



## Erfahrungsbericht zum Lehrlabor-Projekt:

# FormAdTe – Formative Adaptive Tests

---

Projektverantwortliche und Autorinnen und Autoren:

Daniel Moldt, Michael Haustermann

<http://www.informatik.uni-hamburg.de/TGI/>

Fakultät für Mathematik, Informatik und Naturwissenschaften

---

April 2019



## EINFÜHRUNG

Mit der Motivation, die formal-mathematische Grundausbildung in den informatischen Studiengängen zu verbessern, wurden im FormAdTe-Projekt Konzepte zur Integration von Adaptivität in wöchentlichen Selbsttests erprobt und die konzeptionelle und technische Umsetzbarkeit von adaptiven Tests im Kontext der Veranstaltungen *Formale Grundlagen der Informatik I und II* (FGI-I und FGI-II) untersucht.

## AUSGANGSLAGE UND KONZEPT

Der Zugang zu formal-mathematischen Inhalten fällt vielen Studierenden sehr schwer. Um ihnen die Möglichkeit der besseren Selbsteinschätzung zu geben, sind formative Tests sehr gut geeignet. Zudem vertiefen automatisierte Tests Inhalte, da die korrekten Antworten direkt zurückgemeldet werden. Das FormAdTe-Projekt baut auf frühere, durch das Lehlabor des Universitätskollegs geförderte Lehrkonzepte von Dr. Frank Heitmann und Prof. Dr. Michael Köhler-Bußmeier auf, bei denen für einzelne Inhalte formative Tests als Teil eines Inverted-Classroom-Ansatzes in die Veranstaltung FGI-I integriert wurden, und greift eigene Arbeiten zur Bereitstellung derartiger Tests für FGI-II auf. Die Studierenden müssen den jeweiligen Lesestoff der Vorlesung wöchentlich vorbereiten und über einen beliebig wiederholbaren Test in OpenOLAT zeigen, dass sie die grundlegenden „Vokabeln“ dazu beherrschen. Während leistungsschwächere Studierende diese Tests nutzen, um gleichzeitig den Stoff zu erlernen und eher noch mehr einfache Testfragen benötigen, werden leistungsstarke Studierende unterfordert, gelangweilt oder zeitlich unnötig belastet, wenn das Bestehen der Tests obligatorisch ist. Durch die Integration von Ansätzen zum adaptiven Testen sollte im FormAdTe-Projekt diesem Umstand begegnet werden. Die Ziele für das FormAdTe-Lehrprojekt waren zum einen der Einsatz von formativen Tests in FGI-I und II, wobei für FGI-I aufgrund eines Lehrbuchwechsels komplett neue Testfragen erstellt und für FGI-II Vorlesungsinhalte komplettiert werden mussten. Zum anderen standen die Untersuchung für Möglichkeiten der Adaption von Tests zur besseren Selbsteinschätzung und die Erprobung von Adaptionmöglichkeiten in den FGI-Veranstaltungen im Fokus. Wesentliche Bestandteile davon waren die Erstellung, Verbesserung und Auswertung von Testfragen im Hinblick auf die Adaptivität.

## ARTEN DES TESTENS

Bei adaptivem Testen werden Tests an den Kenntnisstand der Getesteten angepasst. Die Adaption kann innerhalb eines Tests erreicht werden, indem die Fragen in einem Test nacheinander gestellt werden und jeweils auf Basis der zuvor beantworteten Fragen (und möglicherweise anderer Parameter) ein Fähigkeitsniveau für die Person festgelegt wird. Die nächste Frage kann dann anhand des Schwierigkeitsgrads (oder anderer Kriterien) ausgewählt werden. Wird eine Frage nicht korrekt beantwortet, wird eine leichtere gewählt und umgekehrt. Wenn der Schwierigkeitsgrad der Fragen

