

Verbesserung der Bestehensquoten durch ein Peer Assessment-Pflichtpraktikum

B. Sc. Simon Roderus, Prof. Dr. Uwe Wienkop
Technische Hochschule Nürnberg, Georg Simon Ohm
Fakultät Informatik
Hohfederstraße 40, 90489 Nürnberg
{simon.roderus, uwe.wienkop}@th-nuernberg.de

Abstract: Peer Assessment ist eine Methode, bei der die Teilnehmer eine gestellte Aufgabe nicht nur bearbeiten und einreichen, sondern—in einer zweiten Phase—diese auch gegenseitig überprüfen, kommentieren und bewerten. Durch diese Methode wird, auch in sehr großen Veranstaltungen, das Üben mit individuellen Bewertungen und individuellem Feedback möglich.

Im Wintersemester 2013/14 wurde dieser Ansatz in der Erstsemesterveranstaltung *Programmieren* an der Technischen Hochschule Nürnberg mit 340 Studierenden als semesterbegleitendes Online-Pflichtpraktikum erprobt. Bei gleichen Leistungsanforderungen wurde bei Studierenden, die erfolgreich am Praktikum teilnahmen, eine Reduzierung der Durchfallquote um durchschnittlich 60 % und eine Verbesserung der Durchschnittsnote um 0,6–0,9 Notenstufen erzielt. Zudem lernten die teilnehmenden Studierenden kontinuierlicher, bereiteten Lerninhalte besser nach und gelangten zu einer überwiegend positiven Einschätzung des Praktikums und der Methode. Im E-Learning System Moodle kann Peer Assessment, mit moderatem Umsetzungs- und Betreuungsaufwand, mit der Workshop-Aktivität realisiert werden. Im Beitrag wird auf die Schlüsselemente des erfolgreichen Einsatzes von Peer Assessment eingegangen.

1 Einführung

Unter Peer Assessment versteht man, dass Studierende von anderen Studierenden (Peers) beurteilt werden und Feedback über die Qualität der eingereichten Arbeit erhalten [Sp12]. Peer Assessment als Forschungsgegenstand erfuhr in den letzten Jahren zunehmend mehr Aufmerksamkeit. Die Zahl der Veröffentlichungen zu diesem Thema hat sich über den Zeitraum von 1998–2010 mehr als verdoppelt [GDO11]. Es wurden unterschiedlichste Fragestellungen untersucht, allerdings nicht immer mit den gleichen Ergebnissen (vgl. [Gi07], S. 202). Die abweichenden Ergebnisse werden auf Unterschiede in der praktischen Durchführung zurückgeführt. Die meisten empirischen Studien zu Assessment nutzen eine vergleichsweise kleine Stichprobe von meist deutlich unter hundert Teilnehmern, zudem wird der Erfolg oft mittels Selbsteinschätzung durch die Teilnehmer ermittelt [Ev13]. Hier kann die vorliegende Arbeit einen Beitrag leisten, denn es wurden mehrere hundert Teilnehmer untersucht und, neben einer Befragung, die Klausurleistung als objektive Messung des Erfolges verwendet.

In der Literatur wird immer häufiger die Nutzung von Peer Assessment als Lernwerkzeug, statt zur Leistungsmessung, diskutiert (vgl. [Gi07]). Die erste und offensichtlichste Funktion ist der Lernerfolg, der durch das Bearbeiten von Aufgaben und somit durch die aktive Auseinandersetzung mit dem Lerninhalt entsteht. Durch das im Rahmen des Peer Assessment vergebene (formative) Feedback kann verhindert werden, dass Ineffektivität durch „Trial & Error“-Lernen entsteht [Sa89]. Nicht zuletzt kann Peer Assessment Auswirkungen auf die Herangehensweise haben, mit der Aufgaben bearbeitet werden: Es soll Studierende in Richtung der erwünschten Lernziele führen und anregen, auf eine tiefe und gründliche Art zu lernen (vgl. [Gi07], S. 21).

2 Motivation für den Einsatz von Peer Assessment

Die Probleme bei der Vermittlung von Lehrinhalten in den ersten Semestern eines Studiums unterscheiden sich nach unseren langjährigen Beobachtungen von denen der höheren Semester. Während die älteren Studierenden meist gelernt haben, dass kurzfristiges Lernen vor einer Klausur nur selten zum Erfolg führt, bleiben die Studierenden der ersten Semester häufig passiv. Übungsaufgaben ohne Pflichtabgabe werden meist nicht oder nur unvollständig bearbeitet. In der Folge können Studierende dem Unterrichtsstoff nur schwer folgen, da das erworbene Wissen nicht durch Übung gefestigt wird. Gerade in sehr

praktischen Fächern, wie z. B. der Programmierausbildung, bleibt – selbst bei Bestehen der Klausur – ohne ausreichende Praxiserfahrung die Fähigkeit zur eigenständigen Lösung von Problemen mittels einer Programmiersprache oft unterentwickelt, was sich dann in Form von Schwierigkeiten in weiterführenden Veranstaltungen zeigt. Zudem ist die Studienabbruchquote im Bereich der Informatik mit 27 % im Vergleich zur durchschnittlichen Abbruchquote, ähnlich wie Ingenieur-Studiengänge, um 8 % höher (Stand: 2010) [He12].

