

# Studentische Interaktion mit automatischen Prüfungssystemen

Michael Striewe, Michael Goedicke  
Paluno - The Ruhr Institute for Software Technology  
Universität Duisburg-Essen, Campus Essen  
Gerlingstraße 16, 45127 Essen  
{michael.striewe,michael.goedicke}@s3.uni-due.de

**Abstract:** Idealerweise werden Studierende dazu angehalten, sich kontinuierlich mit dem Stoff einer Vorlesung zu befassen. Die beschränkte Verfügbarkeit von Ansprechpartnern zur Beantwortung allgemeiner Fragen oder zur Korrektur konkreter Abgaben für Übungsblätter führt jedoch eher zu einer schubweisen Beschäftigung mit einem Thema. Der vorliegende Artikel untersucht als Anwendungsbeitrag an einem konkreten Beispiel, ob und wie ein permanent verfügbares, automatisches Prüfungssystem für Übungsaufgaben eine Änderung der Situation herbeiführen kann und welche Nutzungsstrategien Studierende gegenüber einem solchen System entwickeln.

## 1 Einleitung

Universitäre Lehrveranstaltungen verfolgen in der Regel das Ziel, Wissen und Kompetenzen zu vermitteln, die von den Studierenden gut verinnerlicht und damit langfristig genutzt werden können. Das schnelle, kurzfristige Auswendiglernen von Stoff kurz vor einer Prüfung ist unerwünscht. Stattdessen werden die Studierenden idealerweise dazu angehalten, sich im Laufe des gesamten Semesters kontinuierlich mit dem Stoff der Vorlesung zu befassen und dabei ihren individuellen Lernrhythmus zu finden. Eine solche Aufforderung impliziert, dass auch die Lehrenden kontinuierlich für Rückfragen zur Verfügung stehen sowie Übungen stellen und bewerten sollten. Aus naheliegenden Gründen kann eine solche Betreuung allerdings nicht tatsächlich rund um die Uhr durch das gesamte Semester hindurch geleistet werden. Dozenten sind in der Regel mit mehr als einer Lehrveranstaltung befasst und üben zudem Forschungstätigkeiten aus, so dass sie nicht immer kurzfristig auf individuelle Fragen zum Stoff einer bestimmten Vorlesung antworten können. Tutoren sind in der Regel für eine oder mehrere Gruppen von Studierenden zuständig und bieten schon aus organisatorischen Gründen häufig nur je eine Korrektur und Besprechung pro Übungsblatt oder Übungsaufgabe an. Diese Beschränkungen führen dazu, dass sich die Studierenden nicht völlig frei mit dem Vorlesungsstoff befassen können, sondern zu einer schubweisen Beschäftigung angehalten werden.

Automatisierte Prüfungssysteme, die ohne manuellen Eingriff Rückmeldungen zu Übungsaufgaben erzeugen können, scheinen Abhilfe zu versprechen. Sie können rund um die Uhr betrieben werden und damit für den Bereich der Übungsaufgaben zwei Beschränkungen

umgehen: Studierende sind nicht von den Arbeitszeiten eines Tutors abhängig, um ihre Lösungen korrigieren zu lassen und Studierende sind nicht darauf beschränkt, die Aufgaben zu bearbeiten, die der Tutor als nächstes zu korrigieren und zu besprechen gedenkt. Es kann daher erwartet werden, dass solche Systeme den Studierenden ein Arbeiten in selbstbestimmter Geschwindigkeit erleichtern und damit auch zu einer kontinuierlichen Beschäftigung mit dem Vorlesungsstoff anregen.

Der vorliegende Artikel untersucht nun am Beispiel eines konkreten Prüfungssystems für Programmieraufgaben, welches Nutzungsverhalten durch Studierende tatsächlich beobachtet werden kann. Insbesondere sollen dadurch Erkenntnisse gewonnen werden, wie Studierende an die Bearbeitung von Übungsaufgaben heran gehen, wenn sie sich für die Einholung von Rückmeldungen nicht nach den Arbeitszeiten von Tutoren zu richten haben. Der Artikel ist wie folgt gegliedert: Abschnitt 2 erläutert die Rahmenbedingungen, unter denen das Prüfungssystem eingesetzt wurde und die folglich für die Aussagekraft dieser Studie gelten. Abschnitt 3 betrachtet das Verhalten der Studierenden über die Zeit, d. h. insbesondere die Frage nach den Zeitpunkten, zu denen die Studierenden in Interaktion mit dem System treten. Abschnitt 4 untersucht, ob sich spezielle Verhaltensmuster von Studierenden bei der Bearbeitung einzelner Aufgaben erkennen lassen. Abschnitt 5 fasst die Ergebnisse einer Befragung der Studierenden zusammen, in der die Studierenden nach ihren Nutzungsstrategien für das System gefragt wurden. Abschnitt 6 verweist auf verwandte Untersuchungen und Abschnitt 7 beendet den Artikel mit einem Fazit.

