lehrplan) integriert
-	D2	Prüfungsfragen widerspiegeln nicht das Kompetenzniveau der LV
P3	D3	Unpassende oder schlecht konstruierte Prüfungsfragen
-	D4	Unpassender Einsatz von CBA
P5	D5	Prüfung beinhaltet Fehler
P6	D6	Studierende haben Angst vor der neuen Prüfungsmethode
P7	D7	Mangelndes Interesse von Dozierenden
P8	D8	Studierende lehnen CBA aktiv ab
P9	D9	Fehlende Dokumentation des Prüfprozesses reduziert Effizienz des Prüfungssystems
-	D10	Veraltete Prüfungen im Umlauf
O1	O1	Unrealistischer Zeitplan für CBA
O2	O2	Kurzfristiger Ausfall von Aufsichtspersonal
-	O3	Ausstattung der Prüfungs-Arbeitsstationen für CBA ungeeignet (Bildschirmauflösung, etc.)
O6	O4	Prüfungs-Arbeitsstationen nicht ideal positioniert (Bildschirm einsehbar)
O7	O5	Anzahl Studierenden zu hoch für CBA (maximal verfügbare Anzahl an Computer Arbeitsplätzen)
O14	O6	Studierende lösen die korrekte Prüfung in einer falschen Session
-	O7	Studierende sind nicht vertraut mit dem Prüfungsmodus
O9	O8	Studierende erscheinen zu spät zur Prüfung
O10	O9	Studierende verlassen die Prüfung vorzeitig
-	O10	Student hat sich nicht für die Prüfung angemeldet, deshalb ist keine Prüfung verfügbar
O19	O11	Raumaufsicht ist nicht vertraut mit den Einzelheiten des CBA Systems
O5	O12	Studierenden mit speziellen Handicaps (rot/grün-Blindheit, etc.) werden nicht unterstützt
O13	O13	Studierende welche kein CBA haben, sind beeinträchtigt, da keine Computerarbeitsplätze mehr zu Verfügung stehen
O21	O14	Unklare Rollenverteilung führt zu Fehlinformationen und Fehlentscheidungen
O19	O15	Fehlmanipulationen des CBA Systems aufgrund fehlender Instruktion
O25	O16	Kommunikation zwischen zwei Prüfungssessionen möglich
O26	O17	Keine Prüfung verfügbar bei Prüfungseinsicht
O11	T1	Studierende manipulieren die Prüfungs-Arbeitsstationen
O4	T2	Vorzeitige Veröffentlichung von Prüfungsfragen durch Sicherheitslücken
O12	T3	Unerlaubter Zugriff auf Antworten
-	T4	Studierenden können sich am System nicht anmelden (Passwort vergessen)
W3	T5	Vom Prüfungs-Server kann keine Prüfung bezogen werden
O8	T6	Studierende wählen die falsche Prüfung
T1	T7	Prüfungs-Arbeitsstation stürzt während der Bearbeitung der Prüfung ab
T2	T8	Prüfungs-Server stürzt während einer laufenden Prüfung ab
T7	T9	Prüfungs-Arbeitsstation verliert während Prüfung die Verbindung zum Server
W4	T10	Studierende greifen während einer laufenden Prüfung auf unerlaubte (z.B. web-basierte) Ressourcen zu.
W5	T11	Studierende verwenden verschiedene Betriebssysteme
T5	T12	Mangelnde Usability der Prüfungsclient-Oberfläche
T6	T13	Ein unerwartetes Resultat
W7	T15	Netzwerklast während der Prüfung ist zu hoch
-	T16	Serverlast während der Prüfung ist zu hoch
T3	T17	Antworten werden nicht korrekt gespeichert
T4	T18	Antworten gehen verloren
	T19	Erfolgreicher Rekurs aufgrund nicht ausreichender Beweisführung
F1	F1	Fehlen von Unterstützung durch die Schulleitung
F3	F2	CBA System ist nicht Kosten-effizient
P4	F3	Dozierende wenden sich von CBA ab, da Aufwand zu hoch
	F4	Aufwändige Korrektur aufgrund mangelhafter Prüfungsfragen

Tabelle 4.9: Risikotabelle von CBA. Die R-ID zeigt Risiko-Identifikationsnummer und bezeichnet jedes Risiko eindeutig. Die Lettern D, O, T und F beziehen sich auf die vier Risikogruppen (Abschnitt "4.4.2 Risikoidentifikation"). Z-ID gibt, falls vorhanden, die ID der von Ricketts, Steven und Zakrzewski publizierten Risiken an. Ein "-" in der Spalte Z-ID weist auf ein von uns neu hinzugefügtes Risiko hin.

Um die vorhandenen Ressourcen effizient einsetzen zu können, müssen die Risiken gewichtet werden. Diese Gewichtung erlaubt die Identifizierung der gefährlichsten Risiken. Durch das Festlegen eines *cutoff*-Wertes können zudem jene Risiken erkannt werden, welche vernachlässigt werden können. Dies ist dann der Fall, wenn es eine geringe Eintrittswahrscheinlichkeit aufweist und wenn dessen Auswirkungen auf den Prüfprozess minimal ausfallen. Die Risikogewichtung ist Teil der Risikoanalyse im nachfolgenden Abschnitt.

4.4.4 Risikoanalyse

Mit der Risikoanalyse werden die Risiken auf zwei Arten klassiert:

- 1) Die Wahrscheinlichkeit, dass ein Risiko eintritt, wenn keine geeigneten Gegenmassnahmen getroffen werden.
- 2) Die zu erwartenden Auswirkungen eines Risiko-Ereignisses auf den CBA Prozess.

Diese Klassierung erlaubt das Setzen von Prioritäten, welche helfen sollen, mit den vorhandenen Ressourcen den grössten Effekt zu erzielen. Aus Platzgründen werden hier statt der absoluten Risiko-Beschreibungen die Risiko-IDs verwendet. Die komplette Liste der Risikoanalyse findet sich im Anhang A.3 dieser Arbeit.

Aufgrund von unterschiedlichen Rahmenbedingungen können allerdings die Risikogewichtungen variieren. Diese gilt es in der anschliessenden Bildungsinstitutions-spezifischen Risikomanagement-Strategie zu berücksichtigen.

Eintrittswahrscheinlichkeit der Risiken

Die Eintrittswahrscheinlichkeit eines Risikos bezieht sich auf den Fall, dass keine Massnahmen getroffen werden um die Eintrittswahrscheinlichkeit zu reduzieren. Sie wird auf einer Skala von 0 (nie) bis 1 (immer) angegeben und basiert auf einer Kombination aus publizierten Daten, Erfahrungswerten und realen Zahlen aus dem Unterricht. Die Eintrittswahrscheinlichkeiten der einzelnen Risiken sind in Tabelle 4.10 aufgelistet. Zu jedem Wert ist die Datenquelle angegeben.

Die Eintrittswahrscheinlichkeiten sollten an die Bedingungen der jeweiligen Bildungsinstitution respektive der Art der Leistungskontrolle angepasst werden. Bei formativen, unbenoteten Leistungskontrollen können beispielsweise die Auswirkungen der Risiken tiefer als bei benoteten Schlussprüfungen angesetzt werden.

R-ID	Wer	Auswirkung auf...	Wann?	Kons.	Wahrs.	Quelle	G-Potential	G-Pot (gewicht.)
D1	St/Do	Gesamte Prüfung	B	0.57	0.60	1	0.34	0.34
D2	St/Do	Lehrveranstaltung	B	0.71	0.50	2	0.36	0.36
D3	St/Do/PL/PA	Lehrveranstaltung	B	0.71	0.75	1	0.54	0.54
D4	St/Do/PL	Lehrveranstaltung	B	0.71	0.30	2	0.21	0.21
D5	St/Do/PL/PA	Gesamte Prüfung	D Po	0.57	0.60	2	0.34	0.43
D6	St/Do/PL	Student, Studentin	B	0.14	0.48	3	0.07	0.07
D7	Do/Ma	Lehrveranstaltung	B	0.60	0.95	3	0.57	0.57
D8	St/Ma	Student, Studentin	B	0.14	0.03	3	0.00	0.00
D9	Do/PL/PA/Su	Bildungsinstitution	B	0.71	0.87	1	0.62	0.62
D10	St/Do	Student, Studentin	B	0.14	0.90	2	0.13	0.13
O1	Do/PL/Su	Lehrveranstaltung	B	0.60	0.20	2	0.12	0.12
O2	Do/PL/PA/Su	Session	D	0.36	0.75	1	0.27	0.33
O3	Do/PL/Ma	Raum	B	0.24	0.20	2	0.05	0.05
O4	Do/PL/Ma/Su	Raum	B	0.24	0.90	1	0.21	0.21
O5	Do/PL/Ma	Lehrveranstaltung	B	0.60	0.70	1	0.42	0.42
O6	St/PA	Raum	D	0.29	0.95	3	0.27	0.34
O7	St/Do	Lehrveranstaltung	D	0.71	0.70	2	0.50	0.63
O8	St/PA	Student, Studentin	D	0.14	0.50	2	0.07	0.09
O9	PL/PA	Gesamte Prüfung	D	0.38	0.85	1	0.32	0.40
O10	St/PL/PA	Student, Studentin	D	0.14	0.05	2	0.01	0.01
O11	St/Do/PL/PA	Gesamte Prüfung	B D	0.57	0.50	2	0.29	0.36
O12	St/Do/PL/PA	Session	D	0.43	0.99	1	0.42	0.53
O13	St	Student, Studentin	D	0.14	0.37	1	0.05	0.07
O14	St/PL/PA	Gesamte Prüfung	B D Po	0.57	0.87	1	0.50	0.62
O15	St/PL/PA	Gesamte Prüfung	D	0.57	0.62	1	0.35	0.44
O16	St/Do/PL/PA	Gesamte Prüfung	D	0.57	0.62	1	0.35	0.44
O17	St/Do/Ma	Student, Studentin	Po	0.14	0.37	1	0.05	0.05
T1	St/Do/Ma/Su	Bildungsinstitution	D	0.86	0.37	1	0.32	0.40
T2	Do/PL/Ma	Gesamte Prüfung	B	0.48	0.40	1	0.19	0.19
T3	St/Do/Ma	Gesamte Prüfung	B	0.57	0.35	1	0.20	0.20
T4	St/Ma	Gesamte Prüfung	D	0.57	0.20	2	0.11	0.14
T5	St/PL/PA	Session	D	0.43	0.37	1	0.16	0.20
T6	St/PA	Student, Studentin	D	0.14	0.37	1	0.05	0.07
T7	St/PA	Student, Studentin	D	0.14	0.40	1	0.06	0.07
T8	St/PL/PA/Su	Bildungsinstitution	D	0.86	0.10	1	0.09	0.11
T9	St/PL/PA/Su	Raum	D	0.29	0.10	1	0.03	0.04
T10	St/PL/PA	Bildungsinstitution	D	0.86	0.37	1	0.32	0.40
T11	St/PA	Student, Studentin	D	0.14	0.99	3	0.14	0.18
T12	St/Do/PL	Bildungsinstitution	B	0.86	0.62	1	0.53	0.53
T13	St/Do/PA	Bildungslandschaft	Po	1.00	0.20	1	0.20	0.20
T15	St/PL/PA	Gesamte Prüfung	D	0.57	0.62	1	0.35	0.44
T16	St/PL/PA	Gesamte Prüfung	D	0.57	0.80	2	0.46	0.57
T17	St/Do/PL/Ma	Bildungslandschaft	Po	1.00	0.20	1	0.20	0.20
T18	St/Do/PL/Ma	Bildungslandschaft	Po	1.00	0.40	2	0.40	0.40
T19	St/Do/Ma	Bildungslandschaft	B D Po	1.00	0.50	2	0.50	0.63
F1	Do/Ma	Bildungsinstitution	B	0.71	0.20	2	0.14	0.14
F2	Do/Ma/Su	Bildungsinstitution	B	0.71	0.45	1	0.32	0.32
F3	Do/Ma	Gesamte Prüfung	B	0.48	0.05	2	0.02	0.02
F4	Do/PL	Gesamte Prüfung	Po	0.48	0.40	2	0.19	0.19

Tabelle 4.10: Risikoanalyse von CBA. Angegeben ist die Risiko-ID (R-ID), der betroffene Personenkreis (Wer), der Umfang der Auswirkungen bei einem allfälligen Eintritt des Risikos (Auswirkung auf...), der Eintrittszeitpunkt des Risikos (Wann?), die Schwere der Konsequenzen beim Risikoeintritt (Kons.), die Wahrscheinlichkeit, dass das Risiko eintritt (Wahrs.), die Datenquelle der Risikowahrscheinlichkeit (1=publizierte Daten, 2=Erfahrungswerte, 3=gemessene Werte), das ungewichtete Gefahrenpotential (G-Potential) und das mit einem Korrekturfaktor für zeitkritische Risiken gewichtete Gefahrenpotential (G-Pot. (gewicht.)). Die Abkürzungen der Spalte "Wer" finden sich in Tabelle 4.11. Die Werte beziehen sich auf summative Leistungskontrollen.

Da ein Risiko mit hoher Eintrittswahrscheinlichkeit nicht zwingend gravierende Auswirkungen auf den Prüfprozess hat, müssen für alle Risiken die allfälligen Konsequenzen bestimmt werden. Dies ist Inhalt des nachfolgenden Abschnitts.

Konsequenzen bei Risikoeintritt

Zur Abschätzung der Konsequenzen schlägt Ricketts eine Konsequenzen-Matrix vor (Ricketts & Zakrzewski, 2005). Die ursprünglich für die Universität Plymouth erstellte Formel zur Abschätzung der Konsequenzen wurde durch uns an die strukturellen und hierarchischen Gegebenheiten bei CBA an der ETH Zürich angepasst. Im Vergleich zur Konsequenzen-Matrix von Ricketts weist unser Ansatz je eine Dimension mehr auf, wobei wir zwischen Prüfungsleitung und Prüfungsaufsicht unterscheiden (Ricketts fasste diese als "Academics" zusammen). Ebenso können wir die Auswirkungen der Risiken auf die Bildungslandschaft erfassen. Die Anzahl der Bereiche pro Dimension werden mit A_{max} und W_{max} bezeichnet und beziehen sich auf die Einträge der Tabelle 4.11.

$$\text{Konsequenz} = \frac{A_w \cdot W_A}{A_{max} \cdot W_{max}}$$

$$A_{max} = 7$$

$$W_{max} = 6$$

Tabelle 4.11 zeigt die Konsequenzen-Matrix für CBA. Die Matrix zeigt die Risikogruppen (Wer) und den Umfang der Auswirkungen ("Auswirkung auf..."). Aus der Tabelle lässt sich ablesen, dass z.B. ein Risiko, welches bei *einem Studenten* auftritt und eine Auswirkung auf die *gesamte Prüfung* hat, höher gewichtet wird (Wert=0.57), als ein Ereignis, welches die *Prüfungsleitung* betrifft und Auswirkungen auf *eine Prüfungssession* hat (Wert=0.29). Die Werte aus der Konsequenzen-Matrix sollten als Richtwerte betrachtet werden und müssen den aktuellen Gegebenheiten angepasst werden.

		Wer	Support	Management	P-Aufsicht	P-Leitung	Dozierende	Studierende
			Su	Ma	PA	PL	Do	St
			1	2	3	4	5	6
Auswirkung auf...								
Student	1	0.02	0.05	0.07	0.10	0.12	0.14	
Raum*	2	0.05	0.10	0.14	0.19	0.24	0.29	
Session*	3	0.07	0.14	0.21	0.29	0.36	0.43	
Gesamte Prüfung	4	0.10	0.19	0.29	0.38	0.48	0.57	
Lehrveranstaltung	5	0.12	0.24	0.36	0.48	0.60	0.71	
Bildungsinstitution	6	0.14	0.29	0.43	0.57	0.71	0.86	

Bildungs- landschaft	7	0.17	0.33	0.50	0.67	0.83	1.00
-------------------------	---	------	------	------	------	------	------

Tabelle 4.11: Konsequenzen der Risiken. Ein "Raum" ist definiert als eine Menge Studierende, welche zur gleichen Zeit im gleichen Raum eine Leistungskontrolle ablegen. Eine "Session" bezeichnet eine Menge "Räume", in welchem zur gleichen Zeit eine Leistungskontrolle stattfindet. Informationen zur Berechnung der Tabellenwerte finden sich im Text. Abgeändert nach (Ricketts & Zakrzewski, 2005)

Risikogewichtung

Die Gewichtung der Risiken ist die Grundlage für das Risikomanagement und hat zum Ziel, die Gefahrenpotentiale aller Risiken in der Risikoliste abzuschätzen.

Das Gefahrenpotential eines Risikos (G_p) lässt sich gemäss Ricketts wie folgt abschätzen (Ricketts & Zakrzewski, 2005), wobei die Werte für die Eintrittswahrscheinlichkeit und die Konsequenz aus der Risikoanalysetabelle entnommen werden können:

$$\text{Gefahrenpotential} = \text{Eintrittswahrscheinlichkeit} \cdot \text{Konsequenz}$$

Die Gefahrenpotentiale für die CBA Risiken finden sich in Tabelle 4.10 in der Spalte "G-Potentiale". Die zweitletzte Spalte der Tabelle 2 gibt an, zu welchem Zeitpunkt im Prüfungszyklus das jeweilige Problem auftreten kann (vor einer Prüfung (B), während einer Prüfung (D) oder nach einer Prüfung (Po)). Diese Information ist zur Erstellung von Ablaufstrukturen zur Risikoelimination oder -reduktion wichtig. Zudem hat sich in der Praxis gezeigt, dass die Verhinderung oder Reduktion von Risiken während einer Prüfung oftmals zeitkritisch ist und die Konsequenzen mit zunehmender Risikobehobungsdauer zunehmen (z.B. nicht behobener Netzwerkunterbruch etc.). Deshalb schlagen wir vor, das Gefahrenpotential für zeitkritische Risiken während der Prüfung linear gemäss folgender Formel zu erhöhen (der Gewichtungsfaktor von 1.25 ist von uns geschätzt und beruht auf Erfahrungen aus unserem Unterricht):

$$\text{Gefahrenpotential (gewichtet)} = G_p \cdot 1.25$$

Abbildung 4.24 zeigt die gemäss ihres gewichteten Gefahrenpotentials absteigend sortierten Risiken. Dieses Diagramm dient im folgenden Abschnitt zur Entwicklung von Risikomanagement-Strategien.

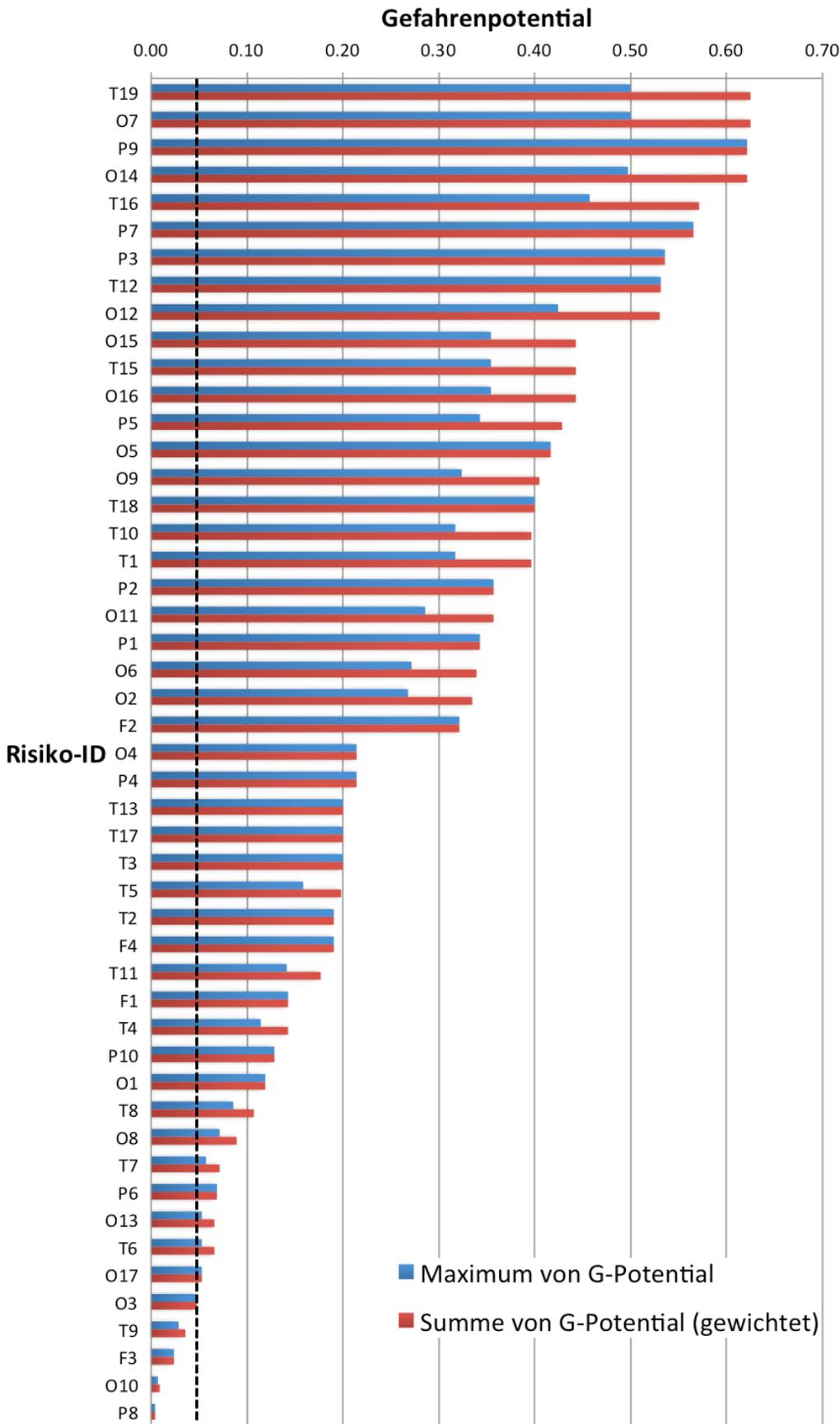


Abbildung 4.24: Rangierte Risiken. Die Risiken sind gemäss ihres Gefahrenpotentials absteigend sortiert. Die gestrichelte Linie bezeichnet den *cutoff*-Wert bei 0.04, unterhalb derer die Risiken als vernachlässigbar eingestuft werden.

4.4.5 Risikomanagement

Anstrengungen zur Reduktion oder Überwachung von Risiken im CBA-Prozess verursachen Kosten. Um die vorhandenen Ressourcen optimal einsetzen zu können, ist es wichtig, die Risiken und deren Auswirkungen auf den Unterricht genau zu kennen. Einerseits können so die Risiken priorisiert bearbeitet und andererseits durch die Festlegung eines *cutoff*-Werts jene Risiken von der weiteren Bearbeitung ausgeschlossen werden, welche entweder sehr tiefe Eintrittswahrscheinlichkeiten haben oder deren Auswirkungen auf den Prüfprozess minimal ausfallen. In der Literatur wird ein auf Erfahrung basierter *cutoff*-Wert für summative Leistungskontrollen von 0.04 angegeben (Zakrzewski & Steven, 2000). Er kann jedoch je nach Art der Leistungskontrolle verändert werden. Bei einer aus legaler Sicht wenig problematischen formativen Leistungskontrolle können beispielsweise Ressourcen eingespart werden, indem der *cutoff*-Wert erhöht und dadurch weitere Risiken als vernachlässigbar taxiert werden.

Abbildung 4.24 zeigt die sortierte Risikoliste und listet 3 von den total 49 Risiken auf, welche bei der Durchführung summativer CBA nicht mehr berücksichtigt werden müssen.

Für die verbliebenen 46 Risiken wird die Bearbeitungspriorität gemäss ihres gewichteten Gefahrenpotentials festgelegt. Die Risikomanagement-Strategie aller Risiken ist Bestandteil unseres Prüfungsablaufs. Sie sind in Form von Checklisten oder Arbeitsabläufe während der Planung und Durchführung von CBA präsent. Im technischen Report, welcher unsere Erfahrungen mit der Computer-basierten Durchführung der Basisprüfung im Jahr 2010 beschreibt, sind unsere Risikomanagement-Strategien festgehalten (Dahinden & Hinterberger, 2010).

Falls bei einem Risiko mit hohem Gefahrenpotential keine geeigneten Gegenmassnahmen getroffen werden können, ist es möglich, dass aus Sicherheitsgründen auf CBA verzichtet werden muss. Diese Entscheidung wird im nachfolgenden Schritt erläutert.

4.4.6 Go/no go-Entscheidung

Die Entscheidung, ob CBA unter Berücksichtigung der vorhandenen Rahmenbedingungen als sinnvoll erachtet werden kann, verstehen wir als Meilenstein im Modell der Risikoabschätzung. An diesem als *Go/no go* bezeichneten Zeitpunkt bietet sich die Möglichkeit, den CB Prüfvorgang abubrechen, sollte das Risiko eines Misserfolgs als zu hoch eingestuft werden.

Als hohes Risiko könnte zum Beispiel ein technisch nicht ausgereiftes Prüfungssystem eingestuft werden, welches aufgrund von unzureichenden Sicherheitsmechanismen den Prüfvorgang nur lückenhaft belegen kann und somit den legalen Anforderungen möglicherweise nicht genügt (technisches Risiko). Die Anforderungen an ein CBA-System werden detailliert im Kapitel 5 besprochen. Als hohes Risiko wäre aber auch ei-

ne CBA Prüfung einzustufen, bei der die zu prüfende Studierendenzahl zu gering ist, um diese effizient abwickeln zu können (finanzielles Risiko). Zur Abschätzung der Effizienz eines CBA-Systems verweisen wir auf die Auflistung des Zeitaufwands bei CBA in Tabelle 5.3.

4.4.7 Durchführung CBA

Die Durchführung von CBA basiert auf den Erkenntnissen der Risikomanagement-Strategien. Grundsätzlich gilt: Je besser man auf die zu erwartenden Risiken vorbereitet ist, desto effizienter und effektiver können allfällig auftretende Probleme gemeistert werden. Um den nachhaltigen Einsatz von CBA sicherzustellen, ist es zudem ratsam, den Prüfungsablauf zu dokumentieren und Erkenntnisse festzuhalten. Diese können wichtige Informationen zur Überarbeitung der Risikomanagement-Strategien liefern.

Eine mehrmalige Durchführung einer CBA kann unter der Voraussetzung einer wiederholten Evaluation der Zielerreichung und Optimierung des CBA-Setups zu einer CBA-Strategie für eine Bildungsinstitution weiterentwickelt werden. Dieses Vorgehen wird im nachfolgenden Abschnitt kurz erläutert.

4.4.8 Risikomanagement als Basis zur Entwicklung einer CBA-Strategie für Bildungsinstitutionen

Ein erfolgreiches Risikomanagement bei CBA hat zum Ziel, möglichst viele Risiken zu eliminieren, deren Wahrscheinlichkeit zu reduzieren oder – sollten sie trotzdem eintreten – deren Auswirkungen abzuschwächen. Es beinhaltet detaillierte Prozeduren zur Risikoverhinderung und -reduktion, deren Implementierung zu einem Qualitätssystem führen kann. Dieses zeichnet sich dadurch aus, dass die zentralen Ablaufstrukturen bei CBA definiert und die Zuständigkeiten aller Beteiligten klar geregelt sind. Dazu müssen die Risikomanagement-Strategien in der Praxis evaluiert und anhand eines Reengineering-Prozesses optimiert werden. Zackrezwski und Steven schlagen vor, dass unter Zuhilfenahme des *Catherine-Wheel*-Prinzips ein initialer Pilot-CBA zu einer bildungsinstitutionsweiten CBA-Strategie weiterentwickelt werden kann (Zakrzewski & Steven, 2000). Sie schlagen folgendes Vorgehen vor, wobei das in den vorangegangenen Abschnitten vorgestellte Risikomanagement in die Schritte 1 und 2 des *Catherine-Wheel*-Prinzips einfließen:

- 1) Planung
- 2) Risiko Analyse und Risiko Management
- 3) Prüfungsdesign
- 4) *Evolutionary Development*
- 5) Evaluation

Die Schritte 4 und 5 stellen zwei zentrale Schritte des Reengineering-Prozesses dar, da sie Evaluationsinstrumente beinhalten, die eine nachhaltige Etablierung von CBA an einer Bildungsinstitution ermöglichen. Zu den in diesem Bereich am weitest verbreiteten Evaluationsinstrumenten zählen Fragebogen zur Befragung der *Stakeholder* sowie der kritische Vergleich der Praxiserfahrungen mit den in der CBA-Strategie vordefinierten Zielen und Ablaufstrukturen. Da die Entwicklung einer CBA-Strategie für die ETH Zürich nicht im Fokus dieser Arbeit liegt, beschränkten wir uns im nachfolgenden Abschnitt darauf, einige Umfrageresultate zu veröffentlichen. Diese können gegebenenfalls zu einem späteren Zeitpunkt für die Entwicklung einer CBA-Strategie verwendet werden.

4.5 Evaluationsresultate

In diesem Abschnitt stellen wir einige ausgewählte Evaluationsergebnisse unserer 9-jährigen Erfahrung mit CBA vor. In dieser Zeit haben wir zwei verschiedene CBA-Systeme eingesetzt: Von 2005-2008 setzten wir eine Eigenentwicklung eines Browser-basierten Online-Prüfungssystems für formative Leistungskontrolle ein (Wyss, Bosia, & Dahinden, 2008). Ab 2008 wurde dieses System durch eine Neuentwicklung, genannt "Sioux" (Secure Interactive Online eXam), abgelöst, da wir während des Reengineering-Prozesses festgestellt haben, dass das Browser-basierte System die für summative Leistungskontrollen erforderlichen Sicherheitsmerkmale nicht erfüllt (mehr zu den Sicherheitsrisiken von Browser-basierten CBA-Systemen finden sich im Kapitel 5). Mit Sioux führen wir seit 2009 benotete Leistungskontrollen durch. Die hier vorgestellten Resultate beruhen ausschliesslich auf den Resultaten mit dem Sioux-Prüfungssystem.

Quantitative Resultate der Studierendenumfragen

Im Folgenden werden Resultate aus der Umfrage unter den Studierenden der beiden Lehrveranstaltungen IBPW ("Informatik für Biologie und Pharmazeutische Wissenschaften") und EvIM ("Einsatz von Informatikmitteln") mit insgesamt rund 550 Studierenden gezeigt und besprochen. Die Teilnahme an den Umfragen war freiwillig. Zum Zeitpunkt der Umfragen hatten die Studierenden die Prüfung absolviert, sie kannten ihre Ergebnisse aber noch nicht.

Die Studierenden der Lehrveranstaltung IBPW wurden nach der ersten benoteten Basisprüfung im August 2010 befragt. Die Rücklaufquote betrug 56.5% (n=152).

Die Studierenden der Lehrveranstaltung EvIM wurden nach dem benoteten Semesterendtest im Dezember 2010 befragt. Die Rücklaufquote betrug 43.9% (n=124).

Aufgrund der hohen Studierendenzahlen beider Lehrveranstaltungen mussten wir die Studierenden auf zwei aufeinanderfolgenden Serien aufteilen. In den beiden Serien wurden die Studierenden in je sechs Räumen geprüft. Die Studierenden der zweiten

Serien wurden im Hörsaal empfangen und von den Assistierenden in die Prüfungsräume begleitet. Dieses Vorgehen verlangte, dass wir den Studierenden ihre individuellen Prüfungstermine kommunizieren mussten. Wir haben diese Information in ihr persönliches Konto auf der Homepage der Lehrveranstaltung aufgeschaltet. Per Email wurden sie aufgefordert, ihren persönlichen Prüfungstermin dort nachzuschauen. Das genaue Vorgehen haben wir im technischen Report beschrieben (Dahinden & Hinterberger, 2010). Abbildung 4.25 zeigt, dass die Massnahmen zur individuellen Kommunikation der Prüfungstermine funktioniert haben.

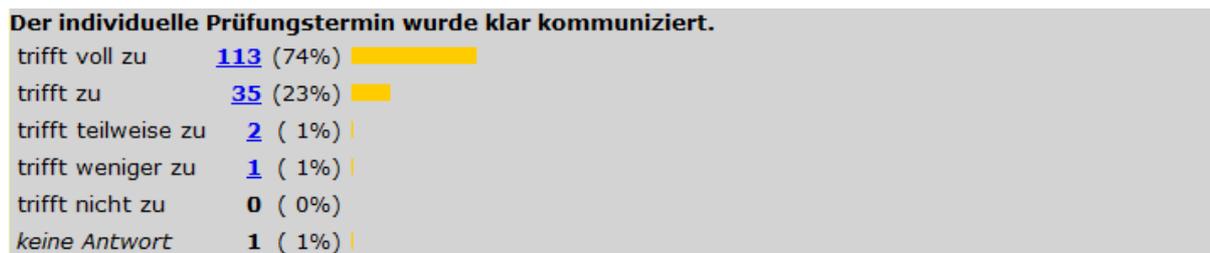


Abbildung 4.25: Kommunikation des individuellen Prüfungstermins (Datenbasis: IBPW FS10).

Aus Gründen der Gleichbehandlung der Studierenden in den verschiedenen Prüfungsgruppen haben wir in allen Räumen standardisierte Lautsprecher-Durchsagen eingesetzt. Die Durchsage gab Auskunft über rechtliche Hinweise zur Prüfung. Diese war während der Durchsage auch in Textform auf dem Bildschirm sichtbar. Wie Abbildung 4.26 zeigt, war die Durchsage klar verständlich. Im Praxiseinsatz zeigte sich, dass durch das Vorlesen der Prüfungsinformation alle Studierenden diese Hinweise mitbekommen haben. Bei früheren Prüfungsdurchführungen ohne Durchsage wurde die Prüfungsinformation oftmals ungelesen weggeklickt, um schneller mit der Prüfung beginnen zu können. Dies konnte so effizient unterbunden werden und führte dazu, dass seither während oder nach der Prüfung kaum Fragen zum Prüfungsmodus mehr gestellt wurden. Es konnte allgemein eine Beruhigung der Atmosphäre im Prüfungsraum festgestellt werden.

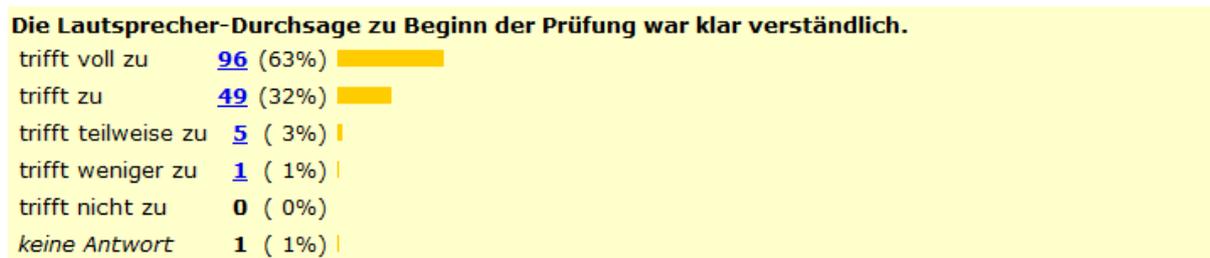


Abbildung 4.26: War die Lautsprecher-Durchsage zu Beginn der Prüfung klar verständlich? (Datenbasis: IBPW FS10).

Eine gute Organisation bei CBA ist die Grundlage für ein zuverlässiges Prüfungsergebnis. Unterschiedliche Betreuung der Prüfungsgruppen oder Abweichungen im Ablauf der Prüfung kann zu einer Ungleichbehandlung der Studierenden führen, was nicht zuletzt das Risiko einer erfolgreichen Anfechtung der Leistungsverfügung erhöht. Unsere

Organisation der CB-Prüfung, basierend auf einer hierarchischen Prüfungsleitung, einer Instruktion der Aufsichts- und Supportpersonen sowie einer permanentem Kommunikationsverbindung zwischen der Prüfungsleitung und den Aufsichtspersonen in den Prüfungsräumen hat sich in den vergangenen Semestern bewährt. Die detaillierte Planung des Prüfungsablaufs gibt uns die nötigen Freiheiten, um auf spezielle Ereignisse reagieren zu können. Dass die Studierenden die detaillierte Planung in der Praxis erkennen, zeigt Abbildung 4.27.

Die Informatik-Prüfung war gut organisiert (Betreuung durch die Assistierenden, Ablauf im Prüfungsraum)

trifft voll zu	92 (61%)	
trifft zu	53 (35%)	
trifft teilweise zu	5 (3%)	
trifft weniger zu	1 (1%)	
trifft nicht zu	0 (0%)	
keine Antwort	1 (1%)	

Abbildung 4.27: Organisation der CB-Prüfung (Datenbasis: IBPW FS10).

Unsere Massnahmen zur Schaffung vergleichbarer Prüfungsbedingungen schliesst die Studierenden mit ein. So versuchen wir den Prüfungsprozess so transparent wie möglich zu machen, indem wir die Studierenden über sicherheitsrelevante Vorgänge (Anmeldung am CBA-System, Raumdurchsagen, etc.) und das Vorgehen bei der Korrektur kommunizieren. Das Ziel dieser Informationsstrategie ist, dass wir damit erreichen wollen, dass alle Studierenden das subjektive Empfinden haben, dass alle vergleichbare Prüfungsbedingungen hatten und somit niemand bevorzugt oder benachteiligt wurde.

Um dies herauszufinden, wollten wir von den Studierenden wissen, ob ihrer Meinung nach alle Studierenden vergleichbare Prüfungsbedingungen hatten. Die Auswertung der Antworten ist in Abbildung 4.28 dargestellt. Die hohe Zustimmung von 95% der Studierenden muss jedoch im Kontext verstanden werden, dass jeder Student tatsächlich nur Einblick in seinen eigenen Prüfungsraum hatte. Dennoch kann aus diesem Resultat die Schlussfolgerung gezogen werden, dass offenbar innerhalb der Prüfungsräume vergleichbare Bedingungen herrschten und dass den Studierenden auch nach dem gegenseitigen Austausch nach Prüfungsende keine Unregelmässigkeiten aufgefallen wären.

Alle Studierenden hatten vergleichbare Prüfungsbedingungen.

trifft voll zu	65 (52%)	
trifft zu	53 (43%)	
trifft teilweise zu	4 (3%)	
trifft weniger zu	1 (1%)	
trifft nicht zu	1 (1%)	
keine Antwort	0 (0%)	

Abbildung 4.28: Subjektives Empfinden der Vergleichbarkeit der Prüfungsbedingungen durch die Studierenden (Datenbasis: EvIM HS10).

Bei der Bedienung der Prüfungssoftware traten fast keine Probleme auf (Abbildung 4.29). Lediglich ein Student oder eine Studentin berichteten von einem Problem. Das gute Abschneiden des Prüfungs-Clients kommt für uns nicht überraschend, denn wir untersuchten bereits in einer früheren Arbeit dessen *Usability* und nahmen gezielt Verbesserungen an der Bedienung und Darstellung der Applikation vor (Dahinden, 2010b). Die noch verbleibenden Probleme sind somit meist nicht Sioux-spezifisch. Dazu gehören zum Beispiel defekte oder volle USB-Sticks. In solchen Fällen verfügt die Raumaufsicht über Ersatzsticks, welche den Studierenden abgegeben werden können.

Ich hatte Probleme bei der Bedienung der Prüfungsapplikation (Sioux)

nein	121 (98%)	
ja	1 (1%)	
keine Antwort	2 (2%)	

Abbildung 4.29: *Usability* der Sioux-Prüfungssoftware (Datenbasis: EvIM HS10).

Die nächste Frage (Abbildung 4.30) erfasst die prospektive Haltung bezüglich eines erneuten Einsatzes von Sioux. Die Antworten umfassen sowohl technische als auch organisatorische Aspekte und positive Antworten lassen somit auf eine Zufriedenheit mit CBA schliessen. Es ist erfreulich, dass sich niemand grundsätzlich gegen den erneuten Einsatz einer Sioux-basierten Leistungskontrolle ausspricht. Es muss aber auch festgehalten werden, dass rund ein Fünftel der befragten Studierenden teilweise kritisch gegenüber einem erneuten Einsatz von Sioux eingestellt sind. Die Gründe dafür sind vielfältig und reichen von einer generellen Abneigung gegenüber Computern bis hin zur Problemen bei der Farbenerkennung aufgrund einer angeborenen Sehschwäche.

Angenommen, Sie hätten wieder einmal eine Online-Prüfung: Ich kann mir vorstellen, eine zukünftige Online-Prüfung wieder mit SIOUX zu absolvieren.

trifft voll zu	32 (26%)	
trifft zu	65 (52%)	
trifft teilweise zu	23 (19%)	
trifft weniger zu	4 (3%)	
trifft nicht zu	0 (0%)	
keine Antwort	0 (0%)	

Abbildung 4.30: Grundsätzliche Zufriedenheit der Studierenden mit dem Prüfungssystem (Datenbasis: EvIM HS10).

Ob diese positive Haltung der Studierenden gegenüber einer Sioux basierten-Prüfung bereits zu Semesterbeginn vorhanden war, konnten wir mit Hilfe einer Vorumfrage klären. Diese führten wir unter den Studierenden der Lehrveranstaltung "Informatik für Biologie und Pharmazeutische Wissenschaften" (Vorlesungsnummer: 552-0432-00; Frühlingsemester 2011; Dozent: Dr. Hans-Joachim Böckenhauer) durch. Die Umfrage wurde vor der ersten Vorlesung im Februar 2011 durchgeführt. Die Studierenden hatten zum Zeitpunkt der Umfrage weder eine Vorlesung noch eine Übung besucht. Das Prüfungssystem Sioux war ihnen noch nicht bekannt. Die Teilnahme an der Umfrage war freiwillig. Die Rücklaufquote betrug 73.9% (n=221).

Abschliessend kann festgehalten werden, dass die Studierenden im Allgemeinen eine positive Meinung gegenüber CBA haben. Die meisten Bedenken bezüglich technischer Probleme werden in Bezug auf die Zuverlässigkeit und Sicherheit der Prüfungsapplikation geäussert. Dass die Bedenken der Studierenden grundsätzlich begründet sind, versuchen wir im nachfolgenden Abschnitt anhand fiktiver sicherheits-kritischer Szenarien zu veranschaulichen.

4.6 Sicherheits-kritische Szenarien aktueller Online-Prüfungssysteme

Dass die Bedenken der Studierenden in Bezug auf die Zuverlässigkeit und Sicherheit von CBA begründet sind, wurde in einer Arbeit von Peter Heinrich aus dem Jahr 2008 gezeigt (Heinrich, 2008). Der Autor kam damals bei der Analyse der bestehenden Systeme zum Schluss, dass sich die führenden Hersteller von CBA-Systemen den anspruchsvollen Sicherheitsaspekten zwar bewusst sind, sich aber mangels Präzedenzfällen und Klagen mit einem relativ geringen Sicherheitsmass zufrieden geben. So begnügt man sich zum Beispiel mit einer Checksumme der Daten, um deren Unversehrtheit zu verifizieren. Dass diese Massnahme einer legalen Überprüfung vermutlich nicht standhält, soll nachfolgend anhand von drei fiktiven Prüfungsszenarien aufgezeigt werden.

Die drei Problem-Szenarien sind so realitätsnah wie möglich formuliert, so dass sie bei traditionellen CBA grundsätzlich so oder in ähnlicher Form auftreten könnten. Mit diesen Szenarien sollen mögliche Sicherheitslücken bestehender CBA-Systeme illustriert werden. Zudem geben sie uns die Möglichkeit, die Notwendigkeit von technischen Neuerungen bei traditionellen Online-Prüfungssystemen anhand konkreter Beispiele zu beschreiben. Für jedes Szenario sind die entsprechenden Risiko-Identifikationsnummern gemäss Risikotabelle (Tabelle 4.9) angegeben.

Das erste Szenario beschreibt den realistischen Ausfall eines Teils der Prüfungs-Infrastruktur und dessen Auswirkungen auf den Prüfungsablauf. Das zweite Szenario schildert den Fall, wenn absichtliche oder technisch-bedingte Manipulationen an den Antworten der Studierenden unter gewissen Umständen nicht erkannt werden können. Das dritte Szenario beschreibt die Problematik, dass bei traditionellen Online-Prüfungen verloren gegangene Antworten nicht festgestellt werden können.

4.6.1 Szenario 1: Teilausfall der Prüfungs-Infrastruktur

Dieses Szenario illustriert die Auswirkungen folgender Risiken: T8, T9, O11, O13.

Ausgangssituation

Eine Basisprüfung mit 300 Studierenden findet am Computer statt. Aus Platzgründen findet die Prüfung in zwei Serien zu je 150 Studierenden statt. Zum Einsatz kommt ein Browser-basiertes Prüfungssystem. Um einen Informationsaustausch zwischen den beiden Serien zu vermeiden, werden die beiden Serien direkt nacheinander geprüft. Die Studieren-

den der ersten Serie versammeln sich in den sieben Prüfungsräumen und melden sich dort am Prüfungssystem an. Nach einer kurzen Instruktion durch die raumverantwortliche Assistenzperson beginnen die Studierenden mit dem Lösen der Prüfung. Nach rund einer halben Stunde Prüfungszeit will eine Studentin eine Antwort an den Server übermitteln. Zuerst reagiert der Prüfungsclient eine Minute lang nicht, dann erscheint im Webbrowser die Meldung "Fehler: Server nicht gefunden".

Nach und nach verbreitet sich die Fehlermeldung, bis schliesslich alle Studierenden vom Serverunterbruch betroffen sind. Da die Raumaufsicht die Dauer des Unterbruchs nicht kennt, unterbricht sie die Prüfung. Die Studierenden warten ungeduldig, bis die Prüfung fortgesetzt werden kann. Nur mit Mühe gelingt es der Raumaufsicht, die Studierenden davon abzuhalten, miteinander zu sprechen. Nach 15 Minuten ist der Server wieder erreichbar und die Prüfung wird fortgesetzt. Für alle Studierenden wird die Prüfungszeit entsprechend verlängert. Diese Verlängerung verursacht für die Studierenden der zweiten Serie eine länger als angekündigte Wartezeit, was Unruhe unter den betroffenen Studierenden verursacht. Zwei Studierende beschwerten sich und verlassen protestierend den Prüfungsraum. Die zweite Prüfungsserie verläuft dann ohne Unterbruch. Die beiden Probanden, welche die Prüfung unter Protest abgebrochen haben, reichen ein Annullierungsgesuch ein. Begründet wird das Gesuch damit, dass sie im Vergleich mit den Studierenden der zweiten Serie benachteiligt wurden. Aufgrund des Unterbruchs und der Wartezeit hätten Sie sich nicht mehr auf die Prüfung konzentrieren können.

Problem

Ein traditionelles Browser-basiertes CBA-System verlangt aus technologischen Gründen eine permanente Verbindung zum Prüfungsserver. Jeder Wechsel der Prüfungsfrage und jedes Abschicken einer Antwort resultiert in einer Anfrage an den Prüfungsserver. Dieser liefert dann eine Antwort in Form von neuen Seiteninhalten zurück. Ein Ausfall des Prüfungsservers oder die Unterbrechung der Verbindung zwischen Prüfungsclient und Prüfungsserver führt deshalb unweigerlich zu einem Betriebsunterbruch. Dieser Unterbruch dauert so lange, bis der Server wieder erreichbar ist. Die letzte Anfrage an den Server geht dabei verloren. Falls es sich dabei um eine Antwort handelte, muss sie von den Studierenden nach der Wiederaufnahme der Prüfung erneut eingetippt und übermittelt werden. Diese Intoleranz von Online-Prüfungssystemen in Bezug auf einen Unterbruch der Kommunikation mit dem Server schränkt die Ausfallsicherheit stark ein.

Ein weiterer Nachteil von Browser-basierten Prüfungssystemen ist die eingeschränkte Skalierbarkeit. Diese macht sich bei den Studierenden in längeren Wartezeiten, zum Beispiel beim Wechsel der Prüfungsfragen oder beim Abschicken von Antworten, bemerkbar. Bei Online-Prüfungen mit hohen Studierendenzahlen kommt es oft vor, dass

das Prüfungssystem in der Schlussphase der Prüfung träge reagiert¹³. Dieser Effekt wird dadurch verursacht, weil die Studierenden zum Schluss der Prüfung noch einmal alle ihre Antworten anschauen wollen und das Nachladen der Seiteninhalte auf dem Server eine hohe Last verursacht. Offenbar ist es in einigen Fällen sogar vorgekommen, dass Antworten nicht gespeichert werden konnten und die Prüfungen nach einigen Minuten automatisch beendet wurden¹⁴.

Lösungsansatz

Bei CBA soll die Bearbeitungsgeschwindigkeit nicht durch das CBA-System, sondern ausschliesslich durch die Studierenden begrenzt sein. Das heisst, es darf nicht vorkommen, dass Studierende während der Prüfungszeit auf das Prüfungssystem warten müssen. Um dies zu erreichen, gibt es zwei Ansätze: entweder wird die technische Infrastruktur redundant und somit überdimensioniert aufgebaut (Brunner & Guretzki, 2009), oder der Client verfügt über eine teilautonome Funktionsweise (Dahinden & Hinterberger, 2008).

Im nachfolgenden Kapitel wird erläutert, wieso wir den Ansatz der teilweisen Autonomie in unserem CBA-System verfolgt haben. Wir werden aufzeigen, warum wir für die Erhöhung der Ausfallsicherheit auf die Entwicklung eines eigenständigen Prüfungsclients setzen. Dieser überträgt zu Beginn der Prüfung alle Prüfungsinhalte vom Server auf den Client und ermöglicht dadurch, dass die Studierenden ohne permanente Verbindung zum Server die Prüfung bearbeiten können. Durch diese Massnahme wird nicht nur die Ausfallsicherheit erhöht, sondern es bieten sich auch eine Reihe weiterer Vorteile:

- Die Belastung des Servers wird reduziert, da alle Prüfungsinhalte nur einmal übertragen werden müssen.
- Wenn die komplette Prüfung nur einmal übertragen wird, werden zum Beispiel auch CBA möglich, wenn die Internetverbindung langsam ist oder häufig Ausfälle auftreten.
- Durch die Autonomie des Prüfungsclients kann nach dem Bezug der Prüfung bis zu deren Abgabe komplett auf den Zugriff auf das Datennetzwerk und das Internet verzichtet werden. Dadurch werden echte *closed book*-Prüfungen möglich, wobei garantiert wird, dass niemand Inhalte austauschen oder nach Informationen im Internet suchen kann. Auch der Einsatz von Drittapplikationen wie zum Beispiel matlab® oder Programmierumgebungen können ohne Sicherheitsbedenken eingesetzt werden.

¹³ schriftliches Feedback, welches uns im Rahmen unserer Unterrichtsevaluation erreichte. Frühjahrssemester 2010.

¹⁴ mündliches Feedback während der Prüfungseinsicht der Lehrveranstaltung IBPW, September 2010.

Dieser Lösungsansatz für die Umsetzung eines zuverlässigen CBA-Systems wird im Abschnitt 5.3.2 "CBA-Designprinzip #2: Ausfallsicherheit dank Teilautonomie der Client-Applikation" ausführlich beschrieben.

4.6.2 Szenario 2: "Das ist nicht meine Antwort!"

Dieses Szenario illustriert die Auswirkungen folgender Risiken: T13, T17, T18, T19.

Ausgangssituation

Eine Studentin schreibt eine Prüfung unter Verwendung eines Browser-basierten CBA-Systems. Nach Abschluss der Prüfung erstellt die Prüfungsleitung eine Sicherheitskopie aller auf dem Server vorhandenen Antworten. Nach einem negativen Notenbescheid vereinbart die Studentin einen Termin für die Prüfungseinsicht. Bei der Durchsicht der Prüfung kommt sie zum Schluss, dass das Resultat der Aufgabe 2.1 unmöglich von ihr sein kann, da sie niemals eine solche Antwort geben würde. Sie ersucht auf dem Rechtsbehelfs-Weg um die Annullierung der Note, da ihrer Überzeugung nach der rechtserhebliche Sachverhalt nicht korrekt festgestellt wurde.

Feststellung des rechtserheblichen Sachverhalts

Da die Bildungsinstitution in der Beweispflicht ist, muss sie die korrekte Feststellung des rechtserheblichen Sachverhalts belegen können. Dazu stellt sie alle Antworten zusammen, welche im Namen des betroffenen Studenten auf der Sicherheitskopie gespeichert sind. Jede Antwort enthält Informationen zum Studenten und dessen Antwort.

Problem

Um die Authentizität der Prüfungsergebnisse belegen zu können, muss das CBA-System alle Antworten mit geeigneten technischen Sicherheitsmerkmalen versehen. Diese Merkmale sollen helfen, die folgenden Fragen zu klären:

- Wer hat die Aufgabe verfasst?
- Wurde sie nachträglich verändert?

Um den von den Gesetzgebern geforderten Beweis zweifelsfrei liefern zu können, müssen diese Sicherheitsmerkmale zusätzlich noch vor Veränderungen geschützt werden. Es wurde festgestellt, dass dieser Schutz bei CBA-Systemen fehlt, wodurch die Sicherheitsmerkmale ihre legale Aussagekraft verlieren dürften (Heinrich, 2008).

Lösungsansatz

Um den rechtserheblichen Sachverhalt eines Prüfungsergebnisses zweifelsfrei nachweisen zu können, müssen die Antworten und deren Sicherheitsmerkmale vor Veränderungen geschützt werden. Ein Schutz vor Veränderungen bietet beispielsweise der Ausdruck des Prüfungsdokuments am Ende der Prüfung und dessen Visierung durch

den Studenten (Küstermann et al., 2005). Dieses Verfahren skaliert jedoch schlecht mit steigenden Studierendenzahlen. Der Abschnitt 5.3.3 "CBA-Designprinzip #3: Belegbarkeit der Prüfungsergebnisse durch client-seitige, digitale Signaturen" wird sich deshalb detailliert mit dieser Anforderung auseinandersetzen und zusätzlich aufzeigen, dass die digitale Signierung der Sicherheitsmerkmale einer Antwort auf dem Client und nicht erst auf dem Server geschehen soll.

4.6.3 Szenario 3: "Wo ist meine Antwort der Prüfungsaufgabe 10?"

Dieses Szenario illustriert die Auswirkungen folgender Risiken: T13, T17, T18, T19, F1.

Ausgangssituation

Ein Student beginnt mit der Bearbeitung einer CBA-Prüfung. Er löst zuerst alle einfacheren Aufgaben. Dann beschliesst er, die aufwändigste Aufgabe, die Aufgabe 10, zu überspringen und stattdessen 15 Minuten zu warten und keine Antwort mehr abzugeben. Er beendet dann die Prüfung, worauf die Prüfung digital auf dem Server gespeichert wird.

Nach der Notenverfügung meldet er sich beim Prüfungsleiter und vereinbart einen Termin für die Prüfungseinsicht. Dort lässt er sich seine Prüfung und die Korrekturen zeigen. Er wird fragen, wo seine Antwort der Aufgabe 10 wäre. Der Prüfungsleiter wird entgegnen, dass dies alle Antworten seien, die von ihm in der Datenbank gespeichert wurden. Der Student wird behaupten, dass er 15 Minuten an der Aufgabe gearbeitet hätte und dass seine Antwort nun offenbar verloren gegangen wäre. Um seine Glaubwürdigkeit zu erhöhen, wird er dem Prüfungsleiter die Lösung der Aufgabe, welche er nach der Prüfung von seinen Mitstudierenden in Erfahrung gebracht hat, kurz mündlich mitteilen.

Feststellung des rechtserheblichen Sachverhalts

Der Student reicht daraufhin innerhalb der gesetzlichen Frist von 30 Tage gegen die Notenverfügung ein Wiedererwägungsgesuch ein. Er begründet sein Gesuch mit der Behauptung, dass seine Antwort verloren gegangen und er somit im Vergleich zu seinen Mitstudierenden benachteiligt worden wäre. Aufgrund der Beweispflicht wird die Bildungsinstitution vom Prüfungsleiter alle Prüfungsunterlagen verlangen. Dieser wird folgende Prüfungsunterlagen bereitstellen können: Log-Datei des Prüfungsablaufs, Zwischenresultate des Studenten, Korrektur der Antworten, Bewertungsschema und eventuell das Ablaufprotokoll der Prüfung.

Problem

Ein traditionelles Browser-basiertes CBA-System schickt die Antworten über eine sichere Verbindung an den Server und dieser speichert sie in einer Datenbank. Nach Abschluss der Prüfung werden die dort gespeicherten Daten wieder herausgelesen und zur Korrektur verwendet.

Das Problem ist, dass Browser-basierte CBA-Systeme, deren gesamte Programmlogik sich auf dem Server befindet, keine geeigneten technischen Möglichkeiten aufweisen, um verloren gegangene Antworten zuverlässig erkennen zu können. Es kann lediglich festgestellt werden, welche Aufgaben existieren, aber es wird nicht möglich sein zu überprüfen, ob Antworten verloren gegangen sind. Eine verloren gegangene Antwort hinterlässt auf dem Server (weder als gespeicherte Antwort noch als Eintrag in einer Log-Datei) keine Spur. Dieser Nachweis wäre jedoch notwendig, um den Vorwurf des Studenten zu entkräften. Eine Reduktion der Gesamtpunktzahl für den betroffenen Studierenden wäre auch keine sinnvolle Lösung, da so dieser Student bevorteilt würde, indem nur die leichteren Aufgaben zur Korrektur und Notengebung verwendet werden.

Lösungsansatz

Beim Nachweis von verlorenen Antworten haben die meisten CBA-Systeme den Ansatz gewählt, dass die Studierenden die Vollständigkeit der Antworten überprüfen können. Im einfachsten Fall können die Studierenden während der Prüfungsbearbeitung eine fehlende Antwort erkennen und diese erneut abschicken. Es gibt aber auch den Ansatz, das Risiko von verlorenen Antworten mittels redundanten Servern zu reduzieren (Möbs, 2008) oder die Prüfungen am Ende auszudrucken und dessen Vollständigkeit per Unterschrift bestätigen zu lassen (Küstermann et al., 2005).

Im Abschnitt 5.3.4 "CBA-Designprinzip #4: Nachweis der Vollständigkeit des Prüfungsdossiers dank durchgehender Nummerierung aller Antworten" wird ein technischer Ansatz beschrieben, der ressourcenschonend implementiert werden kann und auch gut mit steigenden Studierendenzahlen skaliert. Dieser Ansatz basiert auf der client-seitigen Nummerierung und digitalen Signierung aller Antworten.

4.7 Diskussion der Fragestellung 1: Erfüllen aktuelle CBA-Systeme die legalen Anforderungen?

In diesem Kapitel haben wir ein Modell für das Risikomanagement bei CBA vorgestellt. Dieses listet insgesamt 49 verschiedene didaktische, operationale, technische und finanzielle Risiken bei CBA auf und gewichtet sie anhand derer Auswirkungen auf den CBA-Prüfprozess. Diese Gewichtung erlaubt es einerseits, die vorhandenen Ressourcen effizient und effektiv einsetzen zu können, indem beispielsweise vernachlässigbare Risiken nicht weiter berücksichtigt werden. Andererseits gibt sie einen Hinweis auf die grössten Risiken bei CBA. So konnte aufgezeigt werden, dass es aktuell zwei technische Hauptschwierigkeiten bei CBA-Systemen gibt:

1. Aktuelle CBA-Systeme verfügen nur über rudimentäre technische Möglichkeiten, mit denen der Prüfungsablauf protokolliert werden kann (Risiko-ID: T19)
2. Aktuelle CBA-Systeme sind anfällig auf Ausfälle und Überlastung der Infrastruktur (Risiko-ID: T16)

Da bis heute keine rechtsverbindlichen Vorgaben für CBA existieren, ist unklar, ob die bestehenden Sicherheitsmerkmale der aktuellen Prüfungssysteme ausreichen, um den von den Gesetzgebern geforderten rechtserheblichen Sachverhalt zweifelsfrei belegen zu können. Aus unserer Sicht dürfte dies aufgrund des fehlenden physischen oder digitalen Prüfungsdokuments nicht der Fall sein. Dies hätte gravierende Auswirkungen auf den Prüfungsalltag der meisten Bildungsinstitutionen, da per Gerichtsbeschluss mit den aktuellen CBA-Systemen keine summativen Leistungskontrollen mehr durchgeführt werden könnten. In Studiengängen mit vielen Studierenden wäre der Verzicht auf CBA mit einem massiven Mehraufwand für die Korrektur verbunden. Bei Leistungskontrollen mit multimedialen Inhalten wäre der Verzicht auf den Computer gleichbedeutend mit einer Qualitätsreduktion der Leistungskontrolle. Wir haben uns deshalb zum Ziel gesetzt, alternative Ansätze und Möglichkeiten zu suchen, wie zukünftige CBA-Systeme konzipiert sein müssen, um die rechtsverbindliche Vorgaben zu erfüllen.

Eine denkbare Variante zur Umgehung des Problems mit dem fehlenden physischen Prüfungsdokuments wäre das Ausdrucken und Unterschreiben einer digitalen Prüfung, wie es an der Universität Karlsruhe gemacht wird (Küstermann et al., 2005). Dieses Vorgehen ist aber nur eingeschränkt praktikabel, denn es skaliert aufgrund der begrenzten Druckkapazitäten nicht mit hohen Studierendenzahlen. Auch der Einsatz von multimedialen Inhalten ist nur eingeschränkt möglich. Wir sehen deshalb die Lösung in einem elektronischen Prüfungsdokument, das mittels einer digitalen Signatur gegen nachträgliche Veränderungen geschützt wird.

In den folgenden beiden Kapiteln wird mit Sioux ein "Generation R" Prüfungssystem vorgestellt, das unserer Meinung nach über die nötigen technischen Möglichkeiten verfügt, um einen fairen Prüfungsablauf zu ermöglichen und diesen auch rechtsgültig zu protokollieren. Neben technischen Neuerungen verfügt Sioux auch über Eigenschaften, die formative Leistungskontrolle mit sofortiger Korrektur und individualisiertem Feedback ermöglichen. Wie diese Funktionen einen wichtigen Beitrag für den aktuellen Forschungsbereich *learning analytics* liefern können, wird Inhalt des Kapitels 8 sein.

Implementierung des CBA-Systems

Kapitel 5

Designprinzipien eines sicheren CBA-Systems

5.1 Ausgangslage

Nüchtern betrachtet muss festgestellt werden, dass bisher in der Schweiz der Sicherheit von CBA-Systemen nur wenig Aufmerksamkeit geschenkt wurde. So ist es eine weitverbreitete Meinung, dass aktuelle CBA-Systeme den rechtserheblichen Sachverhalt ausreichend sicherstellen und dass der Einsatz von studentischen Laptops die Lösung für begrenzte Raumkapazitäten bei CBA-Prüfungen darstellt. Dabei sollte gerade die Kombination von Browser-basierten-CBA-Systemen und studentischen oder öffentlichen Computern kritisch betrachtet werden. Wie in diesem Kapitel aufgezeigt wird, beinhaltet diese Kombination verschiedene Sicherheitsrisiken, welche aus unserer Sicht ein erhebliches, jedoch grösstenteils vermeidbares Rekursrisiko darstellen.

Die Implementierung von sicheren Softwareanwendungen setzt aus Sicht von Informationssicherheitsspezialisten den Einsatz spezialisierter Hardware voraus (Böck, Huemer, & Tjoa, 2010). Bei Client-Server-Applikationen, zu denen auch CBA-Systeme zählen, stehen die Client-Arbeitsstationen im Zentrum der Sicherheitsüberlegungen. Diese nehmen die Eingaben der Benutzer entgegen und leiten diese zur persistenten Speicherung an einen Server weiter.

Da an den Schweizerischen Bildungsinstitutionen für CBA keine spezialisierte Client-Hardware wie z.B. dedizierte Prüfungsclients oder digitale Signatureinheiten zur Verfügung stehen, wurde für diese Arbeit folgende Zielsetzung definiert:

"Unter Berücksichtigung der vorhandenen Infrastruktur wird ein CBA-System geschaffen, welches sich in der Praxis durch einen für CBA neuartigen Datenschutzmechanismus auszeichnet und gleichzeitig dank Teilautonomie ein hohes Mass an Ausfallsicherheit und Flexibilität speziell in heterogenen Umgebungen bietet. Dieses System wird durch eine kritische Analyse des Restrisikos ergänzt, so dass CBA-Stakeholder individuell entscheiden können, ob sie bereit sind, das Risiko einer CBA zu tragen."

Dieses Kapitel ist wie folgt gegliedert: Im ersten Teil werden die aktuellen CBA-Systeme in Bezug auf deren Sicherheitsmerkmale verglichen und zwei Mängel beschrieben, die

diese Systeme aufweisen. Anschliessend werden fünf Designprinzipien beschrieben, welche unserer Erkenntnis nach essentiell für eine nachhaltige und sichere CBA-Prüfungsdurchführung sind. Zum Schluss des Kapitels werden drei Sicherheitsrisiken besprochen, welche die Designprinzipien bei der Umsetzung in die Praxis zur Folge haben können.