Auch ohne die zuvor genannten Probleme gibt es einige Bereiche in der Lehre, die in typischen Großveranstaltungen meist nicht adressiert werden können:

- *Studierende bekommen tendenziell wenig Feedback zu ihren Arbeiten.*
Aus großen Gruppengrößen resultiert, dass der einzelne Studierende nur selten individuelles Feedback zu seiner Lösung bekommt. Gerade für Anfänger wären eine Anerkennung oder auch weiterführende Kommentare zu den Arbeiten hilfreich.
- *Studierende haben wenig Übung im Verstehen fremder Lösungen.*
Das Verstehen von fremden Lösungsansätzen bedeutet, sich in andere Lösungswege hineindenken und diese nachvollziehen zu können. Diese Kompetenz ist für die spätere Programmierpraxis essenziell, wird aber in der schulischen Ausbildung nur in geringem Maße geübt und kann in größeren Veranstaltungen ebenfalls kaum vermittelt oder trainiert werden.
- *Studierende haben wenig Erfahrung im Bewerten von Leistungen.*
Das Schulsystem ist davon geprägt, dass Leistungseinschätzungen in aller Regel nur von Lehrern gegeben werden. Gerade aber das Bewerten, idealerweise in Kombination mit der Analyse fremder Lösungen, führt zu einem tieferen und gesicherteren Verständnis einer Problemlösung.
- *Stilistische Aspekte können kaum geübt werden.*
Wohl jeder ausgebildete Informatiker weiß um die Bedeutung vom sauber dokumentierten Programmcode. In Lehrveranstaltungen kann jedoch auf diese Wichtigkeit nur hingewiesen und entsprechende Beispiele angegeben werden. Auch bieten Übungen selten ein geeignetes Umfeld, um stilistische Elemente in einer authentischen Situation einzuüben.

Aus diesen Gründen begannen wir bereits vor etwa 6 Jahren damit, uns mit Peer Assessment zu beschäftigen. Hierzu konzipierte Prof. Wienkop ein eigenes Peer-Assessment-System. Als sich dieses als konzeptionell ähnlich zu dem Workshop-Modul, welches zeitgleich in dem Lernmanagementsystem Moodle entwickelt worden war, herausstellte, wurde statt der eigenen Lösung mit der Moodle-Entwicklung weitergearbeitet.

Peer Assessment stellt einen Lösungsansatz für die obigen Problemstellungen durch die folgenden Elemente dar:

- Durch das Prinzip des gegenseitigen Korrigierens erhält jeder Studierende von einem oder mehreren anderen Studierenden Rückmeldung zum eigenen Lösungsansatz.
- Unter Anleitung durch die Dozenten hinsichtlich der Bewertungskriterien kamen erstaunlich gute und hilfreiche Rückmeldungen zustande. Dies deckt sich mit Studien (vgl. [BJ09]), nach denen die Korrektur durch Peers einer Korrektur durch einen Dozenten in der Qualität nicht nachstehen muss.
- Studierende arbeiten sich tendenziell nur ungern in fremde Programme ein, da sie hierfür erst einmal den Lösungsansatz einer anderen Person nachvollziehen müssen. Jedoch wird, wie wohl jeder Dozent bestätigen kann, gerade durch die Korrektur von Aufgaben das Auge in Bezug auf mögliche Fehlersituationen geschult. Im vorliegenden Praktikum haben sich zwei studentische Korrekturen pro Einreichung als brauchbar erwiesen. Entsprechend wurde Moodle so konfiguriert, dass jeder Studierende von zwei Peers je eine Rückmeldung erhält und entsprechend zwei Rückmeldungen selbst abgeben soll.
- Studierende erweisen sich häufig als äußerst strenge Korrektoren, die auch die stilistischen Elemente gut unter die Lupe nehmen, wodurch die Code-Qualität im Mittel sehr viel besser war als in allen Vorjahren.

3 Vorerfahrungen mit dem Einsatz von Peer Assessment

Es wurden über mehrere Jahre Anläufe zum Einsatz von Peer Assessment unternommen. In den ersten Durchführungen konnte mit der Methode kein bzw. kaum Erfolg erreicht werden, und die Ergebnisse waren größtenteils frustrierend. Deshalb sollen im Folgenden die wesentlichen Problemaspekte benannt werden. Dabei wird auch dargestellt, wie sich aus diesen Aspekten Schlüsselfaktoren für den Erfolg der Methode ableiten lassen.

Schneller Rückgang der Teilnahmequote am Peer Assessment.