sich auf einem bestimmten Niveau einpendelt, wird dieses als das Fähigkeitsniveau des/der Getesteten gewertet. Für summative Tests ist die Motivation oft, möglichst effizient zuverlässig entscheiden zu können, auf welchem Kenntnisstand eine geprüfte Person sich befindet [1]. Bei den von uns eingesetzten formativen Tests ist die Motivation eine andere. Die Tests dienen im Wesentlichen nicht der Bewertung der Studierenden, sondern der Selbsteinschätzung [2]. Gleichzeitig können die Studierenden durch die Ausführung der Tests die Inhalte der Veranstaltung erlernen. Die Studierenden sollen auf ihrem individuellen Wissensstand abgeholt werden, um gezielt auf dem Niveau lernen zu können, das ihrem Kenntnisstand entspricht. Im Idealfall erfahren die Studierenden, auf welchem Niveau sie sich befinden und bauen gleichzeitig ihre Kenntnisse aus, um auf ein höheres Niveau zu gelangen. In FormAdTe waren die Tests zwar verpflichtend, konnten aber beliebig oft wiederholt werden, sodass sie letztlich als formative Tests zu sehen sind. Als Adaptionsparameter für Tests können unterschiedlichste Daten herangezogen werden. Es können Ergebnisse aus dem aktuellen Test, Ergebnisse früherer Tests, Ergebnisse des ganzen Kurses oder Metadaten, wie z. B. die Bearbeitungszeit, genutzt werden. Hier kommen zusätzlich Fragen bezüglich Ethik und Datenschutz auf, welche Daten dort zum Einsatz kommen können und sollten.

## **KLASSIFIKATION VON TESTFRAGEN**

Eine elementare Voraussetzung für adaptives Testen ist die Klassifikation von Testfragen nach dem Schwierigkeitsgrad. Mit genügend Vorlauf ist es möglich, Erfolgsstatistiken aus der Vergangenheit für die Einschätzung der Schwierigkeit heranzuziehen. Dafür wurde die Item-Response-Theorie (IRT) herangezogen. Wichtig ist dabei, dass die Fragen fehlerfrei sind, da der so ermittelte Schwierigkeitsgrad für eine im Nachhinein korrigierte Frage nicht mehr zutreffend ist. Auch wenn neue Fragen hinzukommen, muss für diese zuerst der Schwierigkeitsgrad ermittelt werden. Die Chance, dass eine Frage auf dem jeweiligen Schwierigkeitsniveau beantwortet werden kann, sollte auch die vom Lehrenden eingeschätzte Schwierigkeit widerspiegeln. Wenn das nicht der Fall ist, ist das ein Hinweis darauf, dass entweder die Frage schlecht formuliert oder fehlerhaft ist oder der Dozent seine Perspektive hinterfragen muss. Diese Fragen haben in einem adaptiven Test starke Auswirkungen, da die Schwierigkeitsgrade der beantworteten Fragen für die Einschätzung des Niveaus herangezogen werden. Auch die Wahrscheinlichkeit, dass Fragen durch Raten gelöst werden, sollte dafür minimiert werden.

## **UMSETZUNG DES LEHRPROJEKTS**

Die Umsetzung fand während des Sommersemesters 2018 und des Wintersemesters 2018/19 statt. Zum einen wurden Ansätze zur Adaption, die im Kontext des OpenOLAT-Systems sinnvoll erschienen, in den Veranstaltungen FGI-I und II ausgetestet. Begleitend dazu wurde die technische, organisatorische und didaktisch sinnvolle Umsetzbarkeit der Adaptivität untersucht.