5.2 State-of-the-Art aktueller CBA-Systeme

5.2.1 Abgrenzung von CBA gegenüber ähnlichen sicherheitskritischen Client-Server-Applikationen

CBA-Systeme haben Ähnlichkeiten mit technischen und legalen Anforderungen an elektronische Gesundheitskarten und Applikationen für elektronische Abstimmungen (e-Voting).

Die elektronischen Gesundheitskarten unterscheiden sich allerdings von CBA-Applikationen, da dort der *Datenschutz* eine entscheidende Rolle einnimmt. So muss bei der elektronischen Gesundheitskarte unbedingt verhindert werden, dass unberechtigte Personen Zugriff auf sensible Daten erlangen können. Bei CBA hingegen steht die *Datensicherheit* im Zentrum, da jederzeit und unmissverständlich klar sein muss, wer die Urheberschaft der Daten beansprucht und ob sie unverändert vorliegt.

Bei der Referenzimplementierung einer sicheren *e-Voting*-Applikation wird die Datensicherheit in der Regel noch um zwei Aspekte erweitert: Zum einen muss das *e-Voting*-System eine *multi-party* Architektur aufweisen, um Wahlbetrug oder Stimmenfalschzahlungen zu verhindern, zum anderen soll der Stimmenkauf verhindert werden, indem die Stimmpersonen Ihre Wahl nachträglich nicht beweisen können (*receipt-freeness*) (Hirt, 2001, Kapitel 5.1.1).

Im Gegensatz zu *e-Voting*-Applikationen stellt CBA erhöhte Anforderungen an den Funktionsumfang der Clientsoftware und an das Client-System. So muss mittels geeigneter Massnahmen verhindert werden, dass Studierende während der Prüfung auf unerlaubte Ressourcen zugreifen, die Prüfungsapplikation verlassen oder die Prüfungsfragen im Hintergrund mitschreiben. Zudem muss die Möglichkeit bestehen, dass eine begonnene Prüfung nach einem Abbruch fortgesetzt werden kann, ohne dass dabei bereits abgegebene Antworten verloren gehen.

5.2.2 Übersicht über die häufigsten in der Schweiz eingesetzten CBA-Systeme

Zum Zeitpunkt der Drucklegung dieser Arbeit befinden sich zahlreiche elektronische Prüfungssysteme auf dem Markt. In der schweizerischen Prüfungslandschaft dominieren die Open-Source Lösungen. Diese primär als Learning-Management-Systeme (LMS) konzipierten Systeme verfügen über integrierte Prüfungsmodule und werden an den

Bildungsinstitutionen meist von den Informatikdiensten betrieben. In Deutschland sind vielfach auch kommerzielle Prüfungssysteme im Einsatz (Benning, 2008; Bücking, Schwedens, & Laue, 2007). Dabei handelt es sich um reine Prüfungssysteme, welche mittels Schnittstellen an Hochschulinformationssysteme angebunden werden können.

Tabelle 5.12 zeigt eine Übersicht über die bekanntesten, aktuellen Prüfungssysteme, wobei Moodle mit 56% Marktanteil in der Schweiz am weitesten verbreitet ist (Altermatt, Dahinden, Rieder, & Zimmermann, 2013).

	ILIAS	MOODLE	OLAT	QM	LPlus
Lizenzart	OS	OS	OS	kommerziell	kommerziell
Typ	LMS	LMS	LMS	CBA-System	CBA-System
Konzeption	BB	BB	BB	BB	BB
Ausfallsicherheit	Resume	Resume	Resume Clustering	Resume Redundanz	Resume Redundanz
Datensicherheit	https	https	https	https Prüfsummen	https Ausdruck & Unterschrift
Besonderheiten				eigener Secure- Browser	eigener Secure- Browser
Literatur	(Kunkel, 2011)	(HdEL, 2011)	(MELS, 2009)	(Questionmark, 2010)	(Bücking et al., 2007)

Tabelle 5.12: Übersicht über die aktuellen CBA-Systeme. Es gelten folgende Abkürzungen: QM = *Questionmark Perception*, OS = *Open-Source*, BB = *Browser-basiert*, *Resume* = Fortsetzung der Prüfung an einem anderen Computer möglich, https = Verschlüsselte Verbindungen zwischen Client und Server.

Neben diesen weit verbreiteten CBA-Systemen stehen verschiedene spezialisierte Prüfungssysteme im Einsatz. Diese wurden meist spezifisch für einzelne Lehrveranstaltungen entwickelt. Dazu zählt zum Beispiel die *Matlab*-Prüfungsumgebung des D-MATH (Schmucki, 2010, Seite 9) oder die auf dem *DOMjudge* basierte Prüfungslösung des D-INFK (Clifton, 2010). Aufgrund der spezifischen Einsatzgebiete und der begrenzten Ressourcen verfügen diese Systeme meist über keine aufwändigen technischen Sicherheitsmerkmale. So setzen sie zum Beispiel auf *Cronjobs*, um die Daten der Studierenden in regelmässigen Abständen automatisch zu speichern. Die meisten dieser Systeme sind dezentral organisiert und die Antworten der Studierenden müssen anschliessend

manuell oder per Skript eingesammelt werden. Die Authentizität und Integrität wird meist mittels ausgedruckter und unterschriebener Prüfungsdokumente sichergestellt. Aufgrund der Vielfalt an technischen Implementierungen und der geringen Verbreitung dieser Systeme werden im nachfolgenden diese Systeme nicht mehr weiter analysiert.

Wie aus der Tabelle 5.12 ersichtlich ist, sind die aktuellen CBA-Systeme als Browserbasierte Systeme konzipiert. Verglichen mit Client-Server-Applikationen haben diese Systeme folgende Vorteile:

- Moderne Browser unterstützen eine Vielzahl multimedialer Inhalte (Audio, Video, Animationen, etc.), deren Integration in eine Prüfungsfrage einfach ist.
- Die Technologie ist weit verbreitet und es ist einfacher, kompetente Personen für die Softwareentwicklung zu finden.
- Die Entwicklungszeiten können kürzer gehalten werden, da der Browser über mächtige *Rendering*-Funktionen zur Darstellung der Webseite verfügt.
- Die Plattformunabhängigkeit kann unter Einhaltung der html-Spezifikationen und Verwendung eines einheitlichen Browsers grösstenteils sichergestellt werden.

Bei Browser-basierten Systemen benötigt der Client eine permanente Verbindung zum Server, da die User-Interaktionen vom Server verarbeitet und in Form von neuen Bildschirminhalten zurückgeschickt werden. Daraus ergeben sich zwei Hauptprobleme, die speziell bei high-stake Leistungskontrollen ein Problem darstellen: reduzierte Ausfallsicherheit und ungenügende Datensicherheit.

5.2.3 Manko 1: Ausfallsicherheit browserbasierter Systeme

Unterbrechungen der Verbindung zwischen den Clients und dem Prüfungsserver führen dazu, dass die Probanden keine Antworten mehr abschicken und keine neuen Inhalte mehr vom Server übertragen können. Im Fall kurzer Unterbrechungen (<20 Sekunden) führt dies zu verlängerten Reaktionszeiten. Für die Probanden macht sich dies in Form eines trägen Bedienungsverhaltens der Prüfungsapplikation bemerkbar. Bei längeren Unterbrechungen kann je nach Browsereinstellung die Bearbeitung der Prüfung mit dem Hinweis unterbrochen werden, dass die Verbindungen zum Server verloren gingen. Sowohl bei kurzen als auch bei längeren Unterbrechungen wird die Bearbeitung der Leistungskontrolle während der Dauer des Unterbruchs verhindert.

Kurze Unterbrechungen werden im Praxisalltag meist durch eine Überlastung des Prüfungsservers oder des Netzwerks hervorgerufen. Die Lastspitzen treten dabei zu Beginn und am Ende der Prüfung auf, wenn die Mehrheit der Probanden rasch zwischen den Prüfungsfragen wechseln und somit vermehrt Inhalte vom Server auf den Client übertragen werden müssen.

Diese Lastspitzen lassen sich mit technischen Mitteln wie Lastverteilern und redundanten Systemen reduzieren, diese erhöhen allerdings den Aufwand für den technischen Unterhalt und sind teilweise mit hohen Investitionskosten verbunden.

5.2.4 Manko 2: Datensicherheit browserbasierter Systeme

Bei einem Browser-basierten CBA-System beeinträchtigen primär zwei Eigenschaften die Datensicherheit:

- Die Daten werden ausschliesslich auf dem Server gespeichert. Fehlende oder veränderte Resultate können nicht effizient erkannt werden (fehlender Nachweis der Integrität und Identität von Prüfungsergebnissen).
- Die Darstellung der Prüfung wird durch eine Drittapplikation (Browser) übernommen. Unterschiedliche Browser und Betriebssysteme können die Prüfungsinhalte unterschiedlich darstellen, was zum einen legale Fragen aufwirft und zum anderen den Einsatz von neuen Item-Typen erschwert (fehlender Nachweis der objektiven Prüfungsführung).

Nachweis der Integrität und Identität der Prüfungsergebnisse

Bei einer Browser-basierten Applikation werden die Daten auf dem Client bloss zwischengespeichert. Für die persistente Datenspeicherung ist ausschliesslich der Server zuständig. Dieser Speichervorgang verläuft in sechs Schritten (Abbildung 5.33).

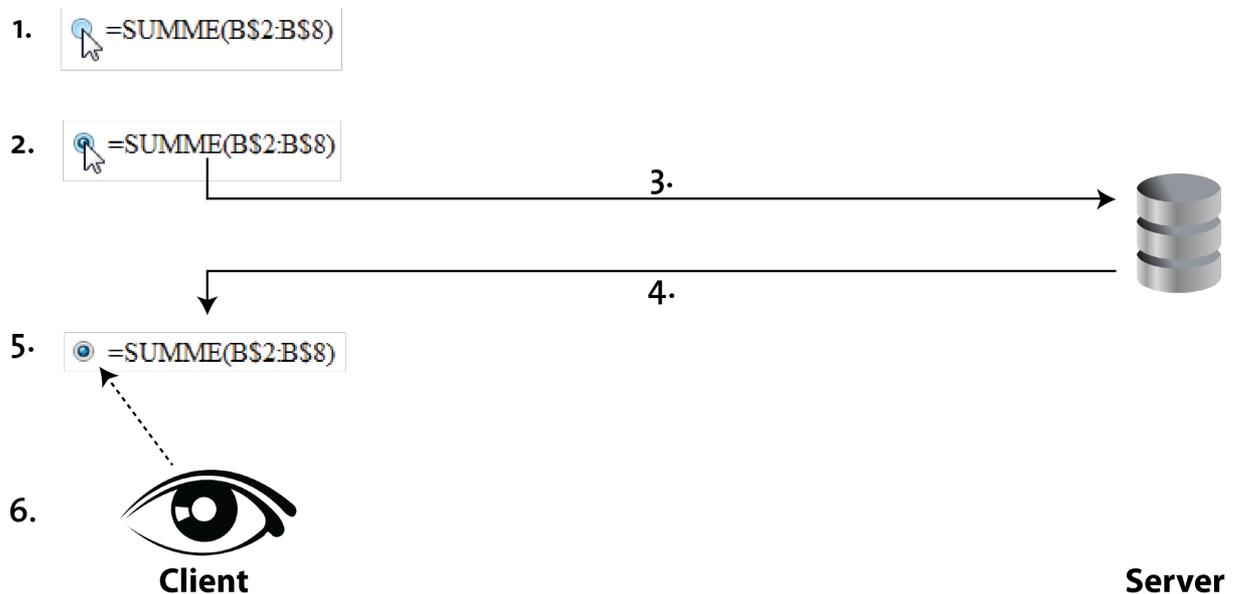


Abbildung 5.33: Speicherung der Antworten bei aktuellen Browser-basierten CBA-Systemen geschieht grob in sechs Schritten (hier illustriert am Beispiel einer *Multiple Choice* (MC) Frage): 1. Die Antwort wird durch Anklicken des Radiobuttons abgeben. 2. Der Radiobutton ändert den Zustand (von "nicht markiert" auf "markiert" oder umgekehrt). 3. Die Antwort wird an den Server übermittelt und dort gespeichert. 4. Der gespeicherte Wert wird aus der Datenbank gelesen und die aktualisierte Webseite auf den Client übertragen. 5. Die Webseite wird auf dem Client gerendert (dargestellt). 6. Die gespeicherte Antwort kann abgelesen werden.

Daraus ergibt sich das Problem, dass die Studierenden erst nach dem Abwarten der Serverantwort überprüfen können, ob ihre Antwort gespeichert wurde. Erschwerend kommt hinzu, dass die Antwort beim Anklicken zwar sofort sichtbar wird (z.B. markierte *Checkbox*), aber abhängig von der Verbindungsgeschwindigkeit noch nicht auf dem Server gespeichert wurde. Erst die Antwort des Servers zeigt an, ob die Daten wie gewünscht gespeichert wurden.

Dies kann zu zwei Problemen führen: Ohne explizite Kontrolle der Serverantwort durch die Studierenden kann es passieren, dass die Resultate entweder nicht oder verändert auf dem Server gespeichert wurden. Technische Lösungen, mit denen auf dem Server versucht wird, fehlende oder veränderte Antworten zu erkennen, sind wenig sinnvoll. So können auf dem Server nur Antworten erfasst und vor Veränderungen geschützt werden, die den Server erreicht haben. Verloren gegangene, nie abgeschickte oder veränderte Antworten können so auf dem Server nicht zweifelsfrei festgestellt werden. Diese Nachweise können ausschliesslich auf dem Client erbracht werden.

Nachweis der objektiven Prüfungsführung

Bei Browser-basierten CBA-Systemen wird die Darstellung (Rendering) der Prüfungs-Webseite von einem Browser übernommen. Es ist allgemeint bekannt, dass die Browser html-Seiten abhängig vom Betriebssystem, der Bildschirmauflösung und der verwendeten html-Elementen unterschiedlich darstellen. Zudem kann clientseitig mittels weit verbreiteter Methoden wie *CSS* und *Javascript* der Inhalt der Webseiten verändert werden (Bouvin, 1999; Paulson, 2005). Dadurch wird es schwierig sicherzustellen, ob die Prüfungsinhalte auf allen Clients nachweisbar identisch dargestellt (gerendert) wurden.

5.3 Anforderungen an ein sicheres CBA-System

Bei der Entwicklung eines CBA-Systems sollte dem Grundsatz gefolgt werden, dass die Fehlerwahrscheinlichkeit eines Softwaresystems mit seiner Komplexität steigt (Rosi, Maccari, & Fantacci, 2007; Valido-Cabrera, 2006). Es gilt also ein System zu entwickeln, welches basierend auf unseren Anforderungen ein Maximum an Sicherheit und Benutzerfreundlichkeit bietet.

Zu unseren Anforderungen zählen (in Klammern angegeben sind jeweils die Nummern der Designprinzipien, in welche diese Anforderung eingeflossen sind):

- Das System muss die durch die Infrastruktur und Ressourcen vorgegebenen Rahmenbedingungen berücksichtigen (Designprinzip: #1).
- Das System muss ausfallsicher sein (Designprinzip: #2).

- Das System muss den Prüfungsprozess nachvollziehbar machen (Designprinzipien: #3 und #4).
- Das System muss einfach zu bedienen sein (Designprinzip: #5).
- Das System muss bei grossen Studierendenzahlen (>150 gleichzeitigen Prüfungen) einsetzbar sein (Designprinzip: #2).
- Das System muss unter heterogenen infrastrukturellen Bedingungen funktionieren (Designprinzip: #1 und #2).
- Das System muss einfach im Unterhalt sein (Designprinzip: #1).
- Das System muss erweiterbar sein, um es an neue Prüfungsmodi anzupassen (Designprinzip: #1).

Bei der Planung und Implementierung unseres CBA-Systems müssen diese technischen und organisatorischen Anforderungen zwingend erfüllt werden. Dies hat zur Folge, dass bei der Sicherheitsanforderung gewisse Abstriche in Kauf genommen werden müssen. Wie später ausgeführt wird, müssen wir beispielsweise auf den Einsatz einer digitalen Signaturerstellungseinheit verzichten, da sie uns im Praxiseinsatz nicht zur Verfügung stehen wird.

Damit wir trotzdem die Anforderungen an ein sicheres, praxistaugliches CBA-System erfüllen können, bedarf es eines Kompromisses zwischen maximaler Sicherheit und maximaler Benutzerfreundlichkeit (*Usability*) und Wartungsfreundlichkeit (*Maintainability*). Die Risikoanalyse in Kapitel 4 dient als Grundlage zur Optimierung dieser Verhältnisse, indem sie hilft, die grössten Risiken zu erkennen, so dass diese priorisiert behandelt werden können.

Das von uns entwickelte Prüfungssystem trägt den Namen Sioux (Secure Interactive Online eXam). 2007 wurde der erste Sioux-Prototyp entwickelt. Im Praxiseinsatz reifte dieser zu einer voll funktionsfähigen CBA-Suite, die den kompletten Prüfungszyklus, bestehend aus Fragepoolverwaltung, Prüfungsdurchführung und Resultatanalyse umfasst. Die Sioux-CBA-Suite wurde unter der Betreuung des Autors als Teil einer Masterarbeit und weiterer, drittmittelfinanzierter Softwareprojekte erstellt.

In den nachfolgenden Abschnitten werden die CBA-Designprinzipien und die technischen Grundlagen für die in Sioux implementierten Lösungen zur Erhöhung der Fallsicherheit und der Datensicherheit beschrieben.

5.3.1 CBA-Designprinzip #1: Berücksichtigung der vorherrschenden Rahmenbedingungen

Die Berücksichtigung der vorherrschenden Rahmenbedingungen ist für die Akzeptanz eines CBA-Systems zentral. Denn nur ein System, welches die vorhandenen Gegebenheiten berücksichtigt, wird später in der Praxis auch Verwendung finden.

Analog zu den vier Risikogruppen der Risikoanalyse (vgl. 4.4.2 "Risikoidentifikation"), stellen folgende Faktoren erhöhte Anforderungen an die Umsetzung einer CBA-Strategie:

- Didaktische Rahmenbedingungen (Fragendesign, Typen der Leistungskontrolle)
- Operationale Rahmenbedingungen (Rechtliche Vorgaben, Raumkapazitäten, Ausbildungsstand der Aufsichtspersonen)
- Technische Rahmenbedingungen (verfügbare Hardware, Schutz der Hardware, Datenschutz)
- Finanzielle Rahmenbedingungen (Menge des zur Verfügung stehenden Personals)

Umsetzung des CBA-Designprinzips #1 im Sioux-System

Aus dem zur Zeit der Drucklegung dieser Arbeit vorherrschende Umfeld ergeben sich für unser CBA-System folgende Rahmenbedingungen:

- Es gibt keine speziellen Computerarbeitsplätze für CBA.
- Die Studierenden verfügen über kein fundiertes technisches Verständnis.
- Die Studierenden verfügen über keine elektronische Signatureinheit (USB-Token, elektronische Studierendenkarte).
- Die im Kapitel 3 beschriebenen Fragetypen (*true-false/hotspot/Lückentext*) sollen unterstützt und später gegebenenfalls noch ergänzt werden.
- Der Prüfungsprozess muss einwandfrei belegt werden können.
- Die Prüfungsclients sind nur rudimentär mittels einer Schliessvorrichtung gegenüber Zugriffen von aussen geschützt.

Diese Voraussetzungen treffen für die Studierenden in unseren Lehrveranstaltungen an der ETH Zürich zu (Hinterberger & Böckenhauer, 2002-2011a, 2002-2011b). Es kann aber beobachtet werden, dass an den meisten Schweizer Bildungsinstitutionen vergleichbare Rahmenbedingungen vorherrschen. Durch den Einsatz von studentischen Laptops oder Funknetzwerken können diese jedoch auch noch heterogener ausfallen, was die Anforderungen an ein sicheres CBA-System zusätzlich erhöht.

Die Berücksichtigung der verfügbaren Rahmenbedingungen werden auch als wichtige Grundvoraussetzung zur Dissemination von Sioux in der Schweizerischen Bildungslandschaft erachtet. Die alleinige Erhöhung der Sicherheit des CBA-Systems würde hierzu nicht ausreichend.

5.3.2 CBA-Designprinzip #2: Ausfallsicherheit dank Teilautonomie der Client-Applikation

Zur Durchführung von CBA in heterogenen Systemumgebungen muss die Client-Applikation über einen gewissen Autonomiegrad verfügen. Nur so kann sichergestellt werden, dass eine Verlangsamung oder ein Ausfall der Verbindung zwischen Prüfungs-server und Prüfungsclient keine Auswirkungen auf die Bearbeitung der Leistungskontrolle hat.

Die Autonomie des Clients soll folgende Funktionen umfassen:

- Antworten müssen auf dem Client zwischengespeichert werden, um Lastspitzen oder kurze Ausfälle zu überbrücken.
- Nach dem Ausfall und der Wiederherstellung der Server-Client-Verbindung muss der Client die zwischengespeicherten Daten automatisch wieder mit dem Server synchronisieren.
- Angefangene Prüfungen müssen an einem anderen Computer fortgesetzt werden können, sogar dann, wenn der Server nicht verfügbar sein sollte.
- Bei der Bearbeitung der Prüfung dürfen keine Wartezeiten beim Wechsel der Fragen oder beim Abgeben der Antworten entstehen.
- Der Client muss sicherstellen, dass die gespeicherten Daten und die angezeigten Bildschirminhalte identisch sind.
- Der Client erstellt für den Fall eines totalen und nicht behebbaren Serverausfalls lokal ein chiffriertes (verschlüsseltes) *Backup* aller Antworten und der Prüfung.
- Der Client kann erst dann beendet werden, wenn alle Resultate auf dem Server und dem *Backup*-Medium gespeichert werden konnten.

Die zur Zeit der Drucklegung dieser Arbeit verfügbaren technischen Möglichkeiten bedeuten für unser Prüfungssystem, auf ein Browser-basiertes System verzichten zu müssen. Gemäss der neusten, als *Working draft* bezeichneten HTML5 Spezifikation (Hickson, 2011), können zwar Webinhalte für den *offline*-Modus verfügbar gemacht werden, diese werden jedoch unverschlüsselt auf dem lokalen Datenträger abgelegt. Zudem wird die als HTML5 Manifest-bezeichnete *offline*-Methode das Zwischenspeichern von Formulardaten vorerst nicht unterstützen. Zum jetzigen Zeitpunkt werden

die geplanten neuen Funktionen von HTML5 als keine Alternative zu einer echten Client-Server-basierten CBA-Applikation erachtet.

Umsetzung des CBA-Designprinzips #2 im Sioux-System

Aus folgenden Gründen haben wir uns für eine auf *Java Enterprise Beans*-basierte Lösung entschieden (Heinrich, 2008):

- Java als leistungsstarke Entwicklungsumgebung
- Vielfach bewährtes *OR-Mapping*, *DB-Pooling* und *Threadpooling*
- Zuverlässige kryptographische *Library*
- Plattformunabhängigkeit der Java-Applikationen
- Einfaches *Deployment* via Java-Webstart

Die technische Infrastruktur und die verwendeten Softwareversionen, die uns für Sioux zur Verfügung standen, sind im Kapitel 6.2.4 "Server Applikationen" aufgeführt.

Sioux wurde als Client-Server Applikation konzipiert. Der Umfang der Autonomie des *ExamClients* ist im Kapitel 6.3 "Sicherheitskomponenten von Sioux" beschrieben.

5.3.3 CBA-Designprinzip #3: Belegbarkeit der Prüfungsergebnisse durch client-seitige, digitale Signaturen

Das Verwaltungsgericht Hannover hält Ende 2008 in einem Beschluss fest, dass eine Online-Prüfung dann als schriftliche Prüfung gilt, wenn sie über ausreichend sichere technische Möglichkeiten verfügt, um die Authentizität der elektronischen Resultate zu belegen (Verwaltungsgerichtsentscheid, 12.2008). Dieser Beleg kann mittels digitaler Signaturen erbracht werden. Eine digitale Signatur ist die chiffrierte Prüfzahl eines digitalen Dokuments (Abbildung 4.34). Damit kann nicht nur die Unversehrtheit (Integrität) des Dokuments überprüft, sondern auch dessen Urheberschaft (Identität) festgestellt werden.



Abbildung 4.34: Digitale Signierung eines Dokuments. Die eindeutige Prüfsumme (*Hash*) eines Dokuments wird mittels eines digitalen Schlüssels chiffriert. So kann nachträglich die Unversehrtheit des Dokuments kontrolliert werden.

Digitale Signaturen basieren auf asymmetrischen Verschlüsselungsverfahren unter Verwendung von Schlüsselpaaren (Rivest, Shamir, & Adleman, 1978). Im Schweizeri-

schen Bundesgesetz über die elektronische Signatur (ZertES, 2003) wird der private Schlüssel als Signaturschlüssel und der öffentliche als Signaturprüfschlüssel bezeichnet.

Bei der digitalen Signierung wird der geheime Signaturschlüssel dazu verwendet, um ein Dokument zu signieren, während mit dem öffentlichen Signaturprüfschlüssel die Signatur überprüft werden kann (Abbildung 5.35). Die Sicherheit dieser Methode beruht darauf, dass der geheime Signaturschlüssel nicht oder nur mit extrem hohem Aufwand aus dem öffentlichen Signaturprüfschlüssel berechnet werden kann.

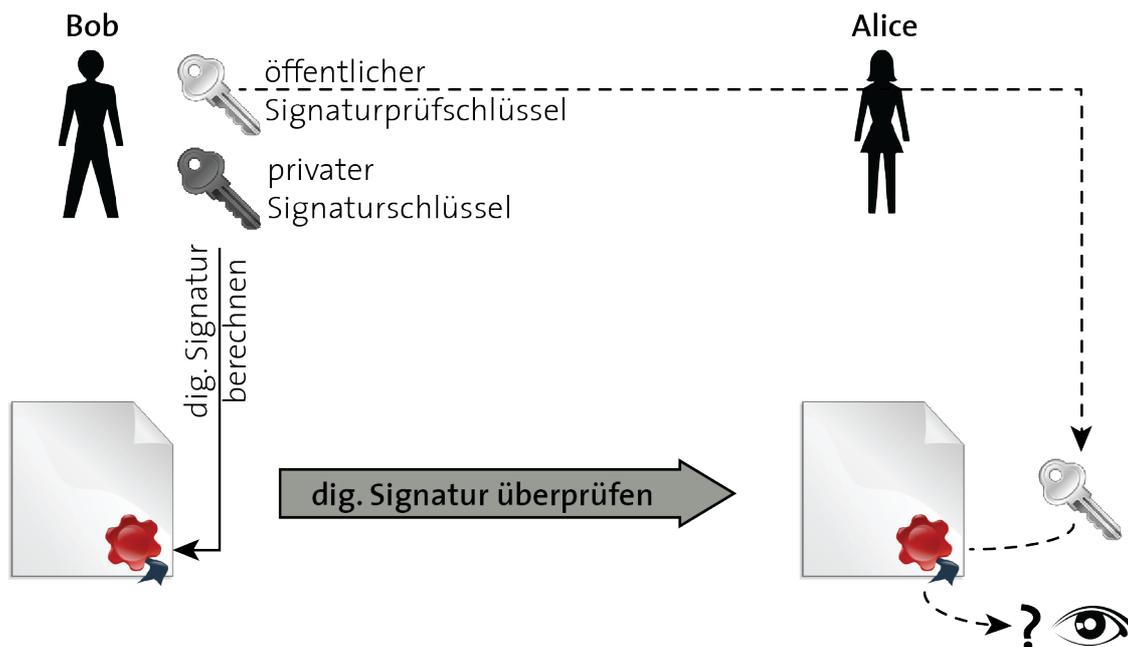


Abbildung 5.35: Prinzip von asymmetrischen Verschlüsselungsverfahren. Diese beruhen auf Einwegfunktionen und sind nur mit enormem Rechenaufwand umkehrbar. Das heisst, eine mit einem Signaturschlüssel signierte Nachricht kann mit dem Signaturprüfschlüssel lediglich verifiziert, nicht aber neu signiert werden.

Der Signaturprüfschlüssel ist bei fortgeschrittenen oder qualifizierten elektronischen Signaturen personifiziert, das heisst, er ist Teil eines Zertifikats, welches zusätzlich zum Schlüssel die Identität des Schlüsselbesitzers oder der Schlüsselbesitzerin beinhaltet (Art 2. des (ZertES, 2003)).

Gemäss Schweizerischem Obligationenrecht (Art. 14 Abs. 2bis) sind diese qualifizierten digitalen Signaturen den eigenhändigen Unterschriften gleichgestellt. Eine digital signierte Datei kann demnach als "Dokument" bezeichnet werden, da dessen Integrität und Authentizität dokumentiert sind. Basierend auf dem Beschluss des Verwaltungsgerichts Hannover würde das bedeuten, dass ein digitales Prüfungsergebnis, dessen Authentizität mittels digitaler Signaturen belegt werden kann, dem physischen Prüfungsdokument einer traditionellen, schriftlichen Prüfung (*paper and pencil test* (PPT)) gleichgestellt wäre.

In der Praxis wird aus Geschwindigkeitsgründen oft nicht das gesamte Dokument, sondern nur dessen *Hash*-Wert (Prüfzahl) chiffriert. Das Vorgehen bei der digitalen Signatur ist im nachfolgenden Abschnitt beschrieben.

Grundprinzip der digitalen Signatur

Bei der digitalen Signierung eines elektronischen Dokuments wird meist das RSA- Verschlüsselungsverfahren eingesetzt, ein asymmetrisches Verschlüsselungsverfahren, welches 1978 publiziert wurde und bis heute als sicher gilt (Rivest et al., 1978).

Die Berechnung einer digitalen Signatur passiert in zwei Schritten:

- *Hash*-Wert berechnen: Vom elektronischen Dokument wird mittels einer *Hash*-Funktion (z.B. SHA-1 oder MD5) ein *Hash*-Wert (Prüfzahl) berechnet (Menezes, van Oorschot, & Vanstone, 1996). Diese Funktion ist nahezu kollisionsresistent, das heisst, es ist praktisch unmöglich, zwei unterschiedliche Dokumente zu finden, deren *Hash*-Werte identisch sind.
- *Hash*-Wert verschlüsseln: Der *Hash*-Wert wird mit dem Signaturschlüssel der Urheberschaft dieses Dokuments chiffriert und zusammen mit dem Dokument gespeichert.

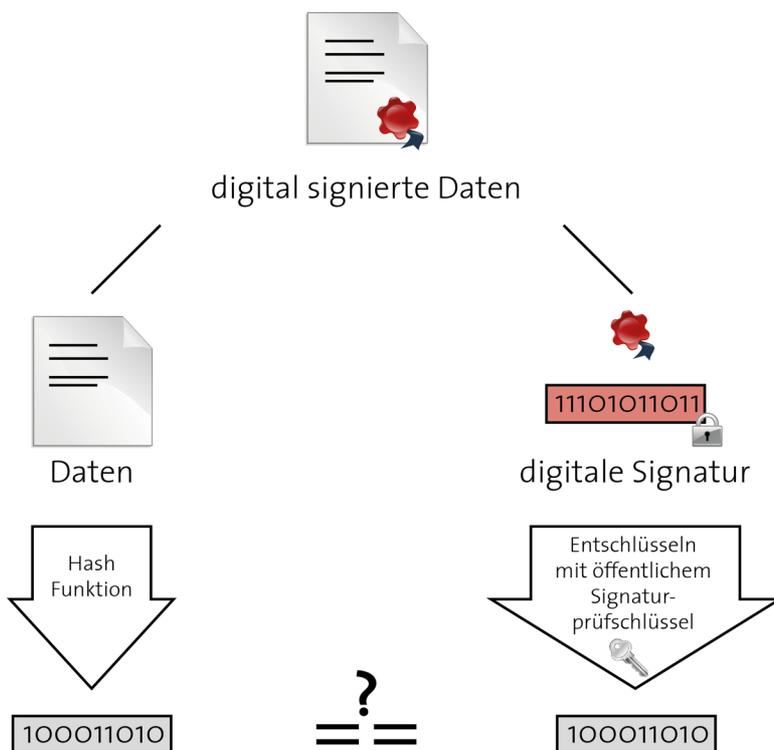


Abbildung 5.36: Überprüfung der digitalen Signatur eines Dokuments. Stimmt der in der digitalen Signatur chiffrierte *Hash*-Wert mit dem basierend auf den Daten neu berechneten *Hash*-Wert überein, ist die digitale Signatur gültig.

Die Überprüfung der digitalen Signatur geschieht in zwei Schritten:

- *Hash-Wert dechiffrieren*: Der chiffrierte *Hash-Wert* wird mit dem öffentlichen Signaturprüf Schlüssel der Urheberschaft dechiffriert.
- *Hash-Wert vergleichen*: Der dechiffrierte *Hash-Wert* wird mit dem separat berechneten *Hash-Wert* des Dokuments verglichen.

Nach der Entschlüsselung des *Hash-Werts* mit dem öffentlichen Signaturprüf Schlüssel der Urheberschaft des signierten Datenpakets können zwei Aussagen gemacht werden:

- Ein passender Signaturprüf Schlüssel gibt Auskunft über die Urheberschaft des Dokuments, weil eine erfolgreiche Entschlüsselung zeigt, dass der personalisierte, öffentliche Schlüssel zum Signaturschlüssel des Urhebers oder der Urheberin passt.
- Der Vergleich des dechiffrierten, mitgelieferten *Hash-Werts* mit dem neu berechneten *Hash-Wert* des empfangenen Dokuments gibt Aufschluss über die Unversehrtheit des Dokuments. Eine Manipulation am Dokument wird erkannt, wenn der dechiffrierte *Hash-Wert* nicht mit dem *Hash-Wert* des empfangenen Dokuments übereinstimmt.

Eine digitale Signierung von Prüfungsergebnissen muss dabei zwingend auf dem Prüfungs-Client durchgeführt werden. Dies hat zwei Gründe:

1. Der private Signaturschlüssel des Studierenden darf sich aus Geheimhaltungsgründen nur auf dem Client, idealerweise sogar nur auf einer direkt am Client angeschlossenen digitalen Signatureinheit befinden.
2. Je enger die Signatureinheit mit der Bedienoberfläche (GUI-Komponente) verknüpft ist, desto unwahrscheinlicher ist, dass durch einen Absturz der Prüfungsapplikation oder ein Netzwerkproblem Prüfungsergebnisse nicht signiert werden und das Fehlen der Antwort unerkannt bleibt.

Umsetzung des CBA-Designprinzips #3 im Sioux-System

Da zur Zeit der Drucklegung dieser Arbeit die Studierenden in der Schweiz noch über keine elektronische Signatureinheit verfügen und dies in naher Zukunft auch nicht geplant ist, muss im Sioux-Prüfungssystem auf den Einsatz von qualifizierten digitalen Signaturen verzichtet werden. Stattdessen werden fortgeschrittene elektronische Signaturen verwendet, die nicht wie die qualifizierten digitalen Signaturen auf einer sicheren Signaturerstellungseinheit beruhen, aber dennoch mit den Prüfungsergebnissen so verknüpft sind, dass eine nachträgliche Veränderung erkannt werden kann (Art. 2, Abs. c des (ZertES, 2003)). Konkrete Angaben hierzu finden sich im Abschnitt "5.4 Auswirkungen der CBA-Designprinzipien".

Das Fehlen einer elektronischen Signatureinheit führt dazu, dass alle Studierenden mit individuellen Schlüsselpaaren ausgerüstet werden müssen. Wir haben zwei Verfahren implementiert und evaluiert:

- Externe Signaturstelle: Für alle Studierenden wird im Prüfungsraum ein individuelles Schlüsselpaar generiert. Die öffentlichen Signaturprüfchlüssel werden durch eine externe Instanz signiert und an den Sioux-Server übertragen. Diese können anschliessend zur Überprüfung der signierten Prüfungsergebnisse verwendet werden. Die externe Instanz führt eine Liste, wann sie welche Schlüssel für welche Identitäten vergeben hat. Dieses Vorgehen entlastet Drittpersonen vom Vorwurf, nachträglich alle Resultate des Studierenden durch das Generieren eines neuen Schlüsselpaares manipuliert und neu signiert zu haben.
- *Backup*-Medium: Für alle Studierenden wird im Prüfungsraum ein individuelles Schlüsselpaar generiert. Die öffentlichen Signaturprüfchlüssel werden an den Sioux-Server übertragen. Diese können anschliessend zur Überprüfung der signierten Prüfungsergebnisse verwendet werden. Die privaten Signaturschlüssel werden auf den USB-Stick der Studierenden gespeichert. Die Studierenden werden angewiesen, die USB-Sticks erst zu löschen, wenn sie das Prüfungsergebnis erhalten haben und damit einverstanden sind. Mit Hilfe des privaten Signaturschlüssels auf dem USB-Stick können die Studierenden entweder aufzeigen, dass die Pakete auf dem Prüfungsserver nicht mit diesem Schlüssel signiert wurden, oder es können Drittpersonen vom Betrugsvorwurf entlastet werden.

Die Details zur Implementierung der digitalen Signaturen finden sich im Abschnitt 6.3.4 "Antworten eingeben".

5.3.4 CBA-Designprinzip #4: Nachweis der Vollständigkeit des Prüfungsdossiers dank durchgehender Nummerierung aller Antworten

Eine korrekte Leistungsbeurteilung setzt die Feststellung des rechtserheblichen Sachverhalts voraus. Darunter versteht man die Nachweiserbringung, dass erstens die unveränderten und zweitens alle Prüfungsergebnisse der Probanden zur Korrektur verwendet wurden.

Wie im obigen Abschnitt erläutert, kann mittels digitaler Signaturen die Authentizität der Antworten belegt werden. So lässt sich nachträglich feststellen, von wem die Antworten nachweislich stammen und ob sie verändert wurden. Zusätzlich muss auch noch die Vollständigkeit der Antworten belegt werden. Bei Papierprüfungen werden dazu die Prüfungsantwortblätter bei der Abgabe meist geheftet, um ein nachträgliches Verschwinden von Antworten zu verhindern. Ein wirklicher Schutz stellt dieses Vorgehen allerdings nicht dar, denn zur Korrektur wird die Heftung meist wieder entfernt und die Prüfung fragenweise unter den Assistierenden verteilt. Die Heftung stellt also mehr eine Erhöhung des subjektiven Sicherheitsempfindens als einen legalen Sicher-

heitsgewinn dar¹⁵. CBA kann diesbezüglich mittels Laufnummern den Nachweis erbringen, dass die Prüfungsergebnisse vollständig zur Korrektur verwendet wurden. Dieser Nachweis ist nicht nur wichtig, um die legale Vorgabe zu erfüllen, sondern auch um das Vertrauen der Studierenden in das CBA-System zu gewinnen, denn bei der Bearbeitung einer Leistungskontrolle verlieren die Studierenden die vollständige Kontrolle über ihr Prüfungsdokument. Sie müssen sich darauf verlassen, dass die Prüfungsapplikation das Dokument korrekt erstellt. Misstrauen gegenüber diesem System kann zu einer erhöhten Anfechtungsrate von Notenverfügungen führen, was die zeitliche Belastung für alle Beteiligten erhöht.

Beim Einsatz von Laufnummern zur Überprüfung der Vollständigkeit des Prüfungsdokuments müssen folgende fünf Vorgaben erfüllt werden:

- Jede Antwort muss mit einer eindeutigen Nummer versehen werden.
- Die Nummerierung muss aufsteigend und lückenlos passieren, denn nur so können fehlende Antworten anhand der Lücken später erkannt werden.
- Die Nummerierung muss zwingend auf dem Prüfungsclient durchgeführt werden, denn nur dort können Antworten erkannt werden, welche den Server nie erreicht haben.
- Bei allen Studierenden muss die letzte und somit höchste Laufnummer klar sein. Sinnvollerweise wird sie dem finalen Prüfungsdokument zugewiesen. Dieses Dokument wird beim Beenden der Prüfung erstellt und als letztes Dokument dem Server übermittelt. Damit dessen Übermittlung die Prüfung auf dem Server als "abgegeben" markiert wird, kann und muss dessen Erhalt von der Prüfungsleitung direkt vor Ort überprüft werden. Mit dieser eindeutigen, letzten Laufnummer soll erreicht werden, dass auch die Laufnummer des letzten Zwischenresultats eindeutig ist. Würde eine solche letzte Laufnummer fehlen, könnte bemängelt werden, dass das letzte Zwischenresultat verloren gegangen sei.
- Zur Integritätsprüfung der Datenpakete müssen diese zwingend digital signiert sein. Nur so kann ausgeschlossen werden, dass diese nachträglich manipuliert wurden. Datenpakete ohne digitale Signatur können jederzeit beliebig verändert werden, ohne dass die Veränderung festgestellt werden kann. So könnten die Laufnummern auf dem Server absichtlich oder aufgrund eines technischen Problems verändert werden. Die Laufnummern ohne Signatur verfügen somit über keine Beweiskraft und können nicht zweifelsfrei zur Überprüfung der Vollständigkeit der

¹⁵ Bei PPT kann nur schwer verhindert werden, dass ein Student eine Prüfungsfrage (bewusst) nicht abgibt und später das Gegenteil behauptet. Der Dozent kann später unmöglich die nicht-Existenz dieser Frage beweisen. Nur eine Vollständigkeitskontrolle beim Einsammeln könnte dies verhindern. Dazu fehlt aber meistens die Zeit.

Prüfungsergebnisse verwendet werden. Des Weiteren könnte eine vorhandene Signatur auch dazu verwendet werden, eine nachträglich veränderte Laufnummer wieder herzustellen (vorausgesetzt, der private Schlüssel ist noch vorhanden).

Umsetzung des CBA-Designprinzips #4 im Sioux-System

Bei Sioux übernimmt der Java-ExamClient die Nummerierung der Antworten. Dazu wird jede Antwort mit einer eindeutigen Laufnummer ergänzt. Diese beginnt bei "1" für die erste abgegebene Antwort. Jede weitere Antwort erhält aufsteigend eine neue Nummer. Korrigieren die Studierenden eine bereits abgegebene Antwort, so erhält die neue Antwort die nächste freie Nummer. Die alte Antwort bleibt bestehen, sie wird aber bei der Korrektur nicht berücksichtigt. Die Nummern werden zusammen mit der eigentlichen Antwort signiert und so gegen nachträgliche Veränderungen geschützt. Alle Antworten werden sowohl lokal im Arbeitsspeicher des Client-Computers und auf einem *Backup*-Medium als auch auf dem Server gespeichert.

Am Schluss der Prüfung erstellt der ExamClient aus allen abgegebenen Antworten das eigentliche Prüfungsdokument. Dieses enthält neben den jeweils zuletzt abgegebenen Antworten auch die kompletten Aufgabenstellungen. Das digitale Prüfungsdokument wird mit der letzten, höchsten Laufnummer versehen und digital signiert. Dieses Dokument stellt die rechtsverbindliche Version der abgelegten Prüfung dar (Lunkeit, 2007) und dient als Grundlage für die Korrektur und Benotung. Im Zweifelsfall kann die Vollständigkeit des Prüfungsdokuments mittels einzeln nummerierter Antworten und der digitaler Signaturen verifiziert werden. Es ist entscheidend, dem digitalen Prüfungsdokument die höchste Laufnummer zu vergeben, da dieses als letztes Dokument signiert und an den Server gesendet wird. Dadurch kann belegt werden, dass vorgängig alle Resultate erfasst und ins Prüfungsdokument integriert wurden. Mit einer lückenlosen Nummerierung aller Antworten kann dies überprüft werden.

Nach Abschluss der Prüfung werden alle Antworten und Prüfungsdokumente vom Server auf einem externen Speichermedium archiviert.

5.3.5 CBA-Designprinzip #5: Reliable Prüfungsergebnisse werden erst durch eine hohe *Usability* der Client-Applikation erreicht

Das letzte CBA-Designprinzip definiert die Anforderungen an die *Usability*, das heißt die Verfügbarkeit und Bedienbarkeit des CBA-Systems. Es ist unbestritten, dass die Bedienoberfläche der CBA-Client-Applikation einen Einfluss auf das Prüfungsergebnis hat (Valenti, Cucchiarelli, & Panti, 2002). Zudem führt eine stabile, einfach zu bedienende Oberfläche zu zuverlässigeren Prüfungsergebnissen als eine Client-Applikation, die oft abstürzt und nicht intuitiv bedient werden kann.

Befürworter von LMS-basierten Prüfungssystemen erwähnen häufig, dass die Studierenden die Prüfungen im selben System ablegen sollen, mit welchen sie auch während

des Semesters gearbeitet haben. Aus Sicht der *Usability* ist dies allerdings fragwürdig, da ein LMS- und ein CBA-System grundsätzlich unterschiedliche Aufgaben zu erfüllen haben, welche nicht kompromisslos mit nur einer Applikation bedient werden können. In letzter Zeit scheinen sich die Hersteller von LMS-basierten Prüfungsumgebungen diesem Sachverhalt immer mehr bewusst zu werden. Sie beginnen Methoden zu implementieren, welche die *Usability* während der Durchführung von CBA optimieren sollen. Diese haben zum Ziel, die überflüssigen Informationen von der Bedienoberfläche zu entfernen, um die Übersichtlichkeit zu erhöhen.

Ein CBA-Client mit einer hohen Usability weist folgende Eigenschaften auf:

- Die Benutzeroberfläche zeigt nur prüfungsrelevante Informationen an.
- Bei Problemen werden zweckmässige Hilfestellungen angezeigt.
- Der Client stürzt nicht ab.
- Antworten können beliebig oft geändert werden.
- Die Reihenfolge der Aufgaben kann frei gewählt werden.
- Die verbleibende Prüfungszeit wird dezent, aber klar angezeigt.
- Wartezeiten bei der Bedienung sind auf ein Minimum reduziert.
- Es werden nur gespeicherte Antworten angezeigt.
- Bei formativen Leistungskontrollen kann nach Bearbeitung der Prüfung ein Feedback angezeigt werden.
- Studierende können Bemerkungen anbringen, wenn Sie bei der Fragenbeantwortung einen Kommentar abgeben möchten. Diese können bei der Korrektur angezeigt werden.
- Eine Navigation mit den verbleibenden und gelösten Aufgaben wird angezeigt.
- Die CBA-Applikation ist Plattform-übergreifend einsetzbar.

Umsetzung des CBA-Designprinzips #5 im Sioux-System

Die Implementierung von Sioux als komplett neu entwickelte Java-Applikation erlaubt uns eine kompromisslose Umsetzung unserer Designprinzipien.

Die Benutzeroberfläche zeigt nur prüfungsrelevante Informationen an. Dazu gehören:

- Aufgaben
- Aufgabenübersicht mit der Angabe der bereits gelösten Aufgaben

- verbleibende Prüfungszeit
- Namen des Probanden oder der Probandin
- Navigationselemente
- Gesendete Antworten

Durch das Laden der kompletten Prüfung zu Beginn und das Zwischenspeichern von gegebenen Antworten auf dem Client sind keine Wartezeiten bei der Bedienung der Prüfungsapplikation zu erwarten. Mit Wartezeiten ist nur zu Beginn und am Ende der Prüfung aufgrund der Datenübertragung zwischen Client und Server zu rechnen. Dies wird den Studierenden mittels Hinweisfenster angezeigt. Auch an anderen Orten, namentlich beim Einlesen des USB-Sticks zur Datensicherung, werden Hilfestellungen angezeigt, falls zum Beispiel der USB-Stick nicht erkannt wurde oder über zu wenig freien Speicherplatz verfügt.

Die Zielerreichung bezüglich der *Usability* von Sioux wurde mittels einer Studie untersucht und die Resultate in einem technischen Bericht festgehalten (Dahinden, 2010c). Eine Zusammenfassung dieser Resultate findet sich auch im Abschnitt 6.4.7 "Usability von Sioux".

5.4 Auswirkungen der CBA-Designprinzipien auf die vier Sicherheitsphasen von CBA-Systemen

5.4.1 Sicherheitsphasen im CBA-Prozess

Bei der Durchführung von CBA können vier Sicherheitsphasen unterschieden werden (Tabelle 5.13). Diese Phasen unterscheiden sich vor allem darin, dass beim Auftreten eines sicherheitsrelevanten Problems die *Stakeholder* unterschiedlich stark involviert sind. So hat zum Beispiel der unerlaubte Zugriff auf den Fragenpool gravierende Auswirkungen für die Dozierenden, während die Studierenden oder die Bildungsinstitution dadurch keine Nachteile erleiden.

Phase	Sicherheitsfokus	Betroffene
Vorbereitungsphase	Schutz der Fragenpools	Dozierende
Durchführungsphase	Rechtsgültige Protokollierung des Prüfprozesses Vergleichbarkeit des Prüfprozesses	Studierende Dozierende evtl. Bildungsinstitution
Korrekturphase	Formalisierte Korrektur Kontrolle der Datenauthentizität	Studierende
Archivierungsphase	Beständige Speicherung der Resultate	Dozierende

Tabelle 5.13: Vier Sicherheitsphasen bei CBA.

Eine Sonderrolle nimmt die Durchführungsphase ein, da diese die primäre Ursache für Beschwerden ist. Erschwerend kommt hinzu, dass diese Phase kritisch ist in Bezug auf zeitliche Verzögerungen, wie sie zum Beispiel durch Betriebsunterbrüche des CBA-Systems oder Client-Abstürze entstehen. Die Durchführungsphase bietet somit eine eingeschränkte Zeittoleranz für Fehlerbereinigungen oder Problembhebungen.

Dass die Durchführungsphase das grösste Gefahrenpotential beinhaltet, zeigt auch Abbildung 5.37, welche den Einfluss der CBA-Designprinzipien auf die vier Sicherheitsphase illustriert.

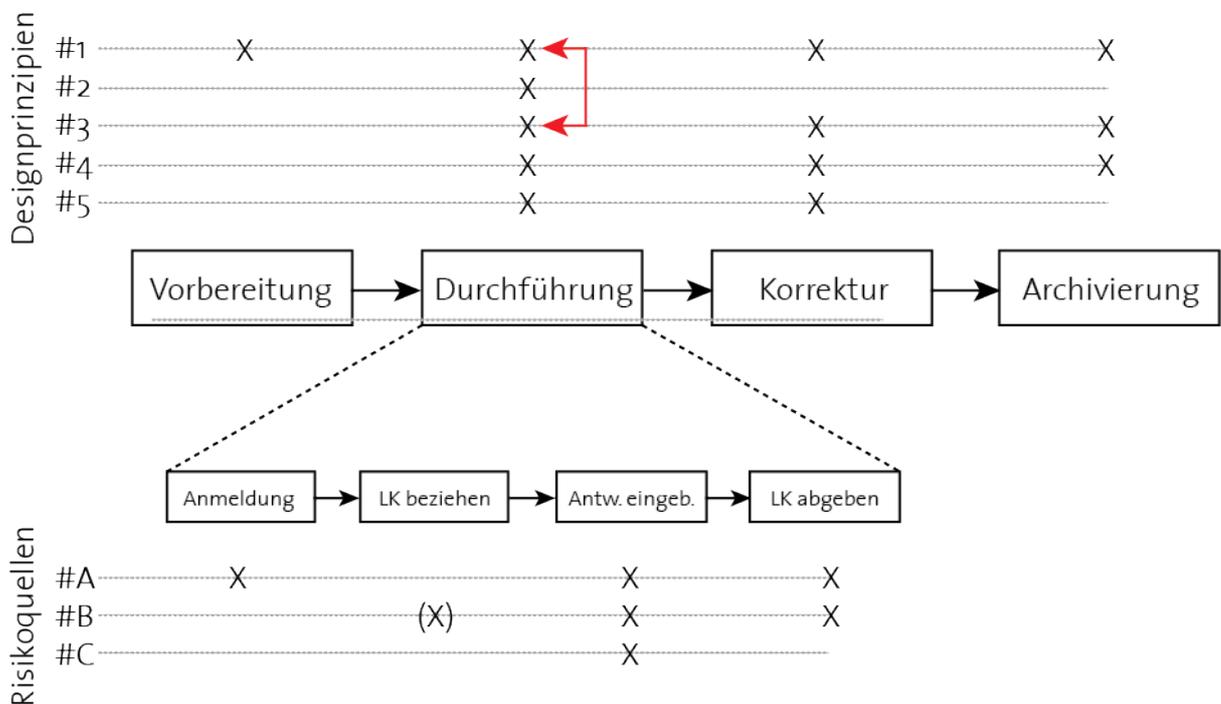


Abbildung 5.37: Auswirkungen der durch die CBA-Designprinzipien verursachten Einschränkungen auf den CBA-Prüfprozess. Falls Designprinzipien auf die gleiche Sicherheitsphase einwirken, können Sie potentiell in Konkurrenz zueinander stehen. Ein solches Beispiel ist mit einem Pfeil markiert und wird im Text erläutert. Die Auswirkungen der durch die Designprinzipien induzierten Risikoquellen (#A-#C) sind mit einem "X" markiert und werden im Text erläutert. Der Arbeitsschritt "Antworten eingeben" birgt das grösste Sicherheitsrisiko, da alle drei Risiken eintreten können.

Bei der Implementierung eines praxisfähigen CBA-Systems, basierend auf den fünf CBA-Designprinzipien, müssen verschiedene Kompromisse eingegangen werden. So verlangt beispielsweise das Designprinzip #1 die Berücksichtigung der vorherrschenden Rahmenbedingungen, welche teilweise in Konkurrenz zu den Sicherheitsprinzipien im Designprinzip #3 stehen (markiert mit einem Pfeil in Abbildung 5.37). In der Praxis könnte das heissen: wenn keine digitalen Signaturerstellungseinheiten zur Verfügung stehen, dann müssen alternative Lösungen für den Schutz der Prüfungsergebnisse gefunden werden.

In den nachfolgenden drei Abschnitten werden die Auswirkungen der CBA-Designprinzipien auf die Implementierung eines CBA-Systems anhand folgender einschränkender Rahmenbedingungen untersucht und alternative Lösungsansätze aufgezeigt:

- #A: Verzicht auf spezialisierte Hardware für die CBA-Clients (Abschnitt 5.4.2)
- #B: Verzicht auf eine qualifizierte elektronische Signatur bei Prüfungsergebnissen (Abschnitt 5.4.3)
- #C: Verzicht auf den erweiterten Schutz der studentischen Antworten und des Fragenpools (Abschnitt 5.4.4)

5.4.2 Verzicht auf spezialisierte Hardware für die CBA-Clients

Die Prüfungsclients nehmen sicherheitstechnisch bei CBA eine Sonderrolle ein. Ihr korrektes Funktionieren ist eine Grundvoraussetzung, dass Prüfungen sicher abgehalten werden können. Wir zählen deshalb diese Clients zur vertrauenswürdigen Zone. Es ist uns jedoch bewusst, dass dadurch ein Sicherheitsrisiko eingegangen wird. So vertrauen wir darauf, dass sich alle Computer in einem einwandfreien, nicht kompromittierten Zustand befinden und zwar genau so, wie sie von den Informatikdiensten eingerichtet wurden. Wir beschränken uns darauf, den Computer vor Prüfungsbeginn mittels des *Safe Exam Browsers* (SEB) in einen Kioskmodus zu versetzen. Dieser Modus verhindert, dass die Studierenden während der Prüfung die Sioux-Applikation verlassen und mittels der im Betriebssystem integrierten Programme mit anderen Studierenden kommunizieren oder auf unerlaubte Ressourcen zugreifen können.

Wir haben darauf verzichtet, dass Sioux die Korrektheit des Clientcomputers untersucht. So lange der Computer über keine kryptographischen *Hardware*-Komponenten verfügt, wie sie z.B. in *Trusted-Computing*-Plattformen (TCP) vorhanden sind, kann weder der Computer noch Sioux überprüfen, ob sich der Prüfungsclient im Originalzustand befindet. Zur Zeit der Drucklegung dieser Arbeit ist es unwahrscheinlich, dass in der Schweiz weder mittel- noch längerfristig spezielle TCP-Prüfungsstationen bei CBA zum Einsatz kommen werden. Im Gegenteil: viel wahrscheinlicher ist, dass aus kosten-

und organisatorischen Gründen portable studentische Rechner eingesetzt werden. Dadurch wird die Kontrolle dieser heterogenen Prüfungsstationen noch schwieriger.

Damit wir trotz fehlender spezialisierter *Client*-Hardware einen gewissen Grad an Sicherheit erlangen können, der mit dem aktuellen Wissensstand für CBA als genügend eingestuft werden kann, wurden folgende Massnahmen getroffen:

Gewaltentrennung: Wir setzen stark auf eine Gewalttrennung bei der Durchführung von CBA. An der ETH Zürich verfügen wir über fünf unabhängige Instanzen, die über unterschiedliche Kernkompetenzen verfügen: die Informatikdienste, das Zentrum für Lehrentwicklung und -technologie (LET), die Prüfungsleitung, die Prüfungsaufsicht und die Studierenden. Wir nutzen diese Tatsache zur Absicherung unseres Prüfungssystems. So sind zum Beispiel die Informatikdienste für die Clientcomputer zuständig. Das LET führt das *Login* an den CBA-Arbeitsstationen durch. Dabei kommen spezielle Prüfungslogins zum Einsatz, mit welchen aufgrund fehlender Berechtigungen weder bewusst noch unbewusst die Computer manipuliert werden können.

Nach dem Starten und Anmelden an der CBA-Arbeitsstation startet die Raumaufsicht den Sioux-Client mittels Link auf der Webseite des LET. Die Studierenden können dann den Prüfungsplatz frei auswählen. Diese Gewalttrennung bietet einen gewissen Schutz davor, dass beispielsweise die Prüfungsleitung oder die Informatikdienste den Prüfungsmisserfolg eines einzelnen Studierenden absichtlich herbeiführen können. Aus legalen Gründen sollten die einzelnen Parteien die zum Einsatz gelangten Programme oder deren Prüfsummen archivieren. So wird belegbar, welche Programme zum Einsatz kamen und gegebenenfalls einer Analyse unterzogen werden.

"Gehärtete" Prüfungsumgebung: Während der Prüfungszeit müssen die eingesetzten Clientcomputer (portable, studentische Computer oder Computer in öffentlichen Computerräumen) in einen gehärteten Prüfungszustand versetzt werden. Dies bedeutet, dass die Prüfungsumgebung den Studierenden den Zugriff auf bestimmte Ressourcen verhindern soll. Zudem muss sichergestellt sein, dass die Prüfungsapplikation einwandfrei mit der eingesetzten Client-Hardware funktioniert. Eine solche "gehärtete" Prüfungsumgebung könnte mittels eines *Boot*-Mediums auf die einzelnen Prüfungsclients übertragen werden.

Ein solches Bootmedium kann mittels *Netboot* oder eines Speichermediums (z.B. USB-Stick, DVD-ROM) verteilt werden. Es bietet zudem den Vorteil, dass es die Computer in einen vordefinierten Zustand versetzt. So können die umfangreichen Tests zu Beginn jeder Prüfungssession reduziert werden, da nicht bei jeder Prüfungsdurchführung das einwandfreie Funktionieren der CBA-Applikation überprüft werden müsste.

Sinnvollerweise verfügt das Bootmedium über eine Möglichkeit zur Überprüfung, ob es in einer virtuellen Maschine gestartet wurde. Technische Lösungen, wie man die Überprüfung technisch umsetzen könnte, wurden unter anderem von Rutkowska beschrieben (Rutkowska, 2006). Interessant ist auch die Möglichkeit der Bootmedien, bereits

auf Betriebssystemebene den Zugriff auf gewisse Hardwareressourcen einzuschränken (WLAN, Festplatten, Netzwerk, etc.). Zudem kann mittels einer *Softwarefirewall* der Netzwerkverkehr auf bestimmte Computer (wie z.B. der Sioux-Server) beschränkt werden.

Noch ungelöst ist allerdings das Problem der sicheren Verteilung per *PXE-Netboot* (Intel Corporation, 1999). So ist es beispielsweise mittels eines ins Netzwerk gehängten DHCP-Servers relativ einfach möglich, ein beliebiges, möglicherweise kompromittiertes Bootmedium zu verteilen. Dieser DHCP-Server könnte vorgängig ins Netz des Prüfungssystems gehängt oder auf einem mitgebrachten studentischen Laptop installiert worden sein. Beim *Booten* jedes Prüfungs-Clients würde dann dieser DHCP-Server aufgrund der kürzesten Antwortzeit die Anfragen beantworten und so möglich machen, den Bootvorgang erheblich zu stören oder den *Netboot*-Vorgang gar auf einen anderen TFTP-Server umzuleiten. Diesbezüglich könnte das iPXE-Protokoll hilfreich sein, weil dieses auch ohne DHCP-Server z.B. von einem *Webserver booten* kann¹⁶. Eventuell wäre sogar ein zertifikatsbasierter Bootvorgang möglich, so dass sichergestellt werden kann, dass alle Clients dasselbe Bootmedium verwenden. In einem aktuellen Projekt untersuchen wir Praxistauglichkeit dieser Technologie im Rahmen von CBA und unter Verwendung von verschiedenen Arten von Client-Computern.

Es ist uns bewusst, dass trotz dieser Massnahmen die spezialisierte Hardware nicht ersetzt werden kann. Wir beobachten aber, dass ungeachtet dieser Risiken die Nachfrage nach CBA mit steigenden Studierendenzahlen stark zunimmt. Wir haben uns deshalb zum Ziel gesetzt, unter Berücksichtigung der vorhandenen Infrastruktur ein Maximum an Sicherheit bei CBA zu erreichen und gleichzeitig auf die möglichen legalen, technischen und organisatorischen Risiken hinzuweisen. Es obliegt dann den einzelnen Prüfungsleitungen, die Risiken zu gewichten und allenfalls auf die Durchführung von CBA zu verzichten.

5.4.3 Verzicht auf eine qualifizierte elektronische Signatur bei Prüfungsergebnissen

Zur qualifizierten elektronischen Signatur wird eine sichere Signaturerstellungseinheit benötigt. Unsere Studierenden verfügen nicht über derartige Hardware. Deshalb waren primär finanzielle Gründe ausschlaggebend, warum auf die qualifizierte elektronische Signierung der Prüfungsergebnisse verzichtet wurde. Aus technischer Sicht wäre der Einsatz einer solchen Signatur ohnehin nur begrenzt sinnvoll, denn ohne den Einsatz von spezialisierter Client-Hardware könnte nicht sichergestellt werden, ob tatsächlich auch jene Daten signiert werden, welche die Studierenden glauben zu signieren.

¹⁶ Da für iPXE (noch) keine Spezifikation vorliegt, kann hier lediglich der Link auf die Projekthomepage angegeben werden: <http://www.ipxe.org> (abgerufen am: 23.05.2014)

Statt qualifizierter elektronischer Signaturen verwenden wir fortgeschrittene Signaturen, die ohne elektronische Signatureinheit verwendet werden können. Das Fehlen einer solchen Signatureinheit hat jedoch den Nachteil, dass der private Signaturschlüssel lesbar sein muss, damit die Applikation die Signierung durchführen kann. Technisch wäre es problemlos möglich, diesen privaten Schlüssel zu kopieren und damit allfällig veränderte Prüfungsdaten neu zu signieren.

Abbildung 5.38 illustriert sechs sicherheitsrelevante Risiken, welche durch den Verzicht auf qualifizierte digitale Signaturen auftreten können. Alle Risiken basieren auf der Tatsache, dass der private Signaturschlüssel der Studierenden nach der Prüfung auf deren USB-Stick gespeichert und elektronisch lesbar ist. Dadurch kann er gelöscht, weitergegeben oder verändert werden.

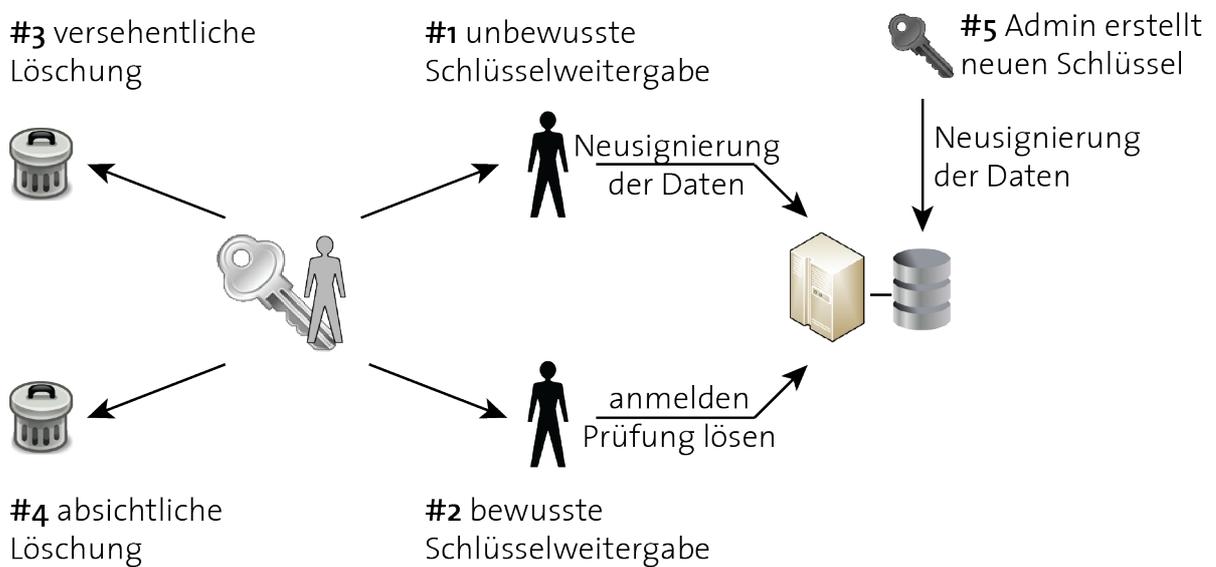


Abbildung 5.38: Sicherheitsrelevante Risiken beim Verzicht auf qualifizierte digitale Signaturen. Die Studierenden sind im Besitz ihres elektronisch lesbaren, privaten Signaturschlüssels. Dieser kann weitergegeben, kopiert, verändert oder gelöscht werden. In der Abbildung sind sechs Möglichkeiten erläutert. Die Nummern (1-6) entsprechen der Nummerierung der Risiken in Tabelle 5.14.