In allen Durchläufen, in denen Studierende sich frei für die Teilnahme am Peer Assessment entscheiden konnten, nahm die Teilnahmequote rapide ab: Von 80 Studierenden einer Veranstaltung nahmen an der ersten Peer Assessment Aufgabe etwa 60 Studierende teil. Davon erhielten aber nur 2/3 wenigstens eine Rückmeldung. Beim nächsten Durchgang nahmen dann nur noch 40 teil. Durch diesen Verlauf reduzierte sich die Teilnehmerzahl mit jeder Aufgabe, so dass meist bereits nach drei Wochen der Punkt erreicht war, an dem zu wenige teilnahmen und das Verfahren keinen Sinn mehr machte.

Hieraus lassen sich zwei Rückschlüsse ziehen:

- 1) Studierende wünschen sich stärker eine Rückmeldung, als sie selbst dazu bereit sind, eine zu geben.
- 2) Es braucht einen starken äußeren Anreiz (z. B. Pflichtteilnahme), um die Studierenden zu einer längerfristigen Teilnahme zu bewegen.

Unspezifisches Feedback frustriert die Teilnehmer.

In den ersten Durchläufen sollten die Studierenden anhand nur weniger Kriterien Rückmeldung geben, dabei handelte es sich meist um die korrekte Funktionalität eines Programms und einen guten Programmierstil. Häufig wurden nur Punkte auf einer Skala von 1-5 vergeben und auf Kommentare weitestgehend verzichtet, so dass die Rückmeldung mangels Nachvollziehbarkeit für den Beurteilten meist nicht hilfreich oder sogar frustrierend war.

Hieraus lassen sich weitere Rückschlüsse ableiten:

- 3) Die Teilnehmer sind schnell frustriert, wenn sie Beurteilungen nicht nachvollziehen können. Zudem ist der Lerneffekt gering, wenn keine Fehler benannt oder Verbesserungsvorschläge aufgezeigt werden. Deshalb sollten Beurteilungen durch Kommentare begründet werden.
- 4) Feedback im Sinne einer gut/schlecht-Checkliste ist nicht ausreichend. Stattdessen ist ein Beurteilungskatalog mit mehreren Abstufungen notwendig, so dass Beurteilungen, selbst wenn diese nicht durch Kommentare begründet werden, immer noch hilfreich sind.

Faires und konsistentes Beurteilen bereitet Probleme.

Studierende wünschen sich detaillierte, strukturierte und vollständige Beurteilungsanweisungen, wie viele Punkte für welche Leistung zu vergeben

sind. Allerdings ist es in der Praxis weder möglich noch sinnvoll, jede nur mögliche Antwort zu antizipieren. Es zeigte sich zudem die Tendenz, dass wider Erwarten die meisten Studierenden nicht zu milden, sondern zu besonders strengen Urteilen tendieren. So wurden, trotz der Aufforderung milde und konstruktiv zu beurteilen, für simple Programmierfehler, wie das Fehlen einer schließenden Klammer, auch bei ansonsten vollständig richtigem Programm, regelmäßig 0 Punkte vergeben.

- 5) Es sollten begünstigende Formulierungen verwendet werden, damit für „fast richtige“ Lösungen noch viele Punkte vergeben werden. Zudem sollte die Abstufung bei der Beurteilung der Kriterien durch Rubriken vorgegeben werden, um differenzierte Urteile zu bekommen.
- 6) Durch die inkonsistente Beurteilung passierte es, dass die Bewertung einer einzelnen Leistung hinter den anderen Ergebnissen eines Teilnehmers zurückfiel bzw. unverhältnismäßig schlecht abschnitt. Entsprechend sollte Teilnehmern die Möglichkeit gegeben werden, Beurteilungen nachprüfen und unfaire Beurteilungen korrigieren zu lassen. Zudem zeigte sich, dass das Mitverfolgen der eigenen Leistungen über ein ganzes Semester hinweg einen stark anspornenden Charakter hatte.

Nach dem Erkennen dieser Erfolgsfaktoren wurden diese im Sommersemester 2013 in einer kleineren Veranstaltung mit 20 Teilnehmern erprobt. Aufgrund des erstmalig guten Verlaufes erfolgte dann die Entscheidung, die Methode in einer Großveranstaltung einzusetzen.