## **2 Rahmenbedingungen der Untersuchung**

Die vorliegende Untersuchung bezieht sich auf die Lehrveranstaltung “Programmierung” an der Universität Duisburg-Essen im Wintersemester 2010/2011. In dieser Lehrveranstaltung werden den Studierenden im ersten Fachsemester die Grundlagen der objekt-orientierten Programmierung in der Programmiersprache Java vermittelt. Die Lehrveranstaltung besteht aus zwei Vorlesungsterminen pro Woche sowie einer Globalübung pro Woche. Darüber hinaus werden Tutorien in Kleingruppen angeboten. Alle zwei Wochen finden Testate statt, bei denen Programmieraufgaben in 45 Minuten annähernd unter Prüfungsbedingungen gelöst werden müssen. Mit jedem Testat werden Punkte erworben, wobei eine bestimmte Mindestpunktzahl Voraussetzung für die Zulassung zur Klausur am Semesterende ist. Jeweils zwei Wochen vor einem Testat wird eine darauf vorbereitende Übungsaufgabe in einem automatischen Tutorensystem zur Verfügung gestellt. Die hier erreichte Punktzahl ist für die Testateilnahme oder die Klausurzulassung unerheblich. Einzige Vorgabe für die Studierenden ist, diese Aufgabe bis zum Testattermin bearbeitet zu haben, d. h. mindestens einen Lösungsversuch online eingereicht zu haben. Die Studierenden verfügen also über weitgehende Freiheit, wann und in welchem Umfang sie die gestellten Aufgaben bearbeiten.

Zur Begleitung der Lehrveranstaltung wird das an der Universität Duisburg-Essen entwickelte automatische Tutorensystem JACK [SBG09] eingesetzt. Das System wird rund um die Uhr betrieben und stand im Wintersemester 2010/2011 durchgängig zur Verfügung. Ein mehrtägiger Serverausfall in den Weihnachtsferien führte dazu, dass Lösungen zwar

hochgeladen werden konnten, aber erst mit mehrtägiger Verspätung geprüft wurden. Das System vergibt für jede Lösung Punktzahlen von 0 bis 100, wobei 100 für eine vollständig korrekte Lösung steht. Die Vergabe der Punkte erfolgt auf Basis von Testfällen und statischen Analysen des Programmcodes, die von den Lehrenden individuell für jede Aufgabenstellung konfiguriert werden können. Zu jedem gefundenen Fehler wird zudem eine textuelle Rückmeldung erzeugt, die den gefundenen Fehler beschreibt und Hinweise zu dessen Behebung liefert [SG10]. Die Prüfung der eingereichten Lösungen nimmt je nach Größe der Aufgabe und der Menge der konfigurierten Tests einige Sekunden bis mehrere Minuten in Anspruch, so dass es bei der gleichzeitigen Benutzung des Systems durch viele Studierende zu Warteschlangen kommen kann. Eine unmittelbare Rückmeldung im Sinne einer Reaktionszeit von wenigen Minuten kann von dem System also nicht garantiert werden.

### **3 Nutzung über die Zeit**

Im Folgenden wird zunächst das Nutzungsverhalten der Studierenden im gesamten Semester unabhängig von einzelnen Aufgaben betrachtet. Dabei wird insbesondere der Frage nachgegangen, ob die größere zeitliche Freiheit beim Einholen von Rückmeldungen von den Studierenden genutzt wird.

#### **3.1 Nutzung im Tagesverlauf**

Einer der hervorstechendsten Aspekte eines automatischen Prüfungssystems ist die Verfügbarkeit rund um die Uhr im Gegensatz zur Erreichbarkeit von Ansprechpartnern nur zu Bürozeiten. Die Erfahrungen im Wintersemester 2010/2011 konnten bestätigen, dass diese erhöhte Verfügbarkeit der Zeiteinteilung den Studierenden entgegen zu kommen scheint. Abbildung 1 gibt an, wie viele Lösungen in Summe pro Stunde hochgeladen wurden. Mit Ausnahme der erwartbaren schwächeren Nutzung in der Nacht, wurde das System sowohl tagsüber als auch in den Abendstunden gleichmäßig stark benutzt. Bemerkenswert daran ist, dass es vormittags deutlich weniger Lösungsversuche gab als in den Abendstunden, obwohl vormittags zu verschiedenen Terminen betreute Tutorien angeboten wurden, während die Studierenden abends bei der Arbeit zu Hause auf sich alleine gestellt waren.