In Tabelle 5.14 werden die sechs plausibelsten Risiken (R), die getroffenen Gegenmassnahmen (G) und deren Effekt auf die Sicherheit des Prüfungsprozess (E) erläutert.

R (#1)	Die Studierenden schreiben die Prüfung mit einem Sioux <i>ExamClient</i> , der die privaten Schlüssel speichert und Drittpersonen zugänglich macht.
G	<p>1. Der ExamClient wird durch die Entwickler mittels qualifizierter Signatur vor Veränderungen geschützt und zusammen mit den Prüfungsdokumenten direkt im Anschluss an die LK archiviert.</p> <p>2. Der Inhalt der Sioux Datenbank wird direkt nach der LK gespeichert und die Prüfsumme dieser Datei wird an <i>Stakeholder</i> der LK geschickt.</p>

E	<p>Durch die Signierung des ExamClients kann verhindert werden, dass betrügerische Versionen des ExamClients in Umlauf kommen.</p> <p>Die Sicherung aller Prüfungsdaten und eingesetzten Programme direkt im Anschluss an die LK hat den Effekt, dass allfällige Kopien der privaten Schlüssel der Studierenden für Datenmodifikationen wertlos werden, da damit nach der Prüfung keine Änderungen mehr an den Prüfungsdaten vorgenommen werden können. Zudem kann anhand der archivierten Programme deren einwandfreies Funktionieren überprüft werden.</p>
R (#2)	<p>Die privaten Schlüssel können durch die Studierenden kopiert und weitergegeben werden. Diese könnten sich dann ebenfalls am Prüfungssystem anmelden und im Namen dieses Studierenden die Prüfung bearbeiten.</p>
G	<p>Die Signaturen werden bei Sioux erst beim Starten der LK generiert. Zu diesem Zeitpunkt befinden sich die Studierenden bereits in der sicheren Sioux Prüfungsapplikation, die das Wegkopieren des privaten Schlüssels verhindert.</p>
E	<p>Die Sicherung aller Prüfungsdaten und eingesetzten Programme hat den Effekt, dass allfällige Kopien der privaten Schlüssel der Studierenden für Datenmodifikationen wertlos werden, da damit keine Änderungen mehr an den Prüfungsdaten vorgenommen werden können.</p>
R (#3)	<p>Der Student verliert oder löscht seinen privaten Signaturschlüssel versehentlich.</p>
G	<ol style="list-style-type: none"> 1. Die Studierenden werden beim Prüfungsbeginn darauf hingewiesen, dass Sie den Inhalt des Prüfungs-USB-Sticks erst löschen sollen, wenn sie das Prüfungsergebnis kennen und damit einverstanden sind. 2. Auf dem Sioux ExamServer werden alle Informationen gespeichert, die benötigt werden, um nachträglich eindeutig die Vollständigkeit und Integrität der Prüfungsergebnisse zu überprüfen.
E	<p>Der Verlust des privaten Signaturschlüssels stellt für die Integrität der Sioux-Datenbank kein Problem dar. Die Resultate der Studierenden sind personalisiert auf dem Sioux-Server gespeichert. Zur Prüfung der Unveränderlichkeit der Prüfungsergebnisse wird der öffentliche Schlüssel des Studierenden benötigt, der auf dem Sioux-Server gespeichert ist.</p>
R (#4)	<p>Der Student löscht seinen privaten Signaturschlüssel bewusst, um ein Prüfungsmisserfolg auf legalem Weg anzufechten. Ohne privaten Signaturschlüssel kann die Urheberschaft der auf dem Server gespeicherten Resultate</p>

	nicht mehr zweifelsfrei belegt werden. Er behauptet dann, die auf dem Server gespeicherten Resultate seien nicht vom ihm.
G	<ol style="list-style-type: none"> 1. Auf dem Server werden die Resultate personifiziert gespeichert und mittels Signatur geschützt. 2. Direkt im Anschluss an die Prüfung wird die Prüfsumme der Sioux-Datenbank an unabhängige <i>Stakeholder</i> übermittelt (zum Beispiel: Prüfungsleitung, Informatikdienste und LET). Eine nachträgliche Änderung an den Daten wird dadurch verunmöglicht. Alternativ kann auch die md5-Prüfsumme durch diese Instanzen vor Ort überprüft und individuell archiviert werden. 3. Alle Personen mit erweiterten Zugriffsberechtigungen auf den Sioux Prüfungsserver müssen schriftlich bestätigen, dass sie niemals direkt oder indirekt die Resultate der Studierenden verändern.
E	Die Kombination dieser drei Massnahmen soll verhindern, dass jemand unbemerkt die Prüfungsergebnisse eines Studenten verändern kann.
R (#5)	Ein nicht integrierter technischer Mitarbeiter des Sioux Teams mit erweiterten Zugriffsberechtigungen auf den Sioux-Server verändert die Resultate eines Studenten und signiert diese mit einem neu erstellten Schlüsselpaar. Der Student erleidet dadurch einen Schaden.
G	<ol style="list-style-type: none"> 1. Der Student hat den privaten Schlüssel, mit dem er seine Resultate ursprünglich signiert hatte, auf seinem USB-Stick gespeichert. 2. Die Prüfsumme der Datenbank wird unmittelbar nach der LK bestimmt und an die <i>Stakeholder</i> kommuniziert. 3. Die an der Prüfung eingesetzte Version der Sioux-Suite wird zusammen mit den Prüfungsergebnissen archiviert. 4. Die technischen Mitarbeiter des Sioux Teams haben keine Administrator-Berechtigungen auf den Prüfungsstationen. Sie können dort keine Programme installieren. 5. Nur der Leiter des Sioux-Teams verfügt über die Möglichkeit, die eingesetzten Sioux-Clients mittels offiziellen Zertifikats zu signieren.
E	Die unmittelbare Bestimmung der Prüfsumme nach der LK beschränkt die Möglichkeiten der Datenmanipulation durch den technischen Mitarbeiter auf die Prüfungszeit. Allfällige Manipulationen am Sioux <i>ExamClient</i> durch diesen Mitarbeiter werden erschwert, indem nur der Leiter des Sioux-Teams die

	<p>ses Programm signieren kann. Zudem werden alle eingesetzten Programme zusammen mit den Prüfungsergebnissen archiviert und können so analysiert werden. Der Student kann mit Hilfe seines privaten Schlüssels zudem aufzeigen, dass der technische Mitarbeiter offenbar ein anderes Schlüsselpaar zur Signierung verwendet hat.</p>
R (#6)	<p>Der Student erstellt einen neuen privaten Schlüssel und behauptet, ein technischer Mitarbeiter des Sioux-Teams hätte seine Daten manipuliert und mit einem neu generierten Schlüsselpaar signiert.</p>
G	<p>1. Zur Chiffrierung und Speicherung der Resultate auf den <i>Backup</i>-Medien der Studierenden wird der Sioux Signaturschlüssel verwendet. Weitere Informationen zu den im Sioux System verwendeten Schlüssel findet sich im Abschnitt 6.2.6 "Zertifikate zur Absicherung von Sioux".</p> <p>2. Für jede LK wird ein neuer Sioux Signaturschlüssel erstellt.</p>
E	<p>Ein neuer privater Schlüssel ist nicht ausreichend, um die Daten auf dem <i>Backup</i>-Medium zu speichern. Dazu braucht der Student noch den Sioux Signaturschlüssel, der zusammen mit der Prüfung vom Sioux ExamServer übertragen und ausschliesslich im Arbeitsspeicher zwischengespeichert wird. Da sich die Studierenden während der gesamten Prüfungszeit in einer abgesicherten Systemumgebung befinden (welche unter anderem das Ausführen von Programmen verhindert), ist es für sie sehr schwierig, den Sioux Signaturschlüssel zu ermitteln.</p>

Tabelle 5.14: Sicherheitsrisiken aufgrund des Verzichts auf qualifizierte digitale Signaturen bei der Speicherung von Prüfungsergebnissen. Es werden folgende Abkürzungen verwendet: R=Risikobeschreibung, G=Gegenmassnahme, E=Effekt der Gegenmassnahme.

Bei Sioux verwenden wir eine fortgeschrittene Signatur, wobei der *ExamClient* die Schlüssel individualisiert auf die Namen der angemeldeten Studierenden ausstellt (der Sioux *ExamClient* als PKI). Es gibt keine zentrale Liste über alle ausgestellten Schlüssel. Dies hat den Nachteil, dass nicht mit Bestimmtheit gesagt werden kann, welche und wie viele Schlüssel für eine bestimmte Person erstellt wurden. Je nach Einsatzgebiet und Sicherheitsanforderung an die CBA wäre es unter Umständen sinnvoll, die PKI an eine externe Stelle auszulagern. Diese Gewaltentrennung – die PKI würde Zertifikate erstellen, hätte aber keinen Zugriff auf die Sioux Infrastruktur – hat den Vorteil, dass weder das Sioux-Entwicklerteam noch die PKI die Prüfungsergebnisse so verändern kann, dass dies die andere Partei nicht feststellen kann.

Es muss allerdings bedacht werden, dass der Einsatz einer externen PKI das CBA-System fehleranfälliger und unflexibler macht. So ist zum Beispiel eine *closed book* Prüfung¹⁷ ohne Internetverbindung nicht mehr möglich, da die externe PKI nur über ein Computernetzwerk erreicht werden kann. Aus Gründen der Flexibilität und aufgrund unklarer legaler Anforderungen verzichten wir momentan auf eine externe PKI. Technisch wäre es jedoch problemlos möglich, die Sioux-CBA-Suite an eine externe PKI anzubinden.

5.4.4 Verzicht auf den erweiterten Schutz der studentischen Antworten und des Fragenpools

Das Schweizerische Datenschutzgesetz schreibt vor, dass nur Daten gesammelt werden dürfen, welche unmittelbar im Zusammenhang mit der Abwicklung einer Sache stehen (DSG, 2011, Art. 13). Diese Vorgabe wird zwar von Sioux erfüllt, aber es verfügt beispielsweise über keine geeigneten technischen Massnahmen, um den Zugriff auf die Prüfungsdaten durch Systemadministratoren zu verhindern. Hoffmann stellte diesbezüglich in seiner Arbeit ein interessantes, auf einem virtuellen, ticketbasierten Dateisystem basiertes CBA-System vor (Hoffmann, 2010). Allerdings wird sich noch zeigen müssen, ob sich dieses Verfahren im CBA Praxiseinsatz bewährt. Schliesslich birgt ein solches System die Gefahr, dass die restriktiven Zugriffsberechtigungen dazu führen können, dass Prüfungsergebnisse zwar gespeichert, aber nicht mehr gelesen und somit nicht korrigiert werden können.

Im nachfolgenden Kapitel wird die Implementierung von Sioux als *proof-of-concept* unserer Designprinzipien für ein sicheres CBA-System präsentiert.

¹⁷ Closed book Prüfung: Eine Prüfung, bei der keine Hilfsmittel erlaubt sind.

Kapitel 6

***Proof-of-concept* der CBA-Designprinzipien: Sioux CBA-Suite**

6.1 Einleitung

Wir haben die Sioux ("Secure Interactive Online eXam") CBA-Suite entwickelt, um die Praxisfähigkeit (*proof-of-concept*) unserer CBA-Designprinzipien im heterogenen CBA-Alltag zu untersuchen. Das Projekt wurde 2008 gestartet und profitierte von den Erfahrungen, die im Rahmen der Computer-basierten Leistungskontrollen seit 2004 gemacht wurden. Sioux wurde durch Fördergelder der AAA-Switch Initiative "*e-infrastructure for e-science*" grosszügig unterstützt.

Im nachfolgenden Abschnitt wird der Funktionsumfang der Sioux CBA-Suite erläutert. Diese umfasst mehrere Programme, welche für die Bearbeitung spezifischer Arbeitsschritte im CBA-Zyklus optimiert sind. Im Abschnitt 5.3 werden mittels Sequenzdiagramms die Funktionsweise der sechs zentralen Sicherheitskomponenten von Sioux illustriert. Anschliessend beschreiben wir unsere Erfahrungen und zeigen Evaluationsresultate aus dem Praxisalltag mit mittlerweile über 4000 Einzelprüfungen. Am Ende dieses Kapitels wird ein Einblick in die aktuellen Weiterentwicklungen von Sioux gegeben und die Erkenntnisse aus der Implementierung der CBA-Designprinzipien zusammengefasst.

6.2 Sioux CBA-Suite: CBA-Designprinzipien im Prüfungsalltag

6.2.1 Die Sioux CBA-Suite

Das Sioux-CBA-System besteht aus mehreren Java-Applikationen, die miteinander verbunden sind und als Sioux-CBA-Suite bezeichnet werden. Tabelle 6.15 zeigt die Übersicht über alle Sioux-Applikationen und beschreibt deren Verwendungszwecke.

Name	Hauptverwendungszweck
<i>AuthorClient</i>	Bewirtschaftung des Fragenpools Zusammenstellung und Aufschaltung der Leistungskontrollen Korrektur und Analyse der Resultate Archivierung der Resultate
<i>ExamClient</i>	Durchführung der Leistungskontrolle
<i>TacticalClient</i>	Überwachung der Prüfungsdurchführung
<i>DocumentServer</i>	Versionierung von Prüfungsfragen Speicherung der Musterlösungen
<i>ExamServer</i>	Auslieferung der Leistungskontrolle Speicherung der Antworten der Studierenden

Tabelle 6.15: Die fünf Applikationen der Sioux-CBA-Suite

Im Zentrum des Systems steht der *AuthorClient*. Dieser dient den Dozierenden und der Prüfungsleitung zur Erstellung neuer Prüfungsfragen und zur Bewirtschaftung des Fragenpools. Die Fragen werden zusammen mit den Musterlösungen auf dem *DocumentServer* gespeichert. Nach der Zusammenstellung einer Prüfung im *AuthorClient* kann diese auf dem *ExamServer* aufgeschaltet werden. Dabei werden auch die Berechtigungen gesetzt, wobei vorgegeben wird, wer in welchem Zeitfenster die Leistungskontrolle bearbeiten kann. Die berechtigten Personen können dann mit dem *ExamClient* die Inhalte der Leistungskontrolle vom *ExamServer* laden und die Antworten abschicken. Diese werden wiederum auf dem *ExamServer* gespeichert und können dort nach Abschluss der Leistungskontrolle mittels *AuthorClients* heruntergeladen und korrigiert werden.

In den nachfolgenden Abschnitten werden die einzelnen Sioux-Applikationen detaillierter erläutert und angegeben, wie die Applikationen auf die Client-Computer verteilt und welche Systemanforderungen an diese gestellt werden.

6.2.2 Systemübersicht

Die Sioux CBA-Suite umfasst drei Client- und zwei Server-Applikationen. Während die Letzteren auf einem einzigen Server an der ETH Zürich installiert sind, können die Sioux-Client auf beliebig vielen Computern gestartet werden. Die Clients greifen über ein funktionierendes Computernetzwerk auf die Funktionen der Server zu. Die Server stellen ihre Funktionen via SOAP-Schnittstelle den Clients zur Verfügung. SOAP ist ein Protokoll zum Austausch XML-basierter Nachrichten und wird regelmässig dort eingesetzt, wo ein direkter Zugriff auf die Serverdatenbank nicht erwünscht wird (Kade, Jokisch, & Küpper, 2010, Seite 7). Die SOAP-Schnittstelle wird vom *Jboss Application Server* zur Ver-

fügung gestellt. Die unerlaubte Nutzung der SOAP-Schnittstelle wird verhindert, indem der *Jboss* nur erfolgreich authentifizierten Benutzern den Zugriff erlaubt.

Die Datenübertragung zwischen den Clients und dem Server wird mittels eines SSL-Tunnels (Schwenk, 2010, Kapitel 4: "Secure Socket Layer") gesichert. Dabei kommt ein offizielles Zertifikat zum Einsatz. Mehr dazu im Abschnitt "6.2.6 Zertifikate zur".

Eine Übersicht der möglichen SSL-Tunnels in Abbildung 6.39 zeigt, auf welche Ressourcen die einzelnen Clients Zugriff haben.

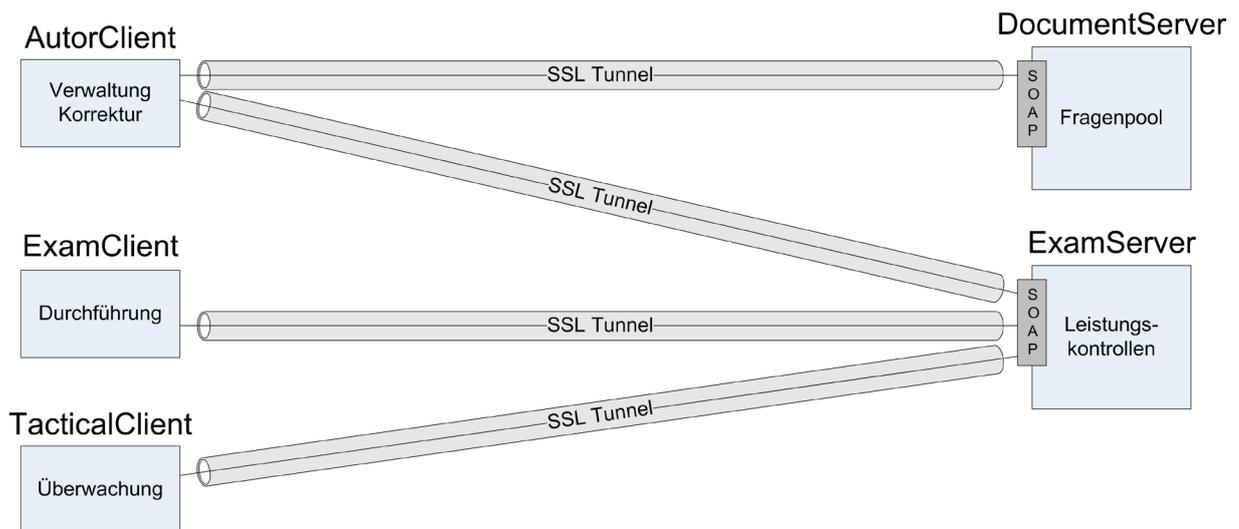


Abbildung 6.39: Systemübersicht der Sioux CBA-Suite. Die SSL-Tunnelverbindungen illustrieren, auf welche Ressourcen die einzelnen Clients Zugriff haben.

Weitere technische Informationen zur Implementierung von Sioux finden sich zusammen mit Quellcode-Beispielen in der Arbeit von Heinrich (Heinrich, 2008).

6.2.3 Client Applikationen

Die Sioux CBA-Suite umfasst drei Client Applikationen: *AuthorClient*, *ExamClient* und *TacticalClient* (Abbildung 6.40). Streng genommen sind es vier Applikationen, wobei die vierte den Namen *SiouxSubmissionViewer* trägt und eine vereinfachte Variante des *ExamClients* darstellt. Erstere kommt im Rahmen der Prüfungseinsicht zum Einsatz und dient dazu, das elektronische Prüfungsdokument anzuzeigen. Sie verfügt aber über keine Verbindung zu einem Sioux-Server, sondern liest das Prüfungsdokument als Datei ein.

In Abbildung 6.40 werden die Einsatzschwerpunkte der drei Sioux-Hauptapplikationen in den fünf Phasen des CBA-Prozesses illustriert. Der zyklische Prozess illustriert den *Reengineering*-Prozess von Leistungskontrollen, wobei die Resultate der Prüfungsanalyse jeweils wieder in die Optimierung der Fragen einfließen. Mehr zu diesem CBA-Entwicklungsprozess findet sich im Kapitel 7.

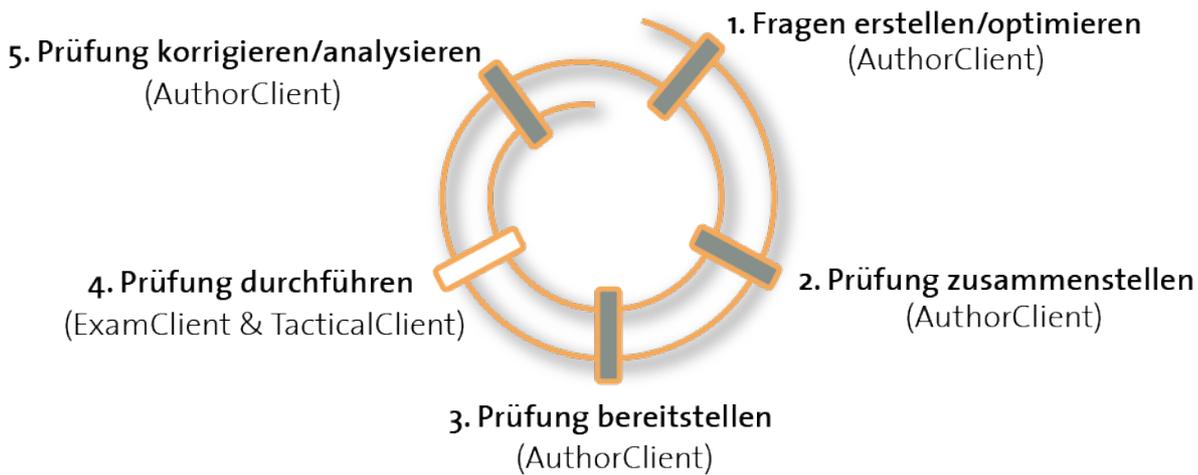


Abbildung 6.40: Funktionsumfang der Sioux Client-Applikationen in Bezug auf den Prüfungsprozess. Die grauen Arbeitsschritte (1,2,3,5) werden vom Sioux *AuthorClient* unterstützt, bei der eigentlichen Prüfungsdurchführung (4) kommen der Sioux *ExamClient* und der *TacticalClient* zum Einsatz.

In den nachfolgenden Abschnitten werden die einzelnen Sioux Client Applikationen kurz vorgestellt. Damit diese auf den Clients gestartet werden können, müssen sie die minimalen Systemvoraussetzungen erfüllen.

Systemvoraussetzungen

Zum Starten der Sioux Client Applikationen wird neben einer funktionierenden Internetverbindung auch die *Java-runtime*-Umgebung benötigt. Diese muss mindestens in der Version 1.6 installiert sein. Da Java eine Abstraktionsschicht zum Betriebssystem darstellt, spielt *Java-runtime*-Umgebung bei der Nutzung von Sioux keine Rolle.

Als Bildschirmauflösung wird 1280x1024 empfohlen. Diese Auflösung entspricht bei Drucklegung dieser Arbeit der Mindestauflösung von TFT-Bildschirmen in den studentischen Computerarbeitsräumen an der ETH Zürich. Im Praxiseinsatz hat sich gezeigt, dass diese Auflösung nötig ist, um die verwendeten Fall-basierten Prüfungsfragen optimal am Bildschirm darstellen zu können, ohne dass die Studierenden den Bildschirminhalt scrollen müssen.

AuthorClient

Der *AuthorClient* wird von den Dozierenden und Prüfungsleitenden bedient. Er bietet folgende Funktionen an: Bewirtschaftung des Fragenpools, Bereitstellen von Leistungskontrollen sowie Korrektur, Analyse und Archivierung von Prüfungsergebnissen.

Fragenpoolbewirtschaftung: Die Fragenpoolbewirtschaftung beinhaltet das Erstellen neuer und Verändern bestehender Prüfungsfragen. Zu jeder Prüfungsfrage können Metadaten, wie zum Beispiel geschätzter Schwierigkeitsgrad oder geprüfte Konzepte, eingegeben werden. Die Metadaten dienen später beim Zusammenstellen der Leistungskontrolle zum Suchen geeigneter Prüfungsfragen. Zusammen mit den Prüfungsfragen

werden auch die Quelldateien abgespeichert. Diese umfassen alle Dateien, die zum Erstellen der Prüfungsfrage verwendet wurden. Dadurch wird eine spätere Überarbeitung der Prüfungsfragen effizienter.

Bei der Fragenerstellung stehen Fall-basierte Fragetypen zur Verfügung. Diese beinhalten auf der linken Bildschirmhälfte eine Aufgabenstellung und auf der rechten Seite eine beliebige Anzahl an Fragestellungen (Items). Bei Sioux stehen drei Item-Typen zur Verfügung: richtig/falsch-Items, Lückentext-Items und *Hotspot*-Items. Die drei Item-Typen werden in Kapitel 7 beschrieben. Pro Item muss die Musterlösung angegeben werden. Es kann ein Feedback erfasst werden, das den Studierenden zusätzlich zur Korrektur angezeigt wird. Dieses Feedback unterscheidet sich je nach Situation, ob die Studierenden eine korrekte oder falsche Antwort abgegeben haben.

Die Fragenpoolbewirtschaftung unterstützt auch den Peer-Review-Prozess, wobei mit dem Prüfungsstoff vertraute Personen Prüfungsfragen erstellen und von erfahrenen Prüfungsfragen-Autoren ein Feedback erhalten. Das Feedback kann als kleine "ToDo"-Mitteilungen direkt im *AuthorClient* dort hinterlegt werden, wo eine Überarbeitung nötig ist (Abbildung 6.41). Erledigte "toDos" können als solche markiert und archiviert werden.

The screenshot shows the 'Aufgabe' (Task) tab in the AuthorClient. The task text reads: 'Die folgende **Permutations-Matrix** visualisiert die Entwicklung des Schweizer Linienverkehrs zwischen 2004 und 2008. Für verschiedene Flughäfen sind die Passagierzahlen für das Total über alle Flughäfen ersichtlich.' A yellow sticky note with a red pushpin is placed over the text, containing the handwritten note: 'ToDo: Bild schlecht lesbar' and 'bitte vergrößern'. Below the text is a data table with columns for years (2004-2008) and rows for cities (Basel, Genève, Zürich, Bern, Total). The table data is as follows:

	2004	2005	2006	2007	2008
Basel	49.7	48.3	46.9	45.5	49.70
Genève	125.7	119.2	112.7	106.3	125.75
Zürich	225.5	220.5	215.6	210.6	225.62
Bern	5.0	4.5	3.3	2.1	5.08
Total	403.1	393.3	383.54	373.74	403.12

On the right side of the interface, there is a 'Multiple Choice' question type with a 'Punkte' (Points) field set to 1.0. Below it, another 'Multiple Choice' question type is partially visible, also with a 'Punkte' field set to 1.0. The interface includes icons for editing, deleting, and marking items as correct or incorrect.

Abbildung 6.41: "ToDo" Mitteilungen unterstützen den *Peer-Review*-Prozess zur Verbesserung der Prüfungsfragen.

Alle Prüfungsfragen werden versioniert, so dass alte Versionen der Fragen angeschaut werden können. Dies ermöglicht es unter anderem, eine Frage in einer Leistungskontrolle einzusetzen und diese parallel dazu weiter zu entwickeln.

Bereitstellen von Leistungskontrollen: Das Bereitstellen von Leistungskontrollen umfasst zwei Schritte: erstens muss die Leistungskontrolle aus den gespeicherten Prüfungsfragen zusammengestellt und zweitens auf dem *ExamServer* aufgeschaltet werden.

Beim Zusammenstellen der Leistungskontrolle unterstützt das Sioux System das Autorenteam, indem anhand der Metadaten geeignete Prüfungsfragen vorgeschlagen werden. Die Prüfung kann themenbasiert gruppiert werden. Die Gruppierung erlaubt später, die Reihenfolge der Prüfungsfragen innerhalb einer Gruppe zufällig anzuzeigen. Dazu kann im *AuthorClient* eine entsprechende Option ausgewählt werden. Um später die Leistungskontrolle zu starten, müssen alle Studierenden erfasst werden. Es steht eine csv-Importfunktion (*column separated value*) zur Verfügung, welche das Einlesen von Studierendenlisten aus LMS oder Lehrapplikationen ermöglicht. Zusätzlich bietet Sioux die Möglichkeit, für die Auslieferung der Leistungskontrolle ein Zeitfenster zu definieren und auf gewisse IP-Adressen zu beschränken. Ausserhalb des Zeitfensters oder des Adressbereichs ist kein Bezug der Leistungskontrolle möglich. Abbildung 6.42 zeigt die Bedienoberfläche zur Festlegung des Modus einer Leistungskontrolle.

Abbildung 6.42: Festlegen des Modus einer Leistungskontrolle im AuthorClient. Festzulegen sind: Art der Leistungskontrolle, Prüfungsdauer (in Minuten), Darstellungsoptionen, Angabe des Zeitfensters und des IP-Bereichs.

Damit eine Leistungskontrolle auf dem *ExamServer* bereitgestellt werden kann, muss deren Modus bestimmt werden: Bei einer formativen Leistungskontrolle wird sofort nach Beendigung der Leistungskontrolle die individuelle Korrektur der Resultate angezeigt. Bei einer summativen Prüfung wird nach Prüfungsschluss der *ExamClient* automatisch geschlossen, ohne dass die Korrekturen angezeigt werden.

Die eigentliche Bereitstellung der Leistungskontrolle geschieht per Mausklick. Hierzu muss der gewünschte *ExamServer* eingetragen werden. Anschliessend überträgt der *AuthorClient* automatisch die Daten der Leistungskontrolle an den *ExamServer*. Im Falle einer formativen Leistungskontrolle werden zusätzlich noch die Musterlösungen mit übertragen. Bei einer summativen Prüfung fehlen diese auf dem *ExamServer*. Zusätzlich werden beim Bereitstellen die Quelldaten entfernt. Diese werden bei der Durchführung der Leistungskontrolle nicht benötigt. Für jede Leistungskontrolle generiert die *AuthorClient* ein individuelles Schlüsselpaar, das dazu verwendet wird, die Daten auf den *Backup*-Speichermedien (zum Beispiel auf USB-Sticks) der Studierenden zu verschlüsseln, um sie vor unerlaubten Zugriffen zu schützen.

Korrektur und Analyse der Leistungskontrollen: Nach Abschluss der Leistungskontrolle kann der *AuthorClient* dazu verwendet werden, die Resultate zu korrigieren und zu analysieren. Der *AuthorClient* bezieht die Resultate vom *ExamServer* und holt die Musterlö-

sungen vom *DocumentServer*. Dieser Vorgang dauert bei einer Leistungskontrolle mit 300 Studierenden rund 2 Minuten. Die meiste Zeit wird für die Übertragung der Resultate verwendet. Die eigentliche Korrektur, das heisst das Vergleichen der Antworten der Studierenden mit der Musterlösung, dauert wenige Sekunden.

Die offenen Antworten, zu denen zum Beispiel die Lückentexte und Hotspot-Items gehören, sollten vor der Notengebung noch manuell überprüft werden. Der *AuthorClient* unterstützt diesen Vorgang, indem auf Wunsch ausschliesslich die falschen Antworten angezeigt werden. So können diese einfach auf offensichtliche Tippfehler oder Fehler in der Musterlösung durchsucht werden. Bei Letzterem kann per Mausclick die Musterlösung ergänzt und das Item neu korrigiert werden. Die Korrekturen der restlichen Items bleiben unverändert. Die Musterlösung kann wahlweise nur für die aktuelle Korrektur verwendet oder persistent auf dem *DocumentServer* gespeichert werden. Bei einem Tippfehler ergänzen wir meistens die lokale Musterlösung, so dass derselbe Tippfehler auch bei anderen Studierenden als korrekt gewertet werden kann. Abbildung 6.43 zeigt die Bedienoberfläche des *AuthorClients* im Korrekturmodus.

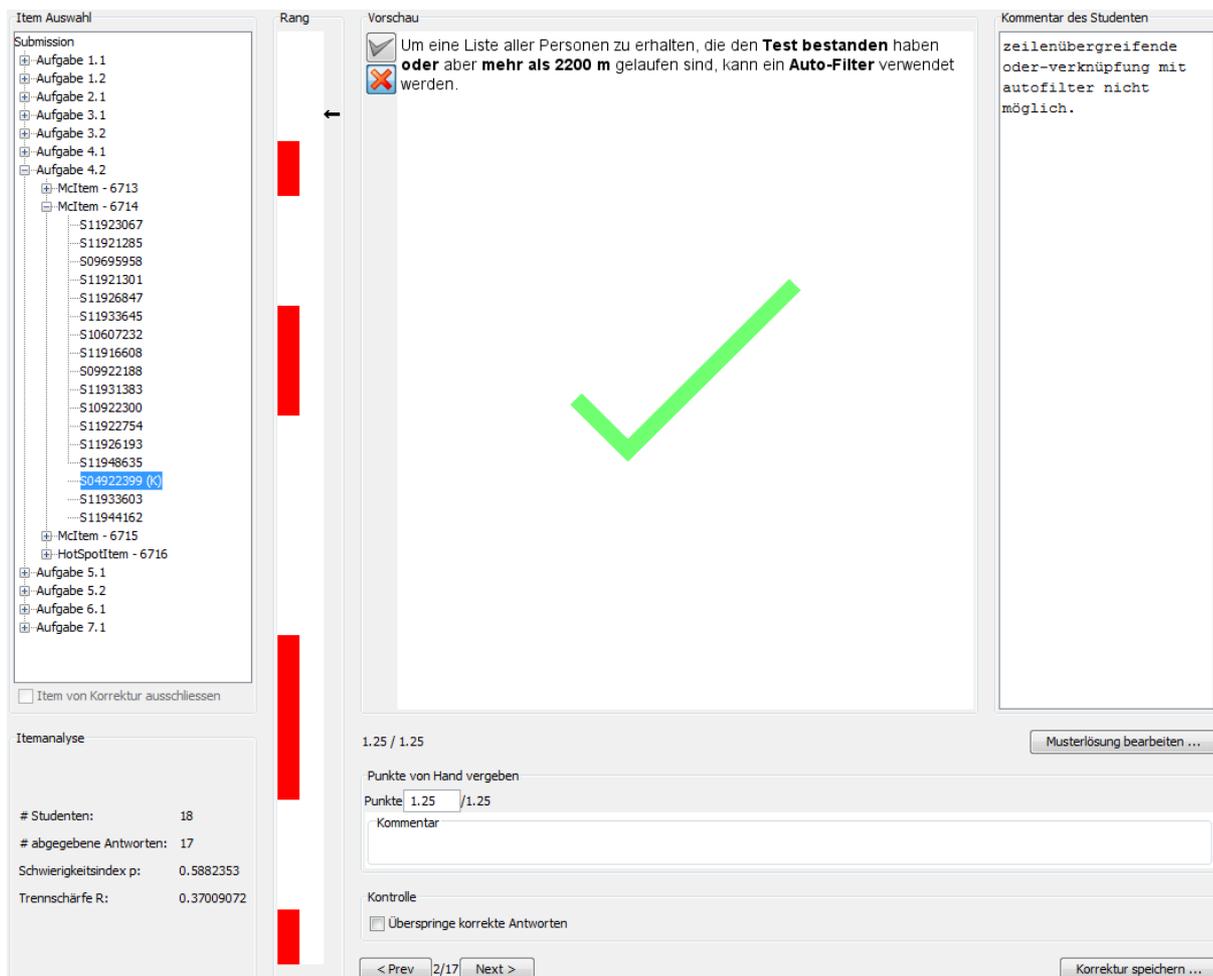


Abbildung 6.43: Der *AuthorClient* im Korrekturmodus. Wahlweise können nur noch die falsch abgegebene Antworten angezeigt werden. Dies erleichtert das Finden von offensichtlichen Tippfehlern oder unvollständigen Musterlösungen. Mittig wird die Fehlerverteilung des angezeigten Items dargestellt (hier abgebildet: Analyse der Frage 4.2, Item 2 der Basisprüfung, Februar 2012, n=17).

Zusätzlich zu den Antworten zeigt der *AuthorClient* auch allfällige Kommentare der Studierenden an. Meist verwenden die Studierenden die Kommentarfunktion, um ihre Antworten zu begründen. Ab und an geben die Kommentare auch Hinweise auf Fehler in der Aufgabenstellung. Der *AuthorClient* bietet die Möglichkeit, die automatische Korrektur zu übersteuern, indem die Punkte für eine Aufgabe manuell gesetzt werden. Dies macht dann zum Beispiel Sinn, wenn ein Student zwar eine falsche Antwort gegeben hat, aber der abgegebene Kommentar darauf schliessen lässt, dass er die Aufgabe eigentlich lösen konnte.

Der Korrekturmodus wird zur Darstellung der Fehlerverteilung und Item-spezifischer Kennzahlen wie Trennschärfe und Schwierigkeitsindex ergänzt. Diese Informationen sind hilfreich bei der Identifikation von auffälligen Items, welche beispielsweise auf eine fehlerhafte oder irreführende Formulierung des Item-Textes zurückzuführen sind. Der *AuthorClient* erlaubt es, per Mausklick ein Item von der Benotung auszuschliessen. Dadurch reduziert sich automatisch die maximal erreichbare Punktzahl der Leistungskontrolle.

Die korrigierten Resultate können als csv-Dateien zur Benotung und weiteren Analysen exportiert werden. Die Benotung ist in der Version zur Zeit der Drucklegung noch nicht Teil der Sioux CBA-Suite. Es steht aber eine Korrekturvorlage zur Verfügung, mit welcher die Sioux-Daten in herkömmliche Tabellenkalkulationssoftware importiert und weiterverarbeitet werden können. Weitere Informationen zur Korrektur und Benotung finden sich im technischen Report von Dahinden (Dahinden & Hinterberger, 2010).

Archivierung der Leistungskontrollen: Zur Archivierung der Leistungskontrolle werden alle Resultate vom *ExamServer* als komprimierte Datei heruntergeladen. Dazu stellt der *AuthorClient* die Funktion "Archiviere *Submissions* von *ExamServer*" zur Verfügung. Die exportierten Resultate beinhalten neben den signierten Prüfungsdokumenten auch alle abgegebenen Antworten. Die Resultate werden nicht verschlüsselt, damit sie später jederzeit einsehbar sind. Zum Schutz der Resultate müssen, wie auch bei Papierprüfungen üblich, diese physisch von unberechtigten Zugriffen geschützt werden.

ExamClient

Der *ExamClient* wird von den Studierenden bedient. Er dient dazu, die Leistungskontrollen anzuzeigen und die Antworten in signierter Form an den *ExamServer* zu übermitteln. Die Oberfläche des *ExamClients* ist bewusst einfach gehalten und zeigt nur die für die Durchführung der Leistungskontrolle notwendigen Informationen (Abbildung 6.44).

Der *ExamClient* stellt die sicherheitsrelevanteste Applikation der Sioux CBA-Suite dar. Nicht nur, weil er zuverlässig funktionieren muss, sondern weil er mit der Signierung der Prüfungsergebnisse den belegbaren Nachweis der Vollständigkeit und Unveränderlichkeit der Prüfungsergebnisse liefert. Da dieser Nachweis nur möglich ist, wenn der *ExamClient* erwartungsgemäss funktioniert, zählen wir ihn neben der Client-Hardware

zur vertrauenswürdigen Zone unseres CBA-Systems. Der nachfolgende Abschnitt listet die Gründe hierzu auf.

Dahinden **S97102362** << vorherige Frage nächste Frage >> 56

Exam

- Publizieren im Internet
 - Aufgabe 1.1 (5.0 Punkte)
 - Aufgabe 1.2 (5.0 Punkte)
- Tabellenkalkulation
 - Aufgabe 2.1 (5.0 Punkte)
- Datenvisualisierung
 - Aufgabe 3.1 (5.0 Punkte)
 - Aufgabe 3.2 (5.0 Punkte)
- Daten verwalten I
 - Aufgabe 4.1 (4.0 Punkte)
 - Aufgabe 4.2 (5.0 Punkte)
- Daten verwalten II
 - Aufgabe 5.1 (5.0 Punkte)**
 - Aufgabe 5.2 (5.0 Punkte)
- Makroprogrammierung
 - Aufgabe 6.1 (5.0 Punkte)
- Vorlesung
 - Aufgabe 7.1 (5.0 Punkte)

Die Datenbank zur Verwaltung von **Spesen**:

```

graph TD
    SpesenReport((SpesenReport)) --- Spesen((Spesen))
    SpesenReport --- Spesen_Typ((Spesen_Typ))
  
```

SpesenReport

ID	Angestellter_ID	Spesen_ID
1	2	4
2	2	3
3	5	1
4	3	2
5	3	1
6	4	7
7	1	8
8	5	6
10	9	5

Spesen_Typ

ID	Typ	Typ_ID
1	Verpflegung	
2	Kunden	
3	Sonstiges	
4	Unterkunft	
5	Mobilität	

HotSpot Item: 1.0 Punkte

Markieren Sie im Beziehungsdiagramm die zwei Tabellen, die in einer **n:n-Beziehung (auch "Unendlich:Unendlich"-Beziehung)** stehen.

Multiple Choice Item: 1.25 Punkte

Wenn Sie folgende **SQL-Abfrage** ausführen

```

SELECT Spesen_Typ.Typ, Angestellter.Nachname
FROM Angestellter, SpesenReport, Spesen, Spesen_Typ
WHERE ((Angestellter.Nachname)="Leimberger");
  
```

erhalten Sie folgendes **Resultat**:

Typ	Nachname
Unterkunft	Leimberger
Unterkunft	Leimberger
Kunden	Leimberger

Multiple Choice Item: 1.5 Punkte

Die **Referentielle Integrität** der Datenbank wird durch einen Eintrag in der Tabelle **SpesenReport** **verletzt**.

Lückentext Item: 1.25 Punkte

Ergänzen Sie folgende **Abfrage**, mit welcher Sie das Resultat unter **(1)** erzielen wollen:

```

SELECT Spesen.ID, Spesen.Text, Spesen.Datum, Spesen.Betrag,
Spesen_Typ_ID
FROM Spesen_Typ
INNER JOIN Spesen ON Spesen_Typ.ID = Spesen_Typ_ID
WHERE ( [ ] > 40
AND [ ] = 'Mobilität' );
  
```

Prüfungsbedingungen anzeigen

Prüfung beenden

ID	Text	Datum	Betrag	Typ_ID
7	Tanken	21.11.2010	65	5
8	Taxi	09.11.2010	43	5

Abbildung 6.44: Bedienoberfläche des ExamClients. Neben den eigentlichen Prüfungsfragen zeigt der ExamClient nur die unmittelbar zur Durchführung der Leistungskontrolle notwendigen Informationen an. Dazu gehört eine Übersicht über die Aufgaben (mit Fortschrittsanzeige), der Name des Probanden oder der Probandin (als Information für die Raumaufsicht und zu Durchführung der Präsenzkontrolle), die verbleibende Zeit sowie Elemente zur Navigation und Abgabe der Leistungskontrolle.

ExamClient als Teil der Vertrauenszone

Als "vertrauenswürdige Zone" werden jene Bereiche eines Computersystems verstanden, von denen erwartet wird, dass diese sich gemäss einer vordefinierten Art und Weise verhalten (Koenig, 2004).

Bei CBA können folgende Punkte zur vordefinierten Funktionsweise eines ExamClients zählen:

- 1) Der ExamClient stellt die Leistungskontrolle am Computerbildschirm dar. Sowohl die Prüfungsleitung als auch die Studierenden müssen sich darauf verlassen können, dass der ExamClient die vom Server bezogene Leistungskontrolle korrekt darstellt.
- 2) Der ExamClient signiert alle Antworten der Studierenden. Es muss darauf Verlass sein, dass er diese vollständig und unverändert signiert. Eine nach der Signatur manipulierte Antwort kann anhand der fehlerhaften Signatur erkannt werden, aller-

dings eine durch den Client manipulierte und erst anschliessend signierte Antwort kann nicht als solche erkannt werden.

- 3) Der *ExamServer* erstellt aus allen abgegeben Antworten das Prüfungsdokument, welches als Basis für die Korrektur und Benotung dient. Zur Überprüfung der Vollständigkeit des Prüfungsdokuments nummeriert der *ExamClient* alle Antworten. Macht er dabei Fehler, kann dieser nachträglich nicht mehr erkannt werden.
- 4) Zum Signieren der Antworten benötigt der *ExamClient* den privaten Signaturschlüssel des Probanden oder der Probandin. Da wir aus organisatorischen Gründen auf den Einsatz von kryptografischen *Tokens* verzichten müssen, liegt während der Prüfung der Signaturschlüssel im Arbeitsspeicher und auf dem *Backup*-Speichermedium der Prüflinge. Käme die Prüfungsleitung in den Besitz dieses privaten Signaturschlüssels, könnte sie nachträglich die Antworten des betroffenen Probanden verändern, ohne dass diese aus technischer Sicht als solche erkannt würden.

Aus pragmatischer Sicht erlaubt der Entscheid, den *ExamClient* als Teil der vertrauenswürdigen Zone zu sehen, die Durchführung von CBA im Praxisalltag unter Berücksichtigung der vorhandenen Infrastruktur. Obwohl in einem solchen Umfeld die korrekte Funktionsweise eines CBA-Arbeitsplatzes (bestehend aus *Client-Hardware* und *ExamClient-Software*) aus technischer Sicht nicht garantiert werden kann, so haben wir dennoch verschiedene Massnahmen getroffen, um die unserer Meinung nach häufigsten Probleme ausschliessen zu können (Tabelle 6.16).

Problem	Massnahme	Überprüfung
<i>Client-Hardware</i> wurde manipuliert	Gehäuse wird mit einem Schloss gesichert	Visuelle Überprüfung zu Beginn jeder Prüfungssession
<i>ExamClient</i> wurde manipuliert	Originalcode wird mittels eines digitalen Codesigning-Zertifikats geschützt (vgl. Abschnitt "6.2.6 Zertifikate zur Absicherung von Sioux")	Visuelle Überprüfung des Herausgebers des Codes. Dieser wird durch die Java-Umgebung bei jedem Start angezeigt.
<i>ExamClient</i> verhält sich nicht gemäss der vordefinierten Funktionsweise	Der <i>ExamClient</i> wird zusammen mit den Prüfungsergebnissen archiviert.	Überprüfung des <i>ExamClient</i> -Codes auf mögliche Fehlerursachen.

Tabelle 6.16: Massnahmen zur Überprüfung der korrekten Funktionsweise der CBA-Arbeitsstation als Teil der vertrauenswürdigen Zone.

Neben diesen drei Massnahmen setzten wir auch auf eine konsequente Gewaltentrennung beim Aufbau der CBA-Arbeitsstationen. Die Zuständigkeiten und Arbeitsschritte wurden in einem technischen Bericht festgehalten (Dahinden & Hinterberger, 2010, Seite 21 ff).

TacticalClient

Der *TacticalClient* wird von der Prüfungsleitung bedient und dient zur Überwachung des Sioux CBA-Systems während der Durchführung einer Leistungskontrolle. Zur Präsenzkontrolle werden folgende Informationen benötigt (und können deshalb im *TacticalClient* abgefragt werden):

- Welche Studierenden haben eine Leistungskontrolle vom *ExamServer* bezogen?
- Welche Studierenden sind nicht zur Leistungskontrolle erschienen?

Zur Erkennung technischer Probleme und zur Einleitung allenfalls notwendiger Hilfeleistungen können folgende Informationen eingesehen werden:

- Wie viele Antworten hat jeder einzelne Student bereits abgegeben?
- Wie hoch ist die verbleibende, individuelle Prüfungszeit der Studierenden?
- An welcher CBA-Arbeitsstation arbeitet ein bestimmter Student?
- Wer hat die Prüfung bereits beendet?
- Wurde das signierte Prüfungsdokument auf dem Server gespeichert?

Neben der reinen Darstellung von Informationen zum Prüfungsablauf bietet der *TacticalClient* auch die Möglichkeit, die Prüfungszeit für bestimmte Studierende zu verlängern. Dies kann dann nötig sein, wenn sie aufgrund eines Zwischenfalls die CBA-Arbeitsstation wechseln mussten oder anderweitig an der Bearbeitung der Leistungskontrolle gehindert wurden. Die Bedienoberfläche des *TacticalClients* wird in Abbildung 6.45 gezeigt.

The screenshot displays the TacticalClients application interface, divided into several sections:

- System Control:** Includes fields for Server (sioux.inf.ethz.ch), Refresh (60), Exam ID (10), and Room ID (0), along with a 'Refresh Now' button.
- Common Control:** A section for general settings.
- Selected Student Information:** Displays details for a selected student: Name (Dahinden), Given Name (Markus), Legi (S97102362), IP Address (129.132.12.184), Host Name (129.132.12.184), Exam Start (checked), Number of Answers (28), Time Left (53), Extra Time (0), and Exam Submitted (unchecked).
- All Registered Students:** A table listing all registered students with columns for LegiNumber, Name, Given Name, and Progress. The student Markus Dahinden (Legi S97102362) is highlighted in yellow, indicating he is the selected student.

LegiNumber	Name	Given Name	Progress
S0802398	Arnsperger	Hajji	42
S08024958	Conrad	Delfina	44
S0802498	Fabrocht	Jelena	44
S1802732	Glasner	Alexandra	42
S1802738	Herzog	Philipp	44
S1818808	Mattner	Hilmar	44
S1821088	Carroll	Claudia	44
S1821301	Dahlenbach	Lorenz	44
S1802754	Dehnbachler	Jana	44
S1802807	Demmel	Florence	43
S1802890	Steinhilber	Susanne	44
S1802847	Pfister	Georg	44
S1803300	Rehmann	Svenja	44
S1803800	Lutz	Jessika	43
S1803845	Pfister	Tim	44
S1844852	Knecht	Tanja	44
S1844858	Stanger	Florian	42
S97102362	Dahinden	Markus	28
S04118857	Pfeifer	Lukas	0
S1818448	Duffy	Adrian	0
S1802318	Dalmaden	Lara	0
S0800328	Pfistermann	Andreas	0
S0802348	Blitzer	Lukas	0
S0803379	Hessli	Florian	0
S1844858	Daniel	Stefan	0
S0802804	Egger	Simon	0
S4800801	Dechenhauser	Hans-Joachim	0
S07818458	Hauser	Mathias	0
S08028158	Preiss	Oskar	0
S0812318	Kasser	Tanja	0

Abbildung 6.45: Bedienoberfläche des TacticalClients. Diese Applikation zeigt während der Durchführung einer Leistungskontrolle zahlreiche Statusinformationen: Wer hat die Prüfung bezogen? Wie viel Prüfungszeit haben die Studierenden noch zur Verfügung (in diesem Beispiel hat der Student Markus Dahinden noch 53 Minuten)? Wie viele Antworten haben die Studierenden bereits abgegeben? An welchem Computer arbeitet ein bestimmter Student? Auf Wunsch kann die verbleibende Prüfungszeit individuell erhöht werden (aus Datenschutzgründen sind gewisse Informationen unscharf dargestellt).

Der *TacticalClient* könnte bei Bedarf auch weitere Überwachungsfunktionen der Prüfungs-Computer übernehmen. Denkbar wäre die Überwachung des Bildschirminhalts der Studierenden oder die Möglichkeit, den Studierenden individuelle Nachrichten zu schicken. Dies könnte speziell dann hilfreich sein, wenn die Prüfungszeit verlängert wurde und man die Studierenden entsprechend informieren möchte.

Wie in Abbildung 6.39 gezeigt, kommunizieren alle Sioux CBA-Clients mit einer oder beiden Sioux-Server Applikationen. Diese werden im nachfolgenden Abschnitt kurz erläutert.

6.2.4 Server Applikationen

Die Funktionalität der drei Sioux CBA-Clients wird durch zwei Server Applikationen sichergestellt: der *DocumentServer* speichert die Fragenpools und der *ExamServer* stellt die zur Durchführung einer Leistungskontrolle notwendigen Funktionen zur Verfügung.

Diese beiden Applikationen funktionieren unabhängig voneinander und können auf unterschiedlichen Rechnern installiert und betrieben werden. Dadurch wird die Sioux-CBA-Suite flexibel einsetzbar, indem beispielsweise die kollaborative Bewirtschaftung von Fragenpools vereinfacht wird, da nicht alle Bildungsinstitutionen eigene *DocumentServer* betreiben müssen.

Ein weiterer Vorteil der Unabhängigkeit der Sioux-Server Applikationen ist die Tatsache, dass die Applikationen für deren jeweiligen Einsatz optimiert und deren Sicherheitsanforderungen individuell definiert werden können. So verfügt beispielsweise der *Exam-Server* über lediglich 800 Zeilen Programmcode, deren Pflege einfacher ausfällt, als wenn er in den um ein Vielfaches grösseren *DocumentServer* integriert wäre.

Verfügbare technische Infrastruktur für Sioux an der ETH Zürich

Der Sioux-Server ist in einem klimatisierten Serverraum an der ETH Zürich in einem 19-Zoll Rack verbaut. Es handelt sich um einen SUN-Microsystem-1HE Server (X2200) mit 8 GB Arbeitsspeicher, zwei Dual-Core AMD Opteron 2.6 GHz Prozessoren und zwei 250GB Festplatten im RAID1-Verbund. Der Server ist nicht-redundant an einen 1Gbit-Switch angehängt. Der Zugang zum Serverraum bleibt autorisierten Personen vorbehalten und wird mittels *Badge* geregelt. Auf dem Server läuft der *jboss Application Server 4.0* unter dem Betriebssystem Ubuntu 12.04 LTS. Sioux wurde mittels eines MySQL-connectors an eine MySQL5 Datenbank angebunden. Mit dieser Infrastruktur wurden Prüfungen von gleichzeitig bis zu 200 Studierenden durchgeführt. Messungen haben ergeben, dass sogar 400 gleichzeitig stattfindende Prüfungen möglich wären. Im Abschnitt "6.4.5 Performancemessungen" finden sich weitere Informationen hierzu.

DocumentServer

Der *DocumentServer* speichert die Fragenpools und verwaltet die verschiedenen Versionen der Prüfungsfragen. Bei der Zugriffsberechtigung zum Fragenpool unterscheidet der *DocumentServer* mehrere Benutzergruppen. Es können neue Fragenpools angelegt und die Berechtigungen individuell festgelegt werden. Dies erlaubt eine kollaborative Bewirtschaftung der Fragenpools.

Beim Abspeichern einer überarbeiteten Frage sichert der *DocumentServer* die Frage unter einer neuen Versionsnummer. Dies erlaubt, dass alle Vorgängerversionen später angezeigt und gegebenenfalls wiederhergestellt werden können. Mit der Versionierung kann aber auch verhindert werden, dass eine für eine Leistungskontrolle vorgesehene Frage ohne Wissen der Prüfungsleitung in der Zeit zwischen dem Zusammenstellen und dem Bereitstellen der Prüfung noch verändert wird. Der *DocumentServer* steuert dies, indem er für eine bestimmte Leistungskontrolle nicht nur die Frage, sondern auch die jeweilige Version abspeichert.

ExamServer

Der *ExamServer* ist massgeblich bei der Durchführung von CBA beteiligt. Er übernimmt zwei Hauptaufgaben: Herausgabe der Leistungskontrolle an berechnigte Personen sowie die persistente Speicherung der Antworten und Prüfungsdokumente.

Bevor der *ExamServer* die Leistungskontrolle herausgibt, überprüft er, ob erstens die angemeldete Person zu Bearbeitung der Prüfung legitimiert ist, ob zweitens das Zeitfenster für die Bearbeitung der Leistungskontrolle gültig ist und ob drittens der Computer, an dem die Person arbeitet, für die Bearbeitung der Leistungskontrolle berechnigt ist. Erst wenn alle drei Bedingungen erfüllt sind, wird die Leistungskontrolle ausgeliefert.

Bei der Speicherung der Antworten und Prüfungsdokumente ist der *ExamServer* weniger restriktiv. So erlaubt der Server beispielsweise das Speichern von Antworten und Prüfungsdokumenten auch ausserhalb des durch die Prüfungsleitung definierten Bearbeitungszeitfensters. So kann sichergestellt werden, dass selbst bei langsamen Netzwerkverbindungen oder Verzögerungen im Prüfungsablauf alle Antworten gespeichert werden können.

Neben den beiden erwähnten Aufgaben übernimmt der *ExamServer* auch die Kontrolle der individuellen Prüfungszeit. Dieses Vorgehen ist sicherer, als wenn die Client-Uhren zur Kontrolle der Prüfungszeit verwendet werden würden. Da diese von den Studierenden verändert werden können, würden diese so die Kontrolle über die Prüfungszeit übernehmen können. Um dies zu verhindern, setzt der Sioux-Server beim Bezug jeder Prüfung einen Zeitstempel und errechnet daraus die individuell verbleibende Prüfungszeit. Diese kann der *ExamClient* beim *ExamServer* in regelmässigen Abständen abfragen. In der Praxis hat sich ein Abfrageintervall von 10 Sekunden bewährt. Ein kürzeres Intervall führt zu einer höheren Netzwerklast, welche bei hohen Studierendenzahlen keineswegs vernachlässigbar ist. Mehr dazu im Abschnitt "6.4.5 Performancemessungen". Wird die Verbindung zum *ExamServer* unterbrochen, muss der Client die Prüfungszeit selber anhand der Differenz aus der Start- und der aktuellen Zeit bestimmen.

6.2.5 Deployment

Das *Deployment*, das heisst das Verteilen der Sioux-Applikationen auf die Prüfungsclients, ist ein zentraler Arbeitsschritt in einem CBA-Prozess. Aus legaler Sicht muss dabei sichergestellt werden, dass ausschliesslich der gewünschte Prüfungsclient verteilt wird. Aus organisatorischer Sicht soll der Zeitaufwand für das *Deployment* minimal gehalten werden.

Sioux verwendet die Java-Webstart-Infrastruktur fürs *Deployment* der Client-Applikationen. Diese hat den Vorteil, dass die Applikationen ohne grossen Aufwand über das Internet verteilt werden können. Das intelligente Java-Webstart-Caching stellt

dabei sicher, dass beim Start der Applikation diese nur dann neu übertragen werden muss, wenn die lokal zwischengespeicherte Version nicht mehr mit der Server-Version übereinstimmt. Ein zweiter Vorteil ist die Möglichkeit, die Applikationen mittels eines *Code-signing* Zertifikats digital zu signieren und so vor Veränderungen zu schützen. Mit Hilfe der Signatur kann - unter der Annahme die Arbeitsstationen sind nicht kompromittiert - beim Start der Applikationen anhand des angezeigten Zertifikatinhabers auf die Authentizität der verwendeten Sioux-Applikation geschlossen werden.

Die beiden Sioux-Clients sind rund fünf Megabytes (Mb) gross. Darin eingebunden sind die rund vier Mb grossen jboss-spezifischen Client-libraries (z.B für die GUI-Darstellung oder die Client-Server-Kommunikation).

6.2.6 Zertifikate zur Absicherung von Sioux

Die Java-Webstart-Applikation wird vom Client per Hypertext Transfer Protokoll (HTTP) angefordert. Das geschieht in zwei Schritten:

- 1) Der Client fordert die jnlp-Datei an, die Informationen über die eigentliche Applikation bereitstellt. Dazu gehören Startparameter und benötigte Bibliotheken. Letztere werden, falls nicht bereits in der aktuellsten Version auf dem Client vorhanden, im zweiten Schritt vom Server nachgeladen.
- 2) Nachladen und Ausführen der eigentlichen java-Programmdateien.

Bei der sicheren Kommunikation des Sioux-Clients mit dem Sioux-Server müssen zwei Bedingungen erfüllt sein: zum einen muss der Client das korrekte Sioux-Programm herunterladen und zum anderen muss die Kommunikation zwischen Client und Server abhörsicher sein, damit Drittpersonen an keine vertraulichen Informationen gelangen können. Abgehörte Kennwörter könnten zum Beispiel dazu verwendet werden, unerlaubterweise Prüfungen einzusehen oder Prüfungsergebnisse abzufragen. Die Veränderung von Prüfungsergebnissen ist allerdings selbst mit einem gültigen Signaturschlüssel eines Studierenden nicht möglich, da Resultate aufgrund der verfügbaren Methoden der Server-Schnittstelle nur hinzugefügt, niemals aber überschrieben werden können.

Bei der Sioux-CBA-Suite kommen insgesamt vier digitale Schlüsselpaare zum Einsatz. Zwei davon sind offizielle Zertifikate, welche von einer externen *Public Key* Infrastruktur (PKI) ausgestellt wurden. Die beiden anderen sind selbst signierte Zertifikate, wobei die Identität des Schlüsselinhabers nicht von einer externen Instanz überprüft, sondern lediglich aus seinen Anmeldedaten am Sioux-System hergeleitet wurden.

Alle Schlüssel weisen eine Länge von 2048 bit auf und gelten zum Zeitpunkt der Drucklegung dieser Arbeit aus Sicht der Verschlüsselungsstärke als sicher. Mehr dazu in Abschnitt "5.3.3 CBA-Designprinzip #3: Belegbarkeit des Prüfungsergebnisse durch client-seitige, digitale Signaturen".

Diese vier Schlüsselpaare ermöglichen einerseits eine sichere Kommunikation zwischen den Sioux-Clients und dem Sioux-Server. Andererseits werden sie verwendet, um nachträgliche Veränderungen der Daten zuverlässig erkennen zu können:

- 1) **Offizielles Code-Signing Zertifikat:** Die Sioux-Client-Applikationen wurden mit einem *Code-Signing* Zertifikat von QuoVadis digital signiert. Das Zertifikat ist auf den Namen "ETH Zürich" ausgestellt. So kann sichergestellt werden, dass die auf den Clients-installierten *Java-Runtime*-Umgebungen die Urheberschaft des Sioux-Clients eruieren können. Die Signierung der Applikation ist aber auch nötig, damit erweiterte Systemfunktionen, zu denen zum Beispiel das Beschreiben und Lesen von Datenträgern gehört, zur Verfügung stehen.
- 2) **Offizielles SSL-Zertifikat:** Damit wird die Kommunikation zwischen den Clients und dem Sioux-Server verschlüsselt. So soll das Abhören vertraulicher Informationen verhindert und sichergestellt werden, dass der Client mit dem gewünschten Server kommuniziert.
- 3) **Sioux-Schlüssel:** Bei der Bereitstellung jeder Leistungskontrolle auf dem *ExamServer* wird vom *AuthorClient* ein digitales Schlüsselpaar generiert. Der öffentliche Schlüssel wird beim Start der Leistungskontrolle an den Client geschickt. Mit diesem Schlüssel werden dann die Daten der Leistungskontrolle alle Resultate auf den *Backup*-Speichermedium der Studierenden verschlüsselt, so dass diese später nicht unberechtigterweise auf die Daten zugreifen können. Der private Schlüssel verbleibt bei der Prüfungsleitung und wird dazu verwendet, um die Daten auf dem *Backup*-Speichermedium zu entschlüsseln. Dies kann beispielsweise nach einem Abbruch und der anschliessenden Wiederaufnahme einer LK oder in einem Rekursverfahren der Fall sein.
- 4) **Student-Zertifikat:** Für alle Studierenden wird bei Prüfungsbeginn ein vom *ExamClient* selbst signiertes und auf den Namen der angemeldeten Person ausgestellt Zertifikat erstellt. Dieses wird verwendet, um alle Antworten zu signieren, wodurch deren Integrität jederzeit überprüft und deren Urheber festgestellt werden kann.

Der Einsatz von digitalen Schlüsseln und Zertifikaten alleine garantiert noch kein sicheres, belegbares Prüfungsergebnis. Entscheidend ist, wie diese Zertifikate in den Prüfungsprozess integriert werden. Dies ist Inhalt des folgenden Abschnitts.

6.3 Sicherheitskomponenten von Sioux

Mit Sioux haben wir die Praxistauglichkeit (*proof-of-concept*) der Sioux Designprinzipien untersucht. Diese dienen dabei als Grundlage für die Implementierung von sechs sicherheitskritischen Komponenten, welche die zentralen Elemente des CBA-Prozesses bilden (Abbildung 6.46).

Diese Sicherheitskomponenten beinhalten jene Sequenzen, die für die Sicherheit der Durchführungsphase einer CBA-Applikation zentral sind (vgl. "5.4.1 Sicherheitsphasen im CBA-Prozess").

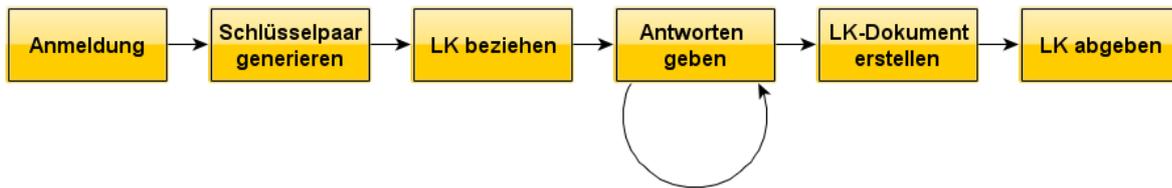


Abbildung 6.46: Sechs sicherheitskritische Vorgänge in der Durchführungsphase von CBA. Eine ausführlichere Beschreibung der einzelnen Vorgänge findet sich in den nachfolgenden Abschnitten.