4 Einsatz von Peer Assessment bei der Lehrveranstaltung Programmieren 1 im Wintersemester 13/14

Im Wintersemester 2013/14 wurde an der Technischen Hochschule Nürnberg die Methode Peer Assessment in der Großveranstaltung Programmieren mit 340 Teilnehmern in drei parallelen Veranstaltungen erneut angewendet. Diese Teilnehmer verteilten sich auf vier Studiengänge, von denen bei dreien die Teilnahme als Pflicht festgelegt worden war (Informatik, Medieninformatik und Wirtschaftsinformatik), während bei dem vierten Studiengang, einer Exportleistung für den Studiengang Mathematik, die Studienordnung noch nicht geändert worden war. Somit befand sich unter den 340 Teilnehmern eine etwa 115 Personen große Teilgruppe, deren Mitglieder wählen konnten, ob sie am

Praktikum teilnehmen wollten oder nicht. Alle Studierenden nahmen an der gleichen Klausur teil. Für das Peer Assessment-Praktikum wurde folgendes Setting verwendet:

- Das Praktikum wurde zur Pflicht erklärt. Dieses Vorgehen erfolgte in Absprache mit den studentischen Vertretern im Fakultätsrat. Es wurde dabei vereinbart, dass durch das Praktikum keine Steigerung des Leistungsniveaus erzwungen werden würde, sondern dass sich die Aufgaben am Schwierigkeitsgrad der früheren freiwilligen Übungsaufgaben orientieren sollten.
- Es gab im Kernzeitraum des Semesters, über neun Wochen hinweg, wöchentlich eine zu bearbeitende Übungsaufgabe, die auch Unteraufgaben enthalten konnte.
- In jeder Woche waren 100 Punkte erzielbar. Insgesamt waren somit 900 Punkte erreichbar. Für das Bestehen des Praktikums waren 700 Punkte notwendig.
- Die 100 Punkte einer Aufgabe verteilen sich auf 70 Punkte für die Abgabe der eigenen Aufgabenlösung und 30 Punkte für die Beurteilung von zwei fremden Einreichungen (je 15 Punkte pro Beurteilung). Diese Aufteilung war so gewählt worden, damit das Praktikum nicht mit dem alleinigen Bearbeiten der Aufgaben bestanden werden konnte, sondern die Studierenden dazu auch andere beurteilen mussten. Zudem waren die Beurteilungspunkte relativ leicht „zu verdienen“.
- Die obigen 70 Punkte (mit Abstufungen von je 10 Punkten) wurden unterteilt in vier Teilkriterien. Zu jedem Kriterium konnten Kommentare abgegeben werden:
 - a) Bis zu 40 Punkte für die korrekte Funktionalität
 - b) 10 Punkte für eine gute Programmstruktur
 - c) 10 Punkte für sinnvolle und angemessene Variablennamen
 - d) 10 Punkte für angemessenes Kommentieren
- Zu Beginn des Praktikums traten bei einigen Studierenden kleinere Anlaufschwierigkeiten auf, die meist durch Bedienfehler wie die Abgabe falscher oder leerer Dateien bedingt waren. Um diese auszugleichen, wurde zunächst diskutiert, die Bestehensgrenze abzusenken. Dann jedoch entschlossen wir uns stattdessen dazu, zwei zusätzliche „Bonusaufgaben“ mit je 100 Punkten anzubieten.

5 Untersuchungsmethoden

Zur Untersuchung des Praktikums kamen unterschiedliche Methoden zum Einsatz. Es wurden die Verbesserung sowohl der Bestehensquote als auch des Notendurchschnitts untersucht. Für jeden Teilnehmer wurden Jahrgang, Studiengang, Praktikumpunkte, Klausurpunkte und ob das Praktikum verpflichtend war, erfasst. Anhand dieser Daten konnten Mittelwerte und Korrelationen verglichen werden. Da es sich bei den Praktikumpunkten und Klausurpunkten um verhältnisskalierte Daten handelt, wurden diese anhand des Korrelationskoeffizienten nach Bravais-Pearson berechnet (vgl. [Bo13]). Die Signifikanz wurde per einseitigem Test auf das 0,05-Niveau geprüft.

Zusätzlich wurde das Abgabeverhalten über Moodle dokumentiert und das Praktikum mittels einer schriftlichen Onlinebefragung mit offenen und geschlossenen Fragen evaluiert. An dieser haben 119 Studierende, also 35 % der Teilnehmer, teilgenommen.

6 Ergebnisse

6.1 Teilnahme und Bestehensquote des Praktikums

Die Bestehensquote ist in Abb. 1 dargestellt. Von den insgesamt 340 Studierenden war für 225 die Teilnahme an dem Praktikum verpflichtend. Von dieser Gruppe haben 90 % ($n = 203$) mindestens 700 Punkte erreicht und das Praktikum somit bestanden. 10 % ($n = 22$) der Pflichtteilnehmer haben die erforderliche Punktzahl nicht erreicht.

Die Auswertung der im Praktikum erzielten Punkte ergab ein deutliches Bild (vgl. Abb. 2): Die Pflichtteilnehmer übertrafen die geforderte Punktzahl meist deutlich. Bei den erfolgreichen Teilnehmern ($n = 207$) lag der Mittelwert bei 918 Punkten. Die größte Teilgruppe befand sich hier im Bereich der maximal möglichen Punktzahlen.

