

DISKUSSIONSBEITRÄGE
AUS DEM
FACHBEREICH WIRTSCHAFTSWISSENSCHAFTEN
DER
UNIVERSITÄT DUISBURG–ESSEN
— CAMPUS ESSEN —

Nr. 133

März 2004

**Zur Wirksamkeit von Practice Tests am
Beispiel der Klausur Deskriptive Statistik**

Michael Westermann und Bianca Krol

Zusammenfassung

In der vorliegenden Untersuchung wird der Einfluss von Practice Tests auf das Klausurergebnis im Fach Statistik untersucht. Dazu wurden die Daten von 442 Studierenden der Universität Duisburg-Essen, Campus Essen, verwendet. Diesen wurden während des Semesters Übungsblätter zur Bearbeitung zur Verfügung gestellt. Da die Bearbeitung der Übungszettel freiwillig ist, kann vermutet werden, dass ein Selbstselektionsprozess vorliegt. Diesem Umstand wird durch die Verwendung des Heckman-Verfahrens Rechnung getragen. Diejenigen Teilnehmer an der Statistik-Klausur im Sommersemester 2002, die Übungszettel bearbeitet haben, erzielten deutlich bessere Ergebnisse in der Klausur. Außerdem war eine deutliche Verbesserung der Ergebnisse mit zunehmender Anzahl der bearbeiteten Übungszettel zu verzeichnen. Ferner konnte ein statistisch signifikanter Einfluss der Nationalität und einer wiederholten Klausurteilnahme auf den Klausurerfolg nachgewiesen werden.

Abstract

The present paper analyses the impact of practice test participation on final exams outcomes using data on 442 students from the University of Duisburg-Essen. Since participation in practice tests is voluntary, the results of such experiments always tend to be biased. Using Heckman's two step procedure we are able to identify an improvement in students' exam results if they participated in practice tests and find that exam results are positively linked to the number of practice tests students participated in. Furthermore, we find socio-demographic factors, such as nationality or to be strongly significantly related to exam results. Moreover, also students who repeated the final exams are more likely to pass.

1 Einleitung*

Die Statistikprüfungen sind für Wirtschaftswissenschaftler im Grundstudium häufig die unbeliebtesten. Die Studierenden scheinen einen gewissen Respekt — manchmal sogar Angst — gegenüber diesem Fachgebiet entwickelt zu haben, der sich über Studierendengenerationen hinweg hartnäckig zu halten scheint. Die Klausurergebnisse der letzten Semester¹ weisen eine durchgängig hohe Durchfallquote auf. Hierdurch scheint diese Angst berechtigt zu sein. Aufgrund dieser Situation bietet der Lehrstuhl für Statistik der Universität Duisburg–Essen, Campus Essen, seit geraumer Zeit Übungszettel für die Studierenden an. Mit Hilfe dieser Übungszettel kann der individuelle Wissensstand im Vorfeld der Klausur getestet werden. Die Übungszettel können von den Studierenden am Lehrstuhl eingereicht werden, wenn sie bearbeitet worden sind. Diese werden korrigiert, mit Musterlösungen versehen und an die Studierenden zurückgegeben. Damit bietet der Lehrstuhl für Statistik ein Instrument zur Verbesserung der Prüfungsleistung an.

Seit der Vorstellung der *teaching–testing machine* von PRESSEY (1926) hat es im Bereich der pädagogischen und psychologischen Forschung eine Vielzahl von Untersuchungen zur Wirksamkeit verschiedener Methoden zur Steigerung des Lernerfolgs gegeben. Zu diesen Methoden gehören u.a. Probeklausuren,² Dozentenreviews³ und permanente Lernerfolgskontrolle durch

*Für hilfreiche Anregungen und Kommentare danken wir Dirk Czarnitzki, Ralf Dewenter sowie Jörg Stank. Weiterhin möchten wir uns bei Michael L. Epstein und Dr. Bernd Jacobs für die freundliche Unterstützung bedanken.

¹Vgl. KLADROBA (2002).

²Vgl. z.B. GRETES/GREEN (2000); SLY (2000); LEE–SAMMONS/WOLLEN (1989).

³Unter Dozentenreviews versteht man Veranstaltungen, in denen die wichtigsten Themenkomplexe und Lösungswege für anstehende Klausuren gesondert vorgestellt werden. Vgl. z.B. JACOBS (2003); BOL/HACKER (2001).