Nachfolgend werden die sechs Sicherheitskomponenten anhand von Sequenzdiagrammen aus technischer Sicht erläutert. Die organisatorischen Massnahmen der Durchführungsphase von CBA werden später im Abschnitt "6.4.1 Organisation der Prüfungsdurchführung" beschrieben.

Sequenzdiagramme

Es wird auf das Abdrucken von Quellcodes verzichtet und stattdessen die sicherheitsrelevanten Methoden mittels Sequenzdiagrammen beschrieben (Herold, Lurz, & Wohlrab, 2010, Kapitel 12). Diese sind nicht spezifisch für Java und können als Idee für die Implementierung neuer oder die Optimierung bestehender CBA-Systeme dienen. Bei Interesse kann der Sioux-Quellcode beim Autor dieser Arbeit bezogen werden. Weitere Informationen hierzu finden sich auf der Projektwebseite von Sioux¹⁸.

Es gilt zu beachten, dass die Methodennamen und Übergabeparameter so gewählt sind, dass sie das Sioux-Funktionsprinzip möglichst intuitiv verständlich aufzeigen. Sie müssen aber nicht zwingend mit den in Sioux implementierten Methoden übereinstimmen. Die für die Sicherheit entscheidenden Methoden sind unterstrichen dargestellt und werden im Text detailliert erläutert.

Absichern der Server-Schnittstellen

Die Sioux-Clients beziehen die Prüfung und speichern die Antworten über die Schnittstelle des *ExamServers*. Diese ist permanent im Computernetzwerk erreichbar und muss gegenüber unberechtigten Zugriffen geschützt werden. Dazu stellt das *Enterprise Java Beans*-System (EJB) ein Rechteverwaltung zur Verfügung, mit dem nur Authentifizierten einen Zugriff auf die Schnittstelle gewährt werden kann. Zudem ermöglicht EJB mittels *Role Annotations* das Einschränken von Methodenaufrufen auf selbst defi-

¹⁸ http://www.cta.inf.ethz.ch/online_assessment (abgerufen am: 23.05.2014)

nierte Benutzergruppen. So können beispielsweise das Herunterladen von Prüfungsergebnissen durch die Benutzergruppen "Studierende" und "Hilfsassistierende" verhindert werden. Weitere Informationen zum Rechteverwaltung mit EJB finden sich in Heinrich (Heinrich, 2008).

6.3.1 Anmelden am System

Bei der Anmeldung am Sioux Prüfungssystem unterstützt Sioux zwei Arten von Benutzerdaten. Einerseits können sich die Studierenden mittels ihres ldap-basierten n.ethz-logins anmelden oder sie erhalten von der Prüfungsleitung ein spezielles Login, das durch Sioux selber authentisiert wird. Letzteres hat den Vorteil, dass auch Bildungsinstitutionen Sioux einsetzen können, welche über keinen ldap- respektive Single-Sign-On-Dienst verfügen oder bei der Durchführung von CBA komplett auf einen Internetzugriff verzichten möchten. In diesem Fall könnte der Sioux-Server direkt mit den Clients im Prüfungsraum vernetzt sein, ohne dass eine Verbindung mit weiteren Ressourcen nötig wäre.

Abbildung 6.47 zeigt das Sequenzdiagramm des Anmeldevorgangs bei Sioux. Bei diesem Vorgang fällt auf, dass das EJB System den Aufruf der Methode erst nach einer erfolgreichen Authentifizierung erlaubt. Nachfolgend werden die einzelnen Schritte bei der Anmeldung am Sioux-System beschrieben.

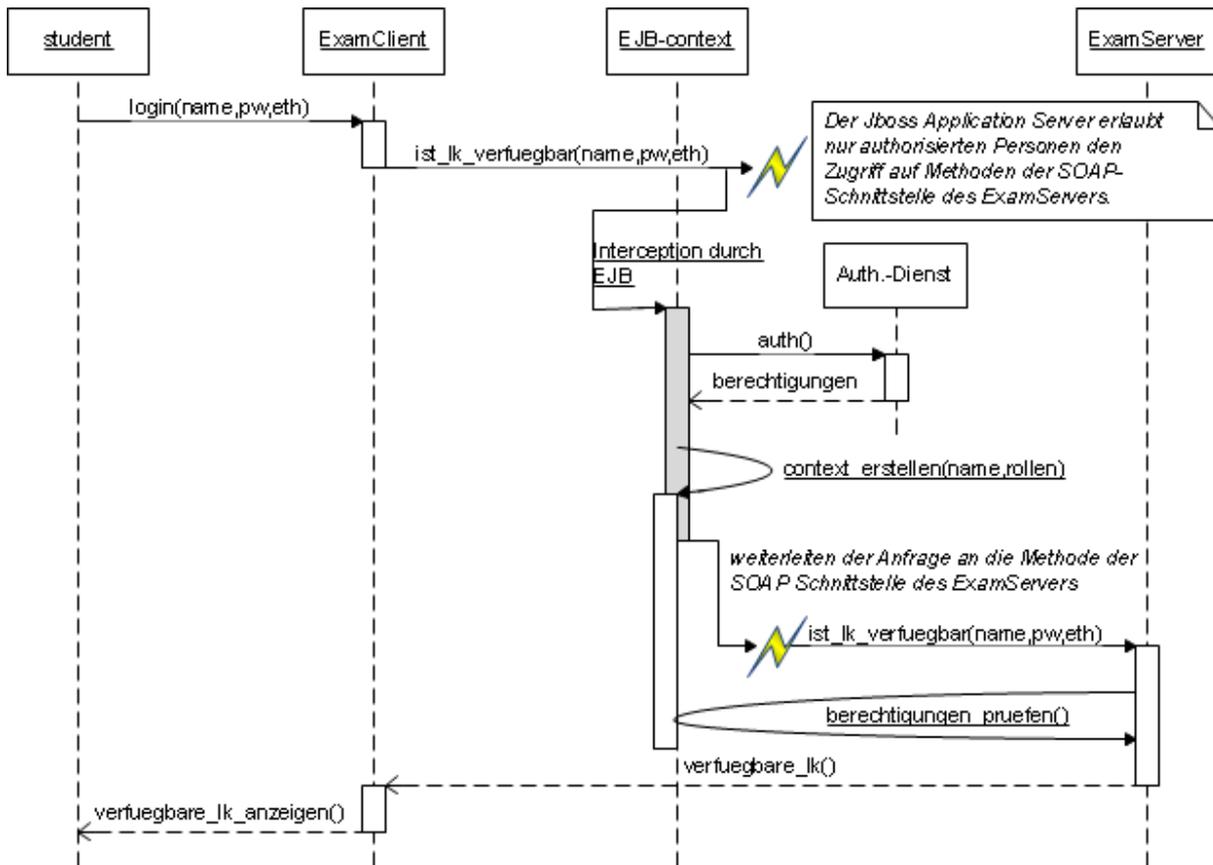


Abbildung 6.47: Anmeldevorgang bei Sioux. Der JBoss *Application Server* stellt sicher, dass nur authentifizierte Benutzer auf die Schnittstelle des ExamServers zugreifen können.

Der *ExamClient* meldet sich unter Verwendung der eingegebenen Logininformationen am ExamServer an. Diese Anfrage wird vom EJB-Kontext abgefangen und gegen den Authentifizierungs-Dienst authentisiert. Dieser ist modular implementiert und unterstützt wie bereits in der Einleitung erwähnt zwei verschiedene Authentifizierungsarten: die Authentifizierung mittels ldap-Server der ETH Zürich oder durch Sioux selber. Der Authentifizierungs-Dienst überprüft die Logininformationen und meldet das Resultat der Überprüfung an den jboss zurück. Der jboss-Kontext wird anschliessend durch benutzerspezifische Eigenschaften ergänzt, so dass anschliessend der ExamServer jederzeit anhand des jboss-Kontexts die Rolle des anfragenden Clients überprüfen kann.

Dies stellt ein wichtiger Sicherheitsgewinn dar, da bei jedem Methodenaufwurf mittels *Java-Annotations* überprüft werden kann, ob der Benutzer die Methode überhaupt aufrufen darf. Mittels des Java-Kontexts können aber auch noch weitere User-spezifische Überprüfungen gemacht werden. So kann z.B. sichergestellt werden, dass nur der Autor eine Prüfungsfrage öffnen und neu speichern darf.

Nach der erfolgreichen Anmeldung am System werden die für diesen Studierenden verfügbaren Leistungskontrollen aufgeführt. Zusätzliche Sicherheitskomponenten zur technischen Absicherung des Anmeldevorgangs, wie zum Beispiel der Einsatz von *Fingerprint*- oder *Iris-Scannern*, stehen nicht zur Verfügung. Stattdessen führt die Raum-

aufsicht nach dem Prüfungsstart visuelle Kontrollen durch, um falsche respektive vertauschte Identitäten erkennen zu können.

6.3.2 Schlüsselpaar generieren

Abbildung 6.48 zeigt das Sequenzdiagramm des Anmeldevorgangs am *ExamClient*. Dazu gehört auch die Generierung eines digitalen Schlüsselpaares mit dessen privatem Signaturschlüssel der *ExamClient* später alle Antworten der angemeldeten Person signiert. Der öffentliche Signaturprüf Schlüssel wird vom *ExamClient* auf den Namen der angemeldeten Person ausgestellt und kann somit auch als selbst signiertes ("self-signed") Zertifikat bezeichnet werden. Damit kann später die Identität der signierten Antworten zweifelsfrei nachgeprüft werden.

Bei der Generierung des Schlüsselpaares kommt die *Java Cryptography Architecture* (JCA) zum Einsatz. Diese umfangreiche Programmierschnittstelle bietet eine Vielzahl verschiedener kryptographischer Funktionen wie digitale Signaturen, Verschlüsselung und Zufallszahlengeneratoren an.

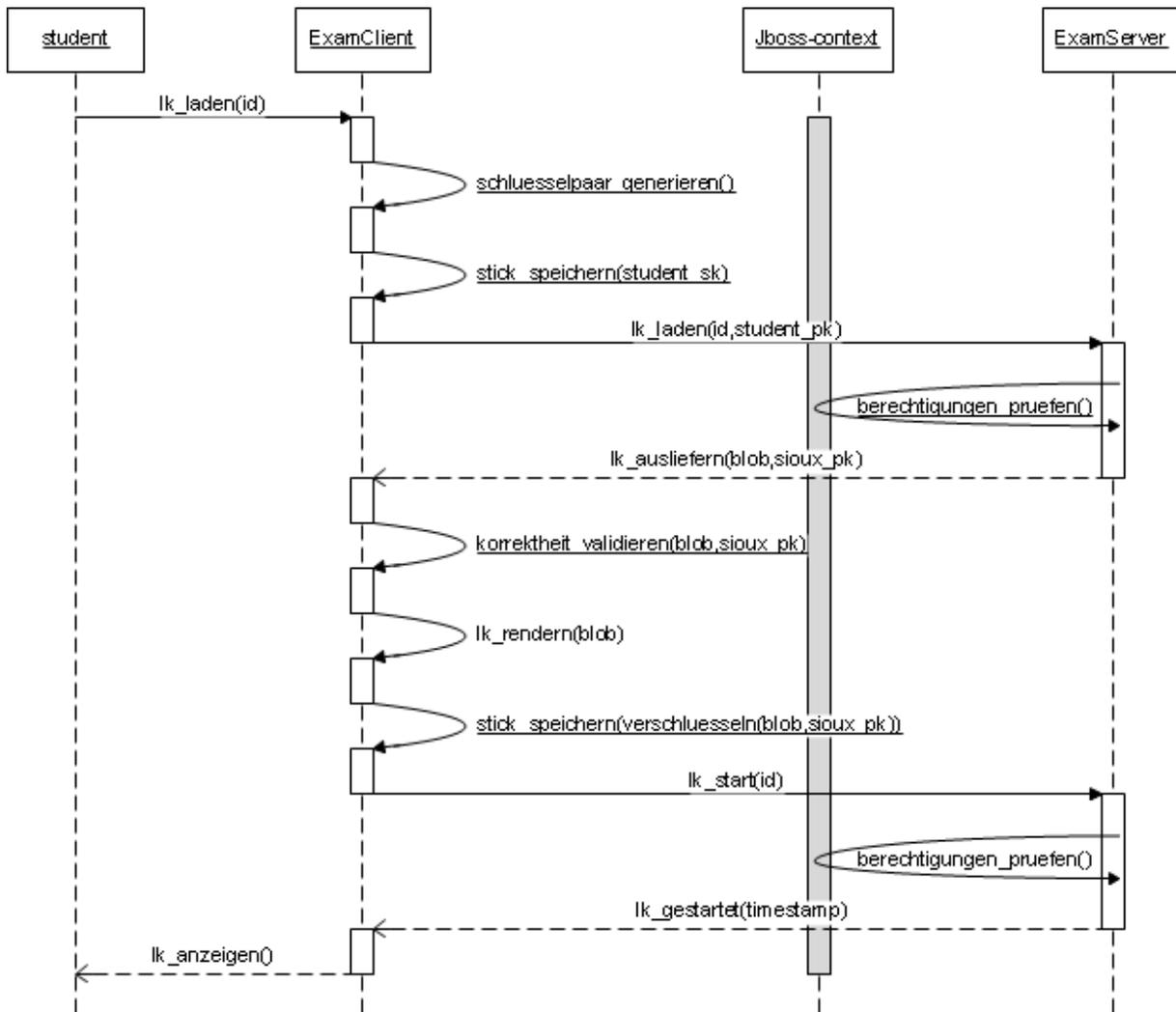


Abbildung 6.48: Schlüsselpaar generieren. Das Schlüsselpaar wird verwendet, um die einzelnen Resultate und das Prüfungsdokument digital zu signieren.

Nach der Generierung des Schlüsselpaars wird der private Signaturschlüssel auf das *Backup*-Speichermedium (z.B. USB-Stick) geschrieben und der öffentliche Signaturprüfchlüssel (d.h. das Zertifikat) wird auf dem ExamServer gespeichert. Das Schreiben auf das *Backup*-Speichermedium ist nötig, um im Zweifelsfalle nachprüfen zu können, ob die signierten Resultate auf dem Server tatsächlich mittels privaten Signaturschlüssels des betroffenen Studierenden erstellt wurden.

Da wir aus finanziellen Gründen auf den Einsatz von digitalen Signatureinheiten verzichten müssen, kann der private Signaturschlüssel auf dem Client nicht hardwaremässig vor unberechtigten Zugriffen geschützt werden. Dies stellt erhöhte Anforderungen an den *ExamClient*. Da dieser zur Signierung der Resultate direkt Zugriff auf diesen Schlüssel benötigt und gleichzeitig sicherstellen muss, dass er diesen Schlüssel vertrauensvoll behandelt und unberechtigten Personen nicht zugänglich macht. Auf weitere Schutzmechanismen, welche z.B. das Auslesen des Schlüssels aus dem Arbeitsspeicher oder vom *Backup*-Speichermedium verhindern soll, wurde aus folgenden Gründen verzichtet:

- 1) Wenn der *ExamClient* einen Teil der Vertrauenszone darstellt, gibt es keinen Grund zur Annahme, dass der private Signaturschlüssel der Studierenden von irgend jemandem wegkopiert werden könnte. Der Signaturschlüssel bleibt nur während der Durchführung der Leistungskontrolle auf dem Client-Computer gespeichert. Durch das Beenden des **ExamClient** wird er aus dem Arbeitsspeicher gelöscht und durch das Entfernen des *Backup*-Speichermediums endgültig vom Client-Computer entfernt.
- 2) Gäbe es Gründe zur Annahme, dass Punkt 1 nicht erfüllt sein könnte, so wäre die sichere Durchführung von CBA ohnehin nicht gewährleistet. Auf einem kompromittierten System können nicht nur die privaten Signaturschlüssel zweckentfremdet, sondern auch Antworten verändert oder gelöscht werden. Dies trifft sowohl bei Browser-basierten als auch bei Client-Server Applikationen zu, wenngleich bei Letzteren die Wahrscheinlichkeit kleiner ist, da die Antworten innerhalb des *ExamClients* mit Laufnummern versehen und signiert werden. Da die Daten auf dem *Backup*-Speichermedium zusätzlich noch mittels öffentlichen Sioux-Schlüssels chiffriert werden, hat ein Angreifer keinen Zugriff auf diese Resultate. Dazu bräuchte er den privaten Sioux-Schlüssel, welcher aber nur der Prüfungsleitung vorliegt. Damit er neue Resultate auf dem *Backup*-Speichermedium erstellen kann, benötigt er den öffentlichen Sioux-Schlüssel. Dieser wird aber nicht auf dem *Backup*-Speichermedium gespeichert und muss somit aus dem Arbeitsspeicher oder aus den Kommunikationsdaten mit dem Server herausgelesen werden. Da diese Verbindung verschlüsselt ist, wäre dies nur mit beträchtlichem Aufwand möglich.

Alternativ zum selbst signierten Zertifikat, das vom *ExamClient* ausgestellt wird, wäre auch der Einsatz eines von einer externen Stelle ausgestellten Zertifikats denkbar. Dabei muss es sich nicht zwingend um ein offizielles Zertifikat handeln, denn es reicht schon der Nachweis, welche Zertifikate ausgestellt wurden. So kann im Zweifelsfall überprüft werden, ob für einen Studierenden mehrere Zertifikate ausgestellt wurden.

Wie wir in einem Praxistest 2008 überprüft haben, würde sich dieses Verfahren grundsätzlich im Praxisalltag bewähren. Wir haben im Test den *Short Lived Credential Service* (SLCS) von Switch verwendet, um Schlüsselpaare zu generieren (SWITCH, 2006). Mit der von uns implementierten Benutzerschnittstelle konnten sich die Studierenden ein Zertifikat generieren lassen und sich damit am Sioux Prüfungssystem anmelden. Wir haben darauf verzichtet, dass die Studierenden zu Hause das Schlüsselpaar erstellen und mit dem auf einem mobilen Datenträger gespeicherten Schlüsselpaar zum Prüfungstermin erscheinen. Dies hatte zwei Gründe: erstens konnten wir nicht davon ausgehen, dass alle Studierenden mit einem gültigen SCLS-Zertifikat erscheinen würden. Zweitens hätte nicht verhindert werden können, dass die Studierenden die Zertifikate vorgängig austauschten oder weitergaben. Im Praxistest wurde deshalb entschieden, die Zertifikate direkt beim Eingang zum Prüfungsraum zu generieren. Dieses Vorgehen funktionierte, allerdings lässt es sich nicht gut skalieren, da das Erkennen des *Backup*-

Speichermediums und Kopieren des Zertifikats auf den Stick pro Student rund 20 Sekunden dauerte. Wir entschieden uns deshalb, die Schlüsselerstellung in den *ExamClient* zu integrieren. Zudem wurde seither auf den Einsatz der SCLS-Zertifikate verzichtet und selbst signierten Zertifikate verwendet. Diese haben den Vorteil, dass sie der *ExamClient* selber ausstellen kann und dadurch auf das Risiko eines Ausfalls des SCLS-Dienstes verzichtet werden kann.

Weitere Überlegungen zum Sicherheitsaspekt von selbst signierten Zertifikaten und dem Fehler einer elektronischen Signatureinheit finden sich in Abschnitt 5.4.3 "Verzicht auf eine qualifizierte elektronische Signatur bei Prüfungsergebnissen".

6.3.3 Leistungskontrolle starten

Nach der Anmeldung am System und der Erstellung des digitalen Schlüsselpaares kann die Prüfung vom ExamServer bezogen werden. Die Leistungskontrolle ist in einem proprietären XML-Format auf dem ExamServer gespeichert und wird zusammen mit dem öffentlichen Sioux-Signaturschlüssel auf die *ExamClients* übertragen. Der Sioux-Signaturschlüssel wird beim Bereitstellen der Leistungskontrolle automatisch durch den *AuthorClient* erstellt. Er dient dazu, die Resultate und das Prüfungsdokument der Studierenden auf ihrem *Backup*-Speichermedium verschlüsselt abzuspeichern. So kann verhindert werden, dass den Studierenden der Inhalt der Leistungskontrolle nachträglich zugänglich wird. Zum Öffnen dieser Daten ist der private Signaturprüfchlüssel nötig, der im Besitz der Prüfungsleitung ist.

Abbildung 6.48 zeigt schematisch die sicherheitsrelevanten Schritte beim Starten einer Leistungskontrolle. Nach dem Ausliefern der Leistungskontrolle wird der Inhalt des XML-Dokuments überprüft. In Sioux geschieht dies implizit, indem das XML in die interne Datenstruktur des *ExamClient* überführt wird. Es wäre aber auch denkbar, die Prüfung mit dem öffentlichen Schlüssel des Studierenden auf dem Server zu verschlüsseln und dann im *ExamClient* zu entschlüsseln. So könnte die Unversehrtheit des Prüfungsdokuments überprüft und sichergestellt werden, dass nur der betroffene Student die Leistungskontrolle öffnen könnte. Allerdings müsste noch abklärt werden, wie gross die Belastung des Servers aufgrund der individuellen Verschlüsselungen der Leistungskontrolle wäre.

Nach der erfolgreichen Überprüfung der XML-Struktur der Prüfungsfrage wird die Leistungskontrolle mit dem öffentlichen Signaturschlüssel von Sioux verschlüsselt und auf dem *Backup*-Speichermedium des Studenten gespeichert. Im Falle eines Programmabsturzes des *ExamClients* kann die Prüfung vom *Backup*-Speichermedium geöffnet und auch ohne Serververbindung fortgesetzt werden. Dazu ist der private Sioux-Schlüssel nötig, in deren Besitz nur die Prüfungsleitung ist.

Nach dem erfolgreichen Rendern der Leistungskontrollen wird dem ExamServer mitgeteilt, dass die Prüfung nun gestartet wird. Danach beginnt die individuelle Prüfungszeit

zu laufen. Mit diesem zweistufigen Vorgehen, Bezug und Start der Leistungskontrolle, kann verhindert werden, dass die Übertragung des Prüfungsdokuments bereits Teil der Prüfungszeit ist. So beginnt bei Sioux erst dann die Prüfungszeit zu laufen, nachdem die Prüfung übertragen, überprüft und gerendert wurde.

6.3.4 Antworten eingeben

Beim Eingeben von Antworten ist es fundamental, dass nur jene Antworten für die Studierenden sichtbar werden, welche tatsächlich gespeichert wurden und somit zur Korrektur verwendet werden können. Es wäre naheliegend, die Antwort erst dann anzuzeigen, wenn sie auf dem Server gespeichert wurde. Dies hätte aber zwei grosse Nachteile:

1. Die Studierenden müssen die Bestätigung des Servers abwarten, was ihre Weiterarbeit verzögert.
2. Ein Unterbruch der Serververbindung führt automatisch dazu, dass die Bearbeitung der Prüfung unterbrochen wird.

Um diese beiden Probleme verhindern zu können, bildet im Sioux *ExamClient* eine Warteschlange ("ExamClient-queue"), das Rückgrat der Serverkommunikation. Die Warteschlange hat den Vorteil, dass sie sich im Hintergrund um die Speicherung der Antworten kümmert, während die Studierenden die Prüfung weiter bearbeiten können. Wie dies implementiert wurde, wird nachfolgend erläutert.

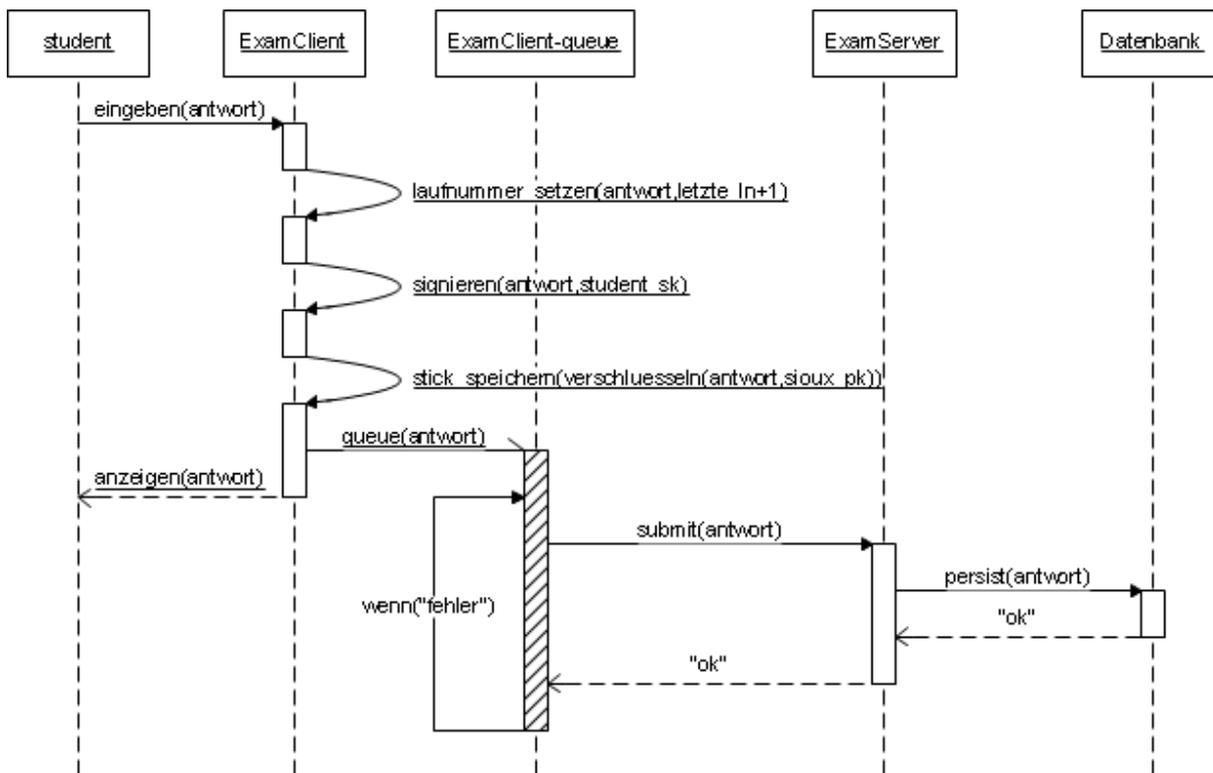


Abbildung 6.49: Eingeben von Antworten. Dieses Sequenzdiagramm illustriert die Schritte, wie Sioux die Antworten sicher speichert, ohne die Bearbeitungsgeschwindigkeit zu reduzieren. Dabei kommt der Warteschleife ("ExamClient-queue") eine entscheidende Bedeutung zu.

Wie aus Abbildung 6.49 ersichtlich ist, wird nach dem Eintippen oder Anklicken einer Antwort eine neue Laufnummer bestimmt und dieser Antwort zusammen mit einem Zeitstempel angehängt. Anschliessend wird diese Antwort mit dem privaten Student-Signaturschlüssel signiert und unter Verwendung des Sioux-Schlüssels verschlüsselt auf dem *Backup*-Speichermedium gespeichert. Die signierte, aber unverschlüsselte Antwort wird dann der Warteschleife übergeben. Diese Schritte werden innerhalb eines *blocking-calls* abgearbeitet, das heisst, die Bearbeitung der Prüfung wird erst dann wieder frei gegeben, wenn diese drei Schritte abgearbeitet wurden. In der Praxis dauert dies einige Millisekunden, wobei die Schreibgeschwindigkeit des *Backup*-Speichermediums den Flaschenhals darstellt.

Während die Studierenden die Prüfung weiter bearbeiten können, kümmert sich die Warteschleife darum, alle Antworten auf den ExamServer zu speichern. In der Regel passiert das innerhalb von Sekundenbruchteilen. Bei einer Überlastung oder Ausfall des Servers versucht die Warteschleife alle zehn Sekunden die Antworten erneut zu speichern. Diese Wartezeit hat sich in der Praxis bewährt: sie ist so kurz, dass keine Wartezeiten beim Beenden der Prüfung entstehen und so lang, dass der ExamServer nicht durch *denial of service* (DOS)-Anfragen durch die *ExamClients* überlastet wird.

Neben der Erhöhung der Arbeitsgeschwindigkeit stellt die Warteschleife sicher, dass der *ExamClient* erst dann beendet werden kann, wenn alle Antworten auf dem Server gespeichert wurden.

Falls der Client unbeabsichtigt beendet wird, kann die Leistungskontrolle mit Hilfe der Daten auf den *Backup*-Speichermedien komplett ohne Server wieder gestartet werden, ohne dass Antworten verloren gehen. Der erwähnte *blocking-call* stellt sicher, dass alle angezeigten Antworten auch auf das *Backup*-Speichermedium geschrieben werden. Es ist allerdings möglich, dass zum Zeitpunkt des Absturzes die Warteschleife nicht leer war und einige Antworten somit nicht auf dem Server gespeichert wurden. Um diesen Datenverlust zu beheben, werden nach dem Fortsetzen der Prüfung alle Antworten auf dem *Backup*-Speichermedium erneut an den Server übermittelt.

Versagt ein *Backup*-Speichermedium während der Bearbeitung einer Leistungskontrolle, so wird dies vom *ExamClient* erkannt. Der Proband wird dann aufgefordert, einen neuen Speicherort für das *Backup*-Speichermedium festzulegen. Anschliessend werden vom *ExamClient* alle prüfungsrelevanten Informationen aus dem Arbeitsspeicher auf das neue *Backup*-Speichermedium geschrieben.

6.3.5 Prüfungsdokument erstellen

Nach Ablauf der Prüfungszeit oder nach Wunsch des Probanden wird die Prüfung beendet. Dabei wird die Benutzeroberfläche für eine weitere Bearbeitung gesperrt und das Prüfungsdokument erstellt. Dieses dient später als Grundlage für die Korrektur der Resultate und Benotung der Leistung des Probanden. Abbildung 6.50 illustriert die Schritte vom Zusammenstellen der Prüfung bis zum Beenden der Prüfungsapplikation.

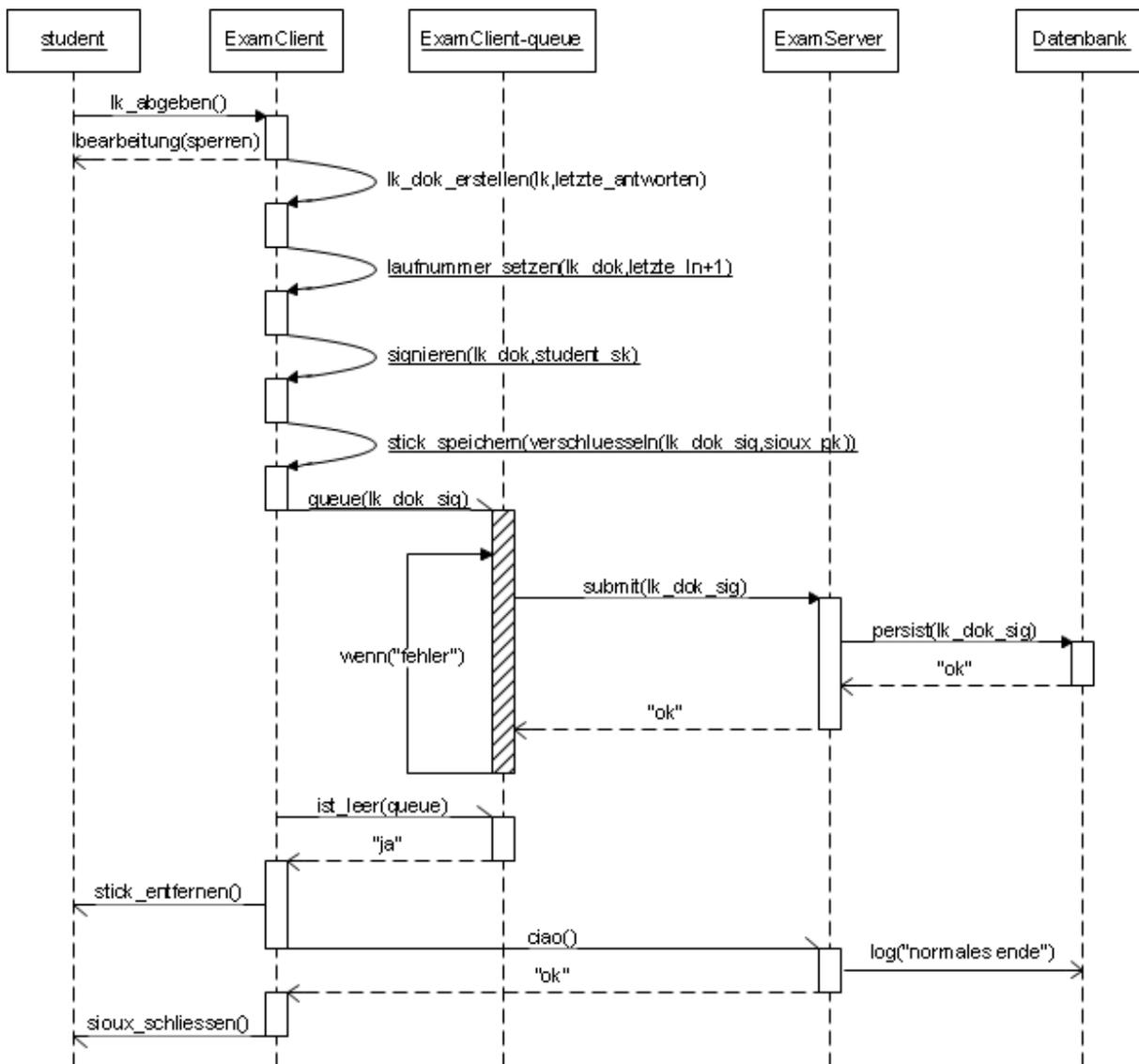


Abbildung 6.50: Erstellen des Prüfungsdokuments und Abgeben der Leistungskontrolle. Das Prüfungsdokument wird mit Hilfe der Warteschleife an den Server übermittelt. Die Warteschleife stellt sicher, dass der *ExamClient* erst geschlossen werden kann, wenn die Warteschleife leer ist.

Das Prüfungsdokument wird aus den Inhalten der Datenstruktur des *ExamClients* zusammengestellt. Als Basis für das Prüfungsdokument dient das vom *ExamServer* übertragene XML-Prüfungsdokument. Ergänzt wird das Dokument durch die für jedes Item zuletzt abgegebene Antwort. Ein Beispiel eines solchen XML-Prüfungsdokuments ist im Anhang A.4 dieser Arbeit abgedruckt.

Das XML-Prüfungsdokument wird mit der letzten, höchsten Laufnummer versehen und mit dem privaten Signaturschlüssel des Probanden signiert. Anschliessend wird es der Warteschleife übergeben und, unter Verwendung des öffentlichen Sioux-Signaturschlüssels, verschlüsselt auf dem *Backup*-Speichermedium gespeichert. Die Verschlüsselung ist nötig, damit die auf dem *Backup*-Speichermedium gespeicherten Daten nachträglich nicht unerlaubterweise durch die Probanden eingesehen werden können. Die verschlüsselte Version des Prüfungsdokuments dient primär als *Backup*

und wird dann benötigt, wenn der Proband sein Prüfungsergebnis nicht anerkennt. Dazu kann die verschlüsselte Version auf dem *Backup*-Speichermedium des Probanden mit dem privaten Sioux-Signaturprüf Schlüssel entschlüsselt und eingesehen werden.

Das Sioux-Prüfungsdokument beinhaltet also neben allen aktuellen Antworten auch die komplette Prüfung inklusive aller Aufgabenstellungen und Medien. Inhaltlich ist es somit identisch mit dem Prüfungsdokument einer schriftlichen Prüfung. Dank der Signatur ist es zudem vor Veränderungen geschützt und mit Hilfe des Schlüsselpaares eindeutig einem Probanden zuzuordnen.

6.3.6 Leistungskontrolle abgeben

Das Abgeben der Leistungskontrolle beinhaltet die Speicherung des Prüfungsdokuments auf dem ExamServer. Dies passiert direkt im Anschluss an das Erstellen des Prüfungsdokuments auf dem *ExamClient*.

Der zentrale, sicherheitskritische Punkt dieses Arbeitsschritts ist die Kontrolle, ob das Prüfungsdokument auf dem ExamServer gespeichert wurde. Sioux bietet zwei Kontrollmöglichkeiten an. Die erste, implizite Kontrolle führt der *ExamClient* selber durch. Er stellt sicher, dass das Programm erst geschlossen werden kann, wenn das Dokument erfolgreich gespeichert ist. Wird die Speicherung aus verschiedensten Gründen verzögert, zeigt der *ExamClient* ein entsprechendes Hinweisfenster. Die zweite, explizite Kontrolle führt die Raumaufsicht durch. Im *TacticalClient* werden alle Studierenden grün markiert, welche die Leistungskontrolle abgegeben haben. Rote ("Prüfung noch nicht bezogen") oder gelbe ("Prüfung noch in Bearbeitung") weisen auf ein fehlendes Dokument hin. Bei summativen, *high-stake* Leistungskontrollen dürfen die Studierenden erst dann den Raum verlassen, wenn die Prüfungsleitung das Vorhandensein aller Prüfungsdokumente bestätigt.

Bei formativen Leistungskontrollen ist nach der Abgabe der Leistungskontrolle für die Studierenden der Prüf-Prozess noch nicht abgeschlossen. Sioux korrigiert nach einer erfolgreichen Abgabe automatisch die Resultate und zeigt die Korrekturen direkt auf dem Bildschirm an. Die Korrektur wird vom *ExamClient* mittels mitgelieferten Musterlösungen durchgeführt. Bei summativen Leistungskontrollen werden die Musterlösungen aus Sicherheitsgründen nicht an den *ExamClient* übertragen. Zusätzlich zu den Korrekturen zeigt der *ExamClient* auch ein individuelles Feedback an und gibt Hinweise, warum eine gewisse Aufgabe falsch oder korrekt war.

Bei summativen Leistungskontrollen wird der *ExamClient* automatisch geschlossen, sobald alle Antworten des Studierenden und das Prüfungsdokument auf dem ExamServer gespeichert wurden. Bei formativen Leistungskontrollen müssen die Studierenden den *ExamClient* manuell schliessen, nachdem sie die Korrekturen studiert haben.

6.4 Sioux im Praxiseinsatz

Mit der Sioux-Suite führen wir seit 2008 regelmässig CBA mit bis zu 300 Studierenden durch. Dabei kommen sieben öffentliche Computerräume an der ETH Zürich zum Einsatz. Die dort fest installierten Computer bieten Prüfungsplätze für rund 150 Studierende. Diese ermöglichen die Prüfung von 300 Studierenden in zwei Serien. Neben der technisch einwandfreien Durchführung der Leistungskontrolle mussten für die Studierenden in allen Prüfungsräumen vergleichbare Prüfungsbedingungen geschaffen werden. Dazu verwenden wir unter anderem einheitliche, vorher aufgezeichnete Raumanweisungen und Funkgeräte zur Kommunikation mit den Prüfungsräumen. Zudem wird für eine klare Instruktion und Aufgabenverteilung der Raumaufsichtspersonen gesorgt.

6.4.1 Organisation der Prüfungsdurchführung

Im Abschnitt "6.3 Sicherheitskomponenten" ist beschrieben, mit welchen technischen Mitteln Sioux die Sicherheit von CBA erhöht. Ein hoher Sicherheitsstandard bei CBA kann aber nicht bloss mittels technischer Lösungen erreicht werden. Dazu sind organisatorische Ablaufstrukturen nötig, welche die verfügbaren didaktischen, operationalen und finanziellen Rahmenbedingungen berücksichtigen.

In einem technischen Report haben wir alle Vorgänge rund um CBA mit Sioux am Beispiel der Basisprüfung der Lehrveranstaltung "Informatik für Biologie und Pharmazeutische Wissenschaften" dokumentiert (Dahinden & Hinterberger, 2010). In diesem Report sind neben den organisatorischen Ablaufbeschreibungen auch detaillierte Anleitungen und Checklisten enthalten, welche die Durchführung von CBA und die Kommunikation mit den Studierenden und Aufsichtspersonen beschreiben.

6.4.2 Zeitaufwand von Leistungskontrolle mit Sioux

In Tabelle 6.17 ist der Zeitaufwand zur Durchführung basierend auf unseren Erfahrungswerten aufgelistet. Sie beziehen sich einerseits auf die Zeit vor 2005, als die Prüfungen noch auf Papier durchgeführt wurden und die Zeit danach, als uns mit Sioux ein CBA-System zur Verfügung stand, mit dem wir auch benotete *high-stake* Prüfungen am Computer durchführen konnten.

Im Sinne der besseren Vergleichbarkeit des Zeitaufwands bei CBA und PPT haben wir folgende Annahmen getroffen:

- 300 Studierende
- Prüfungsdauer: 1 Stunde
- 7 Prüfungsräume bei CBA, 2 Räume bei PPT
- 2 Prüfungsserien bei CBA, 1 Prüfungsserie bei PPT
- Fall-basierte Prüfungsfragen

	CBA			PPT		
	Involv. Personen	Zeit pro Person	Totalzeit	Involv. Personen	Zeit pro Person	Totalzeit
Erstellen der Prüfungsfragen ^{A)}	6	15.00	95.0	6	15.00	95.0
Zusammenstellen der Leistungskontrolle	1	0.50	0.5	1	1.00	1.0
Korrekturlesen der Leistungskontrolle	6	2.00	12.0	6	2.00	12.0
Bereitstellen der Leistungskontrolle ^{B)}	1	0.50	0.5	1	2.00	2.0
Einteilung der Studierenden, Kommunikation d. Termins ^{C)}	1	1.00	1.0	1	0.10	0.1
Instruktion des Aufsichtspersonals	14	0.50	7.0	8	0.10	0.8
Bereitstellen d. Hilfsmaterialien (Listen, Funkgeräte etc.) ^{D)}	2	4.00	8.0	1	0.50	0.5
Testen der Infrastruktur ^{E)}	1	1.00	1.0	0	0.00	0.0
Vorbereiten der Infrastruktur ^{F)}	14	0.75	10.5	8	0.50	4.0
Durchführung der Leistungskontrolle	14	2.00	28.0	8	1.00	8.0
Korrektur der Leistungskontrolle ^{G)}	2	2.50	5.0	8	8.00	64.0
Analyse der Leistungskontrolle ^{H)}	1	0.50	0.5	-	-	-
Notenübermittlung an Studiensekretariate ^{I)}	1	0.50	0.5	2	0.50	1.0
Totalzeit			171.5			189.4

Tabelle 6.17: Geschätzter Zeitaufwand bei CBA und Papier-basierten Prüfungen (PPT). Die Zeiten sind in Stunden angegeben und basieren auf Erfahrungswerten aus unserem Unterricht an der ETH Zürich.

Bemerkungen zur Tabelle 6.17:

A) Durch die Möglichkeit der kollaborativen Fragenpoolbewirtschaftung und der für Fall-basierte Prüfungsfragen dedizierte Sioux Frageneditor (als Teil des *AuthorClients*) dürfte die für die Fragenerstellung benötigte Zeit bei CBA etwas tiefer ausfallen als bei reinen PPT Prüfungen. Bei der Bestimmung des Zeitaufwands zur Erstellung der Prüfungsfragen wurde angenommen, dass eine komplett neue Prüfung erstellt werden muss. Stehen Ankerfragen, das heisst Fragen, die aus statistischen Gründen wiederverwendet werden oder bewährte Fragen, zur Modifikation zur Verfügung, kann der Zeitaufwand für die Prüfungserstellung entsprechend der Anzahl wiederverwerteter Fragen reduziert werden.

B) Das Drucken der Prüfung wird an der ETH Zürich durch die Reprozentrale ausgeführt. Deshalb wird die Druckzeit nicht zum Zeitaufwand bei PPT eingerechnet. Somit beinhaltet dieser Arbeitsschritt für die Prüfungsleitung das Erstellen der Druckvorlage sowie das persönliche Abholen der gedruckten Prüfungen.

C) Die Einteilung der Studierenden ist bei CBA aufwändiger, da diese auf die insgesamt 14 Prüfungsräume aufgeteilt werden müssen und die Einteilung individuell kommuniziert werden muss. Bei PPT wird dieser Schritt von der Prüfungsplanstelle der ETH Zürich übernommen.

D) Bei CBA werden Anmeldedaten, Ersatz-*Backup*-Speichermedien (USB-Sticks), Lautsprecher für die Raumsansage und Funkgeräte benötigt.

E) Bei CBA muss die Softwareinstallation der Arbeitsstationen überprüft werden, da regelmäßig neue *Updates* eingespielt werden und geprüft werden muss, ob Sioux mit diesen *Updates* kompatibel ist.

F) Bei CBA startet die Raumaufsicht die Computer, meldet sich mit speziellen Prüfungslogins an und startet Sioux. Bei PPT müssen die Papierprüfungen auf die einzelnen Arbeitsplätze verteilt werden.

G) Bei CBA findet die Korrektur teilautomatisch statt, wobei alle falschen Antworten noch manuell überprüft werden. Bei PPT rechnen wir damit, dass jede Person pro Stunde sieben Prüfungen korrigieren und doppelt so viele Prüfungen gegenkorrigieren kann.

H) Die Analyse der Prüfungsdaten ist bei PPT nur mit einem erheblichen Zusatzaufwand möglich, da die Resultate nicht in digitaler Form vorliegen. Für eine Itemanalyse werden die Zwischenpunkte aller Aufgaben benötigt. Für eine detaillierte Prüfungsanalyse ist es zudem wichtig, dass auch der Inhalt der falschen Antworten erfasst wird.

I) Bei PPT müssen die Noten zuerst abgetippt und von einer zweiten Person nachkontrolliert werden. Bei CBA liegen die Noten in digitaler Form vor. Dort reicht eine stichprobenartige Nachkontrolle.

Wie in Tabelle 6.17 ersichtlich, nimmt die Erstellung einer neuen Prüfung im gesamten Prüfungszyklus am meisten Zeit in Anspruch. Dieser Zeitaufwand kann bei der wiederholten Prüfungsdurchführung reduziert werden, indem entweder Ankerfragen verwendet oder Varianten bestehender Fragen entwickelt werden.

Bei CBA sind die Arbeiten am Prüfungstag (Vorbereiten der Prüfungsräume und Durchführung der Prüfung) mit 38.5 Personenstunden fast gleich hoch wie alle restlichen Arbeiten zur Vor- und Nachbereitung der Prüfung (Total 37 Personenstunden, ohne Prüfungserstellung). Unsere Bestrebungen zur Reduzierung des Aufwands beziehen sich deshalb primär auf die Arbeiten am Prüfungstag. Konkrete Vorschläge diesbezüglich sind unter "Fazit" am Schluss dieses Kapitels abgedruckt.

Bei PPT nimmt die Korrektur der Prüfung mit Abstand am meisten Zeit in Anspruch. Die dafür benötigte Zeit ist rund einen Faktor zwei höher als alle anderen Arbeiten des Prüfungszyklus (unter Ausschluss der Prüfungserstellung). Diese Korrekturzeit lässt sich durch den Einsatz von *Multiple Choice* Prüfungsfragen und maschinell digitalisier- und korrigierbaren Antwortformularen reduzieren. Dieser Schritt sollte jedoch nur unter-

nommen werden, wenn parallel dazu die Validität und Reliabilität der Prüfungsergebnisse sorgfältig untersucht werden. Die Gefahr ist gross, dass mit mangelhaft formulierten, automatisch korrigierbaren Prüfungsfragen die Qualität der Leistungskontrolle gesenkt wird.

6.4.3 Organisation der Prüfungseinsicht

Die legalen Vorgaben der Prüfungseinsicht sind in der Weisung des Rektorats erläutert (ETH Zürich, 2013b). Sie schreibt vor, dass neben der Prüfung auch die Korrekturen und allfällige Musterlösungen den Studierenden vorgelegt werden müssen.

Bei Sioux basierten Leistungskontrollen findet die Prüfungseinsicht im Büro der Prüfungsleitung direkt am Computer statt. Dabei wird das digitale Prüfungsdokument des Probanden am Computer angezeigt und die Korrekturen sowie die Musterlösung auf Papier abgegeben.

Die Durchfallquote bei unseren Leistungskontrollen liegt erfahrungsgemäss bei ungefähr 10-15%. Dennoch melden sich oft weniger als 2% der Studierenden für eine Prüfungseinsicht an. Die Gründe für diese tiefe Zahl bleiben ungeklärt, aber wir beobachten, dass die automatische Korrektur viele Studierende davon abhält, an der Prüfungseinsicht über die Punktevergabe zu diskutieren. Vielmehr kommen Sie an die Prüfungseinsicht, um die Wahrscheinlichkeit eines erneuten Prüfungsmisserfolgs zu reduzieren.

6.4.4 Dokumentation des Prüfprozesses

Bisher wurde in dieser Arbeit die Dokumentation des Prüfungsprozesses ausschliesslich mit der Notwendigkeit begründet, dass im Falle eines Beschwerdeverfahrens den rechtserheblichen Sachverhalt korrekt und eindeutig belegt werden muss.

Aus technischer, didaktischer und lernpsychologischer Sicht bietet die lückenlose Dokumentation des Prüfprozesses aber auch weitere Vorteile:

- Feststellen der korrekten Funktionsweise der CBA-Applikation
- Erhöhung des subjektiven Gerechtigkeitsempfindens bei den Studierenden
- Beantwortung wissenschaftlicher Fragestellungen im Bereich "*Learning analytics*" (vgl. Kapitel 8)

Diese drei Punkte setzen nicht zwingend das Vorhandensein von digitalen Signaturen voraus. So können beispielsweise auch unsignierte Server-Logdaten verwendet werden, um mögliche Fehler in den Client-Applikationen zu entdecken oder detaillierte Analysen über das Antwortverhalten der Studierenden durchzuführen. Im nachfolgenden Kapitel 7 werden einige Möglichkeiten der Prüfungsanalyse aufgezeigt und wie umfangreiche CBA-Daten zur wissenschaftlichen Arbeit verwendet werden können.

6.4.5 Performancemessungen

Abhängig von der verfügbaren Infrastruktur birgt der Sioux Praxiseinsatz unterschiedliche Risiken. Die Risikoanalyse in Kapitel 4 listet neben organisatorischen (z.B. begrenzte Anzahl an Prüfungsarbeitsplätzen) auch technische Risiken auf. Dazu zählen zum Beispiel Performanceprobleme des Servers (T16) oder des Netzwerks (T15). Beide Risiken verfügen über ein vergleichsweise hohes Gefahrenpotential (vgl. Tabelle 4.10). In diesem Abschnitt werden deshalb die Resultate unserer Performancemessungen präsentiert und eine Schätzung abgegeben, wie viele Studierende an der ETH Zürich unter Verwendung der aktuellen technischen Infrastruktur maximal geprüft werden können.

Resultate der Messung

Zur Evaluation der Leistungsfähigkeit des Sioux Prüfungssystems haben wir mehrfach während der Leistungskontrollen die Leistungs-Kennzahlen des Sioux-Servers mitgeschrieben. Die Daten wurden mittels des Programms `procmon`¹⁹ alle 500 ms erfasst und zusammen mit dem Zeitstempel in eine Datei geschrieben.

Die Auswertung dieser Daten ergab, dass alle Prüfungen mit vergleichbaren Probandenzahlen ähnliche Lastmuster aufwiesen. Wir zeigen hier deshalb die Prüfung mit den bisher meisten Probanden (Abbildung 6.51).

¹⁹ Procmon: Ein von Peter Heinrich geschriebenes kleines Java-Programm, welches die Daten von `/proc/stat` mit-schreibt.

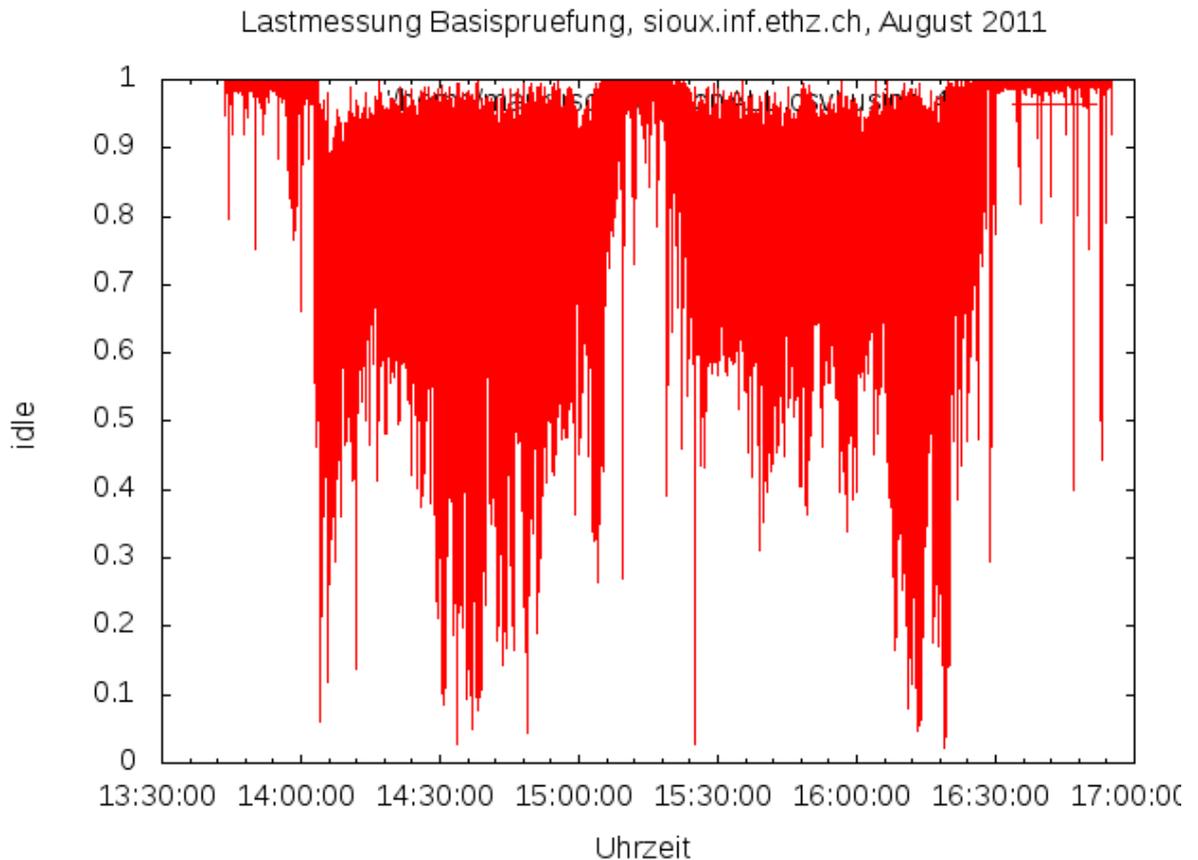


Abbildung 6.51: Lastmessungen während der beiden Prüfungssessionen. Das Diagramm zeigt die Maximalwerte der Leerlauf (*idle*)-Zeit des Sioux-Servers, wobei 1 für 100% Leerlauf und 0 für die komplette Auslastung steht. Die Daten wurden alle 500 Millisekunden erfasst und zur Erstellung des Diagramms verwendet. Die erste Serie wurde um 14:00 Uhr, die zweite um 15:15 Uhr gestartet. Lastspitzen sind jeweils zu Beginn jeder Prüfungssession erkennbar und lassen sich mit der Übertragung der kompletten Prüfungsdaten erklären. Auffallend ist auch eine höhere Last jeweils vor Ende jeder Serie, welche vermutlich auf das eilige Abschicken noch offener Antworten zurückzuführen ist. Datenbasis: Basisprüfung, "Informatik für Biologie und Pharm. Wissenschaften", Dozent: Dr. Böckenhauer, 19.08.2011, n=272.

Die Veränderungen der Serverlast kann anhand folgender Ereignisliste erklärt werden:

	Beschreibung	Zeit
1	Vorbereiten der Prüfungsclients (Starten SEB, Sioux)	13:40 – 14:00
2	Start der ersten Serie (Anmelden am Sioux-System und Beziehen der Prüfung)	14:00 – 14:05
3	Senden von Antworten	14:05 – 15:00
4	Abgeben der Prüfung nach Ablauf der Prüfungszeit und Beenden der ersten Prüfungsserie	15:00 – 15:05
5	Wechsel der beiden Serien	15:05 – 15:15
6	Neustart der Sioux-Client-Applikation	15:05 – 15:15
7	Start der zweiten Serie (Anmelden am Sioux-System und Beziehen der Prüfung)	15:15 – 15:20

8	Senden von Antworten	15:20 – 16:15
9	Abgeben der Prüfung nach Ablauf der Prüfungszeit und Beenden der zweiten Prüfungsserie	16:15 – 16:30
10	Erstellen des Datenbank-Backups	16:45

Schlussfolgerung der Lastmessung

Die Detailauswertung der Serverlast (*load*) hat ergeben, dass diese bei durchschnittlich 1.5 lag. Da der Prozessor über vier Kerne verfügt, wäre somit eine doppelt so hohe Studierendenzahl denkbar, ohne dass die Studierenden längere Wartezeiten in Kauf nehmen müssten. Die zur Zeit der Drucklegung dieser Arbeit verfügbare Anzahl an Prüfungsarbeitsplätzen an der ETH Zürich erlaubt diesbezüglich aber keine weiteren Tests.

6.4.6 Bottleneck des Systems

Die Leistungsfähigkeit eines CBA-Systems ist abhängig von der Netzwerkanbindung sowie der Ausstattung der Client- und Server-Geräte. An der ETH Zürich bietet das Computernetzwerk ausgezeichnete Voraussetzungen für CBA: Die Clients sind ETH-intern mit mindestens 1 Gigabit-Ethernet mit dem Sioux-Server verbunden. Wenn pro Prüfungsstart mit 150 Probanden geschätzt 300 Megabyte an Prüfungsdaten (Fragen und Medien) übertragen werden müssen, erlaubt diese Anbindung deren theoretische Übertragung in 2.4 Sekunden²⁰. Diese Wartezeit wird durch die Studierenden kaum wahrgenommen, zumal es praktisch ausgeschlossen ist, dass alle Studierenden die Prüfung zum gleichen Zeitpunkt vom Server beziehen. Bei der Durchführung von Sioux-basierten Prüfungen hatten wir diesbezüglich auch keinerlei Probleme oder Beschwerden von Studierenden.

Der Einsatz von Sioux ausserhalb der ETH Zürich hat aber auch gezeigt, dass die Kapazität des Computernetzwerks relativ rasch den Engpass darstellen kann. So liegt zum Beispiel beim Einsatz von Funknetzwerken (WLAN) die verfügbare Bandbreite weit unterhalb jener eines kabelgebundenen Netzwerks. Aber auch bei der interkontinentalen Übertragung der Prüfungsdaten, wie sie zum Beispiel beim Einsatz des Sioux *ExamClients* in Riyadh (Saudi Arabien), kann die Bandbreite und Stabilität der Netzwerkverbindung zwischen Client und Server stark schwanken. Dank der teilautonomen Funktionsweise des Sioux-Clients können Netzwerkunterbrüche oder lange Antwortzeiten überbrückt werden, so dass diese für die Studierenden nicht spürbar werden.

Die Client-Arbeitsstationen an der ETH Zürich sind für CBA-Prüfungen eigentlich überdimensioniert. Da der Client abgesehen von den kryptografischen Funktionen keine re-

²⁰ $300\text{Megabyte}/(1000\text{Megabit}/\text{Sekunde}/8\text{bit})=2.4\text{ Sekunden}$

levanten Berechnungen durchführen muss, wären auch *Thin-Clients* ohne eingebaute Festplatte und mit einem leistungsschwächeren Prozessor ausreichend.

Die Ausstattung des Sioux-Servers mit vier Prozessorkernen und acht Gigabyte Arbeitsspeicher reicht theoretisch, wie im vorangehenden Abschnitt "6.4.5 Performancemessungen" beschrieben, für die gleichzeitige Durchführung von 300 CBA-Prüfungen. Die begrenzte Raumkapazität an der ETH Zürich limitiert CBA allerdings auf 150 gleichzeitigen Prüfungen respektive 300 Studierenden in zwei Serien. Sollen noch mehr Studierende geprüft werden, sind zusätzliche Prüfungsserien notwendig. Dies hätte in der Praxis einen der beiden folgenden Nachteile: entweder wird eine zweite, vergleichbare Prüfung erstellt und für eine der drei Serien verwendet, oder die Studierenden der einen Serie müssen in einem überwachten Warteraum die komplette Prüfungszeit einer anderen absitzen, um die Kommunikation mit den Studierenden der anderen beiden Serien zu verhindern.

6.4.7 Usability von Sioux

Ende 2009 haben wir die *Usability* (Benutzungsfreundlichkeit) des Sioux *ExamClients* unter Verwendung von drei *Use-Cases* untersucht und die Resultate in einem technischen Report veröffentlicht (Dahinden, 2010c).

Diese Untersuchung hat unter anderem gezeigt, dass die Benutzerführung des *ExamClients* intuitiv und dass das Ablegen einer CBA-Prüfung mit Sioux ohne zusätzliche Hilfestellungen möglich ist. Untersucht wurden folgende Arbeitsschritte:

- Einloggen ins Prüfungssystem
- Erkennung des *Backup*-Speichermediums
- Navigation durch die Prüfungsfragen
- Geben und ändern von Antworten bei allen drei Fragetypen
- Ablesen der restlichen Prüfungszeit
- Signieren und Abgeben der Prüfung

Neben den Fragen zu den drei *Use-Cases* enthielt der Fragebogen auch Fragen zur Qualität der Prüfungsfragen, zwei offene Fragen zur generellen Zufriedenheit bei der Arbeit mit dem Sioux *ExamClient* sowie einzelne Fragen zur Arbeitsgeschwindigkeit. Bei Letzteren war auffallend, wie die Studierenden die Wartezeit beim Wechseln zwischen den Fragen wahrnahmen. So haben ungefähr 30% der Studierenden angegeben, zwischen 1-3 Sekunden beim Wechsel gewartet zu haben. Diese Wartezeit lässt sich technisch nicht erklären, da beim Prüfungsstart die komplette Prüfung übertragen und gerendert wurde, so dass sie unabhängig von der Netzwerk- oder Serverauslastung unmittelbar, das heisst innerhalb Sekundenbruchteilen, angezeigt werden. Diese Zahlen zeigen somit viel eher die Diskrepanz zwischen der tatsächlichen und der wahrgenommenen Wartezeit. Daraus kann der Schluss gezogen werden, dass die Studierenden in einer

Prüfungssituation offenbar sehr sensibel auf externe Einflüsse reagieren, welche Ihre Arbeit unterbrechen (Abbildung 6.52).

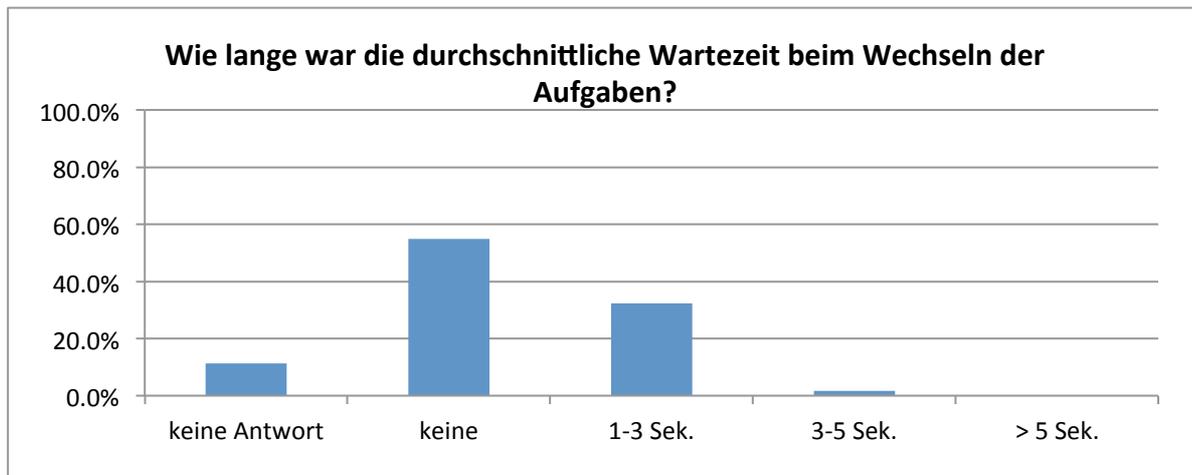


Abbildung 6.52: Geschätzte Wartezeit beim Wechsel der Antworten. Rund 30% gaben an, zwischen 1-3 Sekunden beim Wechseln der Prüfungsfragen gewartet zu haben (abgeändert aus (Dahinden, 2010c) n=62, Probepfegung, "Einsatz von Informatikmitteln" von Prof. Dr. H. Hinterberger, HS2009).

Bei einer anderen Befragung konnte die bemerkenswerte Tatsache festgestellt werden, dass die Studierenden mehrheitlich der Meinung waren, dass sie kein besseres Ergebnis erzielt hätten, wenn Sie die komplette Prüfung auf Papier erhalten hätten (Abbildung 6.53). Hierbei zeigt sich offenbar, dass sich eine grosse Mehrheit der Studierenden durch den Computer bei der Prüfungsbearbeitung nicht benachteiligt fühlt.