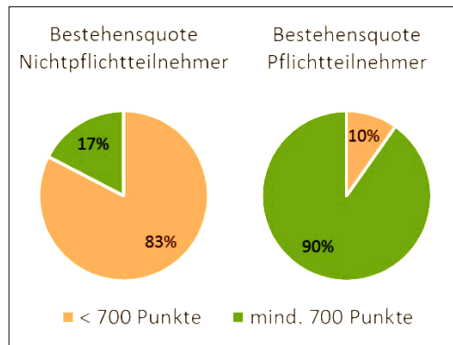


Abbildung 1: Bestehensquoten des Praktikums

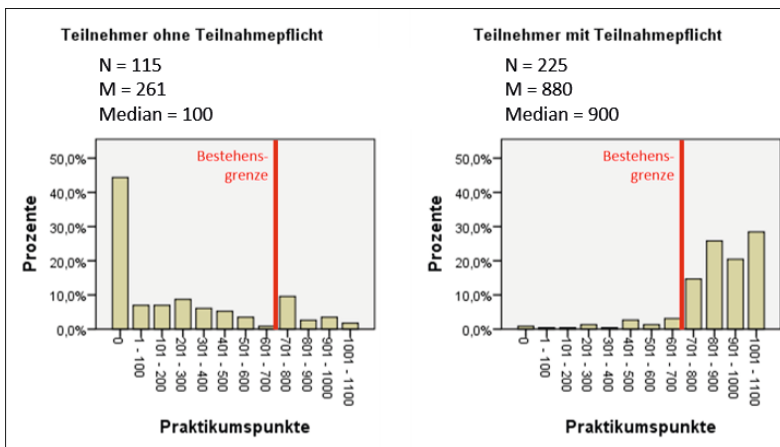


Abbildung 2: Erreichte Punkte im Praktikum in Bezug auf die Gruppen Pflicht und Nichtpflicht.

Ganz anders sah das Bild bei den Teilnehmern ohne Teilnahmepflicht aus. Diese erreichten meist gar keine oder nur sehr wenige Punkte. Leider waren bei einigen Studierenden selbst schlechte Prüfungsleistungen im Erstversuch keine ausreichende Motivation, sich am Praktikum zu beteiligen. Mehrere Studierende aus dieser Gruppe erkundigten sich, ob sie wirklich „unbedingt“ am Praktikum teilnehmen müssten.

6.2 Ein typischer Wochenverlauf

Die Aufgaben wurden jeweils am Montag um 8.00 Uhr freigeschaltet und mussten bis zum nächsten Montag, 8.00 Uhr, bearbeitet werden. Für die Beurteilung war dann bis Donnerstag um 8.00 Uhr Zeit. Die Auswertung der Einreichzeiten (vgl. Abb. 3) zeigt, dass bei den meisten Studierenden das Einreichen erst ziemlich spät (meist: Sonntag) stattfand, die Korrektur erfolgte hingegen typischerweise gleich nach dem Freischalten am Montag der darauffolgenden Woche.

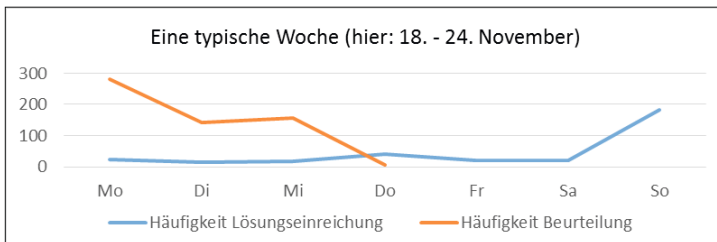


Abbildung 3: Einreichungsverlauf einer typischen Woche (hier: 18.–24. November)

6.3 Steigerung des wöchentlichen Arbeitseinsatzes durch das Praktikum

Vor dem WS 2013/14 war bei den Evaluationen von den Studierenden eine wöchentliche Nachbereitungszeit von etwa 1,5 Std./Woche zurückgemeldet worden, was bei einer 6-stündigen Veranstaltung (4 SWS Vorlesung + 2 SWS Übung) als deutlich zu gering einzustufen ist. Mit dem Pflichtpraktikum stieg der per Onlinebefragung ermittelte durchschnittliche Nachbereitungsaufwand auf 5,2 Std./Woche (vgl. Abb. 4) und liegt damit genau an der Stelle, die auch die Dozenten im Modulhandbuch als angemessen angegeben hatten.

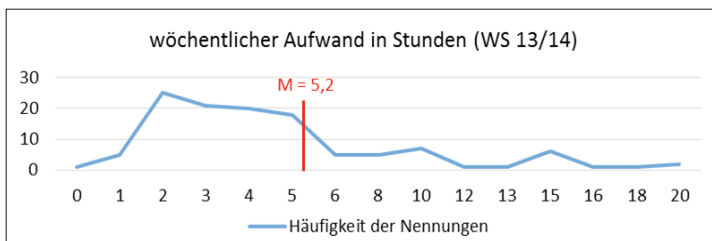


Abbildung 4: Wöchentlicher Arbeitsaufwand für das Praktikum in Stunden im WS 13/14