Diese Beobachtungen belegen, dass die Möglichkeit zur freien Wahl des Bearbeitungszeitpunktes von den Studierenden stark genutzt wird. Ob die Studierenden bei manuell korrigierten Aufgaben ihre Lösungen zu anderen Zeitpunkten angefertigt hätten, muss zwar offen bleiben, aber es ist deutlich erkennbar, dass sich die Studierenden beim Einholen von Rückmeldungen in einem automatisierten System nicht auf die sonst üblichen Bürozeiten beschränken. Ein erheblicher Anteil Studierender zieht offenbar auch die Kombination aus selbst gewähltem Arbeitszeitpunkt und automatisierter Betreuung durch ein technisches System einer Kombination aus vorgegebenen Zeitpunkten mit persönlicher Betreuung in den Tutorien vor.

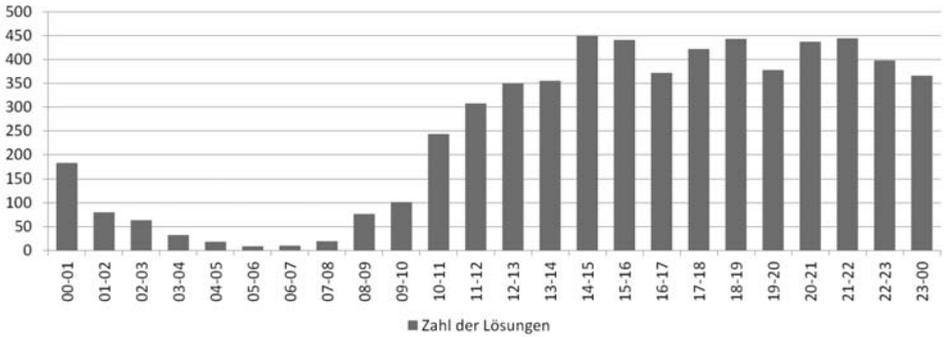


Abbildung 1: Zahl der hochgeladenen Lösungsversuche im Tagesverlauf. Jeder Balken gibt an, wie viele Lösungsversuche in der entsprechenden Stunde in Summe über das gesamte Semester hochgeladen wurden.

### 3.2 Nutzung im Verlauf des Semesters

Trotz der grundsätzlich freien Wahl des Bearbeitungszeitpunktes war durch die feststehenden Testattermine und die darauf basierenden Zeitpunkte für die Veröffentlichung von Übungsaufgaben eine feste Taktung des Semesters gegeben, die sich zwangsläufig auch auf das Nutzungsverhalten niederschlagen musste. Abbildung 2 gibt für jeden Tag des Semesters an, wie viele Lösungen hochgeladen wurden. Die Termine der Testate sind in dieser Darstellung schon vom bloßen Augenschein her deutlich zu erkennen. Unverkennbar (und für Praktiker sicher nicht überraschend) ist, dass es eine erhebliche Menge an Studierenden gibt, die sich trotz des permanenten Angebots automatischer Rückmeldungen erst kurz vor einem Testat mit der jeweils vorbereitenden Aufgabe befasst haben. Die durchschnittliche erreichte Punktzahl an diesen stark frequentierten Tagen liegt allerdings nicht niedriger als an den anderen Tagen, so dass auch bei diesen kurzfristigen Bemühungen von ernsthaften Lösungsversuchen ausgegangen werden kann.

Es ist zudem zu erkennen, dass es im ersten Semesterabschnitt bis Weihnachten eine untere Grenze von mindestens 17 Lösungsversuchen pro Tag gibt. Diese stammen von mindestens 8 verschiedenen Studierenden pro Tag. Im zweiten Semesterabschnitt nach Weihnachten halbieren sich diese Zahlen. Dies ist vor allem darauf zurückzuführen, dass zu diesem Zeitpunkt schon zahlreiche Studierende die nötigen Voraussetzungen für die Teilnahme an der Klausur erfüllten und daher auf eine weitere Teilnahme an den Testaten und an der Bearbeitung der Übungsaufgaben verzichteten. Ein grundlegender Trend zum Verzicht auf "unnötige" Arbeit konnte durch das permanent verfügbare und damit vermeintlich bequemere System folglich nicht gebrochen werden.