Ich hätte ein besseres Ergebnis am Semesterendtest erzielen können, wenn ich die komplette Prüfung auf Papier erhalten hätte.

trifft voll zu	1 (2%)
trifft zu	0 (0%)
trifft teilweise zu	10 (18%) ■
trifft weniger zu	13 (23%) ■
trifft nicht zu	31 (55%) ■■
keine Antwort	1 (2%)

Abbildung 6.53: Medienfrage bei CBA-Prüfungen (n=56, Semesterendtest, "Informatik für Biologie und Pharm. Wiss" von Dr. H.-J. Böckenhauer, FS2011).

Bei der Frage, ob Sie Vorbehalte hätten, weil Sie die Basisprüfung in Informatik mit Sioux ablegen müssen, antworteten 96% mit "nein" (Abbildung 6.54). Die beiden aufgeführten Vorbehalte bezogen sich auf inhaltliche Kritikpunkte zu den Fragestellungen und sind somit in Bezug auf die technische oder organisatorische Qualität irrelevant.

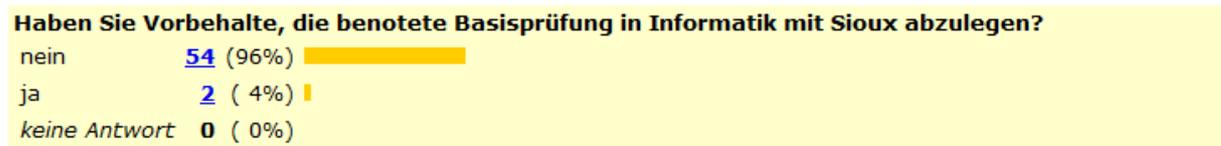


Abbildung 6.54: Prospektive Erwartungshaltung in Bezug zur benoteten CBA-Prüfung (n=56, Semesterendtest, "Informatik für Biologie und Pharm. Wiss" von Dr. H.-J. Böckenhauer, FS2011).

Abschliessend muss auch selbstkritisch festgestellt werden, dass die Meinung der Studierenden bezüglich CBA fragil ausfällt. Eine einzige unsorgfältig formulierte Frage oder ein Fehler in der Musterlösung kann dazu führen, dass die Studierenden sich negativ gegenüber der Prüfung äussern. Dies haben wir z.B. an einer Pilotprüfung an einer Schweizer Fachhochschule erlebt, als nach einem kleinen Fehler im Korrekturmodul²¹ von Sioux die Studierenden die Gelegenheit nutzten, um ihre grundsätzlichen Vorbehalte gegenüber CBA zum Ausdruck zu bringen. Eine erneute Prüfung, mit einem überarbeiteten Korrekturmodul, führte dann an der gleichen Fachhochschule mit dem gleichen Setup einige Monate später dazu, dass CBA von den Studierenden viel positiver bewertet wurde.

6.5 Weiterentwicklungen

Aus technischer Sicht kann die Entwicklung der Sioux-CBA-Suite Ende 2011 als abgeschlossen betrachtet werden. Bis zu diesem Zeitpunkt wurde Sioux während vier Jahren in der Praxis eingesetzt und mehr als 4'000 Einzelprüfungen erfolgreich abgelegt. Speziell bewährt hat sich die Teilautonomie des *ExamClients*, die nicht nur eine verzögerungsfreie Bearbeitung der Prüfung bei grossen Studierendenzahlen ermöglicht, sondern auch die Durchführung von Prüfungen unter erschwerten Bedingungen erlaubt (Prüfungen über drahtlose Netzwerke oder grosse Entfernungen zwischen Client und Server). Auch die Implementierung der Signierung der Prüfungsergebnisse als Teil der zentralen Sioux Sicherheitskomponente funktioniert verlässlich und problemlos. Aus diesem Grund haben wir im Jahr 2010 die im Kapitel 6.3 erwähnten Sicherheitskomponenten modularisiert und als eigenständigen Webservice ("Sioux-Back-End" genannt) implementiert. Die Idee war, dass dadurch bestehende oder neue CBA-Systeme diesen lokalen Webservice einbinden und diesem die für die Datensicherheit und die Kommunikation zwischen Client und Server nötigen Funktionen überlassen konnten.

Im Sinne einer Machbarkeitsstudie wurde in Zusammenarbeit mit dem Institut für medizinische Lehre der Universität Bern das Projekt MEASURED (*Media-rich Electronic Assessment with Secure Delivery*) lanciert. Parallel dazu verfolgen wir die Klärung der legalen Vorgaben für CBA in der Schweiz und sind bemüht, Sioux nachhaltig in der Hoch-

²¹ Ein Teil der Lückentextaufgaben wurde falsch korrigiert, da Leerschläge zu Beginn und am Ende der Antworten bei der Korrektur mit berücksichtigt wurden, jedoch in der Musterlösung nicht enthalten waren.

schulprüfungslandschaft zu integrieren. Diesbezüglich laufen verschiedene Pilotprojekte im In- und Ausland.

6.5.1 MEASURED

Im MEASURED Projekt wird das Ziel verfolgt, CBA in eidgenössischen Prüfungen der Humanmedizin (ehemals "Staatsexamen") aufzunehmen. Aufgrund erhöhter Rekurswahrscheinlichkeit und zeitgleicher Durchführung der Leistungskontrollen an verschiedenen Standorten evaluierte das Institut für Medizinische Lehre der Universität Bern verschiedene CBA-Softwarelösungen. Sie kamen zum Schluss, dass Sioux die Anforderungen an eine hohe Daten- und Ausfallsicherheit erfüllt. Die anvisierten medizinischen Prüfungen stellen jedoch darüber hinaus Anforderungen, welche Sioux nicht leicht (Integration in Parcours-Prüfungen, vollständiger *offline*-Modus für Prüfungsorte ohne Internetanbindung) oder technologiebedingt nicht (gute Multimedia-Unterstützung) erfüllen konnte (Guttormsen et al., 2010).

Für uns ergab sich dadurch die Möglichkeit, den *proof-of-concept* zu erbringen, dass der Entwicklungsaufwand reduziert und die Sicherheitsmerkmale anderer CBA-Systeme durch den Einsatz der modularisierten Sioux Sicherheitskomponenten erhöht werden konnte.

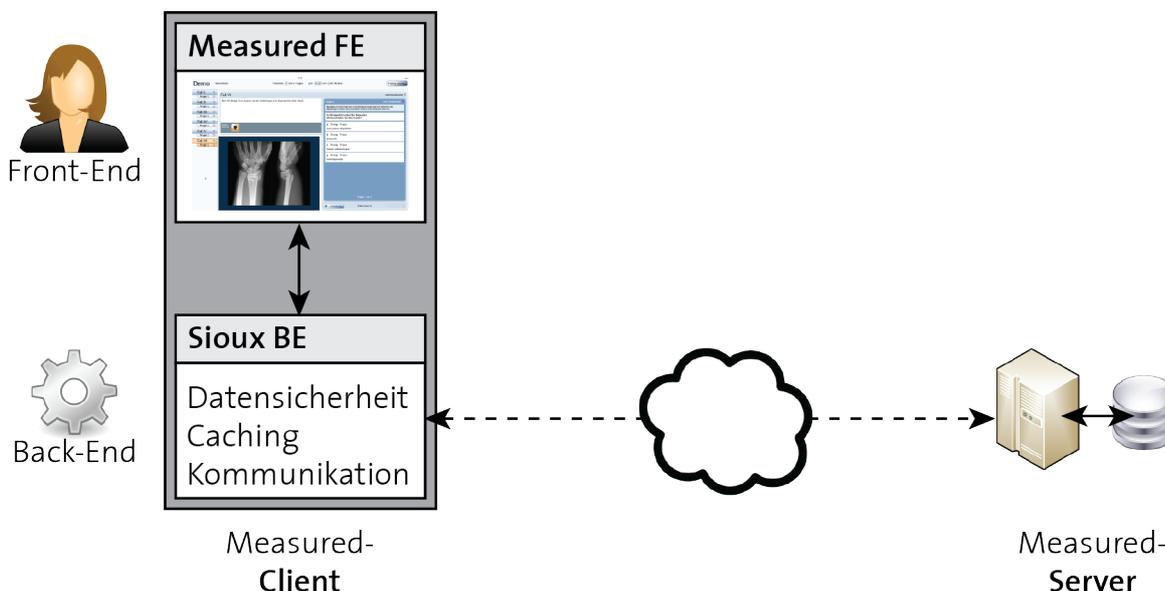


Abbildung 6.55: Design-Prinzip von MEASURED basierend auf dem Sioux-BackEnd. Das Sioux-BackEnd läuft auf dem Client als lokaler Webservice und übernimmt dort vollständig die Kommunikation zwischen Client und Server. Das MEASURED *Front-End* kommuniziert mit dem Sioux-BackEnd, welches alle sicherheitsrelevanten Funktionen wie digitale Signierung und Cachingfunktionen übernimmt.

Das Design-Prinzip von MEASURED ist in Abbildung 6.55 dargestellt. Es zeigt, dass das MEASURED *Front-End* die Benutzerschnittstelle bildet. Dieses kommuniziert mit dem als lokalen Webservice implementierten Sioux *Back-End*, welches die in den CBA-

Designprinzipien beschriebenen Anforderungen bezüglich Belegbarkeit und Ausfallsicherheit des Prüfungsprozesses erfüllt.

Beim MEASURED *Front-End* kam mit der Entwicklungsumgebung Adobe Flex und der Programmiersprache Actionscript eine erfolgsversprechende Technologie zum Einsatz, die über mächtige Werkzeuge für die Integration multimedialer Inhalte wie Filme und Audiodateien verfügt. Durch die Integration des *Sioux-Back-Ends* ergab sich die Möglichkeit, trotz kurzer Entwicklungszeit von bloss einem Jahr ein auf die Prüfungsdurchführung und deren Sicherheit fokussiertes System zu schaffen, welches grosse Flexibilität und Erweiterbarkeit bietet und unter Berücksichtigung der juristischen Vorgaben für verschiedene Prüfungsinhalte und -szenarien verfügbar gemacht werden konnte (Rieder, Dahinden, Nonava, Hinterberger, & Guttormsen, 2012).

Bis Ende 2011 unterstützte MEASURED die beiden Standardfragetypen "Kprim" (Mehrfachantworten) und "A" (Eins-aus-fünf), die mit multimedial angereicherten Fallbeschreibungen ergänzt werden können. Zu den unterstützten Inhalten zählen Filme, Audiodateien und Bilder. Beispiele dieser Fragetypen können auf der MEASURED-Projekthomepage nach dem *Download* des Demo-Programms studiert werden (Rieder & Nonava, 2012).

Der erste produktive Einsatz des MEASURED CBA-Systems fand Anfang September 2011 während einer mehrtägigen, mehrsprachigen und dezentral durchgeführten Parcours-Prüfung mit ca. 750 Studierenden statt. Diese Prüfung hat gezeigt, dass die von Sioux übernommene Technologie vollständig eingebunden und um einen *offline*-Modus (mit Sicherheitsvorbehalten) für die Durchführung an Prüfungslokalen ohne Netzwerk-Konnektivität erweitert werden konnte. Sämtliche fachspezifischen, inhaltlichen und strukturellen Anforderungen (z.B. die Einbindung von Videos und Tonaufnahmen) konnten im MEASURED *Front-End* umgesetzt werden. Ebenso führte die Bedienung bei Kandidierenden, welche zum ersten Mal mit der Prüfungsapplikation konfrontiert waren, zu keinerlei Fragen (Rieder et al., 2012). Ein Nachweis, dass sich Stabilität und Sicherheit im *Sioux-Back-End* mit einem davon unabhängig und frei gestaltbaren *Front-End* kombinieren lässt, welches sich ganz an unterschiedlichen inhaltlichen Bedürfnissen ausrichten kann, ist damit erbracht.

6.5.2 Klärung der legalen Vorgaben bei CBA

Aus technischer, formaler und organisatorischer Sicht erachten wir Sioux basierte Leistungskontrollen als sicher und gehen davon aus, dass unsere Ablaufstrukturen und technischen Sicherheitsmassnahmen einem allfälligen Rechtsverfahren standhalten würden. Absolute Gewissheit haben wir allerdings nicht, denn dazu fehlen uns rechtsgültige Vorgaben. Trotz verschiedentlich Anstrengungen konnten wir bisher keine klaren Aussagen zur Rechtssicherheit von elektronischen Prüfungen, basierend auf digitalen Signaturen, ausarbeiten. Es lässt sich feststellen, dass der Klärungsbedarf zwar akut ist, bei Rechtsdiensten jedoch oft das Bewusstsein für die spezifischen Herausfor-

derungen von elektronischen Prüfungssystemen und das Verständnis für die involvierten Technologien fehlen.

Für die nahe Zukunft schlagen wir deshalb folgendes Vorgehen vor:

- Mittels Checklisten und Ablaufbeschreibungen sollen Anhaltspunkte und Hilfestellungen ausgearbeitet werden, mit denen unter Berücksichtigung der rechtlichen Grundlagen das Risiko einer Beschwerde gegen eine Notenverfügung minimal gehalten werden kann.
- Anhand konkreter Risikobeschreibungen sollen CBA-spezifische Probleme fassbar gemacht und mit möglichst vielen *Stakeholdern* diskutiert werden. Dadurch soll das Bewusstsein geschaffen werden, dass in der Schweiz intensiver über die Sicherheitsanforderungen an CBA diskutiert werden müsste.

Früher oder später wird es in der Schweiz ein Rechtsverfahren geben, aus denen dann Anforderungen an CBA abgeleitet werden können. Mit Sioux haben wir mit der Integration von digitalen Signaturen und Laufnummern proaktiv Sicherheitsvorkehrungen getroffen, es wird aber notwendig sein, stets die neusten Erkenntnisse in die Weiterentwicklung von Sioux einfließen zu lassen.

6.5.3 Nachhaltige Dissemination von Sioux-basierten CBA-Systemen

Zur nachhaltigen Dissemination von Sioux und MEASURED werden wir sogenannte *Cubes* entwickeln. Als ein *Cube* verstehen wir eine technische oder organisatorische Massnahme, mit der die Interoperabilität von Sioux und MEASURED mit den anderen in der Schweiz eingesetzten Systemen gewährleistet werden soll. Mit dem *Sioux-Back-End* haben wir bereits einen *Cube* entwickelt, mit dem bestehende Systeme nachgerüstet werden könnten. Im Falle von webbasierten Applikationen könnte dies zum Beispiel über ein im Browser integriertes Applet erreicht werden, welches via Sioux die Kommunikation mit dem Server übernehmen könnte. Ein wichtiger *Cube* betrifft den Einsatz von portablen studentischen Computern zur Durchführung von CBA. Zur behelfsmässigen Absicherung dieser Systeme erachten wir *Boot-Medien* als vielversprechenden Ansatzpunkt. Unter der Voraussetzung, dass das *Boot-Medium* vom nativen System (und nicht von einer virtuellen Umgebung) gestartet wird, könnte eine solch sichere Systemumgebung eine effiziente Methode sein, um auf den heterogenen studentischen Computern eine homogenere Prüfungsumgebung zu schaffen. Diesbezüglich gilt es noch das Problem der *Deployments* eines solchen Mediums zu lösen. Die grösste Herausforderung wird allerdings sein, ein Bootmedium zu erstellen, welches unabhängig von der Hardware auf den meisten Computern gestartet werden kann. Einige vielversprechende Projekte wurden diesbezüglich bereits gestartet (Aegler et al., 2010; Schneider, 2011).

Weitere nützliche *Cubes* wären zum Beispiel im Bereich der Analyse der Prüfungsergebnisse oder bei der Beschreibung der Ablaufstrukturen sinnvoll. Die Prüfungsanalyse, ein

im Sioux *AuthorClient* integriertes Modul, kann wichtige Hinweise auf die Qualität der Leistungskontrolle und der Korrektur liefern. So können beispielsweise anhand der Analyse des Antwortverhaltens der Studierenden Fehler in der Musterlösung erkannt und Fragen mit einer tiefen Trennschärfe identifiziert werden. Ein solcher *Cube* wäre generisch einsetzbar und könnte unabhängig von den verwendeten Fragetypen dazu beitragen, die Qualität der Prüfungen zu erhöhen. Ein, wie wir feststellen konnten, oft vernachlässigter Punkt bei CBA und mitunter ein Grund, warum *Multiple Choice* Fragen generell einen schlechten Ruf haben. Weitere Informationen zu den Möglichkeiten der Prüfungsanalyse finden sich im Abschnitt 7.4.4 "Auswertung".

Ein *Cube* zur Organisation der Ablaufstrukturen könnte in Form von Checklisten und Flussdiagrammen erstellen werden. Diese hätten zum Ziel, die Planung und Durchführung von CBA zu dokumentieren, um so die zeitliche Belastung für die Prüfungsleitung zu reduzieren und gleichzeitig die organisatorische Sicherheit von CBA zu erhöhen. Diesbezüglich haben die meisten Hochschulen in der Schweiz ihre eigenen Organisationsmodelle entwickelt (Lehre, 2011; Ruedel, Schiefner, Noetzli, & Seiler Schiedt, 2007). Es lässt sich allerdings feststellen, dass üblicherweise der Grossteil der administrativen Aufgaben und Verantwortungen von den Sekretariaten der jeweiligen Dozierenden erledigt werden müssen. Zudem ist oft nicht klar, wer bei einem gravierenden Problem die Verantwortung trägt. Dies zeigte sich beispielhaft 2011 nach einer schriftlichen Prüfung am Institut für *Banking* und *Finance* der Universität Zürich, als aufgrund einer Informationspanne die Studierenden mit unterschiedlichen Prüfungsbedingungen konfrontiert waren. Es war vorerst unklar, wer die Verantwortung für diesen Vorfall tragen musste (Landolt, 2011). Es ist wünschenswert und im Sinne einer Qualitätssteigerung unabdingbar, dass die Organisationsmodelle von CBA auch die Zuständigkeiten und Verantwortlichkeiten beinhalten und von der Schulleitung offiziell verabschiedet werden.

6.6 Diskussion der Fragestellung 2: Welches sind geeignete technische Massnahmen, um in einem heterogenen Umfeld eine sichere Computer-basierte Leistungskontrolle durchzuführen?

Die Vorteile von CBA sind vielfältig. Dank der automatischen Korrektur erhalten nicht nur die Studierenden bei einer formativen Leistungskontrolle ein sofortiges Feedback zu ihrem aktuellen Wissensstand, sondern sie bedeutet auch eine grosse Korrekturzeitersparnis bei summativen Prüfungen. Zudem erlaubt CBA den Einsatz von Medien (Audio, Bilder und Video), welche bei anderen Prüfungsarten nicht oder nur in beschränktem Umfang möglich wären.

Aus legaler Sicht bietet CBA die Vorteile eines protokollierten Prüfungsprozesses und einer formalisierten Korrektur, mit welchen das Prüfungsergebnis jederzeit nachgebildet und überprüft werden kann. Damit diese Überprüfung rechtsgültig sein kann, müs-

sen alle relevanten Prüfungsdaten auf den CBA-Clients mit den digitalen Schlüsseln der Studierenden signiert werden.

Durch den Einsatz der digitalen Signaturen müssen der Sioux-Server und das Netzwerk nicht mehr speziell geschützt werden, da aufgrund der Signaturen nachträgliche Manipulationen (z.B. Übertragungsfehler, Serverfehler) zuverlässig erkannt werden können. Im Einzelnen führen folgende Überlegungen zu diesem Schluss:

- Eine Manipulation, entweder an den Aufgaben oder den Antworten, fällt durch die nicht verifizierbaren Signaturen im Antwortdatensatz zweifelsfrei auf.
- Nachträgliche Manipulationen der Antworten und deren Neusignierung sind nur mit Hilfe der geheimen Signaturschlüssel der Studierenden möglich. Diese befinden sich aber zur Zeit der Prüfung im überwachten Prüfungsraum und nachher ausschliesslich auf dem *Backup*-Speichermedium bei den Studierenden.
- Verloren gegangene Antworten können aufgrund der lückenhaften mitsignierten Laufnummern erkannt werden.

Neben den digitalen Signaturen verfügt der Sioux *ExamClient* über eine Teilautonomie bezüglich der Prüfungsbearbeitung. Diese kommt zustande, indem der *ExamClient* bei Prüfungsbeginn alle für die Prüfungsbearbeitung notwendigen Daten (Fragen, Prüfungsinformationen etc.) vom Server auf den Client überträgt. Dadurch ergeben sich in der Praxis gegenüber Browser-basierten Systemen zwei grosse Vorteile:

1. Durch die auf dem Client enthaltenen vollständigen Prüfungsdaten und die Pufferkapazität der Warteschleife hat ein Ausfall des Servers nach dem Prüfungsstart (vorsätzlich herbeigeführt oder zufällig) für die Durchführung der Prüfung keine unmittelbare Relevanz.
2. Da die Prüfungsdaten nur zum Prüfungsbeginn einmalig übertragen werden müssen, wird die Netzwerkinfrastruktur weniger belastet. Gerade bei grossen Studierendenzahlen oder unter Verwendung von Funknetzwerken kommt dieser Tatsache eine entscheidende Bedeutung zu.

Mit Sioux konnten wir zeigen, dass es dank der konsequenten Umsetzung der fünf CBA-Designprinzipien möglich ist, die Sicherheit von CBA im Praxisalltag zu erhöhen. Dank der Evaluationsresultate ist ersichtlich, dass die neue Prüfungsart von den Studierenden grossmehrheitlich akzeptiert wird und sie sich vom neuen Medium nicht diskriminiert fühlen.

Bei einer kritischen Betrachtung der bestehenden Sicherheitsrisiken müssen wir feststellen, dass das Fehlen von spezialisierter Hardware zwar mit technischen und organisatorischen Massnahmen bis zu einem gewissen Grad kompensiert werden kann, aber dass aus rechtlicher Sicht ein gewisses Restrisiko bestehen bleibt. Dazu zählen:

- Unser System beruht darauf, dass wir den Prüfungsclients vertrauen. Ohne spezialisierte Hardware ist es für uns nicht möglich, im Sioux **ExamClient** die korrekte Funktionsweise des Clients zu überprüfen. Mittels eines gehärteten Bootmediums kann diesem Ziel zwar ein Stück näher gerückt werden, aber gerade in Kombination mit studentischen Computern gilt es, Prüfungsclients als vertrauenswürdige Zone weiterhin als kritisch zu betrachten.
- Das Fehlen von digitalen Signaturerstellungseinheiten führt dazu, dass wir auf den Einsatz von erweiterten Signaturen zum Schutz von Prüfungsergebnissen verzichten müssen. Wir sind aber der Meinung, dass dieses Manko durch eine Gewaltentrennung weitgehend behoben werden kann. Diese organisatorische Massnahme führt dazu, dass bei einer Sioux-basierten CBA-Prüfung verschiedene Parteien involviert sind. Eine Manipulation der Prüfungsergebnisse durch eine einzelne Partei wird dadurch erschwert.

Die in diesem Kapitel beschriebene Implementierung einer sicheren CBA-Prüfungsapplikation ist zwar eine wichtige, nicht aber die einzige Voraussetzung für ein funktionierendes CBA-Prüfungssystem. Die CBA-Applikation legt die Basis für eine hohe Reliabilität einer Prüfung, indem Sie mit einer stabilen und benutzerfreundlichen Bedienoberfläche zuverlässig und wiederholbar das Wissen der Studierenden misst. Für valide Prüfungsergebnisse hingegen spielt das Zusammenspiel von Lernzielen, Lerninhalten und Leistungskontrolle eine entscheidende Rolle. Dieses *Alignment* wird Inhalt des kommenden Kapitels sein. Es wird aufgezeigt, wie valide Leistungsdaten zur Messung und Analyse des *Learning Outcomes* verwendet werden können und einen wichtigen Beitrag zum Forschungsgebiet "*Learning Analytics*" liefern.

Analyse der Leistungsdaten

Kapitel 7

Entwicklung und Evaluation valider Prüfungsfragen für CBA-Systeme

7.1 Übergeordnete Entwicklungszyklen von CBA

In der Literatur existieren verschiedene Modelle für CBA. Wie im Abschnitt 4.4.8 "Risikomanagement als Basis zur Entwicklung einer CBA-Strategie für Bildungsinstitutionen" erläutert, beschreibt beispielsweise das *Catherine-wheel*-Prinzip basierend auf einer iterativen Vorgehensweise die Etablierung von CBA an einer Bildungsinstitution (Zakrzewski & Bull, 1998; Zakrzewski & Steven, 2000). Andere Ansätze beschreiben theoretische Qualitätsmodelle für CBA-Systeme basierend auf ISO-Standards und beschränken sich primär auf die Evaluation der Prüfungsapplikation (Valenti et al., 2002).

Im letzten Teil dieser Arbeit wird ein Entwicklungsmodell für CBA Leistungsmessungen beschrieben. Das Entwicklungsmodell gliedert sich in drei Phasen, wobei die ersten beiden Phasen im aktuellen Kapitel und die dritte Phase im Kapitel 8 beschrieben werden (Abbildung 7.56).

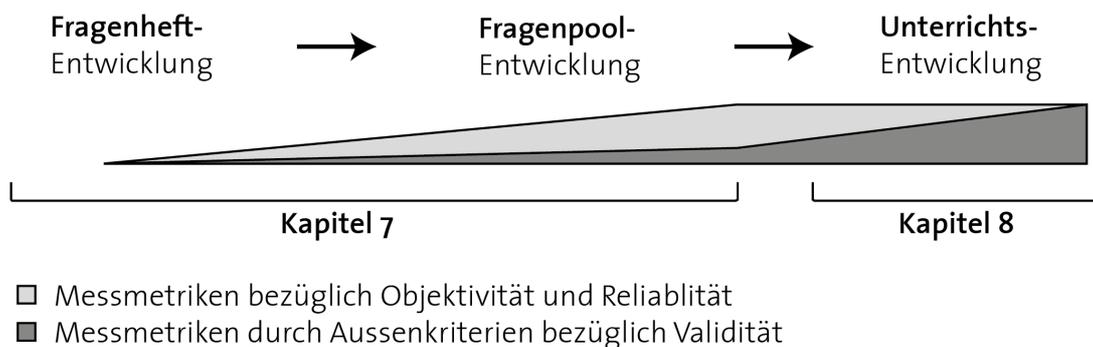


Abbildung 7.56: Die drei Phasen des Entwicklungsmodells für CBA-Leistungsmessungen. Durch die wiederholte Anwendung des Modells steigt die Menge der gesammelten Frage-Metriken, welche zur Weiterentwicklung der Fragenhefte und des Fragenpools verwendet werden können. Gute Fragenpools ermöglichen valide Leistungsmessungen, deren Resultate für die Analyse und Optimierung des Unterrichts verwendet werden können.

Die drei Entwicklungsphasen verlaufen über mehrere Jahre. Bei der erstmaligen Durchführung von computerbasierten Leistungskontrollen steht die Entwicklung von Fragenheften im Zentrum. Ein Fragenheft besteht aus einer Auswahl an Fragen, welche bei einer LK eingesetzt werden sollen. Bei der wiederholten Anwendung des Modells

kann die steigende Anzahl an gesammelten Fragemetriken zur Verbesserung der Fragenqualität und zur Qualitätssteigerung des gesamten Fragenpools verwendet werden. In einem dritten Entwicklungsschritt können die Fragen durch Aussenkriterien validierten Leistungsmessungen nützliche Hinweise zur Unterrichtsentwicklung liefern. Letzteres wird im Kapitel 8 besprochen.

7.2 Entwicklungsmodell für die Fragenheft- und Fragenpool-Entwicklung

Das Entwicklungsmodell für Fragenhefte und Fragenpools ist in Abbildung 7.57 abgebildet und gliedert sich in vier Hauptprozesse:

- I. Planung
- II. Erstellung
- III. Durchführung
- IV. Auswertung

Weil die Qualitätsverbesserung Teil dieses Entwicklungsprozesses ist, sind oft mehrere Durchläufe notwendig, um in der Praxis verlässliche Leistungsmessungen durchführen zu können.

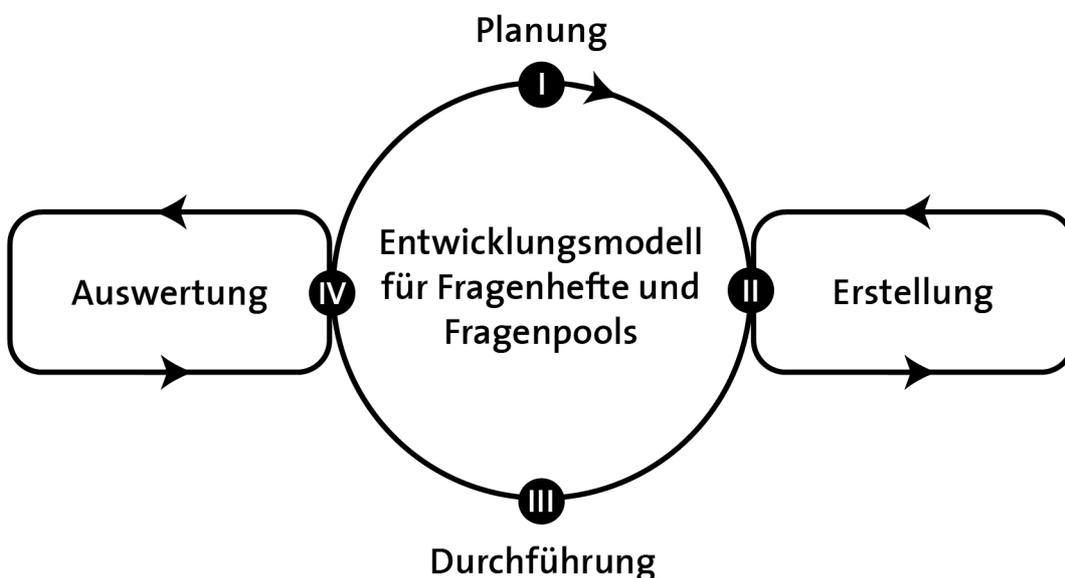


Abbildung 7.57: Die Übersicht über das Entwicklungsmodell für CBA zeigt die vier Hauptprozesse: Planung, Erstellung, Durchführung und Auswertung. Die jeweiligen Subprozesse werden durch eigene Abbildungen illustriert.

Im ersten Teil dieses Kapitels werden die vier Hauptprozesse beschrieben und die Abfolge der jeweiligen Subprozesse in einer eigenen Abbildung illustriert.

Im Mittelteil werden die für die Leistungsmessung verwendeten Fall-basierten Fragen vorgestellt. Diese Fragen lassen sich automatisiert korrigieren und sind für den Einsatz in CBA-Systemen optimiert (Kapitel 7.3).

Im dritten Teil wird das Entwicklungsmodell mit realen Daten aus dem Unterricht veranschaulicht (Kapitel 7.4) und die dritte Fragestellung besprochen (Kapitel 7.5).

7.2.1 Planung

Zieldefinition

Der Ausgangspunkt zur Entwicklung einer CBA ist die Festlegung der didaktischen Ziele dieser LK (Abbildung 7.58, I.A "Zieldefinition"). So wird beispielsweise definiert, ob es sich um eine summative oder formative LK handelt und in welchem Punktebereich die LK selektiv sein soll. Der Selektionsbereich hat Einfluss auf die Schwierigkeit der verwendeten Fragen. Bei formativen LK deckt dieser Bereich meist eine grosse Leistungsbreite ab, um sowohl schwachen als auch starken Studierenden ein Feedback geben zu können. Bei summativen und ganz speziell bei promotionsrelevanten LK ist oft der Leistungsbereich rund um die Bestehensgrenze ausgeprägt, um in diesem Bereich möglichst zuverlässige Messungen vornehmen zu können.

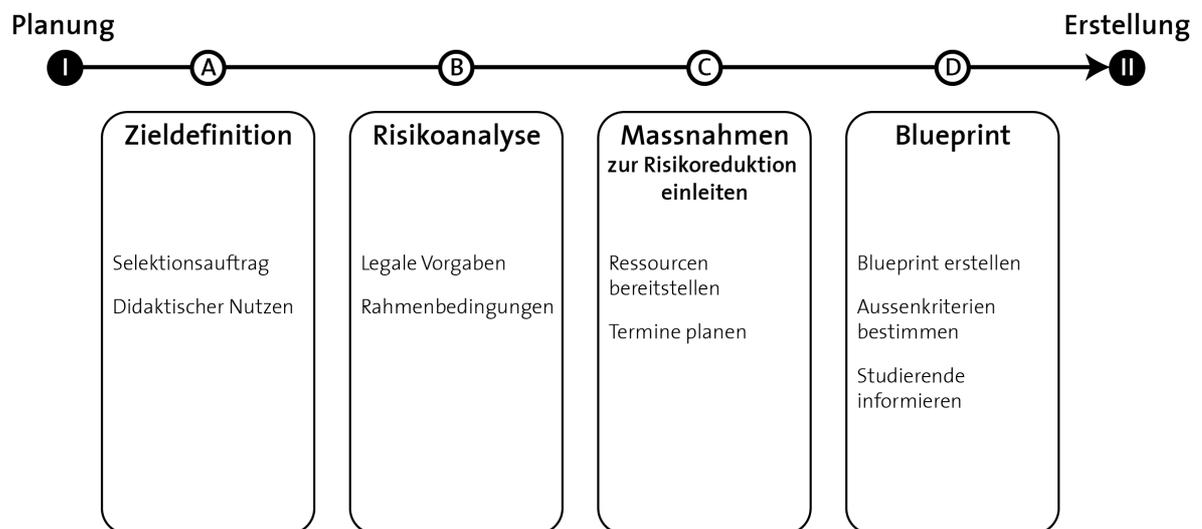


Abbildung 7.58: Detailansicht des Hauptprozesses "Planung". Pro Subprozess sind jeweils die Ziele angegeben. Die römischen Ziffern referenzieren die Hauptprozesse des Entwicklungsmodells in Abbildung 7.57.

Risikoanalyse

Bei der ersten Durchführung von CBA ist eine Risikoanalyse ratsam. Sie hilft, die möglichen didaktischen, operationalen, technischen und finanziellen Risiken zu erkennen, Gegenmassnahmen zu treffen und zu priorisieren. Da das Gefahrenpotential je nach

Ausgangslage unterschiedlich ist, sollten bei der Planung einer CBA die lokalen Gegebenheiten in der Risikoanalyse berücksichtigt werden. Die Erstellung einer Risikoanalyse ist im Kapitel 4.4 beschrieben.

Massnahmen zur Risikoreduktion einleiten

Die meisten Risiken bei CBA können durch rechtzeitig eingeleitete Gegenmassnahmen reduziert werden. Tabelle 4.7 zeigt die nach Gefahrenpotential absteigend sortierten Risiken bei CBA. Es empfiehlt sich, die mit einem höheren Risiko behafteten Massnahmen prioritär zu behandeln.

Als Folge dieses Subprozesses erstellen wir eine Terminplanung, in der die einzelnen Massnahmen und die benötigten Ressourcen eingeplant werden. Im Abschnitt 7.3.1 wird eine derartige Planung an einem praktischen Beispiel aufgezeigt.

Blueprint

Als letzter Schritt der Planungsphase gilt es, den *Blueprint* der CB-LK festzulegen. Dies setzt voraus, dass die Unterrichtsinhalte und die Zeitplanung bekannt sein müssen. Der *Blueprint* dient als Basis für die Zusammenstellung des Fragenhefts. Er soll sicherstellen, dass die geprüften den vermittelten Inhalten in punkto Umfang, Themen und kognitiven Prozessdimensionen entsprechen. Die Grundlage des *Blueprints* wurde im Abschnitt 3.3.5 erläutert und wird im Abschnitt 7.3.1 an einem Praxisbeispiel illustriert.

7.2.2 Erstellung

Die Erstellung der Fragen und des Fragenhefts nimmt über die Hälfte der aufzuwendenden Zeit des gesamten CBA-Prozesses in Anspruch (vgl. Tabelle 6.3). Es lohnt sich deshalb, genug Zeit für diese Arbeiten einzuplanen.

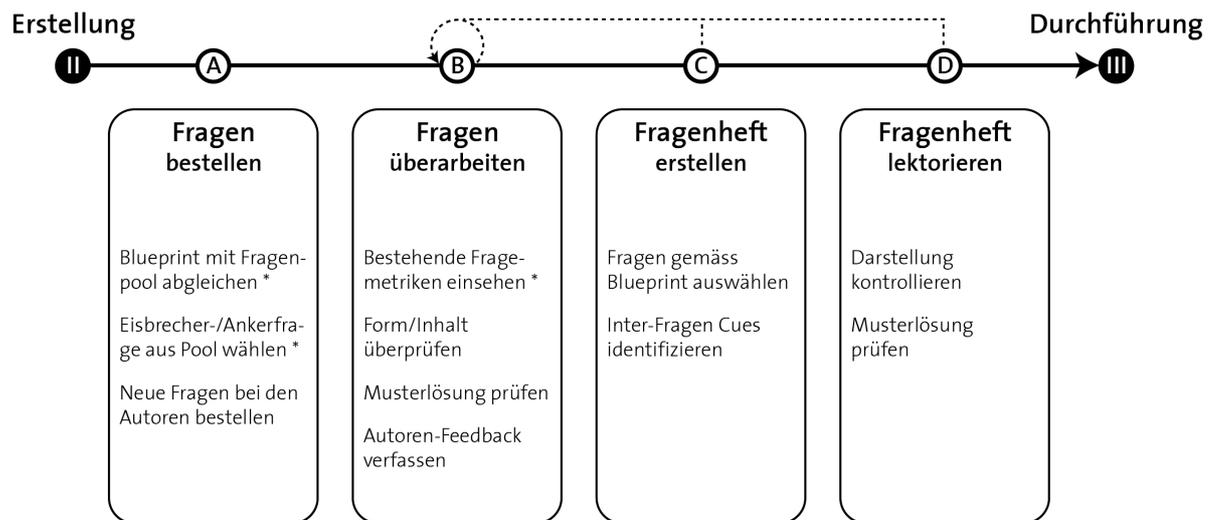


Abbildung 7.59: Detailansicht des Hauptprozesses "Erstellung". Er gliedert sich in vier Subprozesse, wobei die jeweiligen Ziele angegeben sind. Die mit einem Stern (*) markierten Schritte können erst bei wiederholter Anwendung des Modells ausgeführt werden.

Fragen bestellen

Wie aus dem Einsatz von Sioux an anderen Hochschulen bekannt wurde, hängen die Evaluationsresultate direkt mit der Qualität der verwendeten Prüfungsfragen zusammen. Ist diese unzureichend, weil die Fragen unsauber formuliert sind oder sich nicht auf den Unterricht beziehen, wird das gesamte CBA-System als mangelhaft taxiert. Im Sinne einer nachhaltigen Etablierung von CBA lohnt es sich deshalb, nicht nur die technischen, sondern vor allem auch die inhaltlichen Voraussetzungen zu schaffen.

In einem ersten Schritt gilt zu prüfen, welche vom *Blueprint* geforderten Inhalte bereits im Fragenpool vorhanden und ob deren Messmetriken für einen erneuten Einsatz ausreichend sind (Abbildung 7.59, II.A "Fragen bestellen"). Sind Letztere ungenügend, bietet sich eine Überarbeitung an oder die Frage muss durch eine neue ersetzt werden.

Wird die LK erstmalig durchgeführt, ist der Fragenpool noch leer und alle Fragen müssen neu erstellt werden. Wird die LK zum wiederholten Male durchgeführt, kann ein Teil der Fragen aus dem Pool übernommen werden. Dies ist aus drei Gründen sinnvoll:

- Der Aufwand für die Erstellung des Fragenhefts reduziert sich.
- Durch die Wahl bewährter Fragen aus dem Fragenpool können die Prüfungsergebnisse über mehrere Semester miteinander verglichen werden. Diese Fragen werden als Ankerfragen bezeichnet.
- Da die Itemmetriken der Ankerfragen bekannt sind, können diese gezielt ausgewählt werden. Dies ist speziell bei der Bestimmung von sehr leichten Fragen, so genannten Eisbrecherfragen oder bei besonders trennscharfen Fragen zu empfehlen. Eisbrecherfragen sind speziell einfache Fragen zu Beginn einer LK, welche den Pro-

banden den Einstieg erleichtern sollen. Meist werden die Eisbrecherfragen nur mit wenigen Punkten bewertet.

Die benötigten Fragen werden von den Dozierenden und Chefassistierenden erstellt. Wir haben jedoch auch gute Erfahrungen damit gemacht, die Assistenzpersonen bei der Fragenerstellung mit einzubeziehen. Einerseits können die Dozierenden dadurch Zeit sparen und andererseits gelingt es so, den Assistierenden das *hidden Curriculum* nahezubringen und im Übungsbetrieb den Studierenden ein klares Feedback zu geben.

Um sicherzustellen, dass die Frageninhalte dem *Blueprint* entsprechen und die Fragenqualität ausreichend ist, sind die Assistierenden auf Feedback von den Dozierenden oder erfahrenen Chefassistierenden angewiesen. Diesbezüglich hat sich ein Feedbackzyklus bewährt, wie er bei Dahinden et al. beschrieben wird (Dahinden & Hinterberger, 2010). Zusätzlich wird den Assistierenden eine kurze Anleitung abgegeben, welche die Frage-Typen beschreibt und Tipps zum Design der Prüfungsfragen beinhaltet.

Wir haben festgestellt, dass die Qualität der von den Assistierenden erstellten Fragen umso besser ist, je konkreter der Auftrag formuliert wurde. Dazu gehört, dass beispielsweise nicht nur die zu prüfenden Inhalte detailliert vorgegeben werden, sondern auch auf geeignete Rohdatenquellen hingewiesen wird²² oder die Fragetypen vorgeschrieben werden.

Fragen überarbeiten

Bevor die bestehenden oder neu erstellten Fragen in einer LK eingesetzt werden können, bedarf es einer formalen und inhaltlichen Kontrolle der Fragen. Verfügt die Frage bereits über Itemmetriken aus früheren Einsätzen, bieten diese eine gute Ausgangslage für die Überarbeitung. Wurde die LK bereits einmal gemäss unseres Entwicklungsmodells durchgeführt, liegen womöglich für die Fragen bereits Optimierungsvorschläge vor.

Aus motivationalen und nachhaltigen Aspekten wird empfohlen, die Überarbeitung ebenfalls von den Autoren respektive von den Assistierenden vornehmen zu lassen. Dabei wurde auch hier die Erfahrung gemacht, dass gutes *Feedback* die Qualität der Fragen positiv beeinflusst. Unser *Feedback* umfasst entsprechend ein Urteil über die Relevanz der Aufgabenstellung, die Prägnanz der Formulierung, den geschätzten Schwierigkeitsgrad, die Ratewahrscheinlichkeit aufgrund des gewählten Fragetyps, die Vollständigkeit der Musterlösung und die Qualität der eingesetzten Medien.

²² Eine gute Rohdatenquelle ist beispielsweise die Webseite des Bundesamtes für Statistik: <http://www.bfs.admin.ch/> (abgerufen am: 23.05.2014)

Das Überarbeiten der Fragen durch die Autoren aufgrund des *Feedbacks* von Experten ist ein iterativer Prozess. In der Regel sind rund zwei Iterationen notwendig. Dass sich dieser Aufwand lohnt, wurde von den Assistierenden bestätigt. So waren 80% der Assistierenden der Meinung, dass das Experten-*Feedback* geholfen hat, die Fragen zu erstellen und zu optimieren (Abbildung 7.6o).

Das Feedback (Gespräch und Emails) zu meinen Prüfungsfragen hat mir beim Erstellen/Optimieren der Aufgaben geholfen.

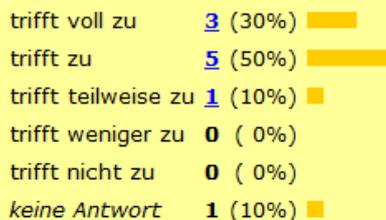


Abbildung 7.6o: Effekt des Feedbacks auf das Erstellen und Optimieren der neu erstellten Fragen. (Datenbasis: Nachumfrage unter den Assistierenden, EvIM, HSo8, n=10)

7.2.3 Durchführung

Die Durchführung der LK gliedert sich in drei Phasen: *Stakeholder* informieren, LK durchführen und Unterlagen archivieren (Abbildung 7.61).

Stakeholder informieren

Das rechtzeitige Informieren der *Stakeholder* ist wichtig, damit die Abläufe und Zuständigkeiten am Prüfungstag klar sind. Unklare Zuständigkeiten können dazu führen, dass während der LK falsche Anweisungen gegeben werden, welche die Gleichbehandlung aller Studierenden in Frage stellen (Landolt, 2011).

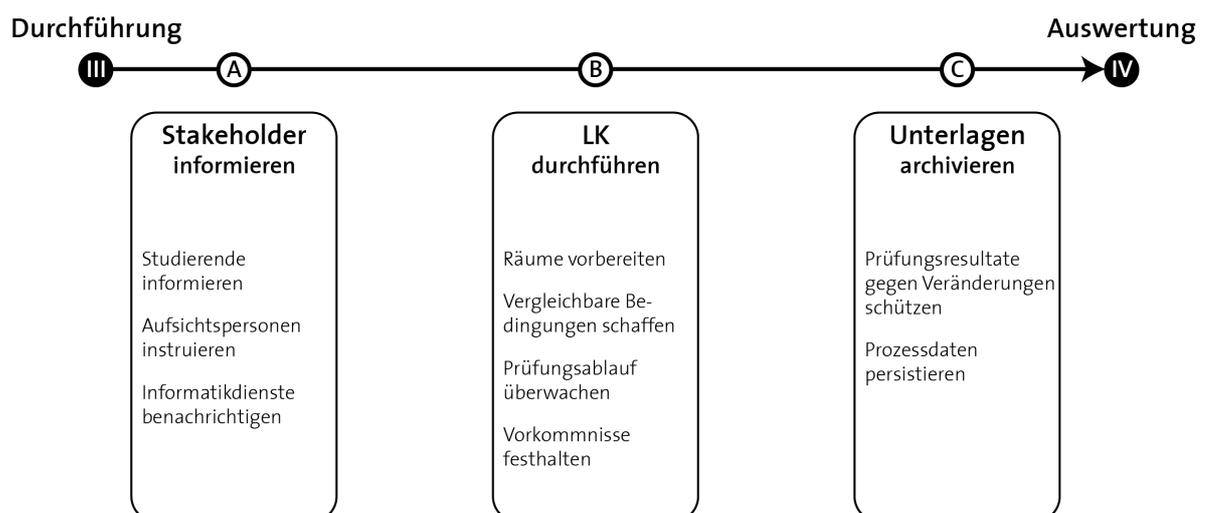


Abbildung 7.61: Detailansicht des Hauptprozesses "Durchführung". Er gliedert sich in drei Subprozesse, wobei die jeweiligen Ziele angegeben sind.

LK durchführen

Das Hauptaugenmerk bei den summativen LK richtet sich auf die Schaffung vergleichbarer Prüfungsbedingungen für alle Studierenden. Dazu gehört, dass die Studierenden in allen Prüfungsräumen mit den gleichen Informationen versorgt werden. Konsequenterweise werden beispielsweise keine inhaltlichen Fragen beantwortet, sondern es wird den Studierenden die Möglichkeit geboten, ihre Bemerkungen auf elektronischem Weg schriftlich mitzuteilen. Diese Kommentare werden später im Rahmen der Korrektur berücksichtigt. Eine Ausnahme bilden Verständnisfragen von fremdsprachigen Studierenden, welche während der LK von den Aufsichtspersonen beantwortet werden.

Unterlagen archivieren

Der Abschluss der Durchführungsphase von LK bildet das Archivieren der Unterlagen. Dazu gehört, dass alle prüfungsrelevanten Daten persistent und vor Veränderungen geschützt gespeichert werden. Die Unversehrtheit dieser Daten spielen später bei der Analyse und Benotung sowie bei einer allfälligen Prüfungseinsicht und einem Rekurs eine zentrale Rolle.

Da zur Zeit der Drucklegung dieser Arbeit noch keine konkreten Vorgaben zur Speicherung von prüfungsbezogenen Daten vorliegen, wurde ein konservativer Ansatz gewählt: Es werden möglichst detaillierte Daten gespeichert und mittels digitaler Signaturen vor Veränderungen geschützt. Der Schutz der Daten ist wichtig, um diese später im Rahmen einer legalen Abklärung auf ihre Integrität hin zu untersuchen und diese als Beweismittel verwenden zu können. Die technischen Details zu diesem Vorgehen sind im Kapitel 6.2.6 "Zertifikate zur Absicherung von Sioux" beschrieben.

7.2.4 Auswertung

Der Schwerpunkt der Auswertung von LK bildet die fehlerfreie Korrektur, die Analyse der Daten zur Feststellung der Qualität der Messung und anschliessend die Benotung aufgrund der korrigierten Resultate (Abbildung 7.62). Der Abschluss des gesamten Prozesses bildet das *Feedback* bezüglich der Resultate und deren Rechtfertigung im Rahmen von Einsichten und allfälligen Rekursen.

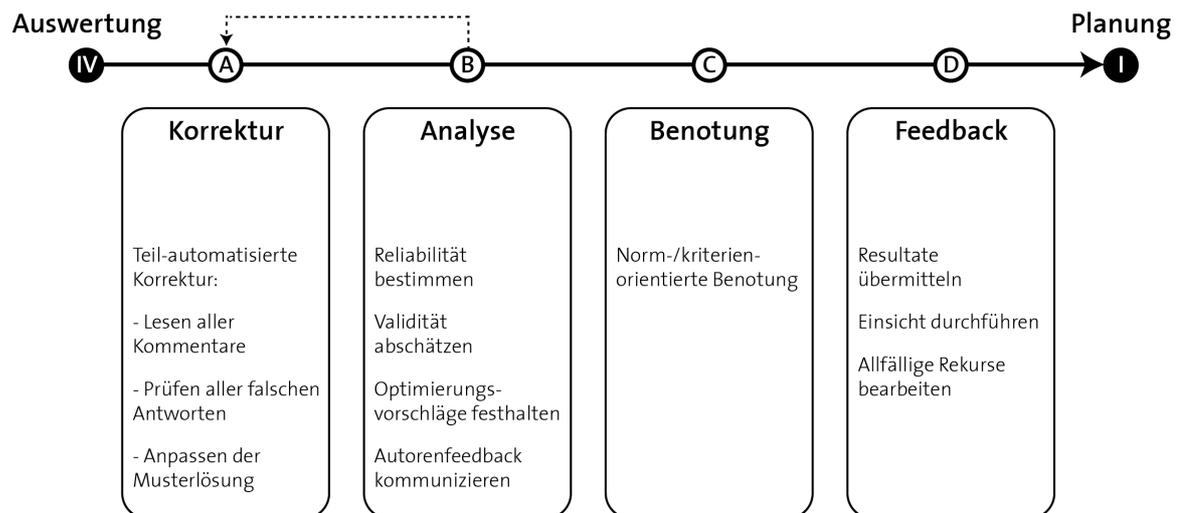


Abbildung 7.7: Detailansicht des Hauptprozesses "Auswertung". Er gliedert sich in vier Subprozesse, wobei die jeweiligen Ziele angegeben sind. Bei der wiederholten Anwendung des Entwicklungsmodells startet der neue Entwicklungszyklus wieder mit der Planung der neuen LK.

Korrektur

CBA-Systeme haben gegenüber anderen Prüfungsformen den grossen Vorteil, dass sich grosse Teile der Korrektur automatisiert durchführen lassen. Durch den Wegfall der Handschrift, der Gruppierung der Fragen nach Aufgaben statt Probanden, sowie der weitgehenden Anonymisierung der Antworten wird ein Maximum an Objektivität erreicht. Aufgrund der formalisierten Musterlösung bietet sich zudem die Möglichkeit, diese jederzeit anzupassen und die LK für alle Probanden erneut korrigieren zu lassen.

Damit unvollständige Musterlösungen erkannt werden, empfiehlt es sich, alle Kommentare zu lesen sowie die falschen Antworten anzuschauen. Gute CBA-Systeme bieten hierfür entsprechende technische Unterstützung, um beispielsweise korrekte Antworten auszublenden oder Kommentare aufgabenbezogen anzuzeigen (vgl. Kapitel 6.2.3 "Client Applikationen").

Analyse

Bei CBA-Systemen liegen die Resultate in digitaler Form vor. Dadurch ist der Aufwand zur Berechnung der Reliabilitätsmetriken wie Schwierigkeitsindex und Trennschärfe minimal. Ebenfalls stehen sie bereits während der Korrekturphase zur Verfügung. So lassen sich rasch unsauber formulierte Aufgabenstellungen oder Fehler in der Musterlösung erkennen. Bei der Beurteilung der Reliabilitätsmetriken wird das in Abbildung 7.62 gezeigte Vorgehen angewendet. Es hilft, die Reliabilität der Fragen einzuschätzen und Fragen für die weitere Verwendung zu klassieren.

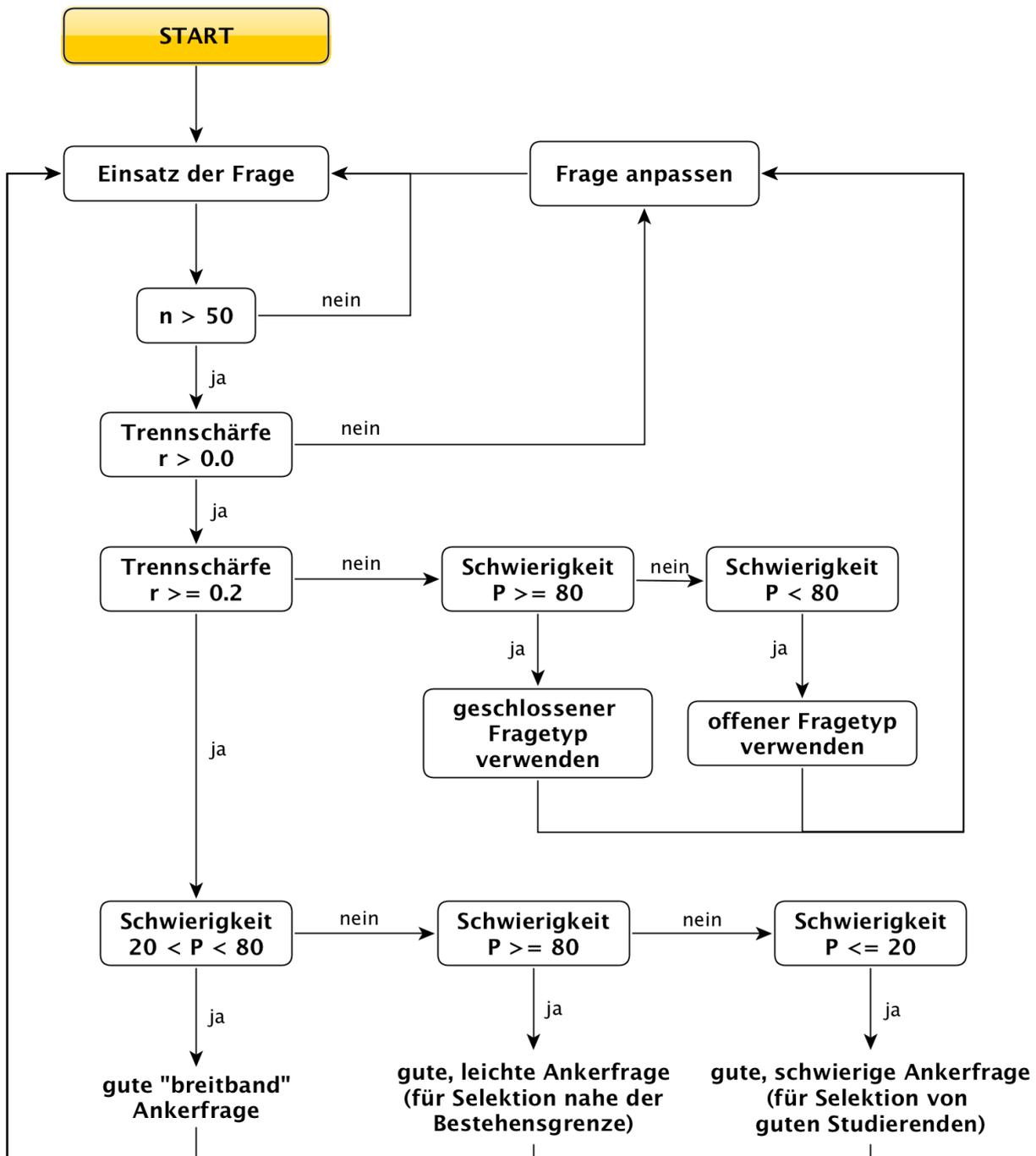


Abbildung 7.62: Flussdiagramm zur Klassierung und Optimierung der Reliabilität von Prüfungsfragen anhand deren Metriken.

Aufwändiger ist die Abschätzung der Validität. Dazu sind alternative Leistungsmessungen nötig, welche als Aussenkriterien zur Korrelation mit den Resultaten verwendet werden können. Wie im Kapitel 3 erläutert, eignen sich dafür beispielsweise Selbstbeurteilungen oder Beurteilungen durch Assistierende. Im Kapitel 7.4.4 "Auswertung" wird dieses Vorgehen mit realen Zahlen aus dem Unterricht illustriert.

Abschluss der Analyse bildet das Festhalten von Optimierungsvorschlägen und das Kommunizieren der Itemmetriken an die Autoren. Obwohl diese beiden Schritte nicht

zwingend sind, so wurden im Unterricht gute Erfahrungen damit gemacht. In der Regel interessiert es die Autoren, wie ihre Fragen funktioniert haben. Entsprechend hoch ist damit deren Bereitschaft, im kommenden Semester die Fragen zu überarbeiten und so die Metriken zu verbessern.

Benotung

Bei der Benotung kommen zwei unterschiedliche Verfahren zum Einsatz: Die kriterien- und die normorientierte Benotung. Die Kriterien-orientierte ist die häufigste Benotungsart (Goubeaud, 2010). Sie zeichnet sich dadurch aus, dass die Note die erreichten Lernziele widerspiegelt. Es sind keine Notenanpassungen aufgrund von Gruppenvergleichen vorgesehen. Das heisst, eine kriterienorientierte Benotung liefert eher absolute Leistungswerte, während die norm-orientierte Benotung Leistungsunterschiede zwischen Gruppen berücksichtigt.

Feedback

Der formale Abschluss einer LK bildet die Übermittlung der Resultate an die entsprechenden Stellen (meist an die Studiensekretariate). Bei Semester-Endprüfungen oder formativen LK können aber auch direkt die Studierenden die Empfänger der Resultate sein. Hier empfiehlt es sich, die Resultate individualisiert freizuschalten und aus Datenschutzgründen auf öffentlich einsehbare Notenlisten zu verzichten. Von der Anonymisierung der Noten durch die Verwendung der Matrikelnummern ist abzuraten, da diese für die Studierenden meist öffentlich zugänglich sind.

Das Vorgehen bei Noteneinsichten und Rekursen ist an den meisten Hochschulen vorgegeben. An der ETH Zürich beispielsweise definiert eine Weisung des Rektorats das Vorgehen bei Prüfungseinsichten (ETH Zürich, 2013b) und ein Merkblatt beschreibt das formale Vorgehen in einem Beschwerdefall (ETH Zürich, 2014).

Bei CBA-LK wird die Prüfungseinsicht insofern erschwert, als dass das Prüfungsdokument in der Regel nur in elektronischer Form vorliegt²³. Gute CBA-Systeme bieten jedoch Möglichkeiten, die Prüfung während der Einsicht in der Originaldarstellung und mit allen von den Studierenden abgegeben Antworten und Kommentaren anzuzeigen (vgl. Kapitel 6.2.3 "Client Applikationen").

²³ Ausnahme bilden CBA-Systeme, welche die Prüfung nach Abschluss ausdrucken. Diese werden von den Probanden unterschrieben.

7.3 Automatisch korrigierbare Prüfungsfragen für den Grundlagenunterricht Informatik

Das im ersten Teil dieses Kapitels beschriebene Entwicklungsmodell für CBA lässt sich unabhängig vom verwendeten CBA-System oder den eingesetzten Fragetypen anwenden. Es ist jedoch unbestritten, dass diese beiden Komponenten die Validität der Leistungsmessungen massgeblich beeinflussen. Bevor wir die Qualität unseres Messverfahrens für kognitive Leistungen empirisch messen, werden die beiden Komponenten spezifiziert.

- Als CBA-System soll die im Rahmen dieser Arbeit entwickelte Sioux CBA-Suite zum Einsatz kommen. Diese eignet sich sowohl für formative als auch für summative LK und ist für den Einsatz in heterogener Umgebung unter Wahrung der legalen Vorgaben bezüglich der Beweisbarkeit des Prüfungsprozesses konzipiert. Die Sioux CBA-Suite und dessen Grundlage wurde in den Kapiteln 4-6 eingehend beschrieben.
- Um das Potential von CBA nutzen zu können, werden automatisch korrigierbare Prüfungsfragen benötigt. Das Prinzip dieser von uns für den Informatik-Grundlagenunterricht entwickelten Fragen werden nachfolgend erläutert.

7.3.1 Anforderungen an Fragen für CBA-Systeme

Grundsätzlich ist es möglich, die gleichen Fragetypen sowohl bei mündlichen, schriftlichen und CBA einzusetzen. In der Praxis unterscheiden sich jedoch die so gemessenen Leistungen erheblich (Lee & Weerakoon, 2001; Ventouras, Triantis, Tsiakas, & Stergiopoulos, 2011). Es gibt kein optimaler, allgemeingültiger Fragetyp. Jeder Prüfungsmodus hat somit seine eigenen, bewährten Fragetypen.

Bei CBA-LK sind im allgemeinen *Multiple Choice* Fragen weit verbreitet. Sie sind formal klar definiert und wurden vielfach empirisch untersucht (Krebs, 2002). Es gibt aber auch Ansätze, wie Freitextaufgaben automatisiert ausgewertet werden können (Noorbehbahani & Kardan, 2011).

Aufgrund der begrenzten Externalisierbarkeit des Wissens (vgl. Kapitel 2.) und der Unschärfe bezüglich der Zuweisung von Fakten-Wissen zur jeweiligen kognitiven Prozessdimension, soll bei unseren Fragen die Handlung der Studierenden im Zentrum stehen.

Unter Einzug der Rahmenbedingungen wie Prüfungszeit, Kohortengrösse, Infrastruktur und Unterrichtsform wurden für unsere Fragen folgende Anforderungen formuliert:

Die Fragen sollen...

- ...anwendungsorientiert sein und das Wissen der Studierenden in der Anwendung auf der Basis einer unbekanntem Problemstellung messen. Dadurch soll primär auf

den kognitiven Stufen K₃+ gemessen werden (vgl. Kapitel 2.5 "K₃+ als Zielsetzung für die Grundlagenausbildung").

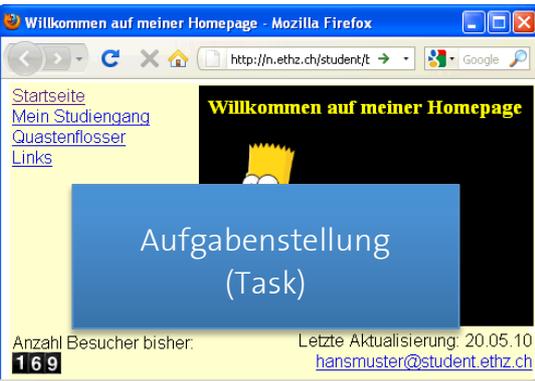
- ...automatisch korrigierbar sein, um einerseits die Zuverlässigkeit der Korrekturen zu erhöhen und die Korrekturzeit zu reduzieren. Andererseits sollen mit dem CBA-System auch formative LK möglich sein, welche idealerweise ein sofortiges *Feedback* durch eine automatisierte Korrektur unterstützen (vgl. Kapitel 3.4 "Leistungskontrolle als Feedbackinstrument").
- ...die Leistungen effizient messen, da die vorgegebene Zeit der summativen Schlussprüfung nur 60 Minuten beträgt.
- ...die Leistungen sollen valide und reliabel sein, damit die Studierenden zeigen können, was sie gelernt haben.
- ...die Prüfungsfragen sollten im Falle eines technischen Defekts auch auf Papier beantwortet werden können.
- ...von Assistierenden erstellt werden können.
- ...die Möglichkeit bieten, dann von ihnen Varianten erstellt werden, wodurch die Erstellungszeit geringer ausfällt, als wenn sie komplett neu erstellt werden müsste.
- ...soweit möglich auf bewährten Elementen basieren.

7.3.2 Grundstruktur des Fragetyps

Ausgehend von obigen Anforderungen wurden folgende Grundstruktur für automatisch korrigierbare Fragetypen entwickelt und in den letzten Jahren im operativen Einsatz optimiert (Abbildung 7.63).

Jede Aufgabe besteht aus einer Aufgabenstellung und meist vier dazugehörigen Fragen. Jede Frage nimmt ausschliesslich Bezug auf die Aufgabenstellung und kann somit unabhängig von den anderen Fragen beantwortet werden.