## UNTERSUCHUNG TECHNISCHER UND ORGANISATORISCHER UMSETZBARKEIT

Für die von der Universität betriebenen Lernplattformen sind bisher keine Funktionen zur Unterstützung von adaptiven Tests verfügbar. Dies erfordert Anpassungen oder Erweiterungen für diese Systeme. Da wir für unsere Veranstaltungen seit Jahren OLAT einsetzen, war das der Fokus unserer Untersuchungen. Aufgrund der zentralen Pflege durch die Firma *frentix* und der zentralen Verwaltung von OpenOLAT an der Universität ist die Entwicklung einer direkten Erweiterung für die Software schwierig. Weiterhin stellt sich die Frage des Datenschutzes, wenn Testanpassungen auf Basis von Daten geschehen sollen, die über den aktuellen Test hinausgehen (Ergebnisse des ganzen Kurses, Ergebnisse des letzten Jahrgangs, Ergebnisse von Personen in anderen durchgeführten Kursen). Um den Datenschutz adäquat zu berücksichtigen, wäre eine Testumgebung mit entsprechend gekapselten Daten sinnvoll. Die standardisierte SCORM-Schnittstelle (Shareable Content Object Reference Model) bietet die Möglichkeit, in verschiedene Lernplattformen eigene Lernmodule zu integrieren. Als SCORM-Modul kann eine grafische Oberfläche zur Testdurchführung implementiert werden, die mit einem externen Server kommuniziert, der die Fragen und die Lernstand-Informationen verwaltet. Dies erfordert jedoch die Neu-Implementierung einer Testoberfläche, da die von OpenOLAT nicht ohne Weiteres verwendet werden kann. Neben dem Engagement im FormAdTe-Projekt sind wir auch an einem weiteren durch das BMBF geförderten Projekt mit dem Namen „CaTS – Computerisiertes adaptives Testen im Studium“ beteiligt. Ziel dieses Projekts ist das adaptive Testen für Prüfungssituationen (summative Tests). Im Kontext des CaTS-Projekts wurde eine Kombination aus SCORM-Modul und Fragenserver in Form eines ersten Prototyps implementiert, den wir zur Verfügung gestellt bekommen haben. Leider ist es uns im Rahmen von FormAdTe nicht gelungen, diesen Prototyp abschließend an unsere Anforderungen und unsere Lernumgebung anzupassen.

## KLAUSURVORBEREITUNGSTEST NACH SCHWIERIGKEIT GRUPPIERT (FGI-I)

Die Veranstaltung *FGI-I* fand im Sommersemester 2018 statt. Semesterbegleitende Selbsttests wurden seit den Lehrlabor-Projekten von Heitmann und Köhler-Bußmeier im Jahr 2013 auch aufgrund des hohen Arbeitsaufwands in *FGI-I* nicht erneut eingesetzt. Im Rahmen des FormAdTe-Projekts sollten diese erneut eingeführt werden. Da die Vorlesungsinhalte auf ein neues Lehrbuch umgestellt wurden, standen keine Testfragen für die Veranstaltung zur Verfügung. Mit den zusätzlichen Ressourcen durch das FormAdTe-Projekt war es möglich, neue Fragestellungen zu entwickeln und durch wechselseitige Überprüfung eine sehr hohe Fragenqualität zu erreichen. Diese neuen Tests wurden semesterbegleitend wöchentlich verpflichtend eingesetzt. Auf Basis der Ergebnisse der semesterbegleitenden Tests haben wir zur Klausurvorbereitung einen Satz von Tests erstellt mit dem Ziel, die Studierenden auf dem Niveau arbeiten zu lassen, auf dem sie sich befinden. Dafür haben wir mehrere Tests ausgearbeitet, die jeweils einem Schwierigkeitsgrad entsprachen. Dieser Satz von Tests sollte einen ähnlichen Effekt haben wie ein Test, der sich an den individuellen Lernstand anpasst. Die Summe aller Fragen der semesterbegleitenden Tests wurde auf vier Tests mit einem Arbeitsum-

fang von jeweils 81 Minuten aufgeteilt. Diese Teilung wurde entsprechend der Erfolgswahrscheinlichkeit der semesterbegleitenden Durchführung vorgenommen. Da OpenOLAT für eine derartige Aufteilung von Fragen keine Unterstützung bietet, wurden für die Gliederung der Tests *Microsoft Excel* und ein selbstgeschriebenes *Java*-Programm zur Hilfe genommen. Die Möglichkeit, diese Tests zum Üben zu verwenden, wurde leider nur von einem kleineren Teil der Studierenden wahrgenommen. Während die semesterbegleitenden Tests im Sommersemester 2018 von 276 Studierenden bestanden wurden, wurde der einfachste Klausurvorbereitungstest lediglich von 59 Studierenden (21 %) und der schwierigste nur noch von 30 Studierenden (11 %) ausgeführt.