6.4 Klausurergebnisse

Zur Auswertung wurden die Klausurergebnisse der Jahre 2010–2013, davon die ersten drei Jahre ohne Pflichtpraktikum und 2013 mit Pflichtpraktikum, in den Studiengängen Wirtschaftsinformatik, Informatik und Medieninformatik verglichen. Zudem wurden die Werte denen des Studiengangs Mathematik gegenübergestellt. Da dort das Praktikum freiwillig war und kaum genutzt wurde, konnte dieser Studiengang als Vergleichsgruppe genutzt werden. Insgesamt ging die Auswertung von 873 Klausuren in die Statistik ein. Bereits bei früheren Jahrgängen fiel auf, dass sich die Durchschnittsnoten in den einzelnen Studiengängen, trotz identischer Klausur und wechselnder Dozenten, regelmäßig unterschieden. Deshalb wurde auch die Auswertung getrennt nach den Studiengängen durchgeführt. Es zeigte sich, dass die Verbesserungen in allen Studiengängen eindeutig bestätigt werden konnte (siehe Tab. 1).

Tabelle 1: Prüfungsteilnehmerzahlen ($N_{\text{Prfg.TN}}$) sowie Durchschnittsnoten (D_{Note}) und Durchfallquoten (DFQ) der erfolgreichen Praktikumsteilnehmer nach Studiengang Mathematik (M), Informatik (IN), Medieninformatik (MIN) und Wirtschaftsinformatik (WIN) sowie deren Veränderung (Δ). Studiengänge mit Pflichtpraktikum sind farbig hinterlegt. Bei der Veränderung der Note (ΔD_{Note}) entsprechen negative Werte einer Verbesserung der Note.

	2010-2012		2013					
Stg.	DFQ	D_{Note}	$N_{\text{Prfg.TN}}$	DFQ	D_{Note}	ΔD_{Note}	ΔDFQ (m.E.)**	ΔDFQ (alle)***
M*	22 %	3,3	36	14 % (33%***)	2,3 (3,3***)			
IN	22 %	2,9	75	9 %	2,3	-0,6	-61 %	-58 %
MIN	26 %	3,2	30	11 %	2,6	-0,6	-58 %	-37 %
WIN	45 %	3,9	98	18 %	2,9	-0,9	-60 %	-53 %
Durchschnittliche Veränderung bei IN, MIN, WIN:							-60 %	-51 %

- *) Den Mathematik-Studiengang gibt es erst seit dem Jahr 2012. Der Vergleich erstreckt sich hier also nur über die Jahre 2012–2013.
- **) Die Spalte ΔDFQ (m. E.) enthält die durchschnittliche Verbesserung der Prüfungsteilnehmer, die das Praktikum erfolgreich absolviert haben.
- ***) Zum Vergleich: die Daten der gesamten Prüfungsteilnehmer.

Die durch Peer Assessment erreichte Verbesserung der Durchfallquote fiel für die erfolgreichen Praktikumsteilnehmer der Studiengänge Informatik, Medieninformatik und Wirtschaftsinformatik mit durchschnittlich 60 % überraschend groß aus. Entsprechend zeigte sich eine deutliche Verbesserung der Durchschnittsnote um 0,6–0,9 Notenstufen. Über die Gesamtheit aller Studierenden dieser drei Studiengänge (inklusive der Nichtpraktikumsteilnehmer) ergab sich eine Verbesserung der Durchfallquote von 51 %.

Die Klausur hatte einen ähnlichen Aufbau hinsichtlich der Themengebiete und Aufgaben wie die Klausuren der Vorjahre. Anhand der Studierenden des Mathematikstudiengangs, die am Praktikum kaum teilnahmen, konnte zudem geprüft werden, ob die Prüfung im Vergleich zum Vorjahr einen ähnlichen Schwierigkeitsgrad hatte. Dies wurde bestätigt. Die Durchschnittsnote lag in diesem Studiengang über alle Teilnehmer im Jahr 2012, ebenso wie im Jahr 2013, mit durchschnittlich 3,3 ganz nah bei der fast gleichen Vorjahresdurchschnittsnote. Auffallend war zudem, dass die Mathematikstudierenden, die freiwillig am Praktikum teilnahmen, deutlich besser abschnitten (vgl. Tab. 1). Die Verbesserung lag hier, im Vergleich zum Vorjahr, bei einer ganzen Notenstufe. Da es sich aber nur um 9 % ($n = 10$) der Mathematikstudierenden handelte und es nicht ausschließbar ist, dass diese Teilgruppe sich in anderen relevanten Faktoren wie Fleiß, Leistungsbereitschaft und Fähigkeit zur Selbstorganisation von den restlichen 91 % unterscheidet, wurde dieser Notenunterschied nicht weiter interpretiert.

Die Autoren gehen davon aus, dass das verpflichtende Bearbeiten der Aufgaben einen großen Einfluss auf das Ergebnis hatte. Interessant wäre zu erfahren, welcher Anteil dieser Ergebnisse durch das Peer Assessment erzielt wurde und ob andere Lernmethoden bei verpflichtender Teilnahme ähnlich gute Ergebnisse liefern.