Aufgabe 1
Hier abgebildet ist die **Browseransicht** einer Studierenden-Homepage.



Willkommen auf meiner Homepage - Mozilla Firefox
http://jn.ethz.ch/student/t

Startseite
Mein Studiengang
Quastenflosser
Links

Willkommen auf meiner Homepage

Aufgabenstellung
(Task)

Anzahl Besucher bisher: **169**
Letzte Aktualisierung: 20.05.10
hansmuster@student.ethz.ch

Ausschnitt aus deren **HTML-Quellcode**:

```

2 <html><head>
3 <title>Willkommen auf meiner Homepage</title>
4 </head><body>
5 <table
6 style="height: 100%; font-family:Arial; background-color:#FFC"
7 border="0" cellpadding="2" cellspacing="0">
8 <tr valign="top">
9 <td><p><a href="Start.html" target="_blank">Startseite</a><br>
10 <a target="_blank" href="Studgang.html">Studiengang</a><br>
11 <a target="_self" href="Quasti.html">Quastenflosser</a><br>
12 <a target="_self" href="links.html">Links</a><br></td>
13 <iframe name="platzhalter" src="Start.html" frameborder="0"
14 height="100%" width="100%"></iframe><br>
15 </td></tr>
16 <tr><td>Anzahl Besucher bisher:&nbsp;&nbsp;&nbsp;<br>
18 <br><br></td>
19 <td align="right">Letzte Aktualisierung: 20.05.10<br>
20 <a href="mailto:hans@muster.ch">hansmuster@student.ethz.ch</a>
21 </td></tr>
22 </table>

```

Multiple Choice Item: 1.0 Punkte
 Beim **erstmaligen Öffnen der Homepage** wird im iFrame die Seite links.html angezeigt.

Lückentext Item: 1.75 Punkte
Sie möchten den HTML-Quellcode so verändern, dass nach dem Klick auf den Link **Quastenflosser** die neue Seite **im iFrame geöffnet** wird. Wie ändern Sie den Code? Ergänzen Sie

Quastenflosser

Multiple Choice Item: 1.0 Punkte
 Nach einem Klick auf den Link "Quastenflosser" wird die Seite links.html im iFrame geladen.
 Sie eine Nachkommastelle aufzurunden.

HotSpot Item: 1.25 Punkte
Die Zelle 3 in der 2. Spalte des obigen HTML-Quellcodes enthält den Titel "Welcome to my homepage". Markieren Sie in der **Browseransicht**, wo sich die Änderung auswirkt.

(Hinweis: Klicken Sie dazu auf das Markierkreuz-Symbol direkt links neben diesem Text und platzieren Sie es im Task an der gewünschten Stelle durch Klicken mit der Maus. Zum **Löschen** einer Markierung klicken Sie auf das rote Symbol links und dann auf die Markierung.)

Fragen
(Items)

Abbildung 7.63: Grundstruktur einer Aufgabe. Links die Problemstellung mit praxisrelevanten Daten und rechts die vier voneinander unabhängigen Fragestellungen. Weitere Informationen und die Beschreibung der drei Typen von Fragestellungen befinden sich im Text.

7.3.3 Aufgabenstellung (Task)

Die Aufgabenstellung besteht aus einem kurzen, prägnanten Einleitungstext und auf einem realen Datensatz.

Einleitungstext

Jede Aufgabenstellung beginnt mit einem prägnant formulierten Einleitungstext. Was ist gegeben? Worum geht es in der Aufgabenstellung? Hier steht die Information, die für alle Items gilt. Auf unnötige Informationen wird verzichtet. Die Einleitung muss selbsterklärend sein, das heisst einen Hinweis auf die Unterlagen der Vorlesung soll vermieden werden. Wichtige Hinweise zur Bearbeitung dieser Aufgabe werden mit "Hinweis:" klar gekennzeichnet und vom Einführungstext visuell abgetrennt.

Realer Datensatz

Auf den Einleitungstext folgt die Präsentation eines realen Datensatzes. Reale Daten helfen, die Aufgabenstellung rasch zu erfassen, was die Einarbeitungszeit reduziert. Der Datensatz stellt das zentrale Element einer Aufgabe dar, indem alle Items darauf

Bezug nehmen. Der Umfang sollte entsprechend so gewählt werden, dass die Aufgabenstellung mehrere Konzepte beinhaltet, die als Basis für die Formulierung der Items dienen. Falls Programmcode verwendet wird, sollte sich dieser kompilieren lassen oder mit einem entsprechenden Hinweise versehen werden.

Alle Abbildungen, Tabellen und Diagramme sind klar beschriftet und gegebenenfalls nummeriert. Nummerierte Bereiche können in den Items referenziert werden. Die Ausschnitte der Daten sind so gewählt, dass ein Scrollen im Aufgabentext nicht nötig ist. Dadurch erhöht sich die Lesbarkeit der Aufgabenstellung.

Bei der Darstellung sollte darauf geachtet werden, dass die verwendeten *Printscreens* gut lesbar sind (Auflösung, Schriftgröße). Die *Printscreens* sollten unskaliert aus den Rohdaten übernommen werden. Dazu ist es eventuell notwendig, die Rohdaten so zu formatieren und die Ausschnitte so zu wählen, dass sie den zu Verfügung stehenden Platz optimal ausnutzen und kein Scrollen nötig ist. Bei Farbcodierungen muss auf Menschen mit einer rot-/grün-Sehschwäche Rücksicht genommen werden. Zusätzlich sollten die Farbcodierungen noch mit Pfeilen oder Nummern ergänzt werden, damit die Datensätze im Notfall auch auf der *Backup*-Papier-Prüfung klar gekennzeichnet sind.

7.3.5 Fragetypen (Items)

Die Fragen auf der rechten Bildschirmseite nehmen direkten Bezug zur Aufgabenstellung. Sie sind voneinander unabhängig und können in beliebiger Reihenfolge beantwortet werden. Die gegenseitige Unabhängigkeit ist wichtig, damit keine Folgefehler auftreten. Ist die Unabhängigkeit nicht gegeben, kann das dazu führen, dass in je nach Bewertungsmodells die Folgefehler nicht gewertet werden können und die maximal mögliche Gesamtpunktzahl reduziert werden muss.

Pro Item kann eine bestimmte Punktzahl erreicht werden. Es gibt nur die volle oder keine Punkte. Teilpunkte für teilweise richtige Antworten werden keine vergeben. Die Punktzahl leitet sich aus der geschätzten Bearbeitungszeit des Items ab. Pro Minute Lösungszeit wird ein Punkt vergeben (d.h. 1.5 Minuten Lösungszeit = 1.5 Punkte).

Die Lösungszeit hängt nicht nur von der Schwierigkeit der Aufgabe, sondern auch vom Fragetypen ab. In unseren Aufgaben werden drei Fragetypen verwendet. Nachfolgende Tabelle gibt eine Übersicht über deren Eigenschaften (Tabelle 7.18).

Fragetyp	true/false	hotspot	cloze
Antwortformat	Checkbox anklicken	Markierung/-en setzen	Text
Antwortmöglichkeiten	2	viele	unbegrenzt
Komplexität der Musterlösung	sehr einfach	einfach	einfach - komplex
Beantwortungsaufwand	gering	mittel	hoch
Korrekturaufwand	sehr gering	gering	gering - hoch
Maximale Anzahl dieses Fragetyps pro Aufgabe	unbegrenzt	1	unbegrenzt
Ideale Einsatzschwierigkeit	$P > 70$	$P < 80$	$P < 50$

Tabelle 7.18: Eigenschaften der drei Fragetypen. Weitere Informationen finden sich im Text im Abschnitt des jeweiligen Fragetyps.

Nachfolgend werden die drei Fragetypen beschrieben und mit je einem Beispiel aus unserem Unterricht illustriert.

Fragetyp: true/false

Sie werden oft als *single-choice* Item bezeichnet und besitzen eine binäre Lösungsmenge (richtig/falsch). Die Wahrscheinlichkeit, die richtige Lösung zu erraten, liegt bei 50%. Sie sollen deshalb mehrheitlich bei leichten Items eingesetzt werden. Bei schwierigen Items nimmt die Wahrscheinlichkeit die richtige Lösung zu erraten zu, was automatisch die Trennschärfe reduziert.

Wann sollen *true/false*-Items verwendet werden?

- Einfache Prüfungsfrage für den Prüfungsanfang (Eisbrecher-Items)
- Bei Items mit klaren, binären Lösungsmengen
- Wenn die Lösungszeit reduziert werden soll

Beispiel eines Items des Typs "*true/false*" (Abbildung 7.64):

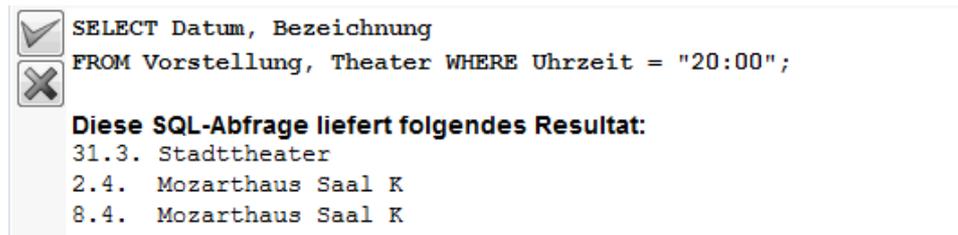


Abbildung 7.64: Beispiel eines Frage-Items des Typs "true/false". Zur Beantwortung wird die Aufgabenstellung benötigt (hier nicht dargestellt). *True/false*-Items können entweder richtig oder falsch sein. Die Antwort wird durch Anklicken des entsprechenden Buttons auf der linken Seite übermittelt. Da der Ratefaktor bei *true/false*-Items mit 50% hoch ist, eignen sie sich besonders für einfache Fragen. (Item wurde eingesetzt in der formativen LK, Modul 5 "Daten verwalten II", Gdl und EvIM, HS 2013)

Fragetyp: hotspot

Der Item-typ "hotspot" gehört wie der Typ "cloze" zu den *offenen Fragetypen*. Die offenen Fragetypen geben keine fixe Lösungsmenge vor. Die Wahrscheinlichkeit, die richtige Lösung zu erraten, wird dadurch reduziert. Die Beantwortungszeit ist länger als bei *true/false* Items. Die offenen Fragetypen können nur teilautomatisch korrigiert werden, das heisst, es müssen alle falschen Antworten nachträglich manuell kontrolliert werden. Eine gute Musterlösung in Kombination mit einer prägnanten Problemstellung reduziert den Kontrollaufwand beträchtlich.

Die *hotspot*-items sind Markier-Aufgaben. Die Antworten dieses Item-Typen werden übermittelt, indem in der Aufgabenstellung (Task) eines oder mehrere Markierkreuze positioniert werden. So kann beispielsweise in einer Visualisierung nach einem einzelnen Datenpunkt gefragt werden, oder die Probanden müssen in einem Quelltext mehrere Stellen markieren. Die Musterlösung wird in Form von UND/XOR verknüpften Bereichen festgehalten.

Das Item wird dann als korrekt bewertet, wenn der Proband die Markierkreuze an den korrekten Stellen in der Aufgabenstellung setzt.

Beispiel eines Items des Typs "hotspot" (Abbildung 7.65):



Abbildung 7.65: Beispiel eines Frage-Items des Typs "hotspot". Zur Beantwortung wird das Markierkreuz an der korrekten Stelle in der Aufgabenstellung positioniert (hier nicht dargestellt). Der Ratefaktor ist tiefer als bei *true/false*-Items, allerdings eignet sich dieser Typ nur für eine begrenzte Anzahl Fragestellungen. Er lässt sich relativ einfach korrigieren. (Item wurde eingesetzt in der formativen LK, Modul 5 "Daten verwalten II", EvIM, HS 2013)

Fragetyp: cloze

Bei *cloze*-Items werden die Antworten in Form von kurzen Texten übermittelt. Ein *cloze*-Item kann aus einer beliebigen Anzahl Textelementen bestehen. Die Spannbreite reicht von einfachen, numerischen Antworten ("Wie viele Zeilen liefert der Spezialfilter?") bis hin zu komplexeren Programmcode-Ergänzungen ("Ergänzen Sie die Schleifenbedingung so, dass die Bildschirmausgabe sich nicht verändert").

Je nach Fragestellung kann das Ausarbeiten der Musterlösung ziemlich aufwändig sein. Deshalb sollte durch geschickt vorgegebene Teillösungen die Anzahl der Freiheitsgrade bei der Beantwortung limitiert werden. Pro Lückentext kann eine beliebige Anzahl an Musterlösungen angegeben werden. Bei der Korrektur müssen alle falschen Antworten manuell nachkontrolliert werden. Das CBA-System sollte die Möglichkeit bieten, allfällig korrekte Antworten in die Musterlösung zu übernehmen und das Item für alle Probanden neu zu korrigieren.

Beispiel eines Items des Typs "*cloze*" (Abbildung 7.66):



Sie führen das Programm **MUSTER** aus. Welcher **Wert** ist in der **Variable :X** gespeichert, wenn das Programm die **Zeile 20** erreicht hat?

In Zeile 20, die Variable :X enthält folgenden Wert:

Abbildung 7.66: Beispiel eines Frage-Items des Typs "*cloze*". Die Antwort wird durch Eintippen übermittelt. Es können mehrere Lücken verwendet werden. Zur Beantwortung dieser Frage wird die Aufgabenstellung benötigt (hier nicht dargestellt). Der Ratefaktor ist tiefer als bei den anderen Fragetypen. Allerdings benötigt das Eintippen und Kontrollieren der Antwort mehr Zeit, was die Bearbeitungszeit dieses Item-Typs erhöht. Es empfiehlt sich deshalb speziell bei schwierigeren Fragen, wo es Sinn macht, die Ratewahrscheinlichkeit auf Kosten einer höheren Bearbeitungszeit zu reduzieren. Das Item ist vergleichsweise aufwändig zu korrigieren, da offensichtliche Tippfehler trotzdem als korrekt gewertet werden sollten.

Weitere Informationen bezüglich der Erstellung von qualitativ hochwertigen CBA-Fragen wurden in einer Broschüre zusammengefasst (Dahinden, 2010a). Diese kann den Assistierenden als Grundlage für die Erstellung von Prüfungsfragen dienen.

7.4 Evaluation

Im zweiten Teil dieses Kapitels soll die Qualität der Prüfungsfragen anhand des im ersten Teil beschriebenen Entwicklungsmodells evaluiert werden. Als Datenbasis dient die Lehrveranstaltung "Einsatz von Informatikmitteln" (252-0839-00, 2 ECTS), welche im Herbstsemester 2013 von 299 Studierenden besucht wurde. Diese LV richtet sich an

Studierende der Naturwissenschaften und vermittelt ihnen die Grundlagen für die Arbeit mit wissenschaftlichen Daten²⁴. Entsprechend ausgewählt sind deshalb die sechs Module:

- 1) Publizieren im Internet
- 2) Simulieren
- 3) Visualisierung mehrdimensionaler Daten
- 4) Datenverwaltung mit Listen und Tabellen
- 5) Datenverwaltung mit einer relationalen Datenbank
- 6) Einführung in die Makroprogrammierung

Die Bearbeitungszeit eines Moduls beträgt rund 6-8 Stunden. Jedes Modul ist gemäss des 4-Schritte-Modells aufgebaut und wird mit der Präsentation der selbständigen Arbeit abgeschlossen (L. E. Fässler, 2007). Unmittelbar im Anschluss an die Präsentation erhält der Student ein individuelles *Feedback* von einer Assistenzperson. Für den Erhalt der Kreditpunkte müssen die Studierenden alle sechs Module präsentieren und die Semester-Endprüfung mit einer genügenden Note bestehen. Abbildung 7.67 zeigt die Semesterübersicht dieser LV.

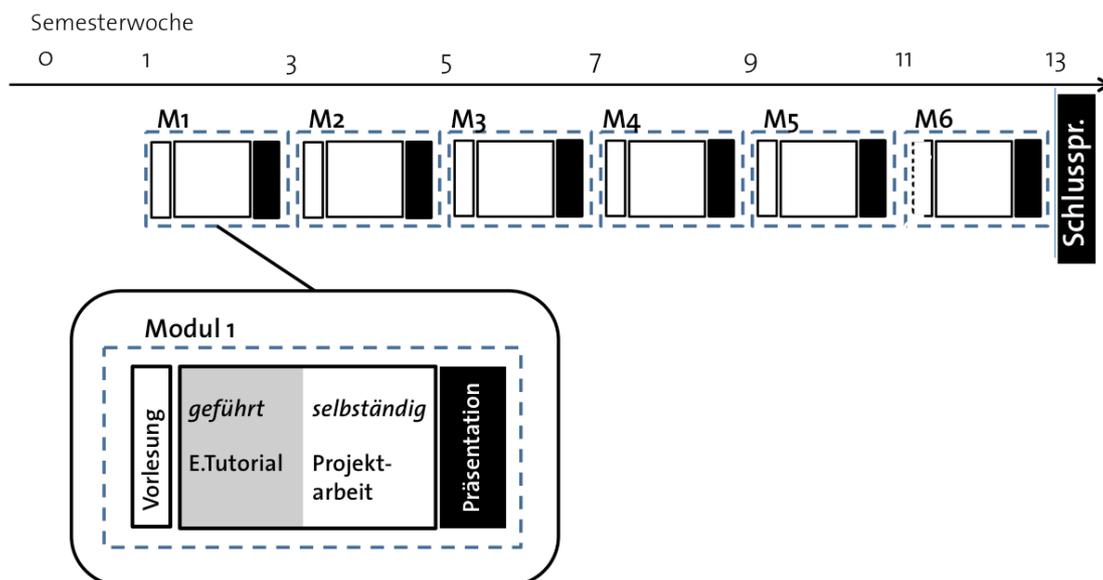


Abbildung 7.67: Semesterübersicht der LV "Einsatz von Informatikmitteln". Um die zwei Kreditpunkte zu erhalten, müssen alle 6 Module bearbeitet und die Semester-Schlussprüfung bestanden werden.

²⁴ Homepage der Lehrveranstaltung: <http://www.evim.ethz.ch> (abgerufen am: 23.05.2014)

In den nachfolgenden vier Abschnitten werden nun die einzelnen Phasen des CBA-Entwicklungsmodells auf der Basis dieser LV durchlaufen. Der Fokus liegt auf der Validierung der summativen LK und wird im Abschnitt 7.4.4 "Auswertung" erläutert.

7.4.1 Planung

Mit der summativen LK der LV "Einsatz von Informatikmitteln" soll gemessen werden, ob die Studierenden die Lernziele erreicht haben oder nicht. Entsprechend zuverlässige Resultate werden deshalb im Bereich der Bestehensgrenze erwartet. Die Prüfungsdauer beträgt gemäss Vorlesungsverzeichnis 60 Minuten und findet am Computer statt. Der LV wird explizit keine Selektionsrolle zugesprochen, das heisst, die Serviceempfangenden Departemente haben sich dagegen ausgesprochen, dass diese LV die Rolle eines "Killer-Fachs" übernehmen soll. Entsprechend können Studierende, welche die Bestimmungen zur Erhaltung der Kreditpunkte nicht erfüllen, die fehlenden Kreditpunkte durch den Besuch einer alternativen Informatik-LV kompensieren.

Die Risikoanalyse für CBA-LK an der ETH Zürich wurde im Kapitel 4 durchgeführt und die Gegenmassnahmen evaluiert.

Als Aussenkriterien sollen die durch die Assistierenden durchgeführten Bewertungen der Präsentationen verwendet werden. Diese Messungen stellen, im Vergleich zur summativen LK, ein alternatives Messverfahren für die individuellen Leistungen der Studierenden dar. Aufgrund der kurzen Dauer der Präsentationen von 5-7 Minuten soll eine reduzierte drei-stufige Lickert-Skala (0 = Lernziele nicht erfüllt, 1 = Lernziele erfüllt und 2 = Lernziele übertroffen) zum Einsatz kommen. Die Assistierenden werden im Rahmen einer Informationsveranstaltung bezüglich dieses Bewertungssystems instruiert. Um den formativen Charakter der Präsentationsbewertungen zu erhalten, haben diese keinen Einfluss auf die Schlussnote der Studierenden. Die Studierenden werden diesbezüglich in der ersten Vorlesung und auf der *Homepage* informiert.

7.4.2 Erstellung

Für die Zusammenstellung wurden neu erstellte Fragen mit bewährten Ankerfragen aus dem Pool gemischt. Die Auswertung des *Blueprints* ergibt folgendes Bild (Tabelle 7.19).

	Kognitive Prozessdimension						Umfang an LK (in %)
	K1	K2	K3	K4	K5	K6	
M1		1	4	3			20%
M2			2	2			10%
M3		2	2	4			20%
M4			2	2			10%
M5			5	3			20%
M6				4			10%
Vorlesung		1	1	2			10%
Total (in %)	0%	10%	40%	50%	0%	0%	100%

Tabelle 7.19: *Blueprint* der summativen LK der LV "Einsatz von Informatikmitteln" im Herbstsemester 2013. Die Prüfungszeit beträgt 60 Minuten. Die Prüfung beinhaltet 40 Fragen, welche sich auf zehn unterschiedliche Aufgabenstellungen beziehen. Der Schwerpunkt der LK liegt auf dem Inhalt der sechs Module. Eine Aufgabenstellung bezieht sich auf ausgewählte Inhalte der Vorlesung.

Die unterschiedliche Gewichtung der Module in der Prüfung erklärt sich mit der Verwandtschaft der Module: Modul 2 und 4 bauen aufeinander auf und sind somit inhaltlich ähnlich. Die Inhalte der Vorlesung nehmen teilweise auch Bezug zu den Modulen und werden dort indirekt mitgeprüft.

7.4.3 Durchführung

Die Durchführung der summativen, *high-stake* LK wurde bereits ausführlich in einem technischen Report beschrieben (Dahinden & Hinterberger, 2010).

Bei der Semester-Endprüfung kam es zu keinen nennenswerten Zwischenfällen und die Studierenden waren zufrieden mit der Organisation (Abbildung 7.68). Es kann davon ausgegangen werden, dass alle Studierenden vergleichbare Prüfungsbedingungen hatten. Alle Unterlagen wurden redundant gespeichert und die sensiblen Daten mittels digitaler Signaturen vor Veränderungen geschützt.

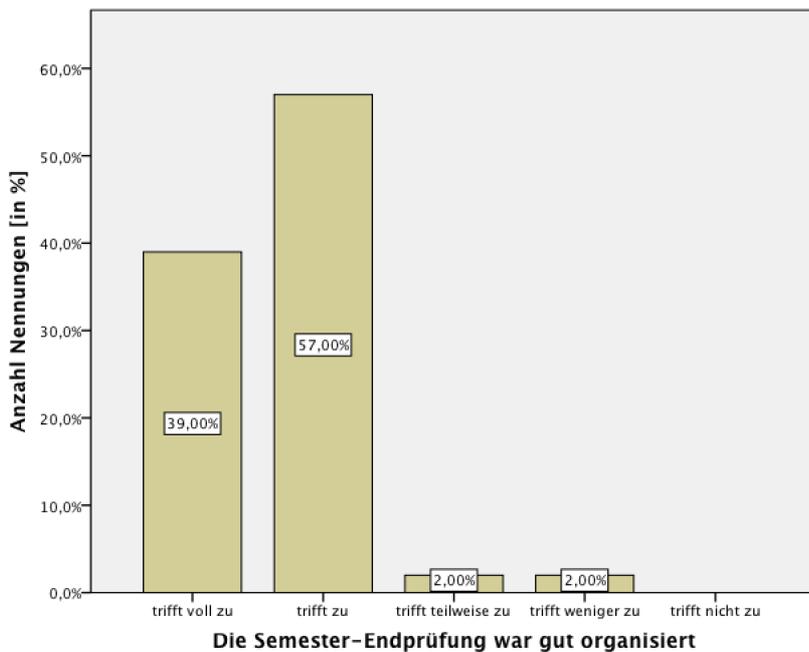


Abbildung 7.68: Umfrage über die Organisation der Semester-Endprüfung. 96% der Studierenden gaben an, dass die Prüfung gut organisiert war. (Datenbasis: Nachumfrage unter den Studierenden, EvIM, HS13, n=100)

Folgende Daten standen zur Analyse der Prüfungsergebnisse zur Verfügung:

- Prüfungsheft mit 40 Fragen und den dazugehörigen zehn Aufgabenstellungen
- Die formalisierten Musterlösungen
- Die individuellen Antworten der Studierenden pro Frage
- Die Bewertung der Präsentationen durch die Assistenten

7.4.4 Auswertung

Die Korrektur wurde im 4-Augen-Prinzip mit dem Sioux *AuthorClient* durchgeführt. Dabei wurden für jede Aufgabe erst die Kommentare gelesen und dann die falschen Antworten angeschaut. Erkannte der Client einen Fehler in der Musterlösung oder einen offensichtlichen Tippfehler in einer Antwort, wurde die Musterlösung entsprechend ergänzt. Anschliessend wurde diese Aufgabe für alle Studierenden neu korrigiert. Die Musterlösung wird im Fragenpool gespeichert und kann später im Falle einer legalen Überprüfung des Korrekturvorgangs erneut abgefragt werden.

Die korrigierten Resultate der Studierenden wurden dann exportiert und in MS Excel und SPSS weiter ausgewertet.

In einem ersten Schritt wurde die Reliabilität der Items mittels der Analyse der Itemmetriken durchgeführt. Dazu wurde eine Fehlerverteilungsmatrix erstellt und die

Trennschärfekoeffizienten und Schwierigkeitsindizes der einzelnen Items berechnet (Abbildung 7.69).

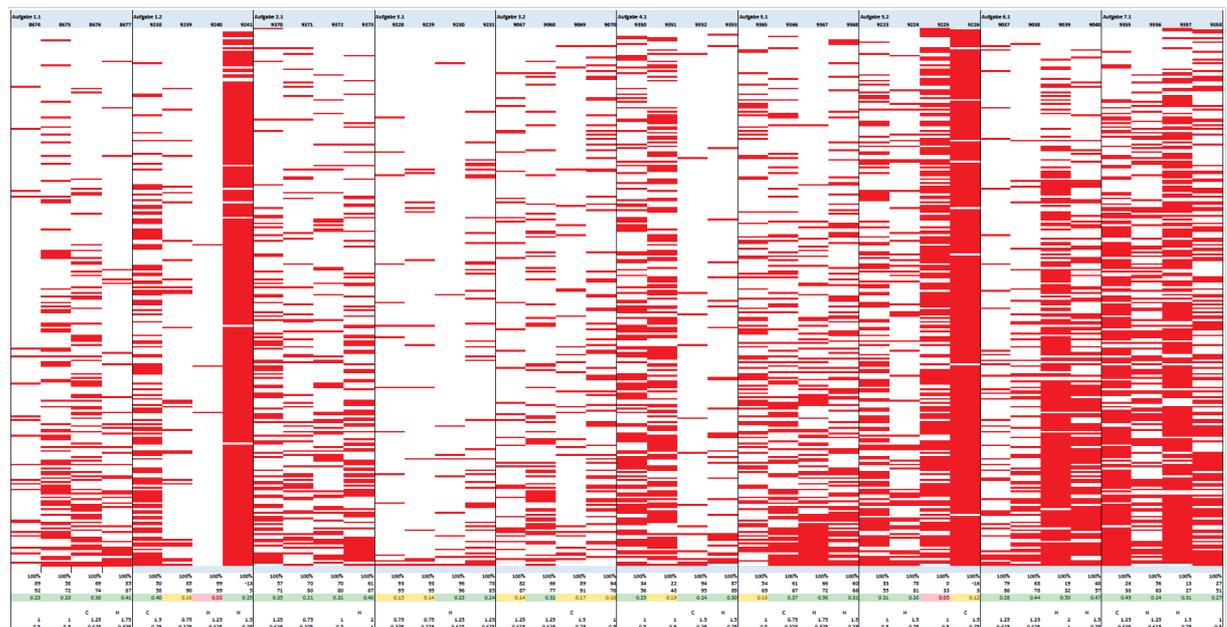


Abbildung 7.69: Fehlerverteilungsmatrix der Rohdaten der Semester-Endprüfung. In jeder Zeile sind die Resultate eines Probanden dargestellt. Weiße und rote Bereiche widerspiegeln korrekte respektive falsche Antworten. Die Probanden sind absteigend nach der erreichten Gesamtpunktzahl sortiert. Zuerst stehend demnach die besten, zuunterst die schwächsten Studierenden. Das vertikale Fehlermuster gibt Auskunft über die Qualität eines Items. Die Itemmetriken sind unterhalb der Fehlerverteilungsmatrix aufgeführt (eine vergrößerte Darstellung dieser Fehlerverteilungsmatrix findet sich im Anhang A.5 dieser Arbeit). Die neunte Frage von rechts musste aufgrund einer mehrdeutigen Formulierung eliminiert werden. (Datenbasis: Semester-Endprüfung, EvIM, HS13, $n=279$)

Die Analyse der Itemmetriken zeigte 2 nicht trennscharfe, 9 schwach trennscharfe und 29 trennscharfe Fragen (Abbildung 7.70). Alle auffälligen Items wurden nachkontrolliert. Dabei wurde festgestellt, dass ein schwach trennscharfes Item aufgrund einer mehrdeutigen Fragenformulierung von der Bewertung ausgeschlossen werden musste. Die beiden nicht trennscharfen Items waren hingegen einwandfrei formuliert.

Die Berechnung der internen Konsistenz der Semester-Endprüfung nach dem Alpha Modell gemäss Cronbach lieferten einen Wert von 0.702. Dieser liegt - wenn auch knapp - innerhalb des in der Literatur empfohlenen Zielbereichs von 0.70 - 0.95. Der Grund für dieses eher knappe Messresultat lässt sich vermutlich auf die Heterogenität der Module zurückführen. Diese bauen nicht strikt aufeinander auf, sondern decken verschiedene Konzepte eines digitalen Datenlabors ab. Entsprechend haben die schwachen Studierenden bei jedem Modul abermals die Chance, dieses erfolgreich zu bearbeiten. In strikt aufbauenden LV, wie beispielsweise bei der Programmier Einführung, dürfte Cronbach Alpha deshalb deutlich höher ausfallen.

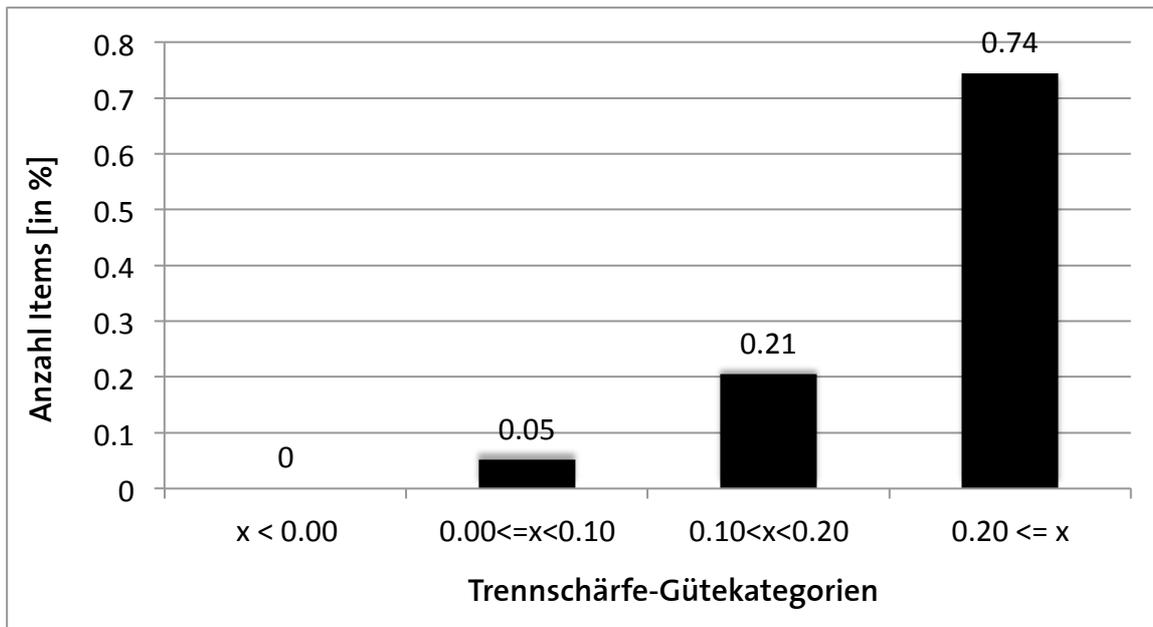


Abbildung 7.70: Anzahl der Items pro Trennschärfe-Kategorie. Positiv ist die klare Häufung der Items der höchsten Gütekategorie. Dieses Resultat lässt eine reliable Messung vermuten. (Datenbasis: Semester-Endprüfung, EvIM, HS13, n=39)

Für die Abschätzung der Validität haben wir die Bewertungen der Präsentationen als Aussenkriterien mit den individuellen Gesamtpunktzahlen korreliert. Wie das Resultat in Tabelle 7.20 zeigt, korrelieren alle sechs Aussenkriterien mit dem Ergebnis der Semester-Endprüfung. Das bedeutet, dass die Assistenten die Leistungen der Studierenden sehr ähnlich beurteilt haben, wie dies die Prüfung getan hat. Das Ergebnis der summativen LK kann also als valide bezeichnet werden.

	fAs_m1	fAs_m2	fAs_m3	fAs_m4	fAs_m5	fAs_m6
e_score	0.248**	0.205**	0.232**	0.190**	0.233**	0.223**
Sig. (2-tailed)	0.000	0.001	0.000	0.001	0.000	0.000
N	278	278	278	278	278	278

** Correlation is significant at the 0.01 level (2-tailed).

* Correlation is significant at the 0.05 level (2-tailed).

Tabelle 7.20: Abschätzung der Validität des Schlussresultats. Die Tabelle zeigt die biseriale Korrelation der Assistentenbeurteilungen mit der Schlusspunktzahl der Studierenden. Abkürzungen fAs_mX = formatives Assessment Modul X (Datenbasis: Bewertungen der Präsentationen durch die Assistenten, EvIM, HS13)

Dieses Ergebnis wird auch durch die Antworten der Studierenden an der Schlussumfrage gestützt (Abbildung 7.71). Die Studierenden scheinen also die statistisch ermittelte Validität auch subjektiv festzustellen.

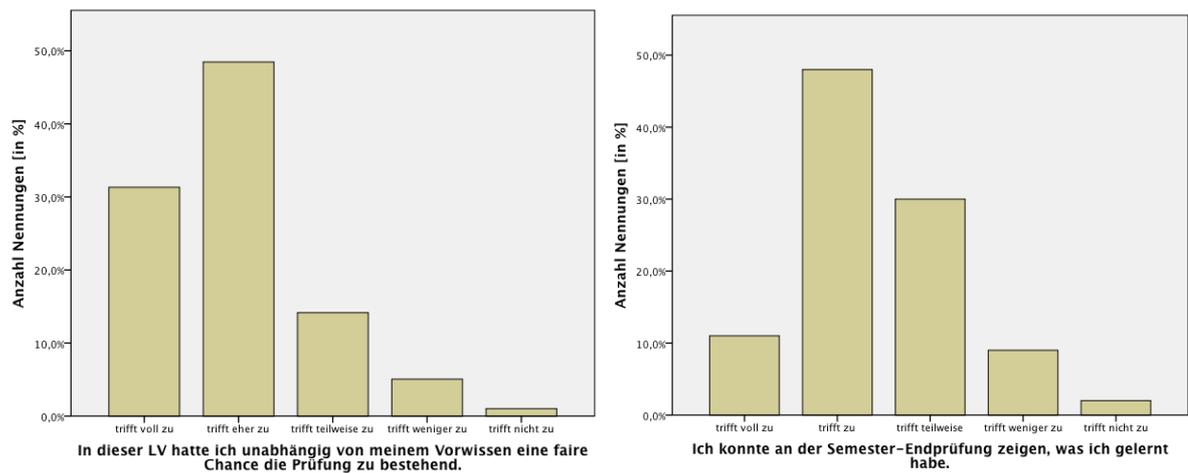


Abbildung 7.71: Umfrage über die Validität der Semester-Endprüfung. Die Resultate lassen vermuten, dass offenbar auch die von den Studierenden subjektiv wahrgenommene Validität die empirisch ermittelte Validität der LK der Semester-Endprüfung stützt. (Datenbasis: Semester-Endprüfung, EvIM, HS13, n=100)

7.4.5 Qualitätszunahme des Fragenpools durch eine wiederholte Anwendung des Entwicklungsmodells

Im bisherigen Verlauf der Evaluation fokussierten sich die Untersuchungen primär auf die Analyse der Items. In diesem abschliessenden Abschnitt soll anhand der Prüfungsdaten vergangener Jahre aufgezeigt werden, dass eine wiederholte Anwendung des Entwicklungsmodells zu einer generellen Verbesserung der Prüfungsreliabilität führt.

Die Analyse der Itemmetriken und das Ausarbeiten von Überarbeitungsvorschlägen für Prüfungsfragen stellt die Grundlage für den übergeordneten Entwicklungsprozess dar. Dieser Arbeitsschritt verläuft gemäss des in Abbildung 7.62 illustrierten Entscheidungsprozesses, mit dem jedes Item klassiert und Überarbeitungsvorschläge ausgearbeitet werden können.

Abbildung 7.72 zeigt die Situation der Itemmetriken am Beispiel des vorgängig erwähnten Semester-Endtests. Jeder Datenpunkt im Diagramm liefert drei Informationen: Trennschärfe, Schwierigkeit und Typ des Items. So lässt sich beispielsweise erkennen, dass die höchsten Trennschärfen mit offenen Fragetypen (*cloze* und *hotspots*) erreicht und dass vermehrt einfachere Fragen eingesetzt wurden. Letzteres begründet sich damit, dass die LK besonders im Bereich der Bestehensgrenze zuverlässig messen soll.

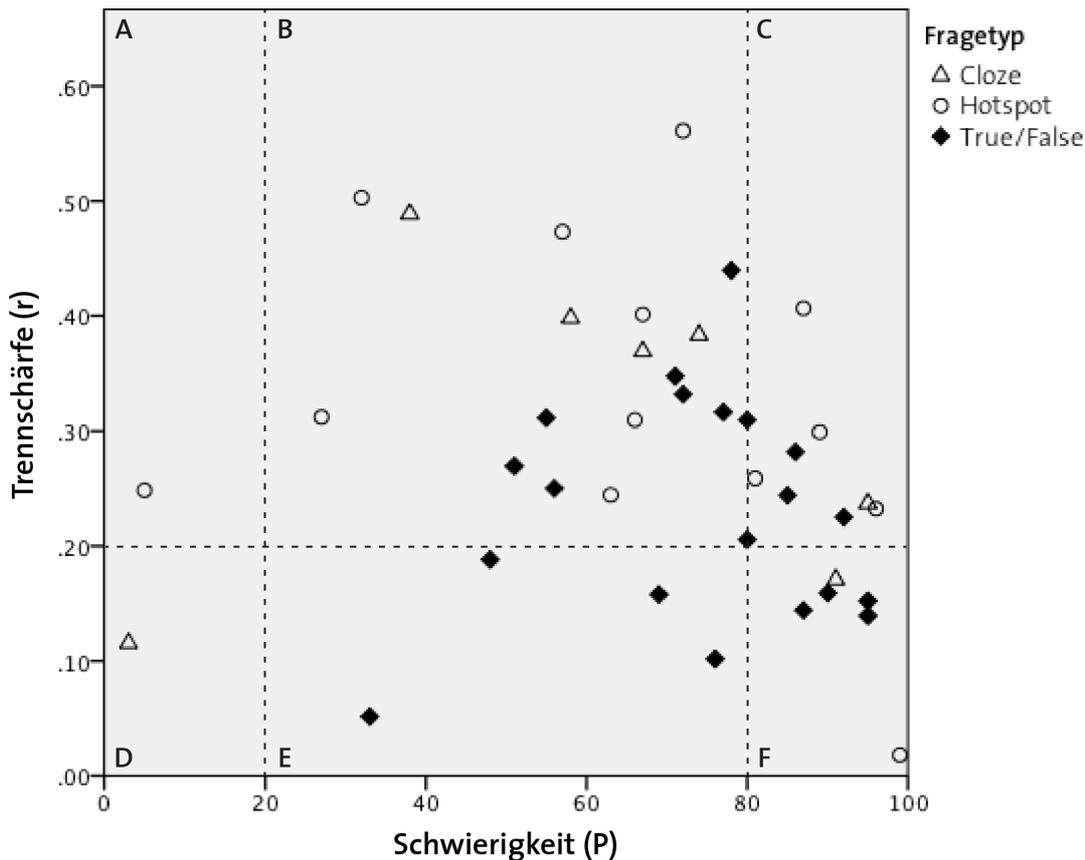


Abbildung 7.72: Verhältnis von Trennschärfe und Schwierigkeit. Sichtbar ist eine Rechtsverschiebung der Schwierigkeitsindizes aufgrund des Messschwerpunkts rund um die Bestehensgrenze (Datenbasis: EvIM, HS13, 40 Items)

Items im Bereich A sind schwierig und trennscharf. Dieser Bereich wird fast ausschliesslich mit *cloze*- oder *hotspot*-Items erreicht, da die Ratewahrscheinlichkeit die Trennschärfe reduziert. Die Items in diesem Bereich verlangen aus messtechnischen Gründen keine Überarbeitung.

Items im Bereich B sind trennscharfe Items und messen zuverlässig die Leistungen der grossen Mehrheit der Studierenden. Eine gute LK verfügt über die meisten Items in diesem Bereich, entsprechend ist eine Überarbeitung der Items aus messtechnischer Sicht nicht notwendig.

Items im Bereich C sind einfach, aber trennscharf. Sie eignen sich für eine zuverlässige Bestimmung der kognitiven Leistungen rund um die Bestehensgrenze²⁵. Eine Überarbeitung dieser Items ist aus messtechnischer Sicht nicht notwendig. Allerdings kann bei offenen Item-Typen ein Wechsel zu *true/false*-Items in Betracht gezogen werden, da die Ratewahrscheinlichkeit aufgrund der geringen Schwierigkeit dieser Items gering ist.

²⁵ Unter der Annahme, dass die Quote der ungenügenden Studierenden ca. 10-15% beträgt. Daraus lässt sich eine Zielschwierigkeit von ca. $P \approx 0.85 - 0.90$ ableiten. Bei anderen Quoten sollte die Zusammensetzung der Frageschwierigkeit entsprechend angepasst werden.

Durch diesen Wechsel wird im Allgemeinen die Antwortzeit verkürzt und ein zusätzliches Item kann der LK hinzugefügt werden.

Items im Bereich D sind sehr schwierig und wenig trennscharf. Möglicherweise ist die Musterlösung falsch oder das Item so schwer, dass keine höhere Trennschärfe erzielt werden kann. Eine Kontrolle dieser Items ist angebracht.

Items im Bereich E sind wenig trennscharf, obwohl die Schwierigkeit im Idealbereich liegt. Meist handelt es sich in diesem Bereich um *true/false*-Items, bei denen mit einem Typenwechsel zu *cloze*- oder *hotspot*-Items die Ratewahrscheinlichkeit reduziert werden kann. Dabei sollte bedacht werden, dass dieser Typenwechsel mit einer Erhöhung der Antwortzeit verbunden ist. Eventuell muss dies durch die Reduktion der Items in der LK oder durch einen Typenwechsel beispielsweise im den Bereichen C und F kompensiert werden.

Items im Bereich F sind klassische Eisbrecherfragen. Fragen in diesem Bereich machen nur Sinn, wenn sie nahe $p=100$ liegen und somit keine höhere Trennschärfe möglich ist. Die restlichen Fragen sollten überarbeitet oder zumindest mit einem Typenwechsel zu *True/False*-Fragetypen hinsichtlich der Reduktion der Antwortzeit optimiert werden.

Dieses Entwicklungsmodell wenden wir seit mehreren Jahren in unserem Unterricht an, wobei jeweils die Itemmetriken vergangener Prüfungen für die Optimierung der Items verwendet wird. Zudem merken wir, dass die Autoren grosses Interesse an den von ihnen erstellten Fragen und deren Einsätzen in der Praxis zeigen. Gerade die Fehlerverteilungsmatrix stösst bei der Rückbesprechung der Items auf grosses Interesse.

Abbildung 7.73 visualisiert den zeitlichen Verlauf der Entwicklung der Trennschärfen verschiedener Prüfungen. Es lässt sich erkennen, dass die Anzahl der auffälligen Items mit Trennschärfen < 0.20 im Verlauf der Jahre kontinuierlich gesunken ist, während die Anzahl der trennscharfen Items gesteigert werden konnte.

Diese positive Entwicklung lässt sich vermutlich auf verschiedene Ursachen zurückführen:

- Durch den Einbezug der Itemmetriken der letzten Verwendung konnten ausgewählte Fragen gezielt optimiert und so deren Qualität sukzessive erhöht werden.
- Durch den Erfahrungsgewinn der Autoren und *Reviewer* stieg die Qualität der neu erstellen Fragen.
- Die Weiterentwicklung der Unterrichtsmaterialien und des verbesserten *Alignments* von Unterricht und formativen Assessment-Elementen konnten die Studierenden individueller gefördert werden, was die Homogenität der Leistungen gefördert hat.

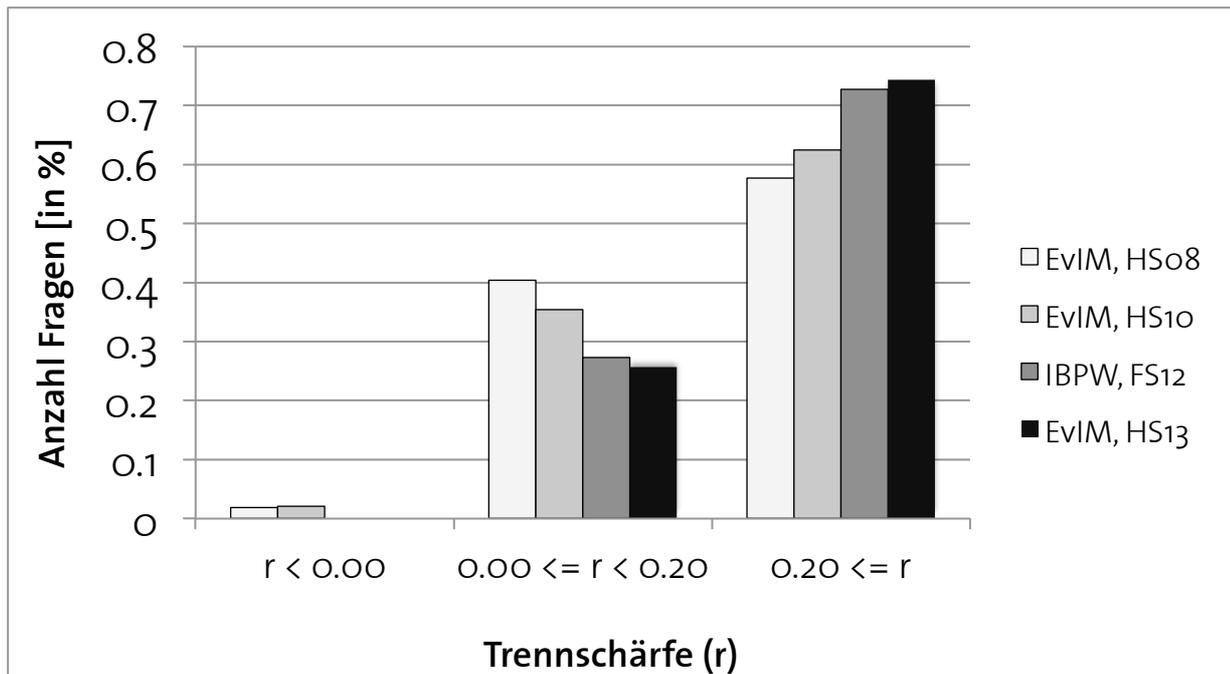


Abbildung 7.73: Zeitliche Entwicklung der Trennschärfen verschiedener summativen LK. Erkennbar ist eine Abnahme der trennschwächeren Items (mit $p < 0.20$) und einer gleichzeitigen Zunahme der trennscharfen Items. Dieser Verlauf ist ein Indiz einer Qualitätssteigerung bei der Reliabilität unserer Prüfungen im Verlauf der letzten Jahre. Möglicherweise wird sich nun der Anteil der trennscharfen Items bei ca. 75% einpendeln. (Datenbasis: wie in der Legende angegeben)

7.5 Diskussion der Fragestellung 3: Können mittels automatisiert korrigierbaren Prüfungsfragen die in der Grundlagenausbildung vermittelten Informatik-Kompetenzen zuverlässig gemessen werden?

Im aktuellen Kapitel wurden Fall-basierte Prüfungsfragen vorgestellt, die auf drei Typen von Fragen basieren. Diese lassen sich automatisiert korrigieren und eignen sich somit für den Einsatz in CBA-Systemen.

Die Fall-basierten Prüfungsfragen wurden gemäss des im aktuellen Kapitels vorgestellten Entwicklungsmodells in der Praxis evaluiert:

Die Objektivität der Leistungsmessung wird durch die automatisierte Korrektur des CBA-Systems erreicht, wodurch Korrekturfehler nahezu ausgeschlossen werden können.

Die Reliabilität der Fall-basierten Prüfungsfragen wurde durch die Analyse der internen Konsistenz der Resultate ermittelt. Dabei kamen mit der Bestimmung der Aufgabenschwierigkeit, der Trennschärfe und des Modells Alpha nach Cronbach drei bewährte Verfahren der Item-Analyse zum Einsatz. Diese Messungen ergaben, dass die Fall-basierten, automatisiert korrigierbaren Prüfungsfragen reliable Leistungsmessungen ermöglichen (vgl. Kapitel 7.4.4 "Auswertung").

Die Validität der Prüfungsfragen wurde anhand von sechs Aussenkriterien bestimmt. Bei diesen Aussenkriterien handelte es sich um sechs unabhängige, kriterienbezogene Leistungsmessungen, welche von Experten zu unterschiedlichen Zeitpunkten durchgeführt wurden. Die ermittelten sechs Validitätskoeffizienten zeigen einen signifikanten Zusammenhang zwischen den Aussenkriterien und den mittels Fall-basierter Fragen gemessenen Leistungen. Zusammen mit der erfolgreichen Prüfung der inhaltlichen Validität mittels eines *Blueprint* kann festgehalten werden, dass die in diesem Kapitel vorgestellten Fall-basierten Fragen eine valide Leistungsmessung von Informatik-Kompetenzen im Grundlagenunterricht ermöglichen.

Es wäre interessant zu untersuchen, inwiefern mit den Fall-basierten Fragen auch in anderen Fachgebieten valide CBA-Leistungsmessungen durchgeführt werden können. Erste Erfahrungen in den Fachbereichen "*Marketing*", "*Management and Accounting*" und "*Controlling*" verliefen durchwegs positiv. Es bedarf jedoch der Anwendung des in dieser Arbeit vorgestellten Entwicklungsmodells für CBA, um die Qualität dieser Leistungsmessungen bestimmen zu können.

Kapitel 8

Learning Analytics auf der Basis valider Leistungsdaten

8.1 Analyse lehrbezogener Daten zur Unterrichtsentwicklung

Das beste Vorgehen zur Evaluation des Unterrichts ist unbestritten der Gang in die Praxis. Durch das Beobachten der Studierenden bei der Arbeit mit den bereitgestellten Lernmaterialien, das Korrigieren von studentischen Arbeiten und das Einholen der Meinungen im persönlichen Kontakt mit den Studierenden können potentielle Verbesserungsmöglichkeiten des Unterrichts zuverlässig erkannt werden.

In modernen Unterrichtsformen mit zeit- und ortsunabhängigen Unterrichtssequenzen (*distant learning*) ist die direkte Evaluation nur eingeschränkt möglich. Die Präsenzdauer in dieser Unterrichtsform ist meist zugunsten des Selbststudiums reduziert und aufgrund der grossen Studierendenzahlen - in Einführungsveranstaltungen nehmen meist mehrere hundert Studierende teil - finden die Präsenzsequenzen je nach Platzverhältnissen in den Vorlesungs- und Unterrichtsräumen meist dezentral statt.

Um den Unterricht dennoch weiterentwickeln zu können, sind effiziente und effektive Massnahmen zur Unterrichtsevaluation gefragt. Traditionelle Verfahren wie Schlussumfragen können helfen, Fehler in den Unterrichtsmaterialien zu finden, organisatorische Ablaufstrukturen zu verbessern oder die Abstimmung zwischen Unterricht und Leistungskontrolle zu optimieren. Für Letztere ist es zudem hilfreich, den *Learning Outcome* mit validen Leistungsmessungen zu bestimmen und die Resultate mit dem im Kapitel 7 beschriebenen Vorgehen zu analysieren. So können grundlegende Probleme im Unterricht zuverlässig erkannt werden. Im Abschnitt 8.3 wird dieser Vorgang anhand eines Praxisbeispiels illustriert.

Die isolierte Analyse der Schlussumfrage und der Resultate der Leistungsmessungen hat jedoch gewisse Grenzen und beschränkt sich auf Zustandsbeschreibungen zu verschiedenen Zeitpunkten. So ist es beispielsweise nicht möglich, einen Entwicklungsprozess zu messen und zu analysieren. Um dies zu bewerkstelligen, ist es nötig, die zu verschiedenen Zeitpunkten und mit verschiedenen Messverfahren ermittelten lehrbezogenen Daten zu aggregieren.

In diesem Schlusskapitel soll das Potential dieses Verfahren anhand von Praxisbeispielen illustriert werden. Dabei geht es nicht darum, dieses Verfahren zu evaluieren oder quantitative Resultate zu liefern. Vielmehr geht es darum aufzuzeigen, wie durch die Aggregation und die Gruppierung lehrbezogener Daten die Lern- und motivationale Prozesse sichtbar gemacht und die Effizienz der Lehrbemühungen festgestellt werden können. Oder um es in den Worten von Tanya Elias auszudrücken:

"Learning is a product of interaction [...] with instructors and tutors, with content and/or with other people. Many educators expend enormous amounts of effort to designing their learning to maximize the value of those interactions. Regardless of the approach taken, a series of questions consistently arises: How effective is the course? Is it meeting the needs of the students? How can the needs of learners be better supported? What interactions are effective? How can they be further improved?" (Elias, 2011, p. p. 1).

Im ersten Teil des aktuellen Kapitels werden die Möglichkeiten und Grenzen aktueller Unterrichtsevaluationen anhand eines Praxisbeispiels illustriert. Anschliessend wird das Prinzip der Aggregation der lehrbezogenen Daten kurz erläutert. Die so aggregierten Daten werden als *educational Datawarehouse* (eDWH) in Anlehnung an den Begriff *Datawarehouse* aus dem Fachgebiet des Informationsmanagements bezeichnet (A. Meier, Dippold, Ringgenberg, Schnider, & Schwinn, 2005). Abbildung 8.74 zeigt die Schritte zum Aufbau und zur Analyse eines eDWH.

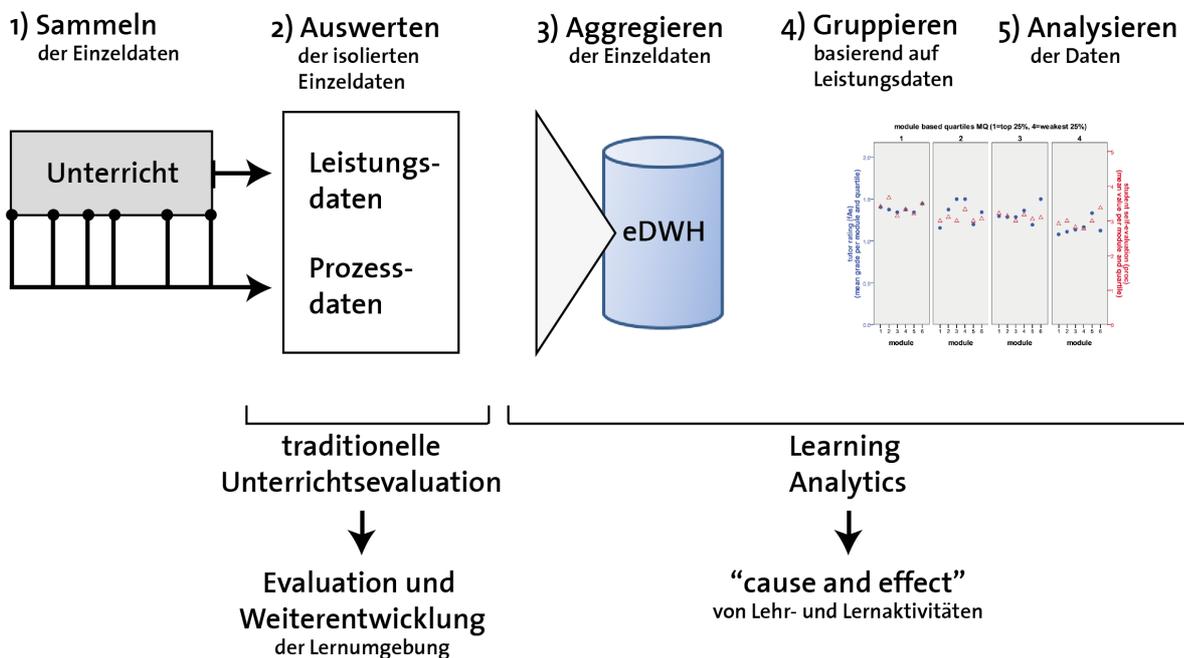


Abbildung 8.74: Die fünf Schritte der Datenanalyse zur Unterrichtsentwicklung auf der Basis von Leistungsdaten. Der vierte Schritt ist fakultativ und abhängig von der gewählten Analyse. Das im Kapitel 7 beschriebene Entwicklungsmodell dient als Grundlage für die Erhebung valider Leistungsdaten, welche als Basis für die Analyse der Daten im "educational Data Warehouse" (eDWH) dienen.

Auf dem eDWH aufbauend werden zwei Analyseansätze untersucht:

- 3) Es soll untersucht werden, welche Aussagen durch die retrospektive Analyse der Resultate der Probe- und der Schlussprüfung gemacht werden können.
- 4) Die Möglichkeiten der prospektiven Analyse lehrbezogener Daten auf der Basis von Leistungsdaten sollen untersucht werden. Dabei werden die Daten von Experten- und Selbstevaluation mit dem Resultat an der Schlussprüfung verglichen. Ziel ist, dass durch die Beobachtung des Prozessverlaufs mögliche Schwächen im Unterricht gefunden werden und individuelle Vorhersagen bezüglich der Lernentwicklung gemacht werden können.

Die in diesem Kapitel vorgestellten Daten stammen aus unserem Unterricht und wurden an zwei internationalen Konferenzen vorgestellt. Die Verweise zu den Publikationen sind nachfolgend jeweils direkt im Text wiedergegeben.

8.2 Möglichkeiten und Grenzen der isolierten Analyse lehrbezogener Daten

Die Analyse isolierter, das heisst nicht aggregierter Daten ist in der Hochschullehre weit verbreitet. Es handelt sich dabei um die Analyse von Umfrageresultaten, Leistungsbeurteilungen oder Prüfungsergebnissen. Sie stellen eine Momentaufnahme dar und erlauben Analyse zu zeitweiligen Zuständen wie Vorwissen, Motivation, Erwartungen und Wissen. Zwar können sich bei diesen Datenerhebungen einzelne Fragen auf die Vergangenheit oder die Zukunft beziehen, jedoch wird die Antwort immer beeinflusst von der persönlichen Situation zum Zeitpunkt der Antwortabgabe. So fällt beispielsweise das Urteil über eine Prüfung unmittelbar danach meist heftiger aus, als wenn die Beurteilung einige Tage danach oder gar in Kombination mit anderen Lehrveranstaltungen erhoben wird.

Die Analyse isolierter lehrbezogener Daten erlaubt vielfältige Aussagen und grundsätzliche Rückschlüsse auf den Unterricht. Die Aussagekraft kann noch erhöht werden, wenn zusätzlich die Daten vergangener Semester zum Vergleich beigezogen werden. Dieses Vorgehen soll nun mit einem Beispiel aus der Praxis illustriert werden. Die hier gezeigten Daten wurden 2005 an der ED-MEDIA Konferenz in Montreal von Dahinden und Fässler präsentiert (L. Fässler, Dahinden, Bosia, & Hinterberger, 2005). Trotz eines unterschiedlichen Fokus bezüglich der Aussage lässt sich aus den Daten auch auf einen interessanten Aspekt bezüglich des *hidden Curriculum*s schliessen.

Im Jahr 2002 haben wir damit begonnen, den klassischen Präsenzunterricht, bestehend aus Vorlesungen und Skript, in einen auf elektronischen Lernmaterialien basierten Unterricht umzustellen. Unter Berücksichtigung des *Learning Outcome* dominierten an der Prüfung die K1 (erinnern) Fragen. Es gab nur wenige K2 (verstehen) und K3 (anwenden) Fragen (Abbildung 8.75, 2002/I&II).

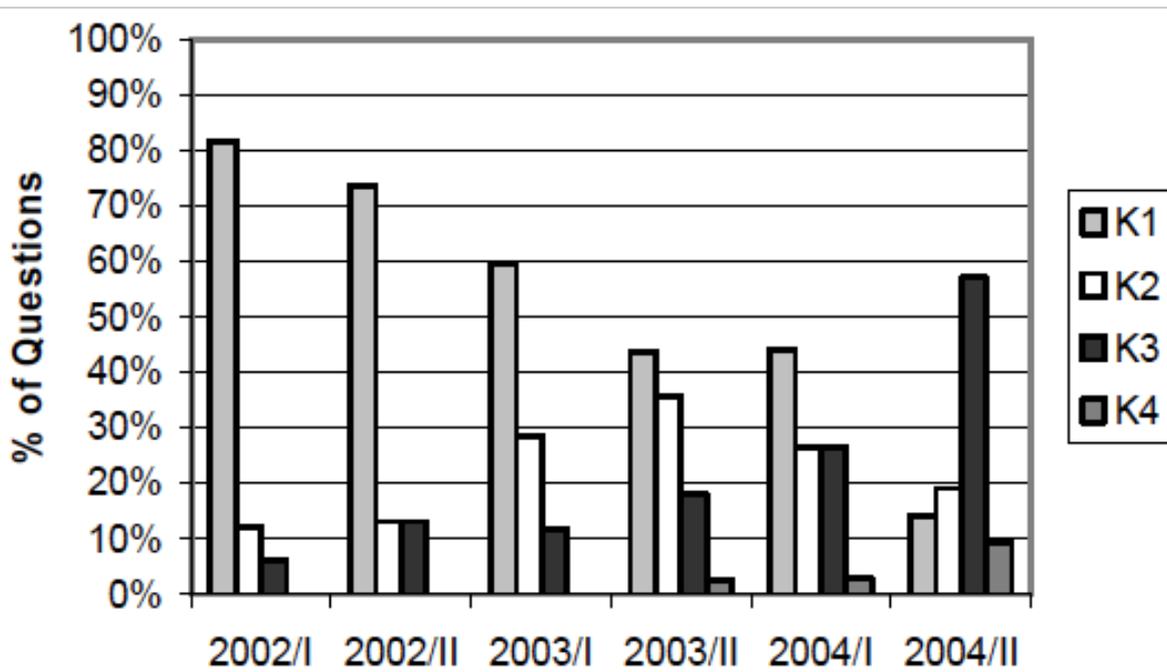


Abbildung 8.75: Zusammensetzung der Prüfungen verschiedener Jahre. Angegeben sind die Anzahl der Fragen pro kognitiver Prozessdimension (K1-4, vgl. Kapitel 2). Es ist erkennbar, dass die Anzahl der K1- und K2-Fragen über die Zeit zugunsten der K3+ Fragen reduziert wurden. (Datenbasis: Schlussprüfung, EvIM und IBPW, Jahre 2002-2004, jeweils zwei Tests pro Jahr, n: 33-42 Fragen) (Abbildung übernommen aus L. Fässler et al., 2005)

Die Umstellung des Unterrichts begann bei den Übungen: durch elektronische Tutorate (heute: *E.Tutorial*²⁶), ein Beiheft zur Vermittlung der für die Übungen benötigten theoretischen Grundlagen sowie auf den Tutoraten aufbauenden selbständigen Aufgaben sollte die Aktivität der Studierenden gefördert werden. Durch die gesteigerte Aktivität der Studierenden erhofften wir uns eine vertiefte, individuelle Auseinandersetzung mit dem Lernstoff und eine Steigerung des *Learning Outcome* auf die Kompetenz- (K3+) statt wie bisher auf der Inhaltsstufe (K1/2).