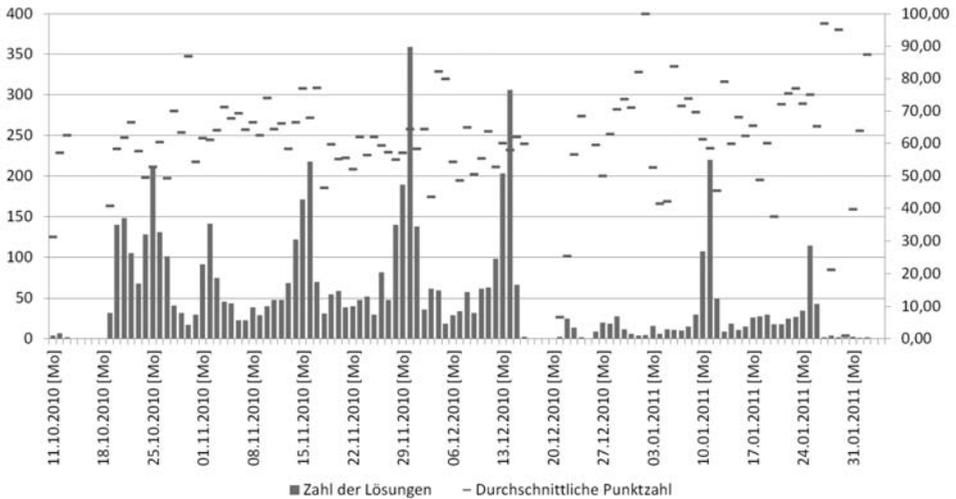


Abbildung 2: Zahl der hochgeladenen Versuche und erreichten durchschnittlichen Punktzahlen im Verlauf des Semesters. Die sechs Testattermine sind an den sechs regelmäßig auftretenden Spitzen gut zu erkennen. Zu Beginn des Semesters wurde ein zusätzliches Probetestat und eine Demoaufgabe im System angeboten, was die besonders hohe Zahl an Einreichungen in der Woche vom 18.10. bis 25.10. erklärt.

## 4 Bearbeitungsstrategien

Aus den Beobachtungen zur Nutzungszeit im vorherigen Abschnitt lässt sich die Frage ableiten, ob Strategien beobachtet werden können, nach denen die Studierenden die Aufgaben bearbeiten und mit dem System in Interaktion treten. Im Folgenden werden generelle und typische Verhaltensweisen charakterisiert, die sich aus der Beobachtung des Systems ableiten lassen.

### 4.1 Abbrüche und wiederholte Versuche

Die durchschnittliche Zahl von hochgeladenen, korrekten und falschen Lösungen ist in Tabelle 1 dargestellt. Im Schnitt wurde jede Aufgabe von 306 Studierenden bearbeitet, d.h. es wurde von dieser Zahl an Studierenden mindestens eine Lösung hochgeladen. Im Schnitt luden 197 Studierende mindestens eine Lösung hoch, die nicht vollständig korrekt war. Daraus ergibt sich, dass im Schnitt 109 Studierende ausschließlich korrekte Lösungen hochgeladen haben. Ob sie diese Lösungen alleine erarbeitet oder im Austausch mit anderen Studierenden bekommen haben, kann an dieser Stelle nicht geklärt werden. Grundsätzlich ist die Prüfung der eingereichten Lösungen auf Duplikate jedoch möglich, um die Zahl an unabhängigen korrekten Lösungen zu ermitteln.

Im Schnitt luden 133 Studierende überhaupt nur eine Lösung hoch. Das heißt, dass im

|  |     |
|--|-----|
| Studierende insgesamt  | 306 |
| Studierende mit mindestens einer falschen Lösung                     | 197 |
| Studierende mit nur einer Lösung                                     | 133 |
| Studierende mit mindestens einer korrekten Lösung                    | 222 |
| Studierende mit ausschließlich korrekter Lösung                      | 109 |
| Studierende, die nach genau einem Versuch aufgegeben haben           | 24  |
| Studierende, die nach mehr als einem Versuch aufgegeben haben        | 60  |
| Studierende mit mindestens einer korrekten und einer falschen Lösung | 113 |

Tabelle 1: Durchschnittswerte für Teilnehmer, korrekte und falsche Lösungen pro Übungsaufgabe.

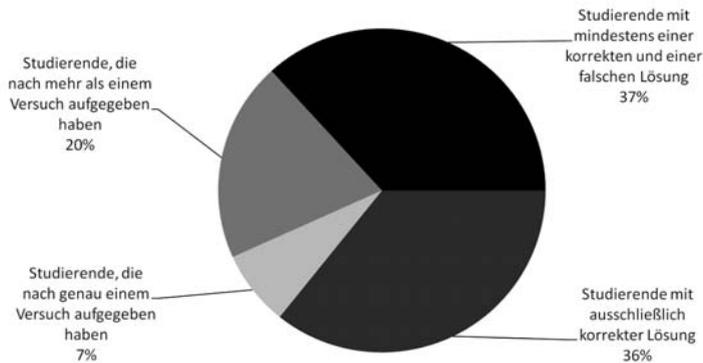


Abbildung 3: Prozentuale Aufteilung des Bearbeitungserfolgs im Schnitt über alle Übungsaufgaben. Es wird angenommen, dass Studierende aufgegeben haben, wenn sie zu einer Aufgabe keine Lösung eingereicht haben, die korrekt war, d.h. 100 Punkte erhalten hat.