Übungsaufgaben.⁴ Zu letzterer Methode zählt die Ausgabe von Übungszetteln an Studierende während des laufenden Semesters. Dabei ist zu beachten, dass die gestellten Aufgaben denen der „richtigen“ Klausur entsprechen, damit der eigene Kenntnisstand durch jeden Studierenden realistisch eingeschätzt werden kann.⁵ Darüber hinaus gibt es in der Literatur eine breite Diskussion über die Wirksamkeit der verschiedenen Formen des Feedbacks an Untersuchungsteilnehmer. Hier kann unterschieden werden zwischen itemspezifischem Feedback, welches sofort nach der Beantwortung einer jeden Frage erfolgt und einer Rückmeldung verschiedener Kennziffern, wie beispielsweise Prozent der richtigen Lösungen oder Noten. Beim itemspezifischen Feedback können unter Umständen durch den Probanden weitere Lösungsversuche unternommen werden. Gibt es bei jedem Versuch eine Rückmeldung und kann der Proband bis zur Ermittlung der richtigen Antwort seine Antwortversuche korrigieren, dann spricht man von einem *answer-until-correct* Prozess.⁶ Ferner können die zeitlichen Zwischenräume von der Bearbeitung der Aufgaben bis zur Erteilung des Feedbacks variieren. Neben dem sofortigen Feedback gibt es das verzögerte Feedback. EPSTEIN hat in verschiedenen Studien mit weiteren Autoren herausgearbeitet, dass sofortiges Feedback gekoppelt mit der Möglichkeit die Antworten so lange zu korrigieren, bis die richtige gefunden worden ist, die besten Ergebnisse erzielte.⁷ Die Überlegenheit des unmittelbaren Feedbacks wurde bereits früher⁸ und teilweise in anderen, wie z.B. motorischen Fähigkeiten, untersucht.⁹

⁴Vgl. z.B. DIHOFF/BROSVIC/EPSTEIN (2003).

⁵Vgl. JACOBS (2001).

⁶Vgl. DIHOFF/BROSVIC/EPSTEIN (2003), S. 534.

⁷Vgl. hierzu DIHOFF/BROSVIC/EPSTEIN (2003), EPSTEIN ET AL. (2002) und EPSTEIN/EPSTEIN/BROSVIC (2001).

⁸Vgl. z.B. DEMPSEY/LITCHFIELD/DRISCOLL (1993), HANNA/LONG (1979) und BEESON (1973).

⁹Vgl. z.B. BROSVIC/COHEN (1988).

Die Studierenden am Lehrstuhl Statistik erhalten nach jedem Übungszettel ein (zeitlich verzögertes) Feedback, aus dem hervorgeht, wie viele Aufgaben korrekt bearbeitet und welche Fehler gemacht worden sind. Die Vorgehensweise am Lehrstuhl Statistik kann damit eingeordnet werden als Practice Test, dem ein itemspezifisches Feedback in Form von *Knowledge of Correct Result* und *elaboriertem Feedback* folgt.¹⁰

Vor diesem Hintergrund ist es das Ziel des vorliegenden Beitrags, die Frage zu klären, ob es Unterschiede in den Klausurergebnissen zwischen Studierenden gibt, die die Möglichkeit des Feedbacks nutzten und solchen, die es nicht nutzten. Dazu werden die Klausurergebnisse des Sommersemesters 2002 untersucht. Nachdem im folgenden Kapitel ein kurzer Überblick über ähnliche Untersuchungen aus nicht-wirtschaftswissenschaftlichen Fachbereichen gegeben wird, schließt sich in Kapitel drei die Beschreibung des Untersuchungsaufbaus an. In Kapitel vier werden die Ergebnisse dargestellt und diskutiert, das abschließende fünfte Kapitel enthält das Fazit der Untersuchung.

2 Lernwirksamkeit von Practice Tests

Die Untersuchung der Auswirkungen von Practice Tests auf den Lernerfolg findet überwiegend in den pädagogischen und psychologischen Fachbereichen statt. In verschiedenen empirischen Untersuchungen ist der Frage nachgegangen worden, welche Auswirkungen Probeklausuren verbunden mit einem Feedback an den Lernenden haben und ob eine Verbesserung der anschließenden Prüfungsleistungen erzielt werden konnte. Probeklausuren sollten nicht zu sehr von der eigentlichen Klausur abweichen. In der Regel könnte eine Probeklausur auch als echte Klausur gestellt werden. Es werden zwar keine auf den Abschluss anrechenbaren Noten vergeben, jedoch erhält der Studierende

¹⁰Vgl. hierzu JACOBS (1998).

im Nachhinein eine Rückmeldung, die seine Leistung eindeutig widerspiegelt.