Eine erste Etappe des Umbaus wurde 2003 erreicht: Alle sechs Übungen wurden auf die Vermittlung von kompetenzorientiertem Wissen umgestellt. Die Prüfung selber wurde nur leicht verändert und wies immer noch eine Mehrheit an K1 und K2 Fragen auf. Nach den positiven Erfahrungen mit den umgebauten Übungen im Unterricht kam die Enttäuschung bei der Analyse der Prüfungsergebnisse: Die Studierenden vermochten die K3+ Fragen nur leicht besser zu beantworten. Wir hatten unser Ziel, Kompetenzen statt Inhalten zu vermitteln, offenbar nur teilweise erreicht. Eine kritische Analyse der Situation ergab, dass der Wechsel zum kompetenzorientierten Unterricht nicht konsequent genug verfolgt wurde (wohl hauptsächlich bedingt durch den damit verbundenen enormen Aufwand). Zum einen gab es viele Vorlesungen, ein Skript und zeitaufwändi-

²⁶ *E.Tutorial* ist ein eingetragenes Markenzeichen der ETH Zürich. Weitere Informationen: www.cta.ethz.ch (abgerufen am: 23.05.2014)

ge Übungen, zum anderen lag der Fokus der Prüfung noch auf der Messung des in der Vorlesung und im Skript vermittelten Wissens. Rückblickend stellt es somit ein eindrückliches Beispiel dar, wie stark das *hidden Curriculum* die Lernaktivitäten und somit den *Learning Outcome* beeinflusst; in unserem damaligen Fall leider nicht in die von uns gewünschte Richtung.

Im Folgejahr gelang uns dann mit einem konsequenten Umbau der Prüfung (Abbildung 8.75, 2004/II) und der klaren Vermittlung des geänderten *hidden Curriculums* der entscheidende Wechsel vom inhalts- zum kompetenzvermittelnden Unterricht (Abbildung 8.76, 2004). Der Mittelwertvergleich nach Kruskal-Wallis (Kruskal & Wallis, 1952) für kleine Stichproben zeigt, dass sich die Ergebnisse der drei Jahre signifikant unterscheiden ($p=0.015$). Der paarweise Vergleich der Mittelwerte nach Dunn (Dunn, 1961) macht zudem deutlich, dass sich insbesondere die Prüfungsergebnisse der Jahre 2003 und 2004 signifikant unterscheiden ($p_{\text{korr}}=0.033$). Die detaillierten Statistiken finden sich im Anhang A.6 dieser Arbeit.

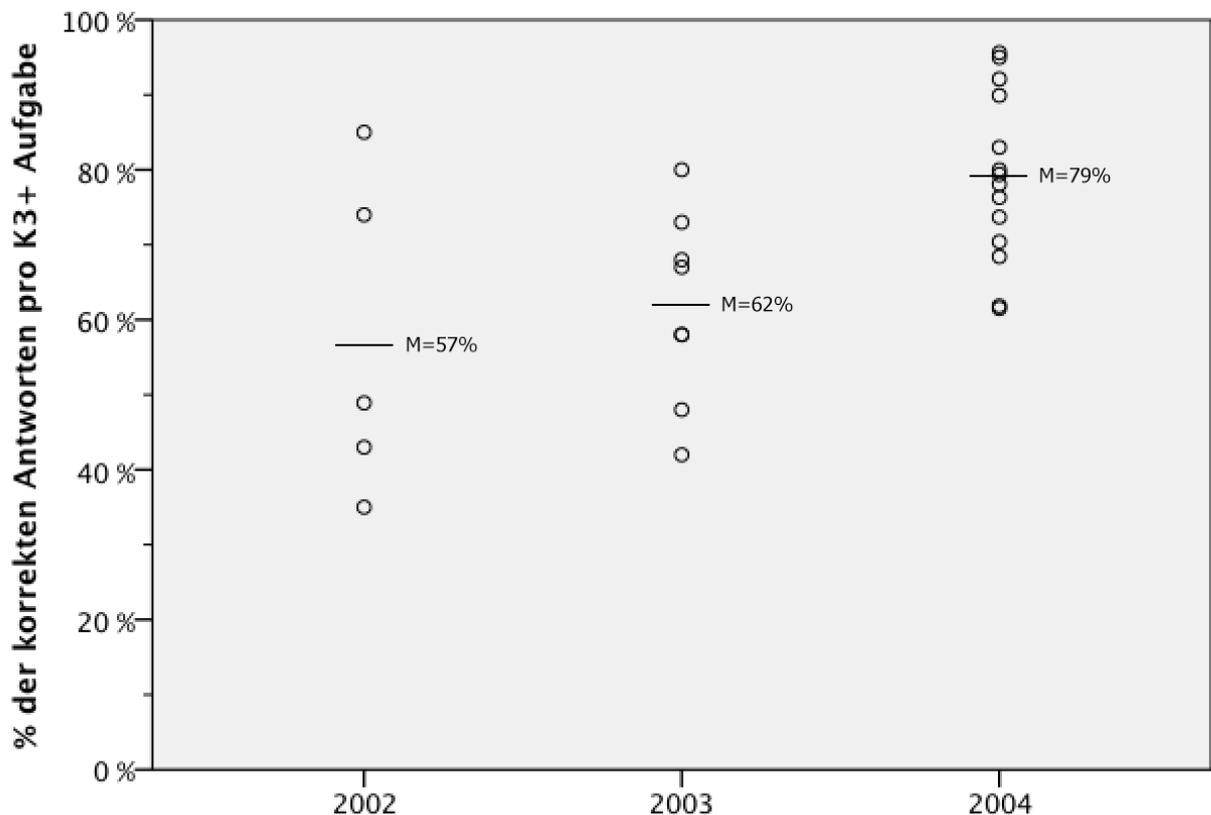


Abbildung 8.76: Anzahl und Durchschnitt der korrekt gelösten K3+ Fragen. Ausgewertet wurden jeweils die Daten 2002/II, 2003/II und 2004/II. Die Resultate zeigen, dass die Anzahl und der Mittelwert der korrekt gelösten K3+ Antworten stetig zugenommen hat. (Datenbasis: Schlussprüfung, IBPW, 2002: N=49, M=57%. 2003: N=94, M=62%. 2004: N=77, M=79%)

Das obige Beispiel zeigt die Analyse von isolierten - das heisst nicht aggregierten - Daten über die Zeit. Damit ist gemeint, dass die Daten der einzelnen Messzeitpunkte nicht miteinander verknüpft (aggregiert), aber dennoch in eine zeitliche Beziehung zueinander gebracht wurden. Diese Methode bietet vielfältige Möglichkeiten zur Analyse des

Unterrichts und zur Visualisierung der Verbesserung des *Learning Outcomes* über die Zeit. Somit ist dieses Verfahren für die Unterrichtsevaluation in den meisten Fällen ausreichend.

Die in den nachfolgenden Abschnitten vorgestellte Aggregation der Daten hat zum Ziel, die Daten einzelner Studierender oder Studierendengruppen so zu verknüpfen, dass eine globale Sicht auf die Daten erreicht wird. Dieses Vorgehen wird mit konkreten Beispielen aus dem Unterricht illustriert.

8.3 Aufbau des educational Data-Warehouses (eDWH)

Wie bereits mehrfach festgestellt wurde, ermöglicht eine lehrbezogene Datensammlung eine Vielzahl von Analysemöglichkeiten (Educause Learning Initiative, 2010; Johnson et al., 2011; Scheuner & Faessler, 2010). Es ist jedoch eine Herausforderung, diese Datenmengen so vorzubereiten, dass eine globale Sicht auf die Daten und übergreifende Analysen möglich werden. Zum einen liegen die Daten dezentral und in unterschiedlichen Datenformaten vor und zum anderen fehlen oft geeignete Attribute, welche eine Aggregation der Daten erlauben und dennoch eine anonymisierte Datenerhebung zulassen.

In diesem Abschnitt wird das Vorgehen beschrieben, das in unserem Unterricht angewandt wird. Dieses basiert auf drei Schritten:

- 1) Definition einer Kennung, welches mit jedem Datensatz miterhoben wird und ihn eindeutig kennzeichnet
- 2) Zusammenführung der Daten in ein *educational Datawarehouse* (eDWH)
- 3) Aggregation der Daten auf der Basis dieser Kennung

8.3.1 Eindeutige Kennung zur Aggregation der Daten

Die Kennung wird benötigt, damit die Datensätze später zu Antwortprofilen verknüpft werden können. Entsprechend muss sie bei jeder Datensammlung miterhoben werden. Bei uns hat sich der Einsatz eines Kürzels basierend auf den ersten beiden Buchstaben des Vornamens der Mutter und des Vaters, sowie der Monatstag des Geburtstags des Studenten bewährt²⁷. Diese aus fünf bis sechs Ziffern bestehende Kennung ist eindeutig genug, um die Daten zu aggregieren und dennoch eine anonymisierte Datenerhebung zu ermöglichen.

²⁷ Beispiel eines Kürzel: Vorname der Mutter (Verena), Vorname des Vaters (Josef) und Monatstag des Geburtstags (14.5.1975) ergibt das Kürzel: VeJo14

Grundsätzlich kann auch eine offizielle Kennung wie beispielsweise die Studierendenummer verwendet werden. Da diese den Dozierenden jedoch meist in Kombination mit den Namen der Studierenden zur Verfügung stehen, ist mit offiziellen Kennungen eine anonymisierte Datenerhebung nicht möglich.

Datenschutzüberlegungen

Sowohl die Verwendung von Kürzeln als auch von offiziellen Kennungen zur Erhebung der lehrbezogenen Daten unterstehen dem Datenschutzgesetz. Da die Auswertung der Daten anonymisiert stattfand und die Teilnahme an der Erhebung der Prozessdaten freiwillig war, haben wir nach Rücksprache mit der Rechtsabteilung der ETH Zürich auf weitere Abklärungen verzichtet. Es ist jedoch ratsam, vor der Erstellung eines eDWH die rechtlichen Grundlagen bezüglich des Datenschutzes und der Datensicherheit abzuklären.

Speicherdauer der Daten

Da die lehrbezogenen Daten für Forschungszwecke auch Jahre später noch relevant sein können, ist das Löschen der Daten in den meisten Fällen nicht sinnvoll. Stattdessen empfiehlt sich die komplette Anonymisierung der Daten unmittelbar nach Ende der Erhebungsdauer der Daten. Dies ist meist nach der Schlussumfrage respektive der Schlussprüfung der Fall.

Bei der Anonymisierung der Daten werden alle identischen Kennungen durch gleiche Zufallszahlen ersetzt (Randomisierung). So bleibt die referentielle Integrität der Daten intakt, jedoch keine Rückschlüsse auf die Identität der Studierenden mehr möglich. Eine Konsequenz davon ist, dass nachträglich keine weiteren Datensätze mehr ins eDWH integriert und mit den bestehenden, randomisierten Daten aggregiert werden können. Diesen Umstand gilt es bei der Wahl des Zeitpunkts der Randomisierung zu beachten.

8.3.2 Speichern und Aggregation der Daten in einem eDWH

Kommt im Unterricht ein *Learning Management System* (LMS) zum Einsatz, liegen viele Daten womöglich bereits zentral vor. In der Praxis kommen jedoch zumindest für die Leistungskontrolle, die auf kurzen Fragebögen durchgeführte Prozessanalyse oder die Übungsbearbeitung Drittapplikationen zum Einsatz. Dann ist es nötig, diese dezentral gesammelten Rohdaten in einer zentralen Datenbank zu speichern, so dass übergreifende Analysen möglich werden. Die Gesamtheit der in der Datenbank gespeicherten lehrbezogenen Daten haben wir als *educational Datawarehouse* (eDWH) bezeichnet.

Unser eDWH basiert auf einem relationalen Datenbanksystem²⁸. Mittels geeigneter SQL-Abfragen wurden die Daten auf der Basis der Kennungen aggregiert und die Analyseresultate in Tabellenform ausgegeben. Zudem wurden *Indizes* und *Views* verwendet, um die Abfragen zu beschleunigen und mittels *Procedures* wurden die Kennungen normalisiert und die *Lickert-Skala* Antworten umcodiert. Anschliessend wurden die Daten in SPSS weiterverarbeitet.

8.4 Kontext der Datenerhebung

Zur Analyse eines eDWH ist es notwendig, den Kontext der Datenerhebung zu kennen. Nur wenn dieser bekannt ist, können die Resultate sinnvoll interpretiert und Entwicklungsvorschläge für den Unterricht ausgearbeitet werden.

Die in den nachfolgenden Abschnitten präsentierten Daten wurden im Rahmen der Lehrveranstaltung "Einsatz von Informatikmitteln" im Herbstsemester 2010 gesammelt und mit dem im Abschnitt 8.3 beschriebenen Vorgehen in einem eDWH gespeichert.

Für diese Lehrveranstaltung waren 299 Erstsemestriige von vier verschiedenen Naturwissenschaftsdisziplinen eingeschrieben: Agrarwissenschaften, Lebensmittelwissenschaften, Erdwissenschaften und Umweltnaturwissenschaften.

Während der Semesterdauer von 14 Wochen galt es, sechs Module zu bearbeiten. Jedes dieser Module war gemäss des 4-Schritte-Modells aufgebaut (L. E. Fässler, 2007). Der Abschluss jedes Moduls bildete die Bearbeitung einer selbständigen Aufgabe, welche im Anschluss im Computerraum einer dort anwesenden Tutoren präsentiert werden musste. Dieses Gespräch dauerte rund fünf bis sieben Minuten und wurde von der Tutoren mit einer vierstufigen Lickert-Skala bewertet²⁹. Diese Bewertung hatte rein formativen Charakter und wurde nicht mit der Schlussnote verrechnet (Tabelle 8.21, Abkürzung: "prozT").

Datenkategorie	Beschreibung	Abkürzung	Anzahl der aggregierten Datensätze
Prozessdaten	Studierende Tutorenbewertung	prozS	177
		prozT	284
Leistungsdaten	Probetest Semester-Endprüfung	test	254
		exam	276

Tabelle 8.21: Übersicht der während des Herbstsemesters 2010 erhobenen und in einem eDWH gespeicherten lehrbezogenen Daten der LV "Einsatz von Informatikmitteln". Eingeschrieben für diese LV waren

²⁸ verwendetes System: MySQL (Version 5.0.67), <http://dev.mysql.com/>

²⁹ Verwendet wurde folgende Bewertung: 0=Lernziel nicht erreicht, 1=Lernziele erreicht, 1+=Lernziele leicht übertroffen, 2=Lernziele klar übertroffen

299 Studierende, wovon 284 aktiv am Unterricht teilnahmen. Die vier Datenquellen werden im Text mittels Abkürzung referenziert und beschrieben.

Nach Abschluss jedes Moduls hatten die Studierenden die Möglichkeit, einen kurzen Fragebogen zum Modul auszufüllen ("prozS"). Unter anderem wurden die Studierenden darin um eine Einschätzung ihrer eigenen Leistung gebeten. Der verwendete Fragebogen findet sich im Anhang A.7 dieser Arbeit.

Zwei Wochen vor der Semester-Endprüfung fand ein freiwilliger Probetest am Computer mit der Sioux CBA-Suite statt. 89% der Studierenden haben daran teilgenommen ("test"). Nach Abschluss des Tests wurden die Korrektur und ein individuelles Korrekturfeedback angezeigt. Die Prüfungszeit und die Anzahl der Fragen waren so gewählt, dass im Anschluss noch eine individuelle Nachbesprechung mit den anwesenden Tutoren möglich war.

Eine Woche vor Semesterende fand die Semester-Endprüfung statt ("exam"). Diese dauerte 60 Minuten, geprüft wurden die Inhalte aller sechs Module. An der Prüfung teilgenommen haben 97% der Studierenden.

Die Validität der Resultate der Semester-Endprüfung wurde mittels neun Aussenkriterien bestimmt. Weitere Informationen zur Datenerhebung und zur Validitätsprüfung können in der Publikation von Dahinden et al. in Erfahrung gebracht werden (Dahinden & Faessler, 2011).

In den nachfolgenden Abschnitten wird das Analysepotential eines eDWH auf der Basis der soeben umschriebenen lehrbezogenen Daten untersucht. Dieses eDWH beinhaltet die aggregierten Daten von 284 Studierenden und umfasst über 50 Dimensionen. Der Fokus bei der Analyse dieses eDWH liegt auf der retro- und prospektiven Datenanalyse, wobei beispielhaft aufgezeigt wird, welche Aussagen bezüglich des Unterrichts gemacht werden können. Zudem soll untersucht werden, welche Auswirkungen die so gewonnenen Erkenntnisse auf die Entwicklung des Unterrichts haben können.

8.5 Retrospektive Analyse des eDWH

Die retrospektive Analyse hat zum Ziel, durch die Gegenüberstellung vergangener Messungen deren Zusammenhänge aufzuzeigen und Datenpunkte für weitere Analysen zu bestimmen.

8.5.1 Vergleich des Resultats am Probetest und der Semester-Endprüfung

Als Beispiel für eine retrospektive Analyse wurden die Resultate des Probetests mit jenen der Semester-Endprüfung verglichen. Beide Dimensionen zeigen die prozentual erreichten Punkte von der Maximalpunktzahl (Abbildung 8.77).

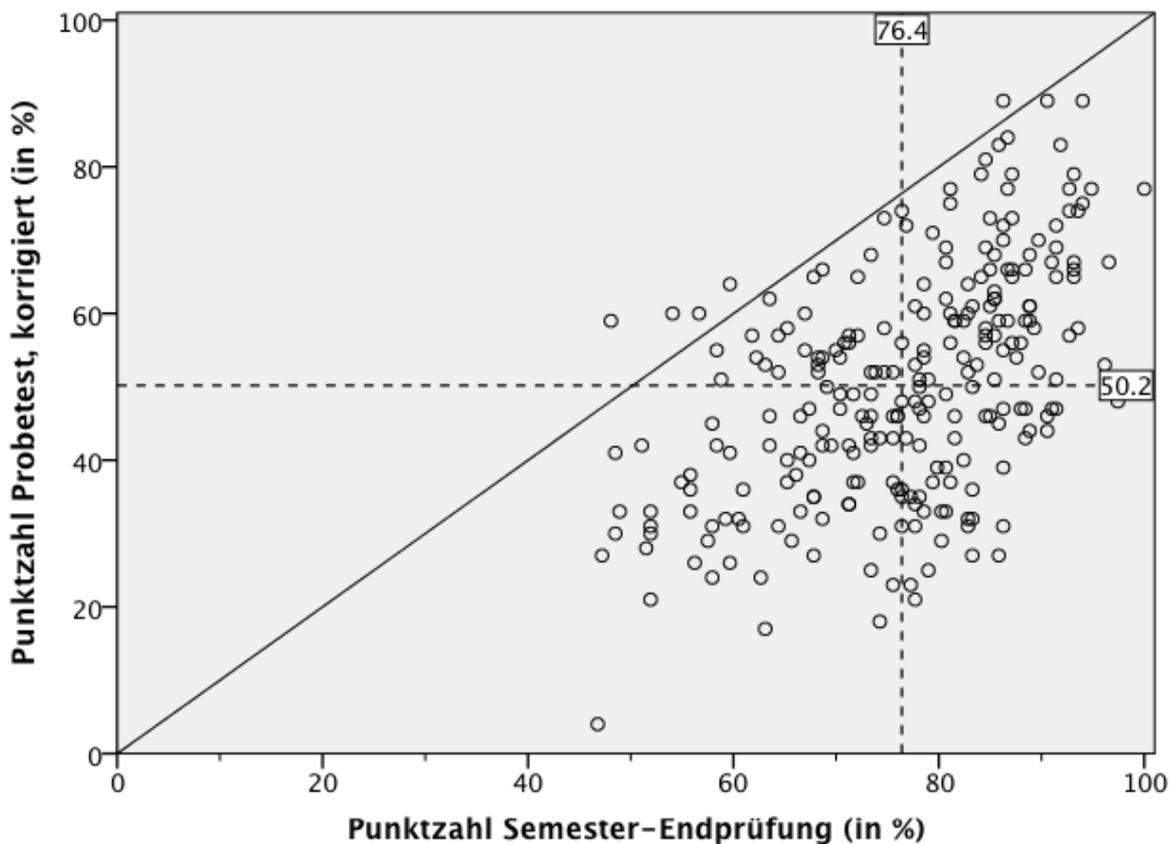


Abbildung 8.77: Vergleich der Resultate des Probetests und der Semester-Endprüfung. Die Diagonallinie unterscheidet zwei Gruppen von Studierenden: jene, welche am Probetest besser abschnitten befinden sich oberhalb der Linie und jene mit einem besseren Resultate an der Semester-Endprüfung unterhalb. Die beiden Mittelwerte sind gestrichelt eingezeichnet. Die Korrelation nach Pearson der beiden Datensätze beträgt $r=0.53$ ($p<0.01$) (Datenbasis: eDWH, EvIM, HS10, $n=254$)

Die Analyse zeigt eine positive Korrelation ($r_{\text{PEARSON}}=0.53$) zwischen dem Resultat des Probetests und der Semester-Endprüfung. Das Ergebnis der Semester-Endprüfung war um 26.2% höher als jenes des Probetests. Es fällt auf, dass fünf Studierende an der Semester-Endprüfung ein schlechteres Ergebnis erzielt haben als am Probetest. Auffallend ist zudem, dass vier dieser fünf Studierenden dem schwächsten Quartil an der Semester-Schlussprüfung angehören.

8.5.2 Diskussion und Ausblick der retrospektiven Analyse

Die hohe Anzahl Studierender am freiwilligen Probetest zeigt, dass diese offenbar motiviert sind, ein formatives Feedback zu ihrem Leistungsstand zu erhalten. Sie erhoffen sich aber auch Informationen bezüglich des *hidden Curriculums*, war doch der Probetest angekündigt als "Probelauf unter regulären Prüfungsbedingungen mit realen Prüfungsfragen". Entsprechend verlief die Durchführung des Tests bis auf wenige Aus-

nahmen gleich wie an der summativen Semester-Endprüfung³⁰. Wie wir in einem späteren Semester ermittelt haben, gaben 80% der Studierenden an, dass sie ruhiger an die summative Prüfung gingen, da sie den Ablauf bereits kannten³¹. Diese Ruhe hilft den Studierenden, sich auf die Bearbeitung der Aufgaben zu konzentrieren und trägt zu einem reliablen Resultat bei.

Die Abbildung 8.77 zeigt die Entwicklung der kognitiven Leistungen in den letzten beiden Unterrichtswochen. Es handelt sich dabei insofern um eine besondere Zeit im Semester, als dass die meisten Unterrichtsinhalte bereits vermittelt sind und den Studierenden nur noch wenig Zeit für die gezielte Repetition der Unterrichtsinhalte bleibt. Entsprechend ist zu erwarten, dass sich die Leistung der Studierenden zwischen des Probetests und der Semester-Endprüfung aufgrund des individuellen Feedbacks zwar absolut verbessert, aber sich relativ in Bezug zu den Mitstudierenden nicht grundlegend verändert. Diese Vermutung wird von den Daten bestätigt: Der Mittelwert der summativen Semester-Schlussprüfung fällt 26.2% höher aus als am Probetest. Der Grund für diese Verbesserung lässt sich jedoch nicht abschliessend beantworten und es braucht noch weiterer Untersuchungen, um zu klären, ob sich die kognitive Leistung der Studierenden tatsächlich so stark verbessert hat oder ob sie den Probetest einfach etwas weniger engagiert bearbeitet haben. Nichtsdestotrotz lässt sich anhand der Verteilung der Rohdaten eine aktive Teilnahme am Probetest ableiten. Ein Hinweis dafür ist die positive Korrelation zwischen den Resultaten des Probetests und der Semester-Schlussprüfung. Zudem fällt auf, dass sich die Leistung einiger Studierender massiv verbessert hat. Dieser Leistungsanstieg kann möglicherweise als *Boost Effekt* ehrgeiziger Studierender verstanden werden. Um dies abschliessend zu klären, sind jedoch weitere Untersuchungen nötig. Ebenso wäre interessant zu klären, wieso vier Studierende des letzten Leistungsquartils ein besseres relatives Ergebnis am Probetest erzielt haben.

Es kann also festgehalten werden, dass mit der retrospektiven Analyse eines eDWH einzelne Studierende identifiziert und genauer untersucht werden können. So lassen sich verschiedene Gruppen von Studierenden mit zum Beispiel unterschiedlicher Lernbereitschaft, Motivation oder Vorwissen identifizieren. Diese Erkenntnisse können wesentlich zum Verständnis einer bestehenden Unterrichtseinheit beitragen und Ansatzpunkte gezielter Verbesserungen liefern. Zudem können die Gruppen in der prospektiven Analyse dazu verwendet werden, um die betroffenen Studierenden bereits während eines laufenden Semesters zu identifizieren und deren Lernaktivitäten mittels individualisierten Feedbacks positiv zu beeinflussen. Im nachfolgenden Abschnitt wird dieses Vorgehen mit zwei Praxisbeispielen illustriert.

³⁰ Im Vergleich zur Semester-Endprüfung dauerte der Probetest nur halb so lang, wurde automatisch korrigiert und die Resultate im Anschluss an die Prüfung mit den in den Prüfungsräumen anwesenden Tutoren nachbesprochen.

³¹ Umfrage im Herbstsemester 2011, EVIM, Von den Studierenden, welche die Prüfung als "schwierig" oder "sehr schwierig" beurteilten, gaben 80% an, dass sie aufgrund des Probetests ruhiger an die Prüfung gingen, weil ihnen der Ablauf bereits bekannt war.

8.6 Prospektives Analysepotential des eDWH

Die prospektive Analyse hat zum Ziel, die Lernaktivitäten und -auswirkungen auf vordefinierte Studierendengruppen oder Einzelpersonen zu untersuchen, um mögliche Interventionsmassnahmen zu erkennen und deren Effekte zu messen. Die prospektive Analyse eines eDWH basiert auf mindestens zwei Datensätzen, welche zu unterschiedlichen Zeitpunkten erfasst wurden. Sie visualisiert somit den Verlauf eines Lernprozesses, wodurch sich die Auswirkungen von Lerninterventionen unmittelbarer messen lassen, als dies bei der retrospektiven Analyse möglich ist.

8.6.1 Vergleich der Chronologie der Expertenurteilen mit dem Resultat an der Schlussprüfung

Zur Abschätzung des prospektiven Analysepotentials eines eDWH wurden die während des Unterrichts erfassten Expertenbeurteilungen mit dem Schlussresultat verglichen (Abbildung 8.78). Die 276 aggregierten Datensätze wurden auf der Basis des Schlussresultats gleichmässig in vier Leistungsquartile (EQ1-EQ4) unterteilt, wobei EQ1 die besten und EQ4 die schwächsten Studierenden beinhaltet. Diese Messpunkte wurden mit den vierstufigen Bewertungen verglichen, welche die Studierenden bei der individuellen Präsentation ihrer selbständigen Aufgaben von den Tutoren erhalten haben.

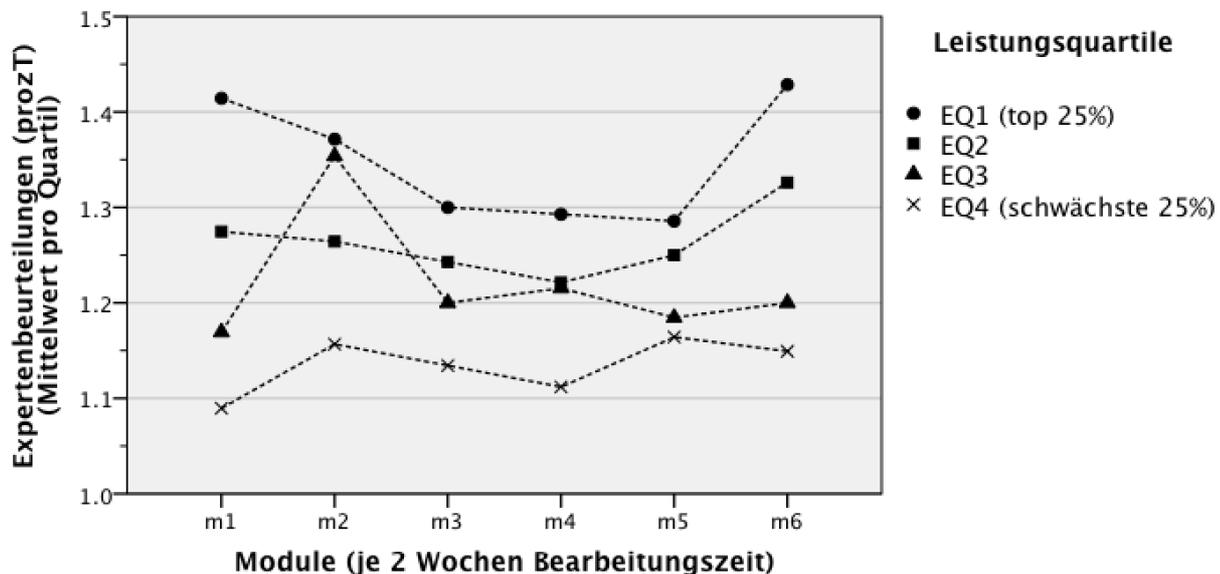


Abbildung 8.78: Vergleich des Resultats an der Semester-Endprüfung mit den Expertenbeurteilungen der Präsentationen der sechs selbständigen Aufgaben (m1-m6). Angezeigt sind die Mittelwerte der Bewertungen der vier Leistungsquartile. Die Leistungsquartile basieren auf dem Ergebnis der Studierenden an der Schlussprüfung. Zur Verdeutlichung des Prozessverlaufs wurden die Werte der Quartile untereinander verbunden. Die Skala der Tutorenbewertungen reicht von 0 bis 2 (vgl. 8.4 "Kontext der Datenerhebung") (Datenbasis: eDWH, EvIM, HS10, n=276)

Die Resultate in Abbildung 8.78 zeigen, dass die Studierenden im schwächsten Quartil (EQ4) von den Tutoren in allen sechs Modulen durchschnittlich die tiefste Bewertung erhielten, während die besten Studierenden (EQ1) im Mittel die höchste Bewertung er-

hielten. Es kann also festgehalten werden, dass die Tutoren die Leistung der Studierenden an der Schlussprüfung während des ganzen Semesters zuverlässig einschätzen können. Dabei ist erstaunlich, dass bereits im ersten Modul die Leistungsunterschiede der Leistungsquartile auch von den Tutoren beobachtet werden. Es bleibt noch zu untersuchen, inwiefern sich dieses Ergebnis auf das heterogene Vorwissen zurückführen lässt, oder ob es nicht eher der Effekt unterschiedlicher Lernvoraussetzungen wie Lernstile, Motivation oder Engagement darstellt.

Ungeachtet dieser Ursachenunschärfe kann festgehalten werden, dass die Expertenurteile der Tutoren für die Studierenden ein wichtiges Feedback bezüglich der Lernzielerreichung darstellen. Grundsätzlich sind aufgrund der Expertenurteile gar prospektive Aussagen bezüglich des Ergebnisses an der Semester-Endprüfung möglich. Diese Erkenntnis haben wir in unserer Schulung der Tutoren einfließen lassen. So fordern wir unsere Tutoren auf, ihre Beobachtungen im Unterricht aktiv zu kommunizieren und Studierenden ein individuelles Feedback zu geben. Dank der in Abbildung 8.78 gezeigten Resultate wissen wir, dass diese Rückmeldungen im Mittel gerechtfertigt und richtig sind.

8.6.2 Vergleich der subjektiven mit den absoluten Leistungseinschätzungen in Abhängigkeit der Leistung an der Schlussprüfung

Abschliessend soll mit der subjektiven Leistungseinschätzung die im letzten Abschnitt vorgestellte prospektive Analyse noch um eine weitere Dimension ergänzt werden. Die subjektive Leistungsbeurteilung beruht auf einer Selbstbeurteilung der eigenen Leistung. Die Studierenden wurden aufgefordert, ihre Leistung nach der Präsentation jeder selbständigen Aufgabe zu beurteilen. Es ist davon auszugehen, dass die meisten Studierenden von den Tutoren ein individuelles Feedback erhalten haben und so ihre Leistungseinschätzung mit diesem Aussenkriterium abgleichen konnten.

Abbildung 8.79 zeigt die Unterschiede der Selbst- und Expertenbeurteilungen auf der Basis der vier Leistungsquartile an der Semester-Schlussprüfung. Es lässt sich erkennen, dass die besten Studierenden jedes Moduls ihre Leistung relativ zur Expertenbeurteilung ziemlich exakt bestimmen konnten. Allgemein fällt auf, dass die Studierenden meist ihre eigene Leistung etwas tiefer einschätzen, als dies die Tutoren tun. Speziell bei der Leistungsgruppe 2 ist diese Tendenz ziemlich ausgeprägt. Bei der Leistungsgruppe 6 hingegen lässt sich eine Tendenz zur Überschätzung der eigenen Leistung vermuten. So beurteilen die Studierenden beispielsweise im Modul 6 ihre eigene Leistung als überdurchschnittlich hoch, während die Tutoren diesen Studierenden im Mittel eine unterdurchschnittliche Bewertung gaben.

Die Gründe für diese Überschätzung bleiben unklar. Möglicherweise kann sie damit erklärt werden, dass sie ihre Leistung aufgrund eines milden Feedbacks seitens der Assistenten oder aufgrund ihrer Teilnahme in leistungsschwachen Übungsgruppen falsch einschätzten. Es ist aber auch denkbar, dass diese Studierenden mit ihrer be-

scheidenen Leistung an der Präsentation tatsächlich zufrieden waren und dass ein möglicherweise undifferenziertes Feedback der Tutoratsperson ihre Fehleinschätzung nicht beeinflussen konnte.

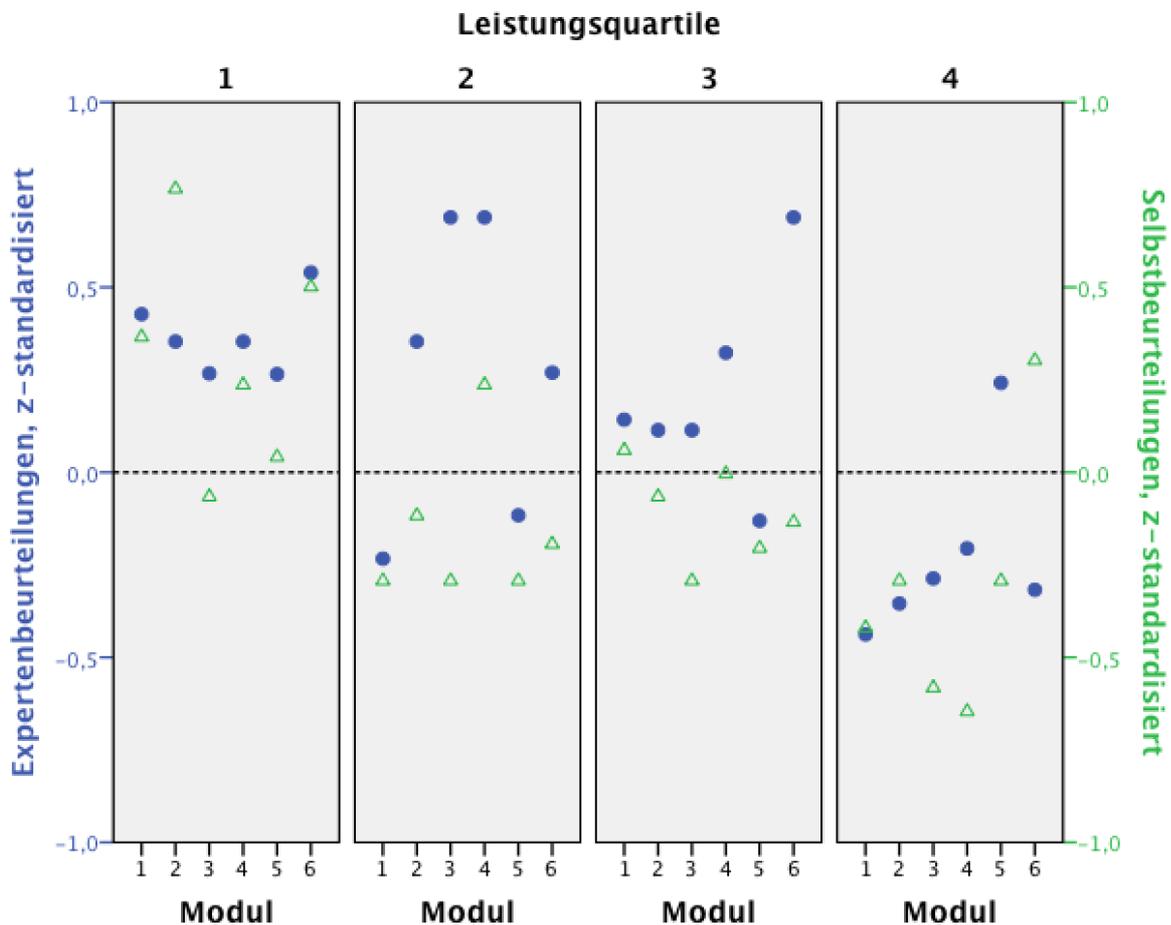


Abbildung 8.79: Vergleich der vier Leistungsquartile bezüglich der Selbstbeurteilung der Studierenden und der Expertenbeurteilung. Die Leistungsquartile basieren auf dem Resultat des Subscores des jeweiligen Moduls an der Semester-Endprüfung. Die besten Studierende sind im Leistungsquartil (LQ) 1, die schwächsten im LQ 4. Die Experten- und Selbstbeurteilungen sind z-standardisiert. Der Gesamtmittelwert der Beurteilungen liegt bei 0,0 und ist gestrichelt eingezeichnet. Die Mittelwerte der Expertenbeurteilungen sind als Punkte (\bullet), jene der Selbstbeurteilungen als Dreiecke (Δ) dargestellt. (Datenbasis: eDWH, EvIM, HS10, die deskriptive Statistik der Häufigkeiten findet sich im Anhang A.8 und variiert zwischen $n=17-134$ (Expertenbeurteilungen) und $n=1-41$ (Selbstbeurteilungen))

Eine weitere interessante Beobachtung lässt sich beim dritten Leistungsquartil machen. In diesem Quartil befinden sich Studierende, welche die Prüfung mit einer knapp genügenden Note bestanden haben. Es fällt auf, dass diese Studierende im letzten Modul unmittelbar vor der Schlussprüfung ihre Leistung stark unterschätzten. Diese Unterschätzung der eigenen Leistung kann möglicherweise dazu geführt haben, dass sie mit ihrer Leistung nicht zufrieden waren und somit die Unterrichtsinhalte noch einmal intensiv repetiert haben. Dank dieser Anstrengung gelang es ihnen möglicherweise, sich vom schwächsten ins zweitschwächste Leistungsquartil zu verbessern. Eine

andere Erklärung wäre, dass es sich dabei um verunsicherte Studierende handeln könnte, welche eigentlich dem zweiten Leistungsquartil angehörten, aber aufgrund von Prüfungsstress verunsichert waren und ihre wahre Leistung an der Semesterabschlussprüfung nicht vollumfänglich zeigen konnten. Weiter könnte untersucht werden, ob allenfalls das Geschlecht bei der Selbstbeurteilung eine Rolle gespielt haben könnte.

Auch wenn über die Ursachen dieser Resultate vorerst nur spekuliert werden kann, so kann dieses Wissen bereits verwendet werden, um Interaktionen zu planen und deren Auswirkungen im Unterricht zu beobachten. In unserem konkreten Fall haben wir unsere Tutoren angewiesen, vor allem auch den schwachen Studierenden ein ehrliches, klares Feedback zu geben. Die Vergabe eines negativen Feedbacks mag sowohl für die Tutoren als auch für den Studenten unangenehm sein, aber es stellt ein ungemein wichtiges Aussenkriterium dar, damit Letztere ihre Leistung realistisch einschätzen können.

8.7 Diskussion der Fragestellung 4: Unter welchen Voraussetzungen bietet der Einbezug von Leistungsdaten einen Mehrwert in der Analyse lehrbezogener Daten?

In modernen E.Learning-Umgebungen fallen mit Daten aus Leistungskontrollen, Fragebögen, Tutoratsgruppen und Log-Daten eine Vielzahl lehrbezogener Datensätze an. Deren Auswertung jedoch ist oft langwierig und anspruchsvoll. Nicht zuletzt, weil mit der Faktorenanalyse zwar mögliche Zusammenhänge aufgedeckt werden können, aber ohne fundiertes Wissen über die Inhalte und das *Alignment* der didaktischen Elemente allfällige Kausalitäten nicht oder falsch identifiziert werden.

Im aktuellen Kapitel wurde auf der Basis valider Leistungsdaten das Analysepotential von aggregierten lehrbezogenen Daten mittels visueller Verfahren untersucht. Die validen Leistungsdaten stellen ein Qualitätsindikator einer Lernumgebung dar und können demnach als *Learning Outcome* verstanden werden. Es konnte gezeigt werden, dass durch Leistungsdaten-basierte Analysen der lehrbezogenen Daten in einem eDWH sowohl retrospektive als auch prospektive Analysen möglich sind.

Mit Hilfe der retrospektiven Analyse lassen sich rückwirkend inhaltliche und organisatorische Probleme erkennen und Gegenmassnahmen für zukünftige Unterrichtssequenzen ausarbeiten. Zudem bildet die retrospektive Analyse die Grundlage für die Identifikation interessanter Datensätze, die weitere Aufschlüsse über die Mikroelemente eines funktionierenden Unterrichts liefern können.

Die prospektive Analyse der lehrbezogenen Daten zielt auf die Identifikation bestimmter Studierender oder Studierendengruppen, um mittels geeigneter Interventionsmassnahmen deren Lernaktivitäten positiven zu beeinflussen und den *Learning Out-*

come zu erhöhen. Durch die prospektive Analyse sollen aber nicht nur die Möglichkeiten eröffnet, sondern die Effekte in der Praxis unmittelbar untersucht werden können. Durch derartige Untersuchungen bietet sich womöglich auch im alltäglichen Hochschulunterricht - wo das Definieren verschiedener empirischer Testgruppen meist aus ethischen Gründen nicht möglich ist - die Gelegenheit, die Ursache und Wirkung von Lerninterventionen unmittelbar zu messen. Und nicht zuletzt können diese Erkenntnisse auch interessante Hinweise für die klassische didaktische oder lernpsychologische Forschung liefern.

Anhang

Anhang

A.1 Auszug aus den Resultaten der Nachumfrage IBPW

Im Frühlingssemester 2012 haben wir die Studierenden der LV IBPW zur freiwilligen Teilnahme an der Nachumfrage aufgefordert. Diese fand im Anschluss an den obligatorischen Semester-Endtest statt. Der Fragebogen wurde von 251 Studierenden beantwortet.

Ich erachte den Semesterend-Test als nutzlos für die Prüfungsvorbereitung und er könnte somit ersatzlos gestrichen werden.

trifft voll zu	<u>2</u> (1%)	
trifft zu	<u>8</u> (3%)	
trifft teilweise zu	<u>20</u> (8%)	■
trifft weniger zu	<u>43</u> (17%)	■
trifft nicht zu	<u>167</u> (67%)	■
keine Antwort	<u>11</u> (4%)	

A.2 Auszug aus den Resultaten der Nachumfrage EvIM

Im Herbstsemester 2011 haben wir die Studierenden der LV EvIM zur freiwilligen Teilnahme an der Nachumfrage aufgefordert. Diese fand im Anschluss an die obligatorische Semester-Endprüfung statt. Der Fragebogen wurde von 67 Studierenden beantwortet.

Die Probepfprüfung hat dazu beigetragen, dass ich ... (mehrere Antworten möglich)

...meine Wissenslücken besser kannte.	<u>54</u> (81%)	■
...ruhiger an die Semesterend-Prüfung ging, da ich den Ablauf der Prüfung bereits kannte.	<u>47</u> (70%)	■
...gewisse Inhalte noch einmal gezielt repetiert habe.	<u>51</u> (76%)	■
...Vertrauen in das Prüfungssystem SIOUX gewonnen habe.	<u>31</u> (46%)	■
...weniger als geplant für die Informatikprüfung gelernt habe.	<u>17</u> (25%)	■
...mehr als geplant für die Informatikprüfung gelernt habe.	<u>11</u> (16%)	■
...der Semesterend-Prüfung besorgt entgegenblickte.	<u>11</u> (16%)	■
...der Semesterend-Prüfung gelassen entgegenblickte.	<u>26</u> (39%)	■
...Angst vor der Semesterend-Prüfung hatte.	<u>5</u> (7%)	■
anderes:	<u>2</u> (3%)	

A.3 komplette Liste der Risikoanalyse

Auf der Folgeseite ist die komplette Liste der Risikoanalyse abgedruckt.

Z-ID	R-ID	Beschreibung	W-Proz. gem. Zakrezesw	Wahrscheinlichkeit	Datenquelle	Wie viel D	Wer D1	Wer D2	Wer D3	Wer D4	Wann?	Wann?	Wie viel	Wer	K (gem. Zakrezesw)	K (berechnet)	Konsequenzen	G-Potential (gewichtet)	G-Potential (gewichtet)
P1	P1	CBA nicht im Curriculum (Lehrplan) integriert	0.6	0.6	1	4	6	5			B	P	Gesamte Prüfung	St/Do	0.33	0.57	0.57	0.34	0.34
P2	P2	Prüfungsfragen widerspiegeln nicht das Kompetenzniveau der LV	0.5	0.5	2	5	6	5			B	P	Lehrveranstaltung	St/Do	0.36	0.71	0.71	0.36	0.36
P3	P3	Unpassende oder schlecht konstruierte Prüfungsfragen	0.75	0.75	1	5	6	5	4	3	B	E	Lehrveranstaltung	St/Do/Pl/PA	0.66	0.71	0.71	0.54	0.54
P4	P4	Unpassender Einsatz von CBA	0.3	0.3	2	5	6	5	4	3	B	P	Lehrveranstaltung	St/Do/Pl	0.3	0.57	0.57	0.21	0.21
P5	P5	Prüfung beinhaltet Fehler	0.6	0.6	2	4	6	5	4	3	D Po	D/K/A	0.3	0.57	0.57	0.34	0.34		
P6	P6	Studierende haben Angst vor der neuen Prüfungsmethode	0.15	0.48	3	1	6	5	4		B	P	Student, Studentin	St/Do/Pl/PA	0.9	0.14	0.14	0.07	0.07
P7	P7	Mangelndes Interesse von Dozierenden	0.7	0.95	3	5	2				B	P	Lehrveranstaltung	Do/Ma	0.8	0.60	0.60	0.57	0.57
P8	P8	Studierende lehnen CBA aktiv ab	0.01	0.03	3	1	6	2			B	P	Student, Studentin	Do/Ma	0.17	0.14	0.14	0.00	0.00
P9	P9	Fehlende Dokumentation des Prüfungsprozesses	0.87	0.87	1	6	5	4	3	1	B	E	Bildungsinstitution	Do/Pl/PA/Su	0.696	0.71	0.71	0.62	0.62
P10	P10	Vergleichte Prüfungen im Umlauf	0.9	0.9	2	1	6	5			B	E	Student, Studentin	St/Do	0.14	0.14	0.14	0.13	0.13
O1	O1	Unrealistischer Zeitplan für CBA	0.4	0.2	2	1	5	4	1	B	P	Lehrveranstaltung	St/Do/Su	0.5	0.60	0.60	0.12	0.12	
O2	O2	Kurzfristiger Aufruf von Aufsichtspersonal	0.75	0.75	2	3	5	4	3	1	D	P	Session	Do/Pl/PA/Su	0.8	0.36	0.36	0.37	0.40
O3	O3	Auswertung der Prüfungs-Arbeitsstationen für CBA ungeeignet (Bildschirmauflösung, etc.)	0.2	0.2	2	2	5	4	2	B	P	Raum	Do/Pl/Ma	0.24	0.24	0.24	0.05	0.05	
O4	O4	Prüfungs-Arbeitsstationen nicht ideal positioniert (Bildschirm einsehbar)	0.9	0.9	1	2	5	4	2	1	B	P	Raum	Do/Pl/Ma/Su	0.8	0.24	0.24	0.21	0.21
O5	O5	Anzahl Studierenden zu hoch für CBA (maximal verfügbare Anzahl an Computer Arbeitsplätzen)	0.7	0.7	1	5	4	2			B	P	Lehrveranstaltung	Do/Pl/Ma	0.3	0.60	0.60	0.42	0.42
O6	O6	Studierende lösen die korrekte Prüfung in einer falschen Session	0.15	0.95	3	2	6	3			D	D	Raum	St/PA	0.17	0.29	0.29	0.27	0.41
O7	O7	Studierende sind nicht vertraut mit dem Prüfungsmodus	0.7	0.7	2	5	6	5			D	D	Lehrveranstaltung	St/Do	0.71	0.71	0.71	0.30	0.75
O8	O8	Studierende erscheinen zu spät zur Prüfung	0.8	0.5	2	1	6	3			D	D	Student, Studentin	St/PA	0.2	0.14	0.14	0.07	0.11
O9	O9	Studierende verlassen die Prüfung vorzeitig	0.85	0.85	1	4	4	3			D	D	Gesamte Prüfung	Pl/PA	0.32	0.38	0.38	0.32	0.69
O10	O10	Student hat sich nicht für die Prüfung angemeldet, deshalb ist keine Prüfung verfügbar	0.05	0.05	2	1	6	4	3	D	D	Student, Studentin	St/Pl/PA	0.14	0.14	0.14	0.01	0.01	
O11	O11	Raumlicht ist nicht vertraut mit den Einzelheiten des CBA Systems	0.5	0.5	2	4	6	5	4	3	B D	E/J	Gesamte Prüfung	St/Do/Pl/PA	0.57	0.57	0.57	0.29	0.43
O12	O12	Studierenden mit speziellen Handycaps (rot/grün-Blindeheit, etc.) werden nicht unterstützt	0.99	0.99	1	3	6	5	4	3	D	D	Session	St/Do/Pl/PA	0.4	0.43	0.43	0.42	0.64
O13	O13	Studierende weiche kein CBA haben, sind beeinträchtigt, da keine Computerarbeitsplätze mehr zu Verfügung stehen	0.37	0.37	1	1	6				D	D	Student, Studentin	St	1	0.14	0.14	0.05	0.08
O14	O14	Unklare Rollenverteilung führt zu Verwirrung	0.87	0.87	1	4	6	4	3	B D Po	E/D/K/A	Gesamte Prüfung	St/Pl/PA	1	0.57	0.57	0.30	0.75	
O15	O15	Fehlmanipulation des CBA Systems aufgrund fehlender Instruktion	0.62	0.62	1	4	6	4	3	D	D	Gesamte Prüfung	St/Pl/PA	0.8	0.57	0.57	0.35	0.53	
O16	O16	Kommunikation zwischen zwei Prüfungssessionen möglich	0.62	0.62	1	4	6	5	4	3	D	D	Gesamte Prüfung	St/Do/Pl/PA	1	0.57	0.57	0.35	0.53
O17	O17	Keine Prüfung verfügbar bei Prüfungsbeginn	0.37	0.37	1	1	6	5	2	1	Po	K	Student, Studentin	St/Do/Ma	0.17	0.14	0.14	0.05	0.05
O18	O18	Studierende manipulieren die Prüfungs-Arbeitsstationen	0.37	0.37	1	6	6	5	2	1	D	D	Bildungsinstitution	St/Do/Ma/Su	0.32	0.86	0.86	0.32	0.48
O19	O19	Vorzugsweise Veröffentlichung von Prüfungsfragen durch Sicherheitslücken	0.4	0.4	1	4	5	4	2		B	E	Gesamte Prüfung	Do/Pl/Ma	0.6	0.48	0.48	0.19	0.19
O20	O20	Unerlaubter Zugriff auf Antworten	0.35	0.35	1	4	6	5	2		B	E	Gesamte Prüfung	St/Do/Ma	0.3	0.57	0.57	0.20	0.20
O21	O21	Studierenden können sich am System nicht anmelden (Passwort vergessen)	0.37	0.37	2	4	6	2			D	D	Gesamte Prüfung	St/Ma	0.57	0.57	0.57	0.11	0.17
O22	O22	Vom Prüfungs-Server kann keine Prüfung bezogen werden	0.37	0.37	1	3	6	4	3		D	D	Session	St/Pl/PA	0.4	0.43	0.43	0.16	0.24
O23	O23	Studierende wählen die falsche Prüfung	0.37	0.37	1	1	6	3			D	D	Student, Studentin	St/PA	0.17	0.14	0.14	0.05	0.08
O24	O24	Prüfungs-Arbeitsstation stürzt während der Bearbeitung der Prüfung ab	0.4	0.4	1	1	6	3			D	D	Student, Studentin	St/PA	0.2	0.14	0.14	0.06	0.09
O25	O25	Prüfungs-Server stürzt während einer laufenden Prüfung ab	0.1	0.1	1	6	6	4	3	1	D	D	Bildungsinstitution	St/Pl/PA/Su	0.5	0.86	0.86	0.09	0.13
O26	O26	Prüfungs-Arbeitsstation verliert während Prüfung die Verbindung zum Server	0.12	0.1	1	2	6	4	3	1	D	D	Raum	St/Pl/PA/Su	0.07	0.29	0.29	0.03	0.04
O27	O27	Studierende greifen während einer laufenden Prüfung auf unerlaubte (z.B. web-basierte) Ressourcen zu	0.37	0.37	1	6	6	4	3		D	D	Bildungsinstitution	St/Pl/PA	1	0.86	0.86	0.32	0.48
O28	O28	Studierende verwenden verschiedene Betriebssysteme	0.62	0.99	3	1	6	3			D	D	Student, Studentin	St/PA	1	0.14	0.14	0.14	0.21
O29	O29	Mangelnde Usability der Prüfungsclient-Oberfläche	0.62	0.62	1	6	6	5	4		B	E	Bildungsinstitution	St/Do/Pl	1	0.86	0.86	0.53	0.53
O30	O30	Ein unerwartetes Resultat	0.2	0.2	1	7	6	5	3	Po	K/A	Bildungsinstitution	St/Do/PA	0.2	1.00	1.00	0.20	0.20	
O31	O31	Netzwerklast während der Prüfung ist zu hoch	0.62	0.62	1	4	6	4	3		D	D	Gesamte Prüfung	St/Pl/PA	0.1	0.57	0.57	0.35	0.53
O32	O32	Serverlast während der Prüfung ist zu hoch	0.8	0.8	2	4	6	4	3		D	D	Gesamte Prüfung	St/Pl/PA	0.57	0.57	0.57	0.46	0.69
O33	O33	Antworten werden nicht korrekt gespeichert	0.2	0.2	1	7	6	5	4	2	Po	K/A	Bildungsinstitution	St/Do/Pl/Ma	0.5	1.00	1.00	0.20	0.20
O34	O34	Antworten gehen verloren	0.3	0.4	2	7	6	5	4	2	Po	K/A	Bildungsinstitution	St/Do/Pl/Ma	0.5	1.00	1.00	0.40	0.40
O35	O35	Erfolgreicher Rekurs aufgrund nicht ausreichender Beweisführung	0.5	0.5	2	7	6	5	2	B D Po	E/D/K/A	Bildungsinstitution	St/Do/Pl/Ma	1.5	1.00	1.00	0.50	0.75	
O36	O36	Fehlen von Unterstützung durch die Schulleitung	0.7	0.2	2	6	5	2			B	P	Bildungsinstitution	Do/Ma	0.6	0.71	0.71	0.14	0.14
O37	O37	CBA System ist nicht kosten-effizient	0.45	0.45	1	6	5	2	1		B	P	Bildungsinstitution	Do/Ma/Su	0.4	0.71	0.71	0.32	0.32
O38	O38	Dozierende wenden sich von CBA ab, da Aufwand zu hoch	0.95	0.95	2	4	5	2			B	P	Gesamte Prüfung	Do/Ma	0.8	0.48	0.48	0.02	0.02
O39	O39	Aufwändige Korrektur aufgrund mangelhafter Prüfungsfragen	0.4	0.4	2	4	5	4			Po	K	Gesamte Prüfung	Do/Pl	0.48	0.48	0.48	0.19	0.19

A.4 XML-Prüfungsdokument

Nachfolgend ist ein Beispiel eines Sioux-XML-Prüfungsdokuments abgedruckt.

```

<examSubmission id="0" clientSequenceNumber="8">
  <exam id="33" name="Testprüfung" duration="60" distributionTimestamp="2014-05-
05T13:58:31.095+02:00" groupName="ICT">
    <examInfo id="33">
      <content id="33">&lt;html&gt;
&lt;head&gt;
  &lt;style type="text/css"&gt;&gt;
    &lt;!--
      body { font-size: 14pt }
    --&gt;
    &lt;/style&gt;
&lt;/head&gt;
&lt;body&gt;
  &lt;p style="margin-top: 0"&gt;&gt;
    &lt;font face="Arial"&gt;Informationen zur Leistungskontrolle k&#246;nne
hier
    eingef&#252;gt werden. Diese werden zu Beginn der Pr&#252;fung angezeigt und
m&#252;ssen von den Studierenden akzeptiert werden.&lt;/font&gt;
    &lt;/p&gt;
&lt;/body&gt;
&lt;/html&gt;
      </content>
    </examInfo>
    <assignmentGroup id="53" name="Fragegruppe 1">
      <assignmentTransport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="taskItemAssignment" name="Testfrage" author="markusd" groupName="ICT" poolId="54"
poolName="test" repositoryId="484" repositoryName="Testfrage" repositoryAuthor="markusd" re-
visionNumber="0" creationDate="2014-05-05T13:42:39+02:00" revisionComment="first import"
id="1875">
        <task xsi:type="simpleTask" id="9000">
          <estimate xsi:type="simpleTaskEstimate" readingTimeAuthor="1" readingTimeReviewer="1"
readingDifficultyAuthor="0" readingDifficultyReviewer="0" id="1875" impressionAuthor="0" imres-
sionRevisor="0"/>
          <content id="9014">&lt;html&gt;
&lt;head&gt;
  &lt;style type="text/css"&gt;&gt;
    &lt;!--
      body { font-size: 14pt }
    --&gt;
    &lt;/style&gt;
&lt;/head&gt;
&lt;body&gt;
  &lt;font face="Arial"&gt;&gt;Hier steht der &lt;b&gt;Inhalt der Aufgabenstel-
lung&lt;/b&gt;.
  Dieser kann auch Medien beinhalten und wird als html-Code interpretiert.&lt;/font&gt;
  &lt;/body&gt;
&lt;/html&gt;
          </content>
        </task>
        <item xsi:type="clozeItem" scoreCorrectAnswer="1.0" scoreFalseAnswer="0.0" selected-
ForTest="false" id="8998">
          <content id="9012">&lt;html&gt;
&lt;head&gt;
  &lt;style type="text/css"&gt;&gt;
    &lt;!--
      body { font-size: 14pt }
    --&gt;
    &lt;/style&gt;
&lt;/head&gt;
&lt;body&gt;
  &lt;p style="margin-top: 0"&gt;&gt;
    &lt;font face="Arial"&gt;Das ist eine Fragestellung des Typs
&lt;i&gt;L&#252;ckentext&lt;/i&gt;. &lt;/font&gt;&lt;br&gt;Und
    hier folgt die L&#252;cke &lt;input name="CLOZE_1"&gt;
&lt;input type="text"&gt;. Es k&#246;nne
    auch mehrere sein.
    &lt;/p&gt;
&lt;/body&gt;
          </content>
        </item>
      </assignmentTransport>
    </assignmentGroup>
  </examInfo>
</exam>

```

```

</html>&gt;
  </content>
    <solution xsi:type="clozeSolution" id="7138">
      <clozeElementSolution xsi:type="clozeTextSolution" ignoreSpaces="true" ignoreTab-
bes="true" ignoreCase="true" id="2699" gapNumber="1">
        <clozeTextSolutionEntry id="7139" normalizedSolution="Das ist die Mus-
terlösung."/>
        <clozeTextSolutionEntry id="7140" normalizedSolution="Und das eine Variante
davon."/>
      </clozeElementSolution>
    </solution>
  </item>
  <item xsi:type="hotSpotItem" scoreCorrectAnswer="1.0" scoreFalseAnswer="0.0" selected-
ForTest="false" id="8999">
    <content id="9013">&lt;html&gt;
&lt;head&gt;
  &lt;style type="text/css"&gt;&gt;
  &lt;!--
    body { font-size: 14pt }
  --&gt;
  &lt;/style&gt;
&lt;/head&gt;
&lt;body&gt;
  &lt;p style="margin-top: 0"&gt;&gt;
    Das ist eine Fragestellung des Typs &lt;i&gt;HotSpot&lt;/i&gt;. &lt;br&gt;Die
L&#246;sung muss
    in der Aufgabenstellung angeklickt werden.
  &lt;/p&gt;
&lt;/body&gt;
&lt;/html&gt;
    </content>
    <solution xsi:type="hotSpotSolution" allAnswersNecessary="true" id="7139">
      <hotSpotSolutionRectangle id="1245" x="157" y="6" width="15" height="13"/>
      <hotSpotSolutionRectangle id="1246" x="77" y="11" width="30" height="21"/>
    </solution>
  </item>
  <item xsi:type="mcItem" scoreCorrectAnswer="1.0" scoreFalseAnswer="0.0" selected-
ForTest="false" id="8997">
    <content id="9011">&lt;html&gt;
&lt;head&gt;
  &lt;style type="text/css"&gt;&gt;
  &lt;!--
    body { font-size: 14pt }
  --&gt;
  &lt;/style&gt;
&lt;/head&gt;
&lt;body&gt;
  &lt;p style="margin-top: 0"&gt;&gt;
    &lt;font face="Arial"&gt;Das ist eine Fragestellung des Typs
&lt;i&gt;true/false&lt;/i&gt;. &lt;/font&gt;
  &lt;/p&gt;
&lt;/body&gt;
&lt;/html&gt;
    </content>
    <solution xsi:type="mcSolution" correctResponse="true" id="7137"/>
  </item>
</assignmentTransport>
</assignmentGroup>
</exam>
<answerTransport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="clozeAnswer"
answerCertain-ty="0" studentComment="Hier verstehe ich die Aufgabenstellung nicht." id="0"
examId="33" item-Id="8998" legiNumber="S97102362" clientSequenceNumber="2" cli-
entTimestamp="2014-05-05T13:59:11.812+02:00">
  <clozeElementAnswer xsi:type="clozeTextAnswer" answer="meine Antwort" id="0" gapNumber="1"/>
</answerTransport>
<answerTransport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:type="hotSpotAnswerSet" answerCer-tainty="0" id="0" examId="33" item-Id="8999" legi-
Number="S97102362" clientSequenceNumber="4" clientTimestamp="2014-05-05T13:59:16.654+02:00">
  <hotSpotAnswer id="0" x="235" y="60"/>
  <hotSpotAnswer id="0" x="95" y="13"/>
</answerTransport>
<answerTransport xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:type="mcAnswer"
answerCorrect="true" answerFalse="false" answerCertainty="0" id="0" examId="33" item-Id="8997"
legi-Number="S97102362" clientSequenceNumber="7" clientTimestamp="2014-05-
05T13:59:25.278+02:00"/>
</examSubmission>

```

A.5 Fehlerverteilungsmatrix

Auf der Folgeseite ist die Fehlerverteilungsmatrix der Resultate der EvIM-Semester-Endprüfung im Herbstsemester 2013 abgedruckt. An dieser Prüfung teilgenommen haben 279 Studierende.

A.6 Vergleich des Lösungsverhaltens der Studierenden über die Zeit

Nachfolgend ist die Ausgabe von SPSS bezüglich der Mittelwertvergleiche für kleine Stichproben nach Kruskal und Wallis abgedruckt. Ausgewertet wurde die Anzahl der korrekt beantworteten K3+-Prüfungsfragen der Jahre 2002-2004.

Ranks

	semester	N	Mean Rank
mittelwerte	2002	5	9,60
	2003	8	9,31
	2004	14	18,25
	Total	27	

Test Statistics^{a,b}

	mittelwerte
Chi-Square	8,346
df	2
Asymp. Sig.	,015

a. Kruskal Wallis Test

b. Grouping Variable: semester

Der paarweise Vergleich der Mittelwerte für kleine Stichproben nach Dunn ergab folgende SPSS-Ausgabe:

Each node shows the sample average rank of semester.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
2003-2002	,288	4,524	,064	,949	1,000
2003-2004	-8,938	3,517	-2,541	,011	,033
2002-2004	-8,650	4,134	-2,092	,036	,109

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is ,05.

A.7 Fragebogen Prozessanalyse

Im Herbstsemester 2011 haben wir die Studierenden der LV EvIM zur freiwilligen Teilnahme an der Prozessanalyse aufgefordert. Dazu wurde nach der Präsentation jedes Moduls automatisiert eine Email mit einem Hyperlink zu einem Fragebogen verschickt. Nachfolgend ist der komplette Fragebogen abgedruckt. Zudem werden die Resultate für Modul 1 angegeben.