Schnitt pro Aufgabe 24 Studierende ( $=7\%$ ) gleich nach dem ersten Versuch aufgaben und offenbar kein Interesse daran hatten, ihre unvollständige Lösung zu verbessern. Insgesamt luden im Schnitt 222 Studierende mindestens eine vollständig korrekte Lösung hoch, so dass sich eine weitere Differenz von 60 Studierenden ( $=20\%$ ) ergibt, die nach mehr als einem Versuch aufgaben. Ferner ergibt sich, dass 113 Studierende ( $=37\%$ ) es schafften, nach einer nicht vollständig korrekten Abgabe in einem späteren Versuch eine korrekte Lösung hochzuladen. Alle Prozentwerte sind in Abbildung 3 dargestellt. Es ist deutlich zu sehen, dass der Anteil derjenigen Studierenden, die nach einem initialen Fehlversuch zu einer korrekten Lösung gekommen sind, deutlich über dem Anteil derjenigen liegt, die nach einem oder mehreren Fehlversuchen aufgegeben haben.

Es kann somit festgestellt werden, dass die Möglichkeit, wiederholt Rückmeldung zu verschiedenen Lösungsversuchen zu erhalten, offenbar stark genutzt wurde und zumindest ein deutlich motivierender Einfluss erreicht wurde. Eine vergleichbar intensive Betreuung mit manueller Korrektur von eingerichteten Lösungen wäre mit den zur Verfügung stehenden Personalressourcen nicht möglich gewesen. Die didaktische Wirksamkeit kann aus diesen Zahlen allerdings nur eingeschränkt beurteilt werden, da nicht in allen Fällen die Kombination aus mehreren falschen und einer korrekten Lösung bedeutet, dass die Aufgabe von

den Studierenden letztlich gelöst wurde. In Stichproben konnten vereinzelt Fälle festgestellt werden, in denen Studierende ganz offensichtlich eine fremde “Musterlösung” hochgeladen haben, für die sie die volle Punktzahl erhielten, und anschließend die Arbeit an ihrer eigenen, unvollständigen Lösung fortgesetzt haben. Dieses Verhalten lässt sich möglicherweise so erklären, dass den betreffenden Studierenden nicht bewusst war, dass ihnen aus dem Einreichen einer vollständig korrekten Lösung keine direkten Vorteile entstehen. Eine bewusste Täuschungsabsicht kann zumindest insofern ausgeschlossen werden, als die Studierenden offenbar auch daran interessiert waren, eine eigene Lösung möglichst weit zu entwickeln.

## 4.2 Bearbeitungsdauer

Wie bereits dargestellt, luden im Schnitt 133 Studierende (=43 %) pro Aufgabe nur eine Lösung hoch. Weitere 20 % unternahmen genau zwei Versuche, weitere 13 % genau drei Versuche. Gut 5 % der Studierenden unternahmen mehr als 10 Versuche. In drei extremen Ausnahmefällen wurden von je einer Person 59, 60 bzw. 65 Lösungen hochgeladen. In einem der Fälle fungierte die betroffene Person allerdings offenbar als Helfer für Mitstudierende und hat über ihren Account sehr verschiedene Lösungen (u.a. auch insgesamt sieben vollständig korrekte Lösungen) hochgeladen. Ein weiterer dieser Fälle fällt genau in die Zeit des oben genannten Serverausfalls und enthält zahlreiche identische Lösungen, die offenbar aus Unsicherheit über den Status des Systems wiederholt hochgeladen wurden. Diese beiden Fälle sind demnach für eine verallgemeinerte Aussage zum Nutzungsverhalten irrelevant. Der verbleibende Fall mit 60 Lösungsversuchen zeigt jedoch ein sehr typisches Verhalten: Der erste Lösungsversuch erhielt 0 Punkte, während der letzte Lösungsversuch zwei Tage später hochgeladen wurde und 88 Punkte erhielt. Kurz zuvor erreichten zwei Versuche mit 96 Punkten fast die volle Punktzahl. Für die gesamte Zeitspanne lassen sich mehrere zeitlich zusammenhängende Phasen feststellen, in denen die Aufgabe offenbar konzentriert bearbeitet wurde und mit aufeinanderfolgenden Lösungsversuchen tendenziell steigende Punktzahlen erreicht wurden. In kleinerem Rahmen mit weniger Lösungsversuchen lässt sich dieses Verhalten bei fast allen Studierenden zumindest bei einigen der sechs Aufgaben des Semesters beobachten. Daraus lässt sich folgern, dass das Hochladen eines Lösungsversuchs für die Studierenden nicht am Ende einer Bearbeitungsphase steht, wie dies bei der Bearbeitung und Einreichung eines klassischen Übungsblattes mit manueller Korrektur der Fall wäre, sondern dass die Einreichung des Lösungsversuchs und die Sichtung der Rückmeldung in den Arbeitsprozess integriert und die Bearbeitung auf Basis der erhaltenen Rückmeldung unmittelbar fortgesetzt wird.