## VOR- UND NACHBEREITUNGSTEST (FGI-II)

Für FGI-II im Wintersemester 2018/19 waren andere Voraussetzungen gegeben. Dadurch, dass die Veranstaltung keinen großen Veränderungen im Vergleich zu den vorherigen Semestern unterlegen war, gab es einen Satz an Fragen und die notwendigen Informationen über die Schwierigkeitsgrade. Diese konnten, dank der zusätzlichen Ressourcen, sorgfältig und entsprechend den Qualitätskriterien überarbeitet werden. Diese Testfragen wurden in den Tests genutzt, die an den Lernstand des kompletten Kurses angepasst sind. Wir haben einen Ansatz ausprobiert, der die Adaption nicht isoliert im Rahmen eines Tests betrachtet, sondern vielmehr eine testübergreifende Adaption ermöglicht. Dafür haben wir den wöchentlichen Test in einen Vor- und einen Nachbereitungstest aufgeteilt, um den Nachbereitungstest an die Ergebnisse des Vorbereitungstests anzupassen. Die Vorbereitungstests dienen der Einweisung in den kommenden Stoff, während die Nachbereitungstests das erwartete Niveau des bisher behandelten und zu erlernenden Stoffs verdeutlichen. Die Nachbereitungstests wurden direkt nach der Vorlesung freigeschaltet und mussten vor der darauffolgenden Vorlesung erfolgreich bestanden sein. Die Nachbereitungstests wurden von uns wöchentlich sowohl auf Basis der Ergebnisse der Vorbereitungstests als auch aufgrund von Daten aus dem letzten Semester erstellt. Dabei haben wir eine sehr einfache datenschutzkonforme Adaptionstrategie verfolgt: Für die Nachbereitungstests wurden nur die schwierigsten Aufgaben (mit der schlechtesten Erfolgsquote der Vorbereitungstests) ausgewählt. Zudem wurden weniger Aufgaben abgeprüft. Dabei wurde implizit überprüft, in wieweit die Studierenden sich den Vorlesungsstoff angeeignet hatten. In der Folge erhielten die Studierenden ein Feedback zu ihrem Lernstand. Diese Art der Testadaption erfordert eine wöchentliche Auswertung und Anpassung der Tests mit entsprechendem Aufwand. Alle Tests wurden nach der Ablauffrist zu beliebigen weiteren, nicht mehr gewerteten Tests zum eigenständigen Üben freigeschaltet. Zur Klausurvorbereitung wurde zudem jeweils eine Kombination aus Vor- und Nachtests freigeschaltet.

## ERGEBNISSE, FAZIT UND AUSBLICK

Als direktes Ergebnis des FormAdTe-Projekts können wir auf einen Satz qualitativ hochwertiger Fragen zurückgreifen, die für die Verwendung von Adaptivität vorbereitet sind. Durch die zusätzlichen Ressourcen konnten neue qualitativ hochwertige Fragen für FGI-I erstellt und die bestehenden Fragen für FGI-II auf ein sehr hohes Qualitätsniveau gebracht werden. Durch den intensiven wöchentlichen Einsatz mit insgesamt vielen Hundert Studierenden liegen ausgezeichnete Bewertungszahlen für die einzelnen Tests und Test-Items vor. Die FGI-I Tests konnten auch im Sommersemester 2019 wieder erfolgreich eingesetzt werden, womit aufgrund der hohen Fallzahl eine IRT-ähnliche Qualitätsstufe erreicht wird. Wir haben festgestellt, dass die Bereitstellung von qualitativ hochwertigen Testfragen einen sehr großen Zeitaufwand erfordert, dieser sich aber rentiert. In früheren Durchläufen wurden die Fragen zwar ebenfalls sehr sorgfältig erstellt, wegen fehlender Ressourcen fand jedoch keine so intensive Qualitätskontrolle statt. Dies führte zu einigen Rückfragen durch Studierende bezüglich missverständlicher oder falscher Fragen. Diese Rückfragen konnten in diesem Projekt auf ein Minimum reduziert werden. Den Evaluationsbögen und den Kommentaren der Studierenden haben wir entnommen, dass es eine hohe Akzeptanz gegenüber den Tests gibt. Während es in früheren Semestern oft noch ein gemischtes Feedback gab, sind die Rückmeldungen in diesen Semestern überwiegend positiv gewesen. Dies führen wir auf die gesteigerte Qualität der Testfragen zurück. Im Freitext der Evaluationsbögen erhielten wir zahlreiche positive Anmerkungen über die wöchentlichen verpflichtenden Tests, da die Studierenden die Verpflichtung als hilfreich empfanden. Freiwillige Tests wie die Klausurvorbereitungstests in FGI-I wurden dagegen nur in geringem Maße genutzt. Eine weitere Beobachtung ist, dass die Studierenden sich sehr leicht von Unregelmäßigkeiten verunsichern lassen. Wenn Fristen verschoben oder verändert wurden (auch mit frühzeitiger Ankündigung), wurde dies häufig negativ kritisiert. Eine Konstanz der Rahmenparameter scheint für die Studierenden daher sehr wichtig zu sein.