6.5 Statistische Analyse

Für alle Teilnehmer der Klausur im WS13/14 ($N = 239$) wurde untersucht, welche Auswirkungen das Praktikum, bzw. die Pflichtteilnahme an diesem, hat. Die Korrelation zwischen den im Praktikum erzielten Punkten und den Klausurpunkten ergab das in Tab. 2 dargestellte Bild:

Tabelle 2: Korrelationen nach Bravais-Pearson. N = 239
(Teilnehmer der Klausur im WS13/14).

	Praktikumspunkte	Pflicht	Klausurpunkte
Praktikumspunkte	1		
Teilnahmepflicht	,732**	1	
Klausurpunkte	,369**	,326**	1

**Die Korrelation ist auf dem Niveau von 0,01 (1-seitig) signifikant.

Die Korrelation zwischen den Punkten im Praktikum und in der Klausur ($r = ,369^{**}$) zeigt, dass die Leistung im Praktikum als mittelstarker Prädiktor für das spätere Klausurergebnis dienen kann. Andere Studien berichten ähnliche Größen der Korrelation zwischen den im Peer Assessment und den in der Klausur erreichten Punktzahlen (vgl. [VSV10]). Die starke Korrelation von Teilnahmepflicht und Praktikumpunkten zeigt den bereits erwähnten Zusammenhang, dass nur bei verpflichtender Teilnahme viele Punkte im Praktikum erreicht werden. Beide Korrelationen sind hochsignifikant auf dem 0,01-Niveau und übertreffen somit das geforderte 0,05-Niveau.

6.6 Ergebnisse der schriftlichen Teilnehmerbefragung

Tabelle 3: Ergebnisse der schriftlichen Evaluation. Abstufungen der Skala Likert₁:
1 = „stimme nicht zu“ bis 5 = „stimme zu“. Abstufungen der Skala Likert₂:
1 = „sehr gering“ bis 5 = „sehr groß“.

Frage	M	SD	Skalentyp
Das Praktikum hilft am Lernen dranzubleiben.	4,2	1	Likert ₁
Schulnote Praktikum	2,3	0,9	Schulnote
Einschätzung des Lernerfolgs	3,7	1	Likert ₂

Wie oben beschrieben, wurde das Praktikum durch eine umfangreiche Befragung gegen Ende des Praktikums evaluiert. Neben den üblichen geschlossenen Fragen, wurden die Studierenden zu Verbesserungsvorschlägen ebenso wie zu positiven Aspekten befragt. Exemplarische Ergebnisse sind in Tab. 3 dargestellt. Die Teilnehmer empfanden, dass ihnen das Praktikum half „am

Lernen dranzubleiben“, auch der Lernerfolg wurde tendenziell als „eher gut“ eingeschätzt. Die Teilnehmer gaben dem Praktikum im Schnitt eine Schulnote von 2,3 – es gibt also auch noch Verbesserungspotential.

Die Verbesserungsvorschläge hatten häufig zum Inhalt, die Kommunikation unter den Dozenten der drei parallelen Veranstaltungen weiter zu verbessern. Es wurde klar, dass an Praktikumsaufgaben, hinsichtlich der vollständigen und zweifelsfreien Formulierung, ähnlich hohe Anforderungen wie für die Formulierung von Klausuraufgaben zu stellen sind.

Aus den zahlreichen positiven Rückmeldungen zur Fragestellung „Was hat Ihnen am Praktikum gefallen?“ sollen im Folgenden einige exemplarisch herausgegriffen werden:

- *Mir selbst hat das Praktikum sehr viel gebracht. Zum einen das Trainieren des Programmierens, zum anderen wurden alle Studenten/innen dazu benötigt, eine Plattform zum gemeinsamen Lösen von Problemen – kurz: zum gemeinsamen Austausch zu nutzen.*
- *Das Feedback von den Mitstudenten, war zumindest bei mir sehr konstruktiv und hilfreich.*
- *Dass jemand auf die Idee kam, dieses Praktikum überhaupt erst einzuführen ;).*
- *Das Praktikum generell ist eine ganz tolle Idee, um Studenten zum kontinuierlichen Mitlernen zu motivieren. Der Umfang war genau richtig. Das Ziel, eine Mindestpunktzahl zu erreichen, war motivierend, aber nicht überfordernd, da diese auch dann noch erreicht werden konnte, wenn bei der ein oder anderen Aufgabe mal Fehler gemacht wurden.*
- *Das Praktikum sollte auf jeden Fall weiterhin durchgeführt werden. Ich würde es sogar für alle Fächer empfehlen, da die wöchentliche Frist sehr gut zum strukturierten und kontinuierlichen Mitlernen anregt.*