Der nachfolgende Literaturüberblick zeigt einen Ausschnitt aus den empirischen Forschungsarbeiten, die zu diesem Themenfeld durchgeführt worden sind.

Übersicht 1: Empirische Arbeiten zur Lernwirksamkeit von Practice Tests

Untersuchungsfokus	Versuchsaufbau	Ergebnisse
BOL/HACKER (2001)		
Unterschiedliche Auswirkungen einer Probeklausur und eines Dozentenreview auf die Klausur.	<ul style="list-style-type: none"> • Zufallsauswahl: Probeklausur oder Dozentenreview • Probeklausur: MC-Aufgaben, nicht identisch mit den Klausuraufgaben, keine individuelle Rückmeldung, allgemeine Erläuterung durch den Dozenten • Dozentenreview: Vorstellung der wichtigsten Themen und Beantwortung von Fragen 	Die Teilnehmer an der Probeklausur schnitten in der Midterm-Klausur deutlich schlechter ab. Zur Abschlussklausur wurde dieser Effekt nicht mehr festgestellt.
DIHOFF/BROSVIC/EPSTEIN (2003)		
Unterschiedliche Auswirkungen einer Probeklausur mit verschiedenen Feedback-Formen.	<ul style="list-style-type: none"> • Fünf Probeklausuren pro Semester • Übungsklausuren mit je 50 Übungsaufgaben. • Jeweils zehn davon werden per Zufallsauswahl für die Hauptklausur ausgewählt 	Klausurteilnehmer, die während des Semesters direktes Feedback erhielten, schnitten deutlich besser ab als die Teilnehmer der anderen Gruppen. Verzögertes Feedback führte zu den schlechtesten Ergebnissen.

Untersuchungsfokus	Versuchsaufbau	Ergebnisse
	<ul style="list-style-type: none"> • Feedback: direkt nach jeder Aufgabe/nach jedem Test/um 24 Stunden verzögert • Kontrollgruppen ohne Feedback 	

GRETES, J.A./GREEN, M. (2000)

Auswirkungen von computer-assisted assessment ¹¹ auf die Ergebnisse von Klausuren.	<ul style="list-style-type: none"> • Ziehung von MC–Aufgaben aus einem Pool für die Hauptklausur, die restlichen Aufgaben standen als Übungsaufgaben zur Verfügung • 13 Übungstests mit jeweils 30–35 Aufgaben, die sich von den Klausuraufgaben komplett unterscheiden • Rückmeldung über die Anzahl der richtigen Lösungen in Prozent • Selbstselektion der Studierenden zur Probeklausur. 	Die Teilnehmer der Übungstests hatten signifikant bessere Klausurergebnisse als die Nicht–Teilnehmer. Ebenso verbesserten sich die Noten mit der Anzahl der gelösten Übungstests.
-----------------------------------------------------------------------------------------------	----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

JACOBS (2003)

Unterschiedliche Auswirkungen einer Probeklausur und einer Seminarsitzung zur Prüfungsvorbereitung auf die Klausurergebnisse und auf die Angst.	<ul style="list-style-type: none"> • Zufallsexperiment: Probeklausur oder Seminarsitzung • Probeklausur: Nicht identisch mit den Klausuraufgaben, aber lehrzieladäquate Aufgaben, umfassendes itemspezifisches Feedback 	Die Probeklausurteilnehmer des Statistikseminars konnten leichte Leistungsvorteile erzielen. Die Angst wurde nicht abgebaut.
-------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------

Untersuchungsfokus	Versuchsaufbau	Ergebnisse
	<ul style="list-style-type: none"> • Seminar: Vorstellung der wichtigsten Themen, Aufgabentypen sowie klausurähnliche Beispielaufgaben 	
SLY (2000)		
Auswirkungen von Practice Tests auf die Ergebnisse eines späteren „richtigen“ CML Tests. ¹²	<ul style="list-style-type: none"> • Untersuchung des Einflusses ähnlicher Inhalte, der Anzahl der Fragen und eines itemspezifischen Feedbacks • Selbstselektion der Studierenden verschiedener Fachrichtungen • 14 einzelne Studien im Zeitraum 1997 bis 1999 • In 12 Studien wurden Practice Tests eingesetzt 	Teilnehmer des Practice Tests schnitten im Mittel besser in der Klausur ab als Nicht-Teilnehmer. Ähnliche Inhalte und die Anzahl der Fragen im Practice Test hatten keinen Einfluss auf den Klausurerfolg. Die Wirkungen eines itemspezifischen Feedbacks konnte nicht eindeutig nachgewiesen werden.