Die Dauer der beobachtbaren Arbeitsphasen schwankt stark. Kurze Phasen, in denen z. B. 4 Lösungen innerhalb von weniger als 10 Minuten hochgeladen wurden, sind ebenso vertreten wie lange Phasen, in denen 6 Lösungen gleichmäßig verteilt über einen Zeitraum von 3 Stunden hochgeladen wurden. Phasen, in denen mehr als 6 Lösungen nacheinander hochgeladen wurden, bilden die Ausnahme. Ebenso sind die Bearbeitungsphasen eher kurz (unter einer Stunde) und nur selten länger als zwei Stunden. Weitergehende Interpretationen dieser Daten sind jedoch nicht möglich, da sich allein aus den Zeitstempeln

der Einreichungen nicht sicher beurteilen lässt, wann eine Arbeitsphase begonnen oder beendet wurde. Ferner ist davon auszugehen, dass Studierende die Bearbeitung einer Aufgabe abgebrochen haben, wenn sie nicht innerhalb einer kurzen Zeitspanne Rückmeldung vom Prüfungssystem erhalten haben. Bei starker Belastung des Systems kann dies wie in Abschnitt 2 beschrieben nicht garantiert werden.

## 5 Eindrücke der Studierenden

Unter anderem um mögliche Gründe für das beobachtete Nutzungsverhalten zu finden, wurde gegen Ende des Wintersemesters eine Umfrage unter allen Studierenden durchgeführt, die an mindestens drei Testaten teilgenommen hatten. Von 317 eingeladenen Studierenden nahmen 61 an der Befragung teil. Die prozentuale Auswertung der Antworten zur Frage nach dem Vorgehen beim Bearbeiten der Übungsaufgaben ist in Tabelle 2 dargestellt. 25 % der Studierenden gaben an, meistens frühzeitig mit der Bearbeitung der Übungsaufgaben begonnen und ihre Lösungsversuche immer sofort durch das Prüfungssystem prüfen lassen zu haben. Dies deckt sich mit der oben beschriebenen Beobachtung von mindestens 8 Studierenden pro Tag, die das System auch weit vor dem jeweiligen Testattermin benutzt haben. 38 % gaben an, meistens frühzeitig mit der Bearbeitung der Aufgaben begonnen zu haben, aber erst spät wenige Lösungsversuche zur automatischen Prüfung eingereicht zu haben. In mehr als der Hälfte der Fälle wurde dies damit begründet, dass die Studierenden nur möglichst korrekte Lösungen im System hochladen wollten. Obwohl die Studierenden zu Beginn des Semesters ausdrücklich darauf hingewiesen wurden, dass das automatische Prüfungssystem vor allem ein Angebot zur Selbstkontrolle sei, scheint trotzdem eine Scheu vor dem Einreichen unvollständiger Lösungen bestanden zu haben. Unter dieser Prämisse muss das Angebot eines solchen Systems als nicht voll akzeptiert betrachtet werden.

Die verbleibenden 37 % gaben an, überhaupt erst spät mit der Bearbeitung der Übungsaufgaben begonnen zu haben. Hier gab etwas mehr als ein Drittel an, nur wenige Lösungsversuche eingereicht zu haben, um nur möglichst korrekte Lösungen hochzuladen. Etwas mehr als ein Viertel gab dagegen an, in der kurzen Bearbeitungsphase sehr viele Lösungen zur automatischen Prüfung eingereicht zu haben.