6.7 Aufwand für die Durchführung des Praktikums

Den drei Dozenten stand für die wöchentliche Korrektur, der studentischen Übungseinreichungen, kein weiteres Personal zur Verfügung. Durch das gegenseitige Korrigieren studentischer Arbeiten war es dennoch möglich, dass wöchentlich alle Aufgaben beurteilt wurden. Neben dem Formulieren der Aufgabenstellungen und dem Beantworten von Nachfragen wurde durch folgende Faktoren weiterer Aufwand verursacht:

Es hat sich auch im vorliegenden Praktikum die zuvor genannte Situation, dass Studierende gerne Feedback erhalten wollen, aber umgekehrt in eher ge-

ringerem Maß andere Studierende bewerten, erneut bestätigt. Trotz der zwei Beurteilungen pro Einreichung kam es immer wieder vor, dass einige Studierende überhaupt kein Feedback erhielten.

Aus prüfungsrechtlichen Gründen musste zudem sichergestellt werden, dass die Teilnahme an der Prüfung nicht alleine durch studentische Beurteilungen versagt wurde. Entsprechend mussten alle Beurteilungen, aufgrund deren Teilnehmer weniger als die erforderlichen 700 Punkte erreichten, durch die Dozenten erneut geprüft und bestätigt bzw. korrigiert werden.

Das gleiche galt für alle Einreichungen, die sehr schlechte Bewertungen, oder zwei inkonsistente Beurteilungen erhalten haben. Alle diese Fälle mussten durch die Dozenten nachgeprüft werden, zudem gab es wöchentlich Beschwerden von Studierenden zu überprüfen, die die erhaltene Beurteilung als ungerecht empfanden. Der Aufwand hierfür lag bei ca. 5 Zeitstunden pro Woche. Im vorliegenden Praktikum verteilten wir diese Nachkorrektur wöchentlich wechselnd auf die drei Dozenten.

7 Zusammenfassung und Ausblick

Im vorliegenden Beitrag wurde gezeigt, wie mit der verpflichtenden Teilnahme an einem mittels Peer Assessment durchgeführten Praktikum die Durchfallquote für erfolgreiche Teilnehmer signifikant um 60 % gesenkt und der Notendurchschnitt um 0,6–0,9 Notenstufen gesteigert werden konnte. Es wurden wesentliche Faktoren für die erfolgreiche Umsetzung des Peer Assessment-Praktikums, wie verpflichtende Teilnahme, strukturierte Bewertungslisten und eine Gesamtpunktetabelle, identifiziert.

Das hier vorgestellte Konzept wird mittlerweile in weiteren Lehrveranstaltungen erprobt. Zudem erforschen wir im nun laufenden Semester, ob der Einsatz von Gamification-Elementen Ansporn zur freiwilligen Teilnahme geben kann, und experimentieren mit Aufgabenformaten, die von Studierenden für Studierende erstellt werden.

Literaturverzeichnis

- [Bo13] Bourier, G.: Beschreibende Statistik. Springer Fachmedien. 11. Auflage. Wiesbaden, 2013; S. 208.
- [BJ09] Bouzidi, L.; Jaillet, A.: Can Online Peer Assessment Be Trusted? In: Educational Technology & Society, Vol. 12, No. 4, 2009; S. 257–268.
- [Ev13] Evans, C.: Making Sense of Assessment Feedback in Higher Education. In: Review of educational research, Vol. 83, No. 1, 2013; S. 77.
- [GDO11] Gielen, S.; Dochy, F.; Onghena, P.: An inventory of peer assessment diversity. In: Assessment and Evaluation in Higher Education, Vol. 36, No. 2, 2011; S. 137–155.
- [Gi07] Gielen, S.: Peer assessment as a tool for learning, 2007.
- [He12] Heublein, U.; Richter, J.; Schmelzer, R.; Sommer, D.: Die Entwicklung der Schwund- und Studienabbruchquoten an den deutschen Hochschulen: statistische Berechnungen auf der Basis des Absolventenjahrgangs 2010. In: HIS: Forum Hochschule, Vol. 3/2012, Hannover, 2012; S. 22.
- [Sa89] Sadler, D.: Formative assessment and the design of instructional systems. In: Instructional Science, Vol. 18, No. 2, 1989; S. 120.
- [Sp12] Spiller, D.: Assessment Matters: Self-Assessment and Peer Assessment, 2012; S. 10–URL: http://www.waikato.ac.nz/tdu/pdf/booklets/9_SelfPeer-Assessment.pdf, (04/2014).
- [VSV10] van Zundert, M.; Sluijsmans, D.; van Merriënboer, J.: Effective peer assessment processes: Research findings and future directions. In: Learning and Instruction, Vol. 20, Issue 4, 2010; S. 270–279. Förderungshinweis: Dieses Vorhaben wird aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL11024 (Projekt QuL) gefördert.

Förderungshinweis: Dieses Vorhaben wird aus Mitteln des Bundesministeriums für Bildung und Forschung unter dem Förderkennzeichen 01PL11024 (Projekt QuL) gefördert.