Das Konzept der Vor- und Nachbereitungstests pro Woche bewerten wir besonders positiv. Die Ergebnisse der ersten FGI-II Klausur waren signifikant besser (ca. eine ganze Note) als in den Jahren zuvor, in denen es keine zweigeteilten Tests gab, bei ansonsten gleichen Voraussetzungen und vergleichbaren Klausuren. Da die Tests sich als erfolgreiche Hilfe erwiesen, war es zudem möglich, die Vorlesungsinhalte an den Kenntnisstand der Studierenden anzupassen, indem eine höhere Interaktivität durch Fragen an die Studierenden erreicht wurde. Waren z. B. keine Tests durchgeführt und vorbereitender Lesestoff angeboten worden (dies war in genau einer Vorlesung in FGI-II der Fall), konnten die Studierenden nicht in der üblichen Weise mit dem Vortragenden interagieren und inhaltliche Querbezüge herstellen. Durch Nachbereitungstests wurde erreicht, dass Übungsaufgaben besser bearbeitet werden konnten. Das Konzept des Vor- und Nachbereitungstest kann als eine Variante der Adaption angesehen werden. Die technische Umsetzung der Adaptivität innerhalb eines einzigen Tests hat sich leider als schwieriger herausgestellt als angenommen. Den verfolgten Ansatz über das SCORM-Modul halten wir jedoch für vielversprechend, da durch den SCORM-Standard eine nachhaltige Lösung für mehrere Lernplattformen geschaffen würde. Die Verwendung eines externen Servers zur Bereitstellung der Fragen ermöglicht das Experimentieren mit Adaptionansätzen und bietet gleichzeitig einen begrenzten Rahmen in Bezug auf die Daten. Die FGI-Veranstaltungen

unterliegen momentan einer Umstrukturierung, sodass FGI-I in zwei verschiedene Veranstaltungen aufgeteilt wird (2. und 4. Semester).

Für die Nachfolgeveranstaltung im zweiten Semester „Einführung in die Theoretische Informatik (ETI)“ wird der erarbeitete Fragenkatalog im Kontext von Vor- und Nachbereitungstest eingesetzt werden. Gleiches gilt für das vorgesehene Modul „Modellierung und Analyse komplexer Systeme (MAKS)“ als Nachfolgeveranstaltung für FGI-II. Weiterhin wollen wir an der technischen Umsetzung arbeiten, um mittelfristig einsatzfähige Softwarelösungen bereitstellen zu können. Gesamtfazit: Tests als Unterstützung unserer Lehrveranstaltungen sind ein hervorragendes Mittel zur Verbesserung der grundständigen Lehre.

## LITERATUR

- [1] Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. *Development of computerized middle school achievement test*, 69 <https://www.rasch.org/memo69.pdf>.
- [2] Mark Gierl, Okan Bulut und Xinxin Zhang (2018): Using Computerized Formative Testing to Support Personalized Learning in Higher Education: An Application of Two Assessment Technologies, in Robert Zheng (Hrsg.): *Digital Technologies and Instructional Design for Personalized Learning*, IGI Global, 2018, S. 99-119 <https://www.igi-global.com/chapter/using-computerized-formative-testing-to-support-personalized-learning-in-higher-education/199535>.