Bemerkenswert ist der Anteil von insgesamt 22 % der Studierenden, die unabhängig vom Beginn der Bearbeitung erst spät wenige Lösungen hochgeladen haben, da sie ohnehin mit keiner schnellen Reaktion des Systems rechneten. Dies ist in zweierlei Hinsicht interessant: Erstens arbeitete das System in jedem Fall schneller als eine klassische manuelle Korrektur von Übungsaufgaben, so dass die Studierenden gegenüber dem automatischen System offenbar eine grundsätzlich andere Erwartungshaltung in Bezug auf die Geschwindigkeit hatten, die nicht erfüllt werden konnte. Zweitens waren die Studierenden offenbar nicht bereit oder nicht in der Lage, bei der Bearbeitung der Aufgaben von einer stark frequentierten Zeit, in der sich Warteschlangen von Lösungen im System aufstauten, zu einer weniger stark frequentierten Zeit zu wechseln. Während Abbildung 1 belegt, dass ein solcher Ausgleich zumindest für den Tagesverlauf offenbar stattgefunden hat, zeigt Abbildung 2 deutlich den mangelnden Ausgleich im Semesterverlauf. Studierende, die diese

|   |      |
|---|------|
| Ich habe meistens frühzeitig mit der Bearbeitung begonnen und meine Lösungsversuche immer sofort prüfen lassen.   | 25 % |
| Ich habe meistens frühzeitig mit der Bearbeitung begonnen, aber erst spät wenige Lösungsversuche prüfen lassen, da mir die Meldungen des Systems ohnehin nicht geholfen hätten. | 7 %  |
| Ich habe meistens frühzeitig mit der Bearbeitung begonnen, aber erst spät wenige Lösungsversuche prüfen lassen, da das System ohnehin nicht schnell genug geantwortet hätte.    | 11 % |
| Ich habe meistens frühzeitig mit der Bearbeitung begonnen, aber erst spät wenige Lösungsversuche prüfen lassen, da ich nur möglichst korrekte Lösungen hochladen wollte.        | 20 % |
| Ich habe meistens spät mit der Bearbeitung begonnen und dann sehr viele Lösungsversuche in kurzem Zeitabstand prüfen lassen.  | 10 % |
| Ich habe meistens spät mit der Bearbeitung begonnen und nur wenige Lösungsversuche prüfen lassen, da mir die Meldungen des Systems ohnehin nicht geholfen hätten.               | 3 %  |
| Ich habe meistens spät mit der Bearbeitung begonnen und nur wenige Lösungsversuche prüfen lassen, da das System ohnehin nicht schnell genug geantwortet hätte.                  | 11 % |
| Ich habe meistens spät mit der Bearbeitung begonnen und nur wenige Lösungsversuche prüfen lassen, da ich nur möglichst korrekte Lösungen hochladen wollte.                      | 13 % |

Tabelle 2: Ergebnisse einer Befragung mit 61 Teilnehmern. Von den vorgegebenen Antwortoptionen konnte genau eine gewählt werden.

Meinung über zu lange Wartezeiten geäußert haben, ziehen möglicherweise eine zusammenhängende, konzentrierte Bearbeitung der Aufgaben ohne zwischenzeitliche Rückmeldungen einer durch Wartezeiten auf Rückmeldung unterbrochenen Arbeitsweise vor.

Nur ein geringer Anteil von insgesamt 10 % der Studierenden scheint die automatisch erzeugten Meldungen für grundsätzlich so nutzlos zu halten, dass diese Studierenden unabhängig vom eigenen Bearbeitungsrythmus der Aufgaben auf eine intensive Nutzung des Systems verzichtet haben. Dies deckt sich mit den weiteren Ergebnissen aus anderen Fragen der Umfrage, nach denen 11 % der Studierenden das Prüfungssystem für nutzlos, aber verbesserungsfähig halten, während 87 % es für nützlich und verbesserungsfähig halten. Lediglich 2 % halten es für so nützlich, dass kein Wunsch nach Verbesserungen besteht. Der häufigste Verbesserungswunsch betrifft die Wartezeiten bis zum Erhalt einer Rückmeldung. In einer Frage nach der Charakterisierung des Systems (bei der Mehrfachnennungen möglich waren) schreiben 85 % der Studierenden dem System die Eigenschaft zu, phasenweise völlig überlastet zu sein. Gleichzeitig gaben jedoch auch je 56 % der Studierenden an, das System sei hilfreich und es ermögliche ihnen selbständiges und unabhängiges Arbeiten.

## 6 Verwandte Arbeiten

Studien über die Arbeitsweise von Studierenden wurden in der Vergangenheit schon für andere Aspekte des Lernverhaltens durchgeführt: In [AGSA09] wird die Nutzung verschiedener Begleitmaterialien zu einer Programmiervorlesung im zeitlichen Verlauf analysiert, allerdings ohne die Beteiligung eines automatisch Prüfungssystems.