Wie lange dauerte die Präsentation dieses Moduls?			
Antwort	Anzahl	Prozent	Graph
<5 min	59	36%	
5-10 min	85	52%	
11-15 min	14	9%	
>15 min	3	2%	
keine Antwort	1	1%	
Total Teilnehmer	162		

Haben Sie die Übung alleine präsentiert?			
Antwort	Anzahl	Prozent	Graph
ja, alleine	131	81%	
nein, zu zweit	30	19%	
keine Antwort	1	1%	
Total Teilnehmer	162		

Ich konnte bei dieser Präsentation zeigen, dass ich die selbständige Aufgabe verstanden habe.			
Antwort	Anzahl	Prozent	Graph
trifft voll zu	39	24%	
trifft zu	84	52%	
trifft teilweise zu	29	18%	
trifft weniger zu	6	4%	
trifft nicht zu	3	2%	
keine Antwort	1	1%	
Total Teilnehmer	162		

Wenn ich mich mit meinen Mits Studierenden vergleiche, dann war mein Wissen in diesem Modul...			
Antwort	Anzahl	Prozent	Graph
...klar überdurchschnittlich	3	2%	
...eher überdurchschnittlich	32	20%	
...durchschnittlich	104	64%	
...eher unterdurchschnittlich	10	6%	
...klar unterdurchschnittlich	2	1%	
keine Antwort	11	7%	
Total Teilnehmer	162		

**Angenommen es gäbe jetzt einen Test über den bisherigen Stoff, welche Note würde ich erwarten?
(Notenskala: 6= sehr gut, 5=gut, 4=genügend)**

Antwort	Anzahl	Prozent	Graph
6	2	1%	
5.5	15	9%	
5	66	41%	
4.5	51	31%	
4	18	11%	
<4	8	5%	
keine Antwort	2	1%	
Total Teilnehmer	162		

Welche Leistung erwarte ich aus der jetzigen Sicht an der Schlussprüfung? (Notenskala: 6= sehr gut, 5=gut, 4=genügend)

Antwort	Anzahl	Prozent	Graph
6	0	0%	
5.5	10	6%	
5	53	33%	
4.5	73	45%	
4	16	10%	
<4	4	2%	
keine Antwort	6	4%	
Total Teilnehmer	162		

A.8 Prospektive Analyse der Selbst- und Fremdbeurteilung

Auf der nachfolgenden Seite ist die deskriptive Statistik ausgewählter und in der Arbeit erwähnter Daten des eDWH abgedruckt.

Statistics

P_m_pktQuart	Modul			Zscore (PA_m_A)	Zscore (PA_m_SE_S)
1	1	N	Valid	98	41
			Missing	0	57
	2	N	Valid	79	24
			Missing	0	55
	3	N	Valid	134	35
			Missing	1	100
	4	N	Valid	108	24
			Missing	1	85
	5	N	Valid	80	19
			Missing	0	61
	6	N	Valid	69	18
			Missing	1	52
2	1	N	Valid	40	16
			Missing	0	24
	2	N	Valid	102	36
			Missing	1	67
	3	N	Valid	17	1
			Missing	0	16
	4	N	Valid	35	6
			Missing	0	29
	5	N	Valid	56	10
			Missing	1	47
	6	N	Valid	76	16
			Missing	1	61
3	1	N	Valid	76	27
			Missing	0	49
	2	N	Valid	49	7
			Missing	0	42
	3	N	Valid	65	14
			Missing	0	51
	4	N	Valid	64	11
			Missing	0	53
	5	N	Valid	79	18
			Missing	0	61
	6	N	Valid	66	10
			Missing	0	56
4	1	N	Valid	59	25
			Missing	0	34
	2	N	Valid	42	9
			Missing	0	33
	3	N	Valid	56	11
			Missing	0	45
	4	N	Valid	65	9
			Missing	0	56
	5	N	Valid	57	6
			Missing	0	51
	6	N	Valid	60	8
			Missing	0	52

Verzeichnisse

Abbildungsverzeichnis

- Abbildung 1.1: Illustration der vier Schwerpunkte dieser Arbeit. Das CBA-System bildet die Grundlage für die sichere Erfassung von qualitativ hochwertigen Leistungsdaten. Die Leistungsdaten dienen einerseits als Selektions- und Feedbackinstrument als auch als Grundlage für deren lehrbezogene Analyse (Learning Analytics). In Klammern findet sich ein Verweis auf das entsprechende Kapitel in dieser Arbeit. 4
- Abbildung 2.2: Als Kompetenz wird die bewusste, selbstorganisierte Anwendung von Wissen auf der Basis einer unbekanntem Problemstellung im Problemlösungsprozess verstanden. Eine Kompetenz kann nur beobachtet, nicht aber direkt gemessen werden (Erpenbeck, 2007). 12
- Abbildung 2.3: Wissen ist an eine Person gebunden und kann nur indirekt gemessen oder zu anderen Personen transferiert werden (Externalisierung von Wissen). 14
- Abbildung 2.4: Von einer kognitiven Leistung kann erst dann gesprochen werden, wenn Wissen in einem kognitiven Prozess zur Anwendung kommt. Wissen alleine ist dimensionslos und kann nicht bewertet werden (F = Fakten-Wissen, K=Konzept-Wissen, P=Prozedurales-Wissen, M=Metakognitives-Wissen). 17
- Abbildung 2.5: Taxonomie der kognitiven Leistungen nach Anderson und Krathwol (L. Anderson et al., 2001). Sie bringt die Wissensarten und die kognitiven Prozessdimensionen in eine Beziehung und ermöglicht so die explizite Adressierung einer kognitiven Leistung. 19
- Abbildung 2.6: Explizites (weiss) und implizites Wissen (grau) übertragen auf die Taxonomie der kognitiven Leistungen nach Anderson. Da sich explizites Wissen über alle kognitive Prozessdimensionen verteilt, können nur beim Messen von implizitem Wissen die Zieldimensionen von K3+ verlässlich erreicht werden. 25
- Abbildung 3.7: Die Leistungskontrollen können aus verfahrenstechnischer und didaktischer Sicht klassiert werden. Diese beiden Klassierungen ergänzen sich. 31
- Abbildung 3.8: Unterscheidung zwischen formativen und summativen Leistungskontrollen. Die formativen LK sind Teil einer Unterrichtseinheit (UE) und liefern den Studierenden und Dozierenden Feedback zu den individuellen Leistungsständen. Das Feedback ist ein wichtiges Element im Lernprozess (vgl. dazu Abschnitt 3.4 "Leistungskontrollen als Feedbackinstrument"). Die summative LK bildet

den Abschluss des Unterrichts und soll den <i>Learning Outcome</i> messen. Er wird benotet und das Feedback ist optional.	32
Abbildung 3.9: Schematische Darstellung des Messfehlers einer Leistungskontrolle. Die beiden Achsen zeigen die tatsächliche (x-Achse) und gemessene Leistung (y-Achse). Die Idealmessung ist als gestrichelte Regressionsgerade dargestellt. Je kleiner die Streuung ausfällt, desto besser ist die Messung. Grau eingezeichnet sind die Bereiche der heiklen Messfehler. Die Bereiche A-D werden im Text erläutert. In der Praxis ist die fiktive tatsächliche Leistung nicht bekannt und kann nur mittels Aussenkriterien annähernd bestimmt werden (vgl. Kapitel 3.3.5 "Validität (Gültigkeit der Messung)").	33
Abbildung 3.10: Die drei Hauptgütekriterien und zwei Nebenkriterien beeinflussen die Qualität des Resultats einer Leistungskontrolle. Diese ist abhängig von drei Hauptgütekriterien: Objektivität, Reliabilität und Validität. Wobei erstere meist durch den Modus und die anderen beiden durch den Inhalt der Leistungskontrolle beeinflusst werden.	35
Abbildung 3.11: Abhängigkeit der Trennschärfe von der Aufgabenschwierigkeit (abgeändert nach Lienert & Raatz, 1998).	39
Abbildung 3.12: Schematische Darstellung der schwierigkeitskorrigierten Trennschärfe durch den Vergleich der korrekt abgegebenen Antwort mit der Leistung der Studierenden. Die Fläche unter der Linie stellt die Schwierigkeit (P) dar, die Steigung entspricht der schwierigkeitskorrigierten Trennschärfe (r_{corr}). Die mit A-G markierten Linien stellen unterschiedlich schwierige und trennscharfe Aufgaben dar. Weitere Informationen finden sich im Text (abgeändert nach Lienert & Raatz, 1998).	40
Abbildung 3.13: Qualitätsmetriken der Leistungskontrolle im Überblick. Für die Messung der Validität werden Aussenkriterien benötigt, die mit den Ergebnissen der LK verglichen werden.	41
Abbildung 3.14: Das erweiterte didaktische Dreieck zeigt die formative Leistungskontrolle als Feedbackelement bezüglich der drei Komponenten: Studierende, Lerninhalte und Dozierende.	44
Abbildung 3.15: Die Lernumgebung dargestellt als Black Box. Je nach Input, sowie der Art und dem Zeitpunkt der Einflüsse verändert sich der Output.	47
Abbildung 3.16: Hierarchie der Computer gestützten Leistungskontrollen in der Übersicht. CAA bildet dabei die Obergruppe und beinhaltet die CBA und OA Systeme. Prozesstechnisch unterscheiden sich die Systeme hauptsächlich in der Art, wie die Prüfung durchgeführt wird. Abbildung abgeändert nach (Conole & Warburton, 2005).	50
Abbildung 3.17: Vereinfachte Funktionsweise einer OA Architektur: Der Client ist über ein Netzwerk permanent mit dem Server verbunden. Der Client speichert keine	

Prüfungsinhalte, sondern lädt bei Bedarf alle Inhalte vom Server. Die Inhalte werden auf dem Client mittels eines Web-Browsers gerendert und angezeigt.....	52
Abbildung 4.18: Verfahren zur Verfügung einer Note. Bei einem Prüfungsblock (bestehend aus mehreren Einzelprüfungen, welche es als Ganzes zu bestehen gilt) beantragt die Notenkonferenz, die Verfügung der Noten beim Rektorat. Bei Einzelprüfungen teilt in der Regel der Examinator die Noten den Studierenden direkt mit. Beschwerdefähige Notenverfügungen werden nur auf Verlangen ausgestellt.....	55
Abbildung 4.19: Juristische Wege zur Anfechtung einer Notenverfügung.	56
Abbildung 3.20: Zwei juristische Herausforderungen bei CBA: Nachvollziehbarkeit des Prüfungsergebnisses und Gleichbehandlung aller Studierenden.....	59
Abbildung 4.21: Bei einer Papier-basierten Prüfung bleibt das "Prüfungsdokument" ganzheitlich bei den Studierenden und ist für diese immer sichtbar. Bei CBA verlassen hingegen die Zwischenresultate den Prüfungsraum und sind für die Studierenden nicht mehr als Ganzes einsehbar.	61
Abbildung 4.22: Stakeholder bei CBA.	64
Abbildung 4.23: Risikomanagement Modell bei CBA. Die Risiken bei CBA variieren je nach Bildungsinstitution. Im Schritt 1 "Statusanalyse" wird den unterschiedlichen Bedingungen Rechnung getragen. Diese fließen in die Schritte 2 bis 5 ein.....	65
Abbildung 4.24: Rangierte Risiken. Die Risiken sind gemäss ihres Gefahrenpotentials absteigend sortiert. Die gestrichelte Linie bezeichnet den <i>cutoff</i> -Wert bei 0.04, unterhalb derer die Risiken als vernachlässigbar eingestuft werden.....	74
Abbildung 4.25: Kommunikation des individuellen Prüfungstermins (Datenbasis: IBPW FS10).	78
Abbildung 4.26: War die Lautsprecher-Durchsage zu Beginn der Prüfung klar verständlich? (Datenbasis: IBPW FS10).....	78
Abbildung 4.27: Organisation der CB-Prüfung (Datenbasis: IBPW FS10).....	79
Abbildung 4.28: Subjektives Empfinden der Vergleichbarkeit der Prüfungsbedingungen durch die Studierenden (Datenbasis: EvIM HS10).	79
Abbildung 4.29: <i>Usability</i> der Sioux-Prüfungssoftware (Datenbasis: EvIM HS10).....	80
Abbildung 4.30: Grundsätzliche Zufriedenheit der Studierenden mit dem Prüfungssystem (Datenbasis: EvIM HS10).....	80
Abbildung 4.31: Bedenken der Studierenden bezüglich CBA (Datenbasis: IBPW FS11).....	81

- Abbildung 4.32: Visualisierung der Freitextrückmeldungen der Studierenden bezüglich ihrer Bedenken bei CBA (Datenbasis: IBPW FS11, Antwort gaben nur jene Studierenden, welche bei der Frage 3.15 Bedenken geäußert haben. n=21). Die Wortwolke wurde mittels Wordle-Applikation unter www.wordle.net erstellt. 81
- Abbildung 5.33: Speicherung der Antworten bei aktuellen Browser-basierten CBA-Systemen geschieht grob in sechs Schritten (hier illustriert am Beispiel einer *Multiple Choice* (MC) Frage): 1. Die Antwort wird durch Anklicken des Radiobuttons abgegeben. 2. Der Radiobutton ändert den Zustand (von "nicht markiert" auf "markiert" oder umgekehrt). 3. Die Antwort wird an den Server übermittelt und dort gespeichert. 4. Der gespeicherte Wert wird aus der Datenbank gelesen und die aktualisierte Webseite auf den Client übertragen. 5. Die Webseite wird auf dem Client gerendert (dargestellt). 6. Die gespeicherte Antwort kann abgelesen werden.95
- Abbildung 4.34: Digitale Signierung eines Dokuments. Die eindeutige Prüfsumme (*Hash*) eines Dokuments wird mittels eines digitalen Schlüssels chiffriert. So kann nachträglich die Unversehrtheit des Dokuments kontrolliert werden. 100
- Abbildung 5.35: Prinzip von asymmetrischen Verschlüsselungsverfahren. Diese beruhen auf Einwegfunktionen und sind nur mit enormem Rechenaufwand umkehrbar. Das heisst, eine mit einem Signaturschlüssel signierte Nachricht kann mit dem Signaturprüf Schlüssel lediglich verifiziert, nicht aber neu signiert werden.101
- Abbildung 5.36: Überprüfung der digitalen Signatur eines Dokuments. Stimmt der in der digitalen Signatur chiffrierte *Hash*-Wert mit dem basierend auf den Daten neu berechneten *Hash*-Wert überein, ist die digitale Signatur gültig.102
- Abbildung 5.37: Auswirkungen der durch die CBA-Designprinzipien verursachten Einschränkungen auf den CBA-Prüfprozess. Falls Designprinzipien auf die gleiche Sicherheitsphase einwirken, können Sie potentiell in Konkurrenz zueinander stehen. Ein solches Beispiel ist mit einem Pfeil markiert und wird im Text erläutert. Die Auswirkungen der durch die Designprinzipien induzierten Risikoquellen (#A-#C) sind mit einem "X" markiert und werden im Text erläutert. Der Arbeitsschritt "Antworten eingeben" birgt das grösste Sicherheitsrisiko, da alle drei Risiken eintreten können.... 109
- Abbildung 5.38: Sicherheitsrelevante Risiken beim Verzicht auf qualifizierte digitale Signaturen. Die Studierenden sind im Besitz ihres elektronisch lesbaren, privaten Signaturschlüssels. Dieser kann weitergegeben, kopiert, verändert oder gelöscht werden. In der Abbildung sind sechs Möglichkeiten erläutert. Die Nummern (1-6) entsprechen der Nummerierung der Risiken in Tabelle 5.14. 113
- Abbildung 6.39: Systemübersicht der Sioux CBA-Suite. Die SSL-Tunnelverbindungen illustrieren, auf welche Ressourcen die einzelnen Clients Zugriff haben. 121

- Abbildung 6.40: Funktionsumfang der Sioux Client-Applikationen in Bezug auf den Prüfungsprozess. Die grauen Arbeitsschritte (1,2,3,5) werden vom *Sioux AuthorClient* unterstützt, bei der eigentlichen Prüfungsdurchführung (4) kommen der *Sioux ExamClient* und der *TacticalClient* zum Einsatz..... 122
- Abbildung 6.41: "ToDo" Mitteilungen unterstützen den *Peer-Review*-Prozess zur Verbesserung der Prüfungsfragen..... 123
- Abbildung 6.42: Festlegen des Modus einer Leistungskontrolle im *AuthorClient*. Festzulegen sind: Art der Leistungskontrolle, Prüfungsdauer (in Minuten), Darstellungsoptionen, Angabe des Zeitfensters und des IP-Bereichs..... 125
- Abbildung 6.43: Der *AuthorClient* im Korrekturmodus. Wahlweise können nur noch die falsch abgegeben Antworten angezeigt werden. Dies erleichtert das Finden von offensichtlichen Tippfehlern oder unvollständigen Musterlösungen. Mittig wird die Fehlerverteilung des angezeigten Items dargestellt (hier abgebildet: Analyse der Frage 4.2, Item 2 der Basisprüfung, Februar 2012, n=17). 126
- Abbildung 6.44: Bedienoberfläche des *ExamClients*. Neben den eigentlichen Prüfungsfragen zeigt der *ExamClient* nur die unmittelbar zur Durchführung der Leistungskontrolle notwendigen Informationen an. Dazu gehört eine Übersicht über die Aufgaben (mit Fortschrittsanzeige), der Name des Probanden oder der Probandin (als Information für die Raumaufsicht und zu Durchführung der Präsenzkontrolle), die verbleibende Zeit sowie Elemente zur Navigation und Abgabe der Leistungskontrolle. 128
- Abbildung 6.45: Bedienoberfläche des *TacticalClients*. Diese Applikation zeigt während der Durchführung einer Leistungskontrolle zahlreiche Statusinformationen: Wer hat die Prüfung bezogen? Wie viel Prüfungszeit haben die Studierenden noch zur Verfügung (in diesem Beispiel hat der Student Markus Dahinden noch 53 Minuten)? Wie viele Antworten haben die Studierenden bereits abgegeben? An welchem Computer arbeitet ein bestimmter Student? Auf Wunsch kann die verbleibende Prüfungszeit individuell erhöht werden (aus Datenschutzgründen sind gewisse Informationen unscharf dargestellt). 131
- Abbildung 6.46: Sechs sicherheitskritische Vorgänge in der Durchführungsphase von CBA. Eine ausführlichere Beschreibung der einzelnen Vorgänge findet sich in den nachfolgenden Abschnitten. 136
- Abbildung 6.47: Anmeldevorgang bei Sioux. Der *JBoss Application Server* stellt sicher, dass nur authentifizierte Benutzer auf die Schnittstelle des *ExamServers* zugreifen können. 138
- Abbildung 6.48: Schlüsselpaar generieren. Das Schlüsselpaar wird verwendet, um die einzelnen Resultate und das Prüfungsdokument digital zu signieren. 140

- Abbildung 6.49: Eingeben von Antworten. Dieses Sequenzdiagramm illustriert die Schritte, wie Sioux die Antworten sicher speichert, ohne die Bearbeitungsgeschwindigkeit zu reduzieren. Dabei kommt der Warteschleife ("*ExamClient-queue*") eine entscheidende Bedeutung zu.....144
- Abbildung 6.50: Erstellen des Prüfungsdokuments und Abgeben der Leistungskontrolle. Das Prüfungsdokument wird mit Hilfe der Warteschleife an den Server übermittelt. Die Warteschleife stellt sicher, dass der *ExamClient* erst geschlossen werden kann, wenn die Warteschleife leer ist.....146
- Abbildung 6.51: Lastmessungen während der beiden Prüfungssessionen. Das Diagramm zeigt die Maximalwerte der Leerlauf (*idle*)-Zeit des Sioux-Servers, wobei 1 für 100% Leerlauf und 0 für die komplette Auslastung steht. Die Daten wurden alle 500 Millisekunden erfasst und zur Erstellung des Diagramms verwendet. Die erste Serie wurde um 14:00 Uhr, die zweite um 15:15 Uhr gestartet. Lastspitzen sind jeweils zu Beginn jeder Prüfungssession erkennbar und lassen sich mit der Übertragung der kompletten Prüfungsdaten erklären. Auffallend ist auch eine höhere Last jeweils vor Ende jeder Serie, welche vermutlich auf das eilige Abschicken noch offener Antworten zurückzuführen ist. Datenbasis: Basisprüfung, "Informatik für Biologie und Pharm. Wissenschaften", Dozent: Dr. Böckenhauer, 19.08.2011, n=272.153
- Abbildung 6.52: Geschätzte Wartezeit beim Wechsel der Antworten. Rund 30% gaben an, zwischen 1-3 Sekunden beim Wechseln der Prüfungsfragen gewartet zu haben (abgeändert aus (Dahinden, 2010c) n=62, Probeprüfung, "Einsatz von Informatikmitteln" von Prof. Dr. H. Hinterberger, HS2009).....156
- Abbildung 6.53: Medienfrage bei CBA-Prüfungen (n=56, Semesterendtest, "Informatik für Biologie und Pharm. Wiss" von Dr. H.-J. Böckenhauer, FS2011).156
- Abbildung 6.54: Prospektive Erwartungshaltung in Bezug zur benoteten CBA-Prüfung (n=56, Semesterendtest, "Informatik für Biologie und Pharm. Wiss" von Dr. H.-J. Böckenhauer, FS2011).157
- Abbildung 6.55: Design-Prinzip von MEASURED basierend auf dem *Sioux-BackEnd*. Das *Sioux-BackEnd* läuft auf dem Client als lokaler Webservice und übernimmt dort vollständig die Kommunikation zwischen Client und Server. Das *MEASURED Front-End* kommuniziert mit dem *Sioux-BackEnd*, welches alle sicherheitsrelevanten Funktionen wie digitale Signierung und Cachingfunktionen übernimmt..... 158
- Abbildung 7.56: Die drei Phasen des Entwicklungsmodells für CBA-Leistungsmessungen. Durch die wiederholte Anwendung des Modells steigt die Menge der gesammelten Frage-Metriken, welche zur Weiterentwicklung der Fragenhefte und des Fragenpools verwendet werden können. Gute Fragenpools ermöglichen valide Leistungsmessungen, deren Resultate für die Analyse und Optimierung des Unterrichts verwendet werden können..... 167

- Abbildung 7.57: Die Übersicht über das Entwicklungsmodell für CBA zeigt die vier Hauptprozesse: Planung, Erstellung, Durchführung und Auswertung. Die jeweiligen Subprozesse werden durch eigene Abbildungen illustriert.....168
- Abbildung 7.58: Detailansicht des Hauptprozesses "Planung". Pro Subprozess sind jeweils die Ziele angegeben. Die römischen Ziffern referenzieren die Hauptprozesse des Entwicklungsmodells in Abbildung 7.57. 169
- Abbildung 7.59: Detailansicht des Hauptprozesses "Erstellung". Er gliedert sich in vier Subprozesse, wobei die jeweiligen Ziele angegeben sind. Die mit einem Stern (*) markierten Schritte können erst bei wiederholter Anwendung des Modells ausgeführt werden. 171
- Abbildung 7.60: Effekt des Feedbacks auf das Erstellen und Optimieren der neu erstellten Fragen. (Datenbasis: Nachumfrage unter den Assistierenden, EvIM, HSo8, n=10)173
- Abbildung 7.61: Detailansicht des Hauptprozesses "Durchführung". Er gliedert sich in drei Subprozesse, wobei die jeweiligen Ziele angegeben sind.....173
- Abbildung 7.62: Flussdiagramm zur Klassierung und Optimierung der Reliabilität von Prüfungsfragen anhand deren Metriken. 176
- Abbildung 7.63: Grundstruktur einer Aufgabe. Links die Problemstellung mit praxisrelevanten Daten und rechts die vier voneinander unabhängigen Fragestellungen. Weitere Informationen und die Beschreibung der drei Typen von Fragestellungen befinden sich im Text..... 180
- Abbildung 7.64: Beispiel eines Frage-Items des Typs "*true/false*". Zur Beantwortung wird die Aufgabenstellung benötigt (hier nicht dargestellt). *True/false*-Items können entweder richtig oder falsch sein. Die Antwort wird durch Anklicken des entsprechenden Buttons auf der linken Seite übermittelt. Da der Ratefaktor bei *true/false*-Items mit 50% hoch ist, eignen sie sich besonders für einfache Fragen. (Item wurde eingesetzt in der formativen LK, Modul 5 "Daten verwalten II", Gdl und EvIM, HS 2013) 183
- Abbildung 7.65: Beispiel eines Frage-Items des Typs "*hotspot*". Zur Beantwortung wird das Markierkreuz an der korrekten Stelle in der Aufgabenstellung positioniert (hier nicht dargestellt). Der Ratefaktor ist tiefer als bei *true/false*-Items, allerdings eignet sich dieser Typ nur für eine begrenzte Anzahl Fragestellungen. Er lässt sich relativ einfach korrigieren. (Item wurde eingesetzt in der formativen LK, Modul 5 "Daten verwalten II", EvIM, HS 2013) 183
- Abbildung 7.66: Beispiel eines Frage-Items des Typs "*cloze*". Die Antwort wird durch Eintippen übermittelt. Es können mehrere Lücken verwendet werden. Zur Beantwortung dieser Frage wird die Aufgabenstellung benötigt (hier nicht dargestellt).

Der Ratefaktor ist tiefer als bei den anderen Fragetypen. Allerdings benötigt das Eintippen und Kontrollieren der Antwort mehr Zeit, was die Bearbeitungszeit dieses Item-Typs erhöht. Es empfiehlt sich deshalb speziell bei schwierigeren Fragen, wo es Sinn macht, die Ratewahrscheinlichkeit auf Kosten einer höheren Bearbeitungszeit zu reduzieren. Das Item ist vergleichsweise aufwändig zu korrigieren, da offensichtliche Tippfehler trotzdem als korrekt gewertet werden sollten.184

Abbildung 7.67: Semesterübersicht der LV "Einsatz von Informatikmitteln". Um die zwei Kreditpunkte zu erhalten, müssen alle 6 Module bearbeitet und die Semester-Schlussprüfung bestanden werden.185

Abbildung 7.68: Umfrage über die Organisation der Semester-Endprüfung. 96% der Studierenden gaben an, dass die Prüfung gut organisiert war. (Datenbasis: Nachumfrage unter den Studierenden, EvIM, HS13, n=100).....188

Abbildung 7.69: Fehlerverteilungsmatrix der Rohdaten der Semester-Endprüfung. In jeder Zeile sind die Resultate eines Probanden dargestellt. Weisse und rote Bereichen widerspiegeln korrekte respektive falsche Antworten. Die Probanden sind absteigend nach der erreichten Gesamtpunktzahl sortiert. Zuoberst stehend demnach die besten, zuunterst die schwächsten Studierenden. Das vertikale Fehlermuster gibt Auskunft über die Qualität eines Items. Die Itemmetriken sind unterhalb der Fehlerverteilungsmatrix aufgeführt (eine vergrößerte Darstellung dieser Fehlerverteilungsmatrix findet sich im Anhang A.5 dieser Arbeit). Die neunte Frage von rechts musste aufgrund einer mehrdeutigen Formulierung eliminiert werden. (Datenbasis: Semester-Endprüfung, EvIM, HS13, n=279)189

Abbildung 7.70: Anzahl der Items pro Trennschärfe-Kategorie. Positiv ist die klare Häufung der Items der höchsten Güteklasse. Dieses Resultat lässt eine reliable Messung vermuten. (Datenbasis: Semester-Endprüfung, EvIM, HS13, n=39)190

Abbildung 7.71: Umfrage über die Validität der Semester-Endprüfung. Die Resultate lassen vermuten, dass offenbar auch die von den Studierenden subjektiv wahrgenommene Validität die empirisch ermittelte Validität der LK der Semester-Endprüfung stützt. (Datenbasis: Semester-Endprüfung, EvIM, HS13, n=100).....191

Abbildung 7.72: Verhältnis von Trennschärfe und Schwierigkeit. Sichtbar ist eine Rechtsverschiebung der Schwierigkeitsindizes aufgrund des Messschwerpunkts rund um die Bestehensgrenze (Datenbasis: EvIM, HS13, 40 Items)192

Abbildung 7.73: Zeitliche Entwicklung der Trennschärfen verschiedener summativen LK. Erkennbar ist eine Abnahme der trennschwächeren Items (mit $p < 0.20$) und einer gleichzeitigen Zunahme der trennscharfen Items. Dieser Verlauf ist ein Indiz einer Qualitätssteigerung bei der Reliabilität unserer Prüfungen im Verlauf der letzten Jahre. Möglicherweise wird sich nun der Anteil der trennscharfen Items bei ca. 75% einpendeln. (Datenbasis: wie in der Legende angegeben).....194

Abbildung 8.74: Die fünf Schritte der Datenanalyse zur Unterrichtsentwicklung auf der Basis von Leistungsdaten. Der vierte Schritt ist fakultativ und abhängig von der gewählten Analyse. Das im Kapitel 7 beschriebene Entwicklungsmodell dient als Grundlage für die Erhebung valider Leistungsdaten, welche als Basis für die Analyse der Daten im "educational Data Warehouse" (eDWH) dienen.....198

Abbildung 8.75: Zusammensetzung der Prüfungen verschiedener Jahre. Angegeben sind die Anzahl der Fragen pro kognitiver Prozessdimension (K1-4, vgl. Kapitel 2). Es ist erkennbar, dass die Anzahl der K1- und K2-Fragen über die Zeit zugunsten der K3+ Fragen reduziert wurden. (Datenbasis: Schlussprüfung, EvIM und IBPW, Jahre 2002-2004, jeweils zwei Tests pro Jahr, n: 33-42 Fragen) (Abbildung übernommen aus L. Fässler et al., 2005)..... 200

Abbildung 8.76: Anzahl und Durchschnitt der korrekt gelösten K3+ Fragen. Ausgewertet wurden jeweils die Daten 2002/II, 2003/II und 2004/II. Die Resultate zeigen, dass die Anzahl und der Mittelwert der korrekt gelösten K3+ Antworten stetig zugenommen hat. (Datenbasis: Schlussprüfung, IBPW, 2002: N=49, M=57%. 2003: N=94, M=62%. 2004: N=77, M=79%).....201

Abbildung 8.77: Vergleich der Resultate des Probetests und der Semester-Endprüfung. Die Diagonallinie unterscheidet zwei Gruppen von Studierenden: jene, welche am Probetest besser abschnitten befinden sich oberhalb der Linie und jene mit einem besseren Resultate an der Semester-Endprüfung unterhalb. Die beiden Mittelwerte sind gestrichelt eingezeichnet. Die Korrelation nach Pearson der beiden Datensätze beträgt $r=0.53$ ($p<0.01$) (Datenbasis: eDWH, EvIM, HS10, $n=254$) 206

Abbildung 8.78: Vergleich des Resultats an der Semester-Endprüfung mit den Expertenbeurteilungen der Präsentationen der sechs selbständigen Aufgaben (m1-m6). Angezeigt sind die Mittelwerte der Bewertungen der vier Leistungsquartile. Die Leistungsquartile basieren auf dem Ergebnis der Studierenden an der Schlussprüfung. Zur Verdeutlichung des Prozessverlaufs wurden die Werte der Quartile untereinander verbunden. Die Skala der Tutorenbewertungen reicht von 0 bis 2 (vgl. 8.4 "Kontext der Datenerhebung") (Datenbasis: eDWH, EvIM, HS10, $n=276$)208

Abbildung 8.79: Vergleich der vier Leistungsquartile bezüglich der Selbstbeurteilung der Studierenden und der Expertenbeurteilung. Die Leistungsquartile basieren auf dem Resultat des Subscores des jeweiligen Moduls an der Semester-Endprüfung. Die besten Studierenden sind im Leistungsquartil (LQ) 1, die schwächsten im LQ 4. Die Experten- und Selbstbeurteilungen sind z-standardisiert. Der Gesamtmittelwert der Beurteilungen liegt bei 0.0 und ist gestrichelt eingezeichnet. Die Mittelwerte der Expertenbeurteilungen sind als Punkte (•), jene der Selbstbeurteilungen als Dreiecke (Δ) dargestellt. (Datenbasis: eDWH, EvIM, HS10, die deskriptive Statistik der Häufigkeiten findet sich im Anhang A.8 und variiert zwischen $n=17-134$ (Expertenbeurteilungen) und $n=1-41$ (Selbstbeurteilungen)).....210

Tabellenverzeichnis

Tabelle 1.1: Informatik-Einführungsveranstaltungen des Departement Informatik für Studierende der Naturwissenschaften. Typ O: Obligatorische Lehrveranstaltung, Typ Z: Zusatzangebot zum Vorlesungsverzeichnis, AGR: Agrar- und Lebensmittelwissenschaften, ERDW: Erdwissenschaften, HST: Gesundheitswissenschaften und Technologie, INF: Informatik, CHAB: Chemie und angewandte Biowissenschaften, UWIS: Umweltnaturwissenschaften (Informationen zusammengestellt gemäss Vorlesungsverzeichnis der ETH Zürich, Dezember 2013). 8

Tabelle 2.2: Die vier Wissensarten nach Anderson (L. Anderson et al., 2001). Diese vier Wissensarten sind nicht hierarchisch. Die Externalisierbarkeit hat Einfluss auf die Messbarkeit dieser Wissensart und wird im Abschnitt "2.6.1 Verbalisierbarkeit des Wissens" erläutert. Generell lässt sich festhalten, dass explizites Wissen einfacher messbar ist. 15

Tabelle 2.3: Adaptation der Bloom'schen Taxonomie nach Fuller (Fuller et al., 2007). Im Hinblick auf die Leistungskontrolle ist vor allem die Unterscheidung der beiden Dimensionen "*producing*" und "*interpreting*" interessant. Sie sind ergänzend und nicht hierarchisch zu verstehen und zeigen, dass beide Verfahren zur Leistungsmessung gleichbedeutend verwendet werden können. 23

Tabelle 3.4: Funktionen und Zielsetzungen von Leistungskontrollen (abgeändert nach (Flehsig, 1974)). In der Hochschulpraxis sind primär die ersten beiden Funktionen wichtig. 30

Tabelle 3.5: Die verschiedenen Prüfungsmethoden messen bezüglich der Gütekriterien unterschiedlich gut, was zu mehr oder weniger objektiven oder reliablen Prüfungen führt (abgeändert nach Krebs, 1999). * = geringe Güte, ** = mittlere Güte, *** = hohe Güte. Da die Validität primär vom Inhalt der LK abhängig ist, kann hier keine pauschale Aussage gemacht werden. 36

Tabelle 3.6: Metriken zur Bestimmung der Reliabilität einer LK. Diese Werte können anhand der Prüfungsergebnisse berechnet werden. 37

Tabelle 3.7: vierdimensionaler Blueprint zur Überprüfung der Validität der Schweizerischen Facharztprüfung "Allgemeinmedizin" (aus Krebs, 1999). Jede Prüfungsfrage wird genau einer Kategorie pro Dimension zugeteilt. Die prozentualen Angaben dienen als Zielvorgabe für die Gewichtung des Inhalts der LK. Die kognitive Kompetenzstufe ist in diesem Blueprint nicht berücksichtigt. 42

Tabelle 4.8: Übersicht über die im Rahmen von Leistungskontrollen relevanten Dokumente.....	55
Tabelle 4.9: Risikotabelle von CBA. Die R-ID zeigt Risiko-Identifikationsnummer und bezeichnet jedes Risiko eindeutig. Die Lettern D, O, T und F beziehen sich auf die vier Risikogruppen (Abschnitt "4.4.2 Risikoidentifikation"). Z-ID gibt, falls vorhanden, die ID der von Ricketts, Steven und Zakrzewski publizierten Risiken an. Ein "-" in der Spalte Z-ID weist auf ein von uns neu hinzugefügtes Risiko hin.....	69
Tabelle 4.10: Risikoanalyse von CBA. Angegeben ist die Risiko-ID (R-ID), der betroffene Personenkreis (Wer), der Umfang der Auswirkungen bei einem allfälligen Eintritt des Risikos (Auswirkung auf...), der Eintrittszeitpunkt des Risikos (Wann?), die Schwere der Konsequenzen beim Risikoeintritt (Kons.), die Wahrscheinlichkeit, dass das Risiko eintritt (Wahrs.), die Datenquelle der Risikowahrscheinlichkeit (1=publizierte Daten, 2=Erfahrungswerte, 3=gemessene Werte), das ungewichtete Gefahrenpotential (G-Potential) und das mit einem Korrekturfaktor für zeitkritische Risiken gewichtete Gefahrenpotential (G-Pot. (gewicht.)). Die Abkürzungen der Spalte "Wer" finden sich in Tabelle 4.11. Die Werte beziehen sich auf summative Leistungskontrollen.....	71
Tabelle 4.11: Konsequenzen der Risiken. Ein "Raum" ist definiert als eine Menge Studierende, welche zur gleichen Zeit im gleichen Raum eine Leistungskontrolle ablegen. Eine "Session" bezeichnet eine Menge "Räume", in welchem zur gleichen Zeit eine Leistungskontrolle stattfindet. Informationen zur Berechnung der Tabellenwerte finden sich im Text. Abgeändert nach (Ricketts & Zakrzewski, 2005).....	73
Tabelle 5.12: Übersicht über die aktuellen CBA-Systeme. Es gelten folgende Abkürzungen: QM = <i>Questionmark Perception</i> , OS = <i>Open-Source</i> , BB = <i>Browser-basiert</i> , <i>Resume</i> = Fortsetzung der Prüfung an einem anderen Computer möglich, https = Verschlüsselte Verbindungen zwischen Client und Server.....	93
Tabelle 5.13: Vier Sicherheitsphasen bei CBA.....	109
Tabelle 5.14: Sicherheitsrisiken aufgrund des Verzichts auf qualifizierte digitale Signaturen bei der Speicherung von Prüfungsergebnissen. Es werden folgende Abkürzungen verwendet: R=Risikobeschreibung, G=Gegenmassnahme, E=Effekt der Gegenmassnahme.....	116
Tabelle 6.15: Die fünf Applikationen der Sioux-CBA-Suite	120
Tabelle 6.16: Massnahmen zur Überprüfung der korrekten Funktionsweise der CBA-Arbeitsstation als Teil der vertrauenswürdigen Zone.....	129
Tabelle 6.17: Geschätzter Zeitaufwand bei CBA und Papier-basierten Prüfungen (PPT). Die Zeiten sind in Stunden angegeben und basieren auf Erfahrungswerten aus unserem Unterricht an der ETH Zürich.....	149

Tabelle 7.18: Eigenschaften der drei Fragetypen. Weitere Informationen finden sich im Text im Abschnitt des jeweiligen Fragetyps.....	182
Tabelle 7.19: <i>Blueprint</i> der summativen LK der LV "Einsatz von Informatikmitteln" im Herbstsemester 2013. Die Prüfungszeit beträgt 60 Minuten. Die Prüfung beinhaltet 40 Fragen, welche sich auf zehn unterschiedliche Aufgabenstellungen beziehen. Der Schwerpunkt der LK liegt auf dem Inhalt der sechs Module. Eine Aufgabenstellung bezieht sich auf ausgewählte Inhalte der Vorlesung.	187
Tabelle 7.20: Abschätzung der Validität des Schlussresultats. Die Tabelle zeigt die biseriale Korrelation der Assistentenbeurteilungen mit der Schlusspunktzahl der Studierenden. Abkürzungen fAs_mX = formatives Assessment Modul X (Datenbasis: Bewertungen der Präsentationen durch die Assistierenden, EvIM, HS13)	190
Tabelle 8.21: Übersicht der während des Herbstsemesters 2010 erhobenen und in einem eDWH gespeicherten lehrbezogenen Daten der LV "Einsatz von Informatikmitteln". Eingeschrieben für diese LV waren 299 Studierende, wovon 284 aktiv am Unterricht teilnahmen. Die vier Datenquellen werden im Text mittels Abkürzung referenziert und beschrieben.....	204

Literaturverzeichnis

- Aegler, M., Kraft, Y., Utiger, M., & Hassenstein, G. (2010). secureXM - Sichere Online-Prüfungen auf Notebooks *Bachelor Thesis* (pp. 1). Bern, Schweiz: Bern University of Applied Sciences.
- Altermatt, T., Dahinden, M., Rieder, L., & Zimmermann, P. (2013). Computer-Based Assessment: Aktuelle Situation und Trends in der Schweizer Hochschullehre. Bern: Institut für Medizinische Lehre (IML), Universität Bern.
- Ames, C., & Archer, J. (1988). Achievement goals in the classroom: Students' learning strategies and motivation processes. *Journal of educational psychology*, 80(3), 260.
- Anderson, L., Krathwohl, D., & Bloom, B. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*: Longman.
- Anderson, R. C. (1961). The Role of Educational Engineer. *Journal of Educational Sociology*, 34(8), 377-381.
- Apostolopoulos, N. (2007). Prüfung 2.0 - Computergestützte Prüfungen an der Freien Universität Berlin. (24.06.2009). <http://www.uni-protokolle.de/nachrichten/text/141545/>
- Bacon, D. (2011, 2011). *Effective Feedback with Adaptive Questions*.
- Bangert-Drowns, R. L., Kulik, C.-L. C., Kulik, J. A., & Morgan, M. (1991). The instructional effect of feedback in test-like events. *Review of educational research*, 61(2), 213-238.
- Bauer-Messmer, B., Fässler, L., & Wyss, M. (2006). *Einführungskurse ins Programmieren - eine didaktische Herausforderung Elektronische Daten Entwicklung, Einsatz und erste Erfahrungen mit E.Tutorials im Programmierunterricht*. Zürich: ETH, Eidgenössische Technische Hochschule Zürich, Department Informatik.
- Baumgartner, P. (2011). *Taxonomie von Unterrichtsmethoden*: Waxmann.
- Bennett, R. E. (2002). Inexorable and Inevitable: The Continuing Story of Technology and Assessment. *The Journal of Technology, Learning, and Assessment*, 1(1).
- Benning, M. (2008). Schummeln ist nicht. *duz MAGAZIN*, 1-2/09, (24.06.2009). http://www.duz.de/docs/artikel/m_01_09schummeln.html
- BFS. (2008). Schüler und Studierende, Entwicklung 1980-2008. *Bundesamt für Statistik BFS*. <http://www.bfs.admin.ch/bfs/portal/de/index/themen/15/02/data/blank/01.html>

- Black, P., & Wiliam, D. (1998a). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-74. doi: 10.1080/0969595980050102
- Black, P., & Wiliam, D. (1998b). Inside the Black Box: Raising Standards Through Classroom Assessment. *Phi Delta Kappa International*, 139-148.
- Bloom, B. S. (1956). *Taxonomy of educational objectives; the classification of educational goals* (1st ed.). New York,: Longmans, Green.
- Böck, B., Huemer, D., & Tjoa, A. M. (2010). *Towards More Trustable Log Files for Digital Forensics by Means of "Trusted Computing"*. Paper presented at the Advanced Information Networking and Applications (AINA), 2010 24th IEEE International Conference on.
- Bouvin, N. O. (1999). *Unifying strategies for Web augmentation*. Paper presented at the ACM Conference on Hypertext and hypermedia, Darmstadt, Germany.
- Brunner, P., & Guretzki, C. (2009). OLAT - scalability. *OLAT CommunityDay 2009*. Retrieved 14.5.2010, from http://www.olat.org/website/en/download/coco/mi/cg_skalierbarkeit.pdf
- Bücking, J., Schwedens, K., & Laue, H. (2007). *Compuergestützte Klausuren an der Universität Bremen: Zentrum für Multimedia in der Lehre (ZMML)*.
- Bull, J., & McKenna, C. (2003). *Blueprint for computer-assisted assessment*: RoutledgeFalmer.
- Butler, A. C., & Roediger, H. L. (2008). Feedback enhances the positive effects and reduces the negative effects of multiple-choice testing. *Memory & Cognition*, 36(3), 604-616.
- Chan, P. W., & Yan, K. K. (2013). Alternative Assessment: Developing e-Portfolio for Final Year Project. In S. S. Cheung, J. Fong, W. Fong, F. Wang & L. Kwok (Eds.), *Hybrid Learning and Continuing Education* (Vol. 8038, pp. 90-101): Springer Berlin Heidelberg.
- Clifton, J. (2010). *A Simple Judging System for the ACM Programming Contest*. Paper presented at the Midwest Instruction & Computing Symposium (MICS) 2010, Wisconsin, USA.
- Conole, G., & Warburton, B. (2005). A review of computer-assisted assessment. *Research in Learning Technology*, 13(1), 17 - 31.
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/BF02310555
- Dahinden, M. (2010a). *Anleitung zur Erstellung von problembasierten Prüfungsfragen für den ICT-Unterricht.*: ETH Zürich, Information Technology and Education.
- Dahinden, M. (2010b). *Usability Kurzbericht: Sioux Exam-Client*: ETH Zürich, Information Technology and Education.
- Dahinden, M. (2010c). *Usability-Kurzbericht: Sioux Exam-Client* (pp. 25): ETH Zürich, Information Technology and Education.

- Dahinden, M., & Faessler, L. (2011, July 20–26). *Monitoring Blended Learning Environments Based On Performance Data*. Paper presented at the IADIS International Conference e-Learning 2011, Rome, Italy.
- Dahinden, M., & Hinterberger, H. (2008). SIOUX e-Assessmentsystem, Kurzbeschreibung. <http://www.switch.ch/aaa/projects/detail/ETHZ.1>. Retrieved 31.10.2008
- Dahinden, M., & Hinterberger, H. (2010). Computer-basierte high-stake Leistungskontrolle mit Sioux: Planung, Durchführung und Auswertung einer Basisprüfung mit 269 Studierenden: ETH Zürich, Information Technology and Education.
- DSG. (2011). *Bundesgesetz über den Datenschutz (DSG)*. Bundesversammlung der Schweizerischen Eidgenossenschaft.
- Duder, J. C., & White, G. R. (1971). A reliable multiple choice test marking system. *Phys. Educ.*, 6, 165-166.
- Dunn, O. J. (1961). Multiple comparisons among means. *Journal of the American Statistical Association*, 56(293), 52-64.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American psychologist*, 41(10), 1040.
- Educause Learning Initiative. (2010). 7 Things you should know about ANALYTICS. *ELI 7 Things You Should Know*.
- Einstein, A. (1905). Zur Elektrodynamik bewegter Körper. *Annalen der Physik*, 322(10), 891-921. doi: 10.1002/andp.19053221004
- Elias, T. (2011). Learning Analytics: Definitions, Processes and Potential. <http://learninganalytics.net/LearningAnalyticsDefinitionsProcessesPotential.pdf>
- Erpenbeck, J. (2007). *Handbuch Kompetenzmessung Erkennen, Verstehen und Bewerten von Kompetenzen in der betrieblichen, pädagogischen und psychologischen Praxis* (2., überarb. und erw. Aufl. ed.). Stuttgart: Schäffer-Poeschel.
- ETH Zürich. (2012). Verordnung der ETH Zürich über Lerneinheiten und Leistungskontrollen an der ETH Zürich (Leistungskontrollverordnung ETH Zürich). *SR 414.135.1*.
- ETH Zürich. (2013a). Ausführungsbestimmungen des Rektors zur Leistungskontrollverordnung der ETH Zürich.
- ETH Zürich. (2013b). Weisung: Akteneinsicht und Aktenweitergabe im Rahmen von Leistungskontrollen.
- ETH Zürich. (2013c). Weisungen zum Prüfungsplan. (Dez. 2013).
- ETH Zürich. (2014). Merkblatt: Rechtsmittel und Rechtsbehelfe bei Prüfungen und weiteren Leistungsbewertungen. (11.02.2014), v.a. Ziffer 2.
- Fässler, L., Dahinden, M., Bosia, L., & Hinterberger, H. (2005, June 27-July 2). *Assessment as an Instrument to Evaluate Quality of Instruction* Paper presented at the World

- Conference on Educational Media and Technology (ED-MEDIA), Montreal, Canada.
- Fässler, L. E. (2007). *Das 4-Schritte-Modell: Grundlage für ein kompetenzorientiertes E-Learning*. (Thesis, Eidgenössische Technische Hochschule ETH Zürich, Nr 17521, 2007), ETH, Zürich. Retrieved from <http://e-collection.ethbib.ethz.ch/show?type=diss&nr=17521>
- Flechsig, K.-H. (1974). *Prüfungen und Evaluation*: Interdisziplinäres Zentrum für Hochschuldidaktik, Univ.
- Forsythe, G. E., & Wirth, N. (1965). Automatic grading programs. *Commun. ACM*, 8(5), 275-278. doi: 10.1145/364914.364937
- Fuchs, L. S., & Fuchs, D. (1986). Effects of systematic formative evaluation: a meta-analysis. *Exceptional children*.
- Fuller, U., Johnson, C. G., Ahoniemi, T., Cukierman, D., Hern\, I., \#225, . . . Thompson, E. (2007). *Developing a computer science-specific learning taxonomy*. Paper presented at the Working group reports on ITiCSE on Innovation and technology in computer science education, Dundee, Scotland.
- Gesellschaft für Softwarelösungen mbH. (2007). EvaExam - Expansion package for EvaSys *Produktprospekt*. Lüneberg, Germany.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports students' learning. *Learning and teaching in higher education*, 1(1), 3-31.
- Goubeaud, K. (2010). How is science learning assessed at the postsecondary level? Assessment and grading practices in college biology, chemistry and physics. *Journal of Science Education and Technology*, 19(3), 237-245.
- Guttormsen, S., Rieder, L., Hinterberger, H., & Dahinden, M. (2010). MEASURED (Media-rich Electronic Assessment with Secure Delivery). <http://www.switch.ch/aaa/projects/detail/UNIBE.5>. Retrieved 26.04.2010, from <http://www.switch.ch/aaa/projects/detail/UNIBE.5>
- HdEL, Fachstelle Hochschuldidaktik & E-Learning (Ed.). (2011). *Lernkultur im Wandel: Die Lernplattform Moodle im Praxis-Einsatz II* (Vol. 9.11): Berner Fachhochschule.
- Heinrich, P. (2008). *Design und Implementierung einer Software für online-Prüfungssysteme*. (Masterarbeit, Eidgenössische Technische Hochschule, 2008), ETH, Zürich. Retrieved from <http://e-collection.ethbib.ethz.ch/show?type=dipl&nr=410>
- Herding, D. C. (2013). *The Tutor-in-the-Loop Model for Formative Assessment*. RWTH Aachen.
- Herold, H., Lurz, B., & Wohlrab, J. (2010). *Grundlagen der Informatik praktisch, technisch, theoretisch* (3rd. print. ed.). München: Pearson Studium.
- Hey, J. (2004). The data, information, knowledge, wisdom chain: the metaphorical link. *Intergovernmental Oceanographic Commission*.

- Hickson, I. (2011). HTML5 - A vocabulary and associated APIs for HTML and XHTML. Retrieved 03.11.2011, from <http://www.w3.org/TR/html5/>
- Hinterberger, H. (2010). Making Informatics Work for Everyone *Technical Report Nr. 686*: ETH Zürich, Information Technology and Education.
- Hinterberger, H., & Böckenhauer, H.-J. (2002-2011a). Einsatz von Informatikmitteln (252-0839-00). Retrieved 04.11.2011, from <http://www.evim.ethz.ch>
- Hinterberger, H., & Böckenhauer, H.-J. (2002-2011b). Informatik für Biologie und Pharmazeutische Wissenschaften (252-0850-00). Retrieved 01.11.2011, from <http://www.ibpw.ethz.ch>
- Hinterberger, H., & Dahinden, M. (2008). Secure Interactive Online eXam (SIOUX). Retrieved 13.11.2008, from http://www.cta.ethz.ch/computerbased_assessment
- Hirt, M. (2001). *Multi-Party Computation: Efficient Protocols, General Adversaries, and Voting*. ETH Zurich.
- Hoffmann, A. (2010). *Ein Sicherheitskonzept für elektronische Prüfungen an Hochschulen auf Basis eines virtuellen, ticketbasierten Dateisystems*. Universität Siegen, Siegen. Retrieved from <http://dokumentix.ub.uni-siegen.de/opus/volltexte/2010/458/>
- Hollingsworth, J. (1960). Automatic graders for programming classes. *Commun. ACM*, 3(10), 528-529. doi: 10.1145/367415.367422
- Hottinger, U., Krebs, R., Hofer, R., Feller, S., & Bloch, R. (2004). Strukturierte Mündliche Prüfung für die ärztliche Schlussprüfung: Institut für Medizinische Lehre, Universität Bern.
- Preboot Execution Environment (PXE) Specification, Version 2.1 C.F.R. (1999).
- James, M., & Brown, S. (2005). Grasping the TLRP nettle: preliminary analysis and some enduring issues surrounding the improvement of learning outcomes. *The Curriculum Journal*, 16(1), 7-30. doi: 10.1080/0958517042000336782
- Johnson, L., Smith, R., Willis, H., Levine, A., & Haywood, K. (2011). The 2011 horizon report *The New Media Consortium*. Austin, Texas.
- Joint declaration of the European Ministers of Education. (1999). The Bologna Declaration. http://www.ond.vlaanderen.be/hogeronderwijs/bologna/documents/MDC/BOL OGNA_DECLARATION1.pdf
- Kade, D., Jokisch, T., & Küpper, S. (2010). Jack: Your business contact butler. https://www.fbi-h-da.de/fileadmin/personal/b.harriehausen/NLP/NLP_WS1112/JACK_Doku_JCU.pdf
- Klauer, K. J. (1987). *Kriteriumsorientierte Tests*: Verlag für Psychologie, Hogrefe.
- Knight, P., & Yorke, M. (2003). *Assessment Learning and Employability*: Open University Press.
- Koenig, C. (2004). *Trusted computing*: Verlag Recht und Wirtschaft.

- Koschnick, W. J. (1984). *Standard dictionary of the social sciences*: Saur.
- Krebs, R. (1999). Kompetent Prüfen.
- Krebs, R. (2002). Anleitung zur Herstellung von MC-Fragen und MC-Prüfungen.
- Krebs, R. (2005). Itemschwierigkeit und -trennschärfe. Bern: Institut für Medizinische Lehre, Abteilung für Assessment und Evaluation (AAE), Universität Bern.
- Kruskal, W. H., & Wallis, W. A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*, 47(260), 583-621.
- Kunkel, M. (2011). Join the free spirit (ILIAS-Broschüre). http://www.ilias.de/docu/repository.php?ref_id=1855&cmd=sendfile&rep_frame=1
- Küstermann, R., Melcher, J., Toussaint, F., & Seese, D. (2005). Online-Klausur mit ILIAS. *RZ-News*, 8-12.
- Landolt, C. (2011). Uni-Prüfung: In einem Raum mit, im anderen ohne Taschenrechner. *Tages Anzeiger*. Retrieved from <http://www.tagesanzeiger.ch/zuerich/stadt/UniPruefung-In-einem-Raum-mit-im-anderen-ohne-Taschenrechner/story/12779465>
- Lee, G., & Weerakoon, P. (2001). The role of computer-aided assessment in health professional education: a comparison of student performance in computer-based and paper-and-pen multiple-choice tests. *MEDICAL TEACHER*, 23, 152-157.
- Lehre, M. (2011). Anleitung Durchführung von Online-Prüfung. Retrieved 23.03.2012, from <http://www.let.ethz.ch/pruefungen/onlinepruefungen/dienstleistungen/dienstleistungen/durchfuehrungsanleitung>
- Lienert, G. A., & Raatz, U. (1998). *Testaufbau und Testanalyse* (6 ed.). Weinheim: Beltz: Psychologie Verlagsunion.
- Lunkeit, A. (2007). Elektronische Signaturen in PDF/A – Konzepte, Einsatz und technische Interoperabilität. Retrieved 30.1.2009, from http://www.pdfa.org/doku.php?id=elektronische_signaturen_in_pdfa
- Marzano, R. J. (2007). *The art and science of teaching: A comprehensive framework for effective instruction*. USA: Association for Supervision and Curriculum Development (ASCD).
- Mayer, R. E. (2002). A taxonomy for computer-based assessment of problem solving. *Computers in Human Behavior*, 18(6), 623-632. doi: 10.1016/S0747-5632(02)00020-1
- Meier, A., Dippold, R., Ringgenberg, A., Schnider, W., & Schwinn, K. (2005). Unternehmensweites Datenmanagement. *Von der Datenbankadministration bis zum Informationsmanagement*. Braunschweig, Wiesbaden, 4.
- Meier, C. (2008). Qualitätsunterstützung für Prüfungen und Assessments (QualiAss). Retrieved 20.02.2009, from <http://www.switch.ch/aaa/projects/detail/UNISG.1>

- MELS, Multimedia & E-Learning Services. (2009). OLAT - Online Learning And Training: Universität Zürich.
- Menezes, A. J., van Oorschot, P. C., & Vanstone, S. A. (1996). Hash Functions and Data Integrity *Handbook of Applied Cryptography*: CRC Press.
- Merkel, K. (2013). *Eine Informatik-didaktische Methode für das Erlernen von Projektarbeit in Schule und Hochschule*. RWTH Aachen. (4736)
- Miller, C., Doering, A., & Scharber, C. (2010). No such thing as failure, only feedback: Designing innovative opportunities for e-assessment and technology-mediated feedback. *Journal of Interactive Learning Research*, 21(1), 65-92.
- Miller, C. M. L., & Parlett, M. (1974). *Up to the Mark: A Study of the Examination Game*. London: Society for Research into Higher Education.
- Möbs, D. (2008). Das mobile elektronische Klausuren- und Evaluationssystem. Retrieved 24.03.2010, from http://www.his.de/publikation/seminar/Workshop_E-Pruefung
- Möltner, A., Schellberg, D., & Jünger, J. (2006). Grundlegende quantitative Analysen medizinischer Prüfungen. *GMS Zeitschrift für Medizinische Ausbildung*, 23 (3).
- Newble, D. I., Hoare, J., & Elmslie, R. G. (1981). The validity and reliability of a new examination of the clinical competence of medical students. *Medical Education*, 15(1), 46-52. doi: 10.1111/j.1365-2923.1981.tb02315.x
- Nonaka, I., & Takeuchi, H. (2012). *Die Organisation des Wissens*: Campus.
- Noorbehbahani, F., & Kardan, A. A. (2011). The automatic assessment of free text answers using a modified BLEU algorithm. *Computers & Education*, 56(2), 337-345. doi: DOI: 10.1016/j.compedu.2010.07.013
- Paulson, L. D. (2005). Building rich web applications with Ajax. *Computer*, 38(10), 14-17.
- Peat, M., & Franklin, S. (2002). Supporting student learning: the use of computer-based formative assessment modules. *British Journal of Educational Technology*, 33(5), 515-523. doi: 10.1111/1467-8535.00288
- Polanyi, M. (1985). *Implizites Wissen*: Suhrkamp.
- Posteraro, R. H., Blackwell, D. E., & Huddleston, A. L. (1986). TECHSCORE: A program for tabulating the results of multiple choice questions and correcting multiple choice examinations. *Computers in Biology and Medicine*, 16(4), 259-265. doi: Doi: 10.1016/0010-4825(86)90009-0
- Pressman, R. S. (2010). *Software engineering*: McGraw-Hill Higher Education.
- Questionmark. (2010). Questionmark assessment solutions (brochure). London: Questionmark Computing Ltd.
- Rahs, O. (2009). Safe Exam Browser (SEB). Retrieved 30.09.09, from <http://www.switch.ch/aaa/projects/detail/ETHZ.3>

- Rektorenkonferenz der Schweizer Universitäten. (2012). Bologna-Monitoring - Zweiter Zwischenbericht. Bern.
- Reznick, R. K., Blackmore, D., Dauphinee, W. D., Rothman, A. I., & Smee, S. (1996). Large-scale high-stakes testing with an OSCE: report from the Medical Council of Canada. *Academic Medicine*, 71(1), 519-21.
- Ricketts, C., & Zakrzewski, S. (2005). A risk-analysis approach to implementing web-based assessment. *Assessment & Evaluation in Higher Education*, 30(6), 603 - 620.
- Rieder, L., Dahinden, M., Nonava, P., Hinterberger, H., & Guttormsen, S. (2012). MEASURED - Abschlussbericht AAA-Switch: Universität Bern.
- Rieder, L., & Nonava, P. (2012). MEASURED - Projekthomepage. Retrieved 16.2.2012, from <https://github.com/measured/measured>
- Rivest, R. L., Shamir, A., & Adleman, L. (1978). A method for obtaining digital signatures and public-key cryptosystems. *Communications of the ACM*, 21, 120-126.
- Rosi, M., Maccari, L., & Fantacci, R. (2007). *STRESS: Stress Testing and Reverse Engineering for System Security*.
- Ruedel, C., Schiefner, M., Noetzli, C., & Seiler Schiedt, E. (2007). Risikomanagement für eAssessment. In M. Merkt (Ed.), *Studieren neu erfinden - Hochschule neu denken: Gesellschaft für Medien in der Wissenschaft*, Waxmann Verlag.
- Rutkowska, J. (2006). Introducing Blue Pill. Retrieved 12.5.2009, from <http://theinvisiblethings.blogspot.com/2006/06/introducing-blue-pill.html>
- Scheuner, B., & Faessler, L. (2010, July 22-29). *Log-File Evaluation of a Problem Based e-Learning Unit on Visual Literacy*. IADIS International Conference e-Learning 2010. Paper presented at the IADIS International Conference e-Learning, Freiburg, Germany.
- Schmucki, B. (2008, 11.03.2009). Online-Prüfung mit ILIAS. from <http://blogs.ethz.ch/net/2008/06/11/online-prufung-mit-ilias/>
- Schmucki, B. (2010). Abschlussbericht "Online-Prüfungen an der ETH Zürich": LET - Lehrentwicklung und -technologie, ETH Zürich.
- Schneider, S. (2011). sebian - Projekthomepage. Retrieved 16.2.2012, from <http://www.eqsoft.org/sebian/doku/doku.php>
- Schulmeister, R. (2007). *Grundlagen hypermedialer Lernsysteme : Theorie - Didaktik - Design*. München: München : Oldenbourg.
- Schwenk, J. (2010). *Sicherheit und Kryptographie im Internet*: Vieweg+Teubner Verlag / GWV Fachverlage GmbH, Wiesbaden.
- Scott, I. (2011). The Learning Outcome in Higher Education: Time to think again? *Worcester Journal of Learning and Teaching*.
- Snyder, B. R. (1971). *The hidden curriculum*. Cambridge, Mass. [u.a.]: MIT Press.

- Snyder, L. (2008). *Fluency with information technology*: Pearson Addison Wesley.
- SWITCH. (2006). Short Lived Credential Service (SLCS). Retrieved 16.08.2010, from <http://www.switch.ch/grid/slcs/about/>
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International journal of medical education*, 2, 53-55.
- Valenti, S., Cucchiarelli, A., & Panti, M. (2002). Computer Based Assessment Systems Evaluation via the ISO9126 Quality Model. *Journal of Information Technology Education*, 1(3), 19.
- Valido-Cabrera, E. (2006). Software reliability methods: Technical Report, Technical University of Madrid.
- Ventouras, E., Triantis, D., Tsiakas, P., & Stergiopoulos, C. (2011). Comparison of oral examination and electronic examination using paired multiple-choice questions. *Computers & Education*, 56(3), 616-624. doi: DOI: 10.1016/j.compedu.2010.10.003
- Verwaltungsgerichtsentscheid. (12.2008). Verwaltungsgericht Hannover, Urteil Az.: 6 B 5583/08.
- Weinert, F. E. (1999). Concepts of competence. *Contribution within the OCED-Project Definition and Section of Competencies: Theoretical and Conceptual Foundations (DeSeCo)*. Neuchâtel: Bundesamt für Statistik.
- Wolf, K. (2007). E-Assessment an Hochschulen: Organisatorische und rechtliche Rahmenbedingungen. *Proceedings of a workshop at the Swiss Centre for Learning Innovations, St. Gallen*, 27-40.
- Wyss, M., Bosia, L., & Dahinden, M. (2008). K3+ basierte, elektronisch gestützte Leistungskontrolle. Zurich: ETH.
- Zakrzewski, S., & Bull, J. (1998). The Mass Implementation and Evaluation of Computer-based Assessments. *Assessment & Evaluation in Higher Education*, 23(2), 141 - 152.
- Zakrzewski, S., & Steven, C. (2000). A Model for Computer-based Assessment: The catherine wheel principle. *Assessment & Evaluation in Higher Education*, 25(2), 201 - 215.
- Zakrzewski, S., Steven, C., & Ricketts, C. (2009). Evaluating computer-based assessment in a risk-based model. *Assessment & Evaluation in Higher Education*, 34(4), 439 - 454.
- ZertES. (2003, 1. August 2008). Bundesgesetz über die elektronische Signatur, ZertES. Retrieved 30.07.2009, from http://www.admin.ch/ch/d/sr/943_03/index.html

Lebenslauf

Lebenslauf

Personalien

Name: DAHINDEN
Vorname: Markus
Geburtsdatum: 14. Mai 1975
Heimatort: Luzern

Ausbildung

2009 - aktuell Doktorat
 ETH Zürich, Departement Informatik
2002 - 2003 Master of Science in Bioinformatik
 Universität Lausanne und Universität Genf
 Masterarbeit am Schweizerischen Bioinformatik Institut, Epalinges
1997 - 2001 Master of Science in Biologie mit Schwerpunkt Mikrobiologie
 Universität Bern
 Diplomarbeit am Institut für Immunologie, Inselspital, Bern
1991 - 1996 Primarlehrdiplom
 Städtisches Lehrerseminar Luzern

Berufliche Tätigkeiten

2012 - aktuell Projektleiter Informatik, Institut für Medizinische Lehre, Universität Bern
2012 - aktuell Lehrspezialist, Departement Informatik, ETH Zürich
2011 - 2012 Mittelschullehrer Informatik, Teilpensum, Mathematisch-
 Naturwissenschaftliches Gymnasium (MNG) Rämibühl, Zürich
2004 - 2012 Wissenschaftlicher Mitarbeiter, Departement Informatik, ETH Zürich
1996 - 1997 Klassenlehrperson Primarschule, Maihof, Luzern

Stand: September 2014