Für einzelne Prüfungssysteme liegen ausgewählte detailliertere Daten vor: Für die Systeme BOSS [HJBG05] und CourseMaker [HHST03] wurde beispielsweise die Zahl wieder-

holter Einreichungen verbesserter Lösungsversuche untersucht, für letzteres zudem auch die zeitliche Verteilung der Einreichungen über den Tag und durchschnittliche erreichte Punktzahlen. Die Beziehung zwischen erreichten Punktzahlen und Zahl der wiederholten Einreichungen wird auch in [Che04] untersucht. Anders als im vorliegenden Artikel fehlt jedoch bei den genannten Publikationen die Analyse, ob Studierende bei der Bearbeitung der Aufgaben aufgegeben haben oder zu einem erfolgreichen Abschluss kamen. Ohne genauere Kenntnis der Bewertungskriterien ist der Vergleich erreichter Punktzahlen zwischen verschiedenen Systemen zudem schwierig. Einen Anhaltspunkt liefert [Tho03], wo die Ergebnisse automatischer und manueller Korrektur verglichen werden.

## 7 Ergebnisse und Fazit

In diesem Artikel wurde durch verschiedene statistische Auswertungen untersucht, wie Studierende mit einem automatisierten Prüfungssystem für Programmieraufgaben in Interaktion treten. Es konnten zwei Anzeichen entdeckt werden, die auf eine bewusste Integration des Systems in die studentische Arbeitsweise sprechen: Erstens treten die Studierenden gleichmäßig von morgens bis in die späten Abendstunden mit dem System in Interaktion und damit über einen wesentlich längeren Zeitraum pro Tag, als dies mit persönlicher Betreuung durch Tutoren möglich wäre. Zweitens stellt das Einholen automatischer Rückmeldungen nicht das Ende des Bearbeitungsprozesses einer Übungsaufgabe dar, sondern wird in den Bearbeitungsprozess integriert.

Dass die permanente Verfügbarkeit eines automatischen Prüfungssystems keine grundlegend andere Arbeitsweise bewirken kann, konnte ebenfalls gezeigt werden. Obwohl durch das Prüfungssystem jederzeit Rückmeldungen zu Lösungsversuchen eingeholt werden konnten, reichte ein erheblicher Teil der Studierenden erst kurz vor der Abgabefrist eine Lösung ein. Auch die Beobachtung, dass das System in solchen Phasen überlastet wird, konnte dabei keine Veränderung des Verhaltens bewirken.

In welchem Umfang die bewusste Nutzung des Systems zu didaktischen Vorteilen führt, kann alleine aufgrund der vorliegenden Zahlen nicht beurteilt werden. Hierzu sind weitergehende, detailliertere Analysen notwendig. Es konnte lediglich beobachtet werden, dass aufeinanderfolgende Lösungsversuche der Studierenden für eine Aufgabe in der Regel mit tendenziell steigenden Punktzahlen bewertet wurden. Daraus kann zumindest geschlossen werden, dass eine rasche automatische Rückmeldung zu einer weiteren Beschäftigung mit der Aufgabe und einem neuen Versuch anregt.

## Literatur

- [AGSA09] Eva Altenbernd-Giani, Ulrik Schroeder und Mostafa Akbari. Programmierungslehreveranstaltung unter der Lupe. In Andreas Schwill und Nicolas Apostolopoulos, Hrsg., *DeLFI*, Jgg. 153 of *LNI*, Seiten 55–66. GI, 2009.
- [Che04] Peter M. Chen. An Automated Feedback System for Computer Organization Projects.

- [HHST03] Colin Higgins, Tarek Hegazy, Pavlos Symeonidis und Athanasios Tsintsifas. The CourseMarker CBA System: Improvements over Ceilidh. *Education and Information Technologies*, 8(3):287–304, 2003.
- [HJBG05] PeyShan Heng, Mike Joy, Russell Boyatt und Nathan Griffiths. Evaluation of the BOSS Online Submission and Assessment System. Bericht RR-415, Department of Computer Science, University of Warwick, Coventry, UK, 2005.
- [SBG09] Michael Striewe, Moritz Balz und Michael Goedicke. A Flexible and Modular Software Architecture for Computer Aided Assessments and Automated Marking. In *Proceedings of the First International Conference on Computer Supported Education (CSEDU)*, 23 - 26 March 2009, Lisboa, Portugal, Jgg. 2, Seiten 54–61. INSTICC, 2009.
- [SG10] Michael Striewe und Michael Goedicke. Feedback-Möglichkeiten in automatischen Prüfungssystemen. In *DeLFI 2010 - 8. Tagung der Fachgruppe E-Learning der Gesellschaft für Informatik e.V.*, number 169 in LNI, Seiten 85–96. GI, 2010.
- [Tho03] Pete Thomas. The evaluation of electronic marking of examinations. In *ITiCSE '03: Proceedings of the 8th annual conference on Innovation and technology in computer science education*, Seiten 50–54, New York, NY, USA, 2003. ACM.