Der kurze Ausschnitt der empirischen Studien in diesem Bereich zeigt, dass aufgrund der teilweise sehr unterschiedlichen Untersuchungsausrichtungen nur wenige einheitliche Ergebnisse in Bezug auf die Wirksamkeit der Teilnahme an Probeklausuren auf die Ergebnisse in der realen Klausur abgeleitet werden können. Die Wirkungsweise und die Effekthöhe scheint damit situationsabhängig zu sein. Selbst ein einheitlich positiver Effekt von Probeklausuren kann nicht konstatiert werden. So haben BOL/HACKER (2001) herausgefunden, dass die Teilnehmer einer Probeklausur sogar schlechter abschnitten als die Teilnehmer einer anderen Art der Prüfungsvorbereitung

¹¹Unter computer-assisted assessment versteht man laut JACOBS (2003) Übungstests mit Multiple-Choice-Aufgaben, die unbenotet sind.

¹²CML = Computer Managed Learning.

(Dozentenreview). Als Begründung wird angegeben, dass die Probeklausurteilnehmer anscheinend identische Aufgaben in der Hauptklausur erwartet hätten. Hier hat die Teilnahme an der Probeklausur also dazu geführt, dass sich die Studierenden in einer falschen Sicherheit wogen.

JACOBS (2003) hingegen konnten in einer Untersuchungsgruppe (Teilnehmer des Statistikseminars) Leistungsvorteile der Probeklausurteilnehmer gegenüber einer dozentengesteuerten zusätzlichen Seminarsitzung als Prüfungsvorbereitung nachweisen. Sie haben zusätzlich die Auswirkungen auf einen Abbau der Angst vor der Hauptprüfung untersucht. Diese ließ sich aber auch mit der Probeklausur nicht verringern (sie nahm sogar in einem Fall zu).

Auch die Auswirkungen von verschiedenen Feedback-Formen ist nicht einheitlich belegt worden. In zwei der Studien innerhalb der Untersuchung von SLY (2000) führten Practice Tests ohne itemspezifisches Feedback sogar zu einem größeren positiven Effekt in der Hauptklausur. DIHOFF/BROSVIC/EPSTEIN (2003) konnten in ihrer Untersuchung jedoch eindeutig positive Auswirkungen eines itemspezifischen Feedbacks feststellen. Die Höhe des Effektes hing allerdings stark von der Art des Feedbacks ab. So führte ein direktes Feedback zu den besten Ergebnissen in der Hauptklausur, gefolgt von einem um 24 Stunden verzögerten und einem Feedback am Ende des jeweiligen Tests.

GRETES/GREEN (2000) überprüften ferner die Auswirkungen einer mehrfachen Teilnahme an Probeklausuren und konnten eine zunehmende Verbesserung mit einer steigenden Anzahl von Probeklausuren feststellen. Auch die allgemein in Practice Tests als zielführend anerkannten Faktoren wie der Vorteil einer gezielten Vorbereitung, die Transparenz über die prüfungsrelevanten Inhalte und Aufgabenstellungen sowie der Übungseffekt als solcher werden nicht durchgehend bestätigt.¹³ Eine Berücksichtigung von soziodemo-

¹³Vgl. SLY (2000).

graphischen Faktoren, wie beispielsweise Nationalität, Alter, Geschlecht etc. hat dagegen nicht in allen Studien stattgefunden hat. In anderen Studien wurde deren Einfluss als nicht signifikant identifiziert.¹⁴

In der vorliegenden Untersuchung wird keine Probeklausur im klassischen Sinn betrachtet. Vielmehr handelt es sich um eine kontinuierliche Prüfungsvorbereitung zu fünf Zeitpunkten — eine Vorgehensweise, welche der, wie sie in DIHOFF/BROSVIC/EPSTEIN (2003) beschrieben wird, sehr ähnlich ist. Es werden die Effekte einer Bearbeitung von Übungsblättern auf die Ergebnisse der Klausur *Deskriptive Statistik für Wirtschaftswissenschaftler* ermittelt. Diese Ergebnisse werden dann mit denen der oben angeführten Studien verglichen. Um den Einfluss der Anzahl der bearbeiteten Übungsblätter sowie soziodemographischer Faktoren auf das Klausurergebnis ermitteln zu können, werden Logit- und Probit-Schätzungen durchgeführt. Da die Teilnahme an diesen lernfördernden Maßnahmen freiwillig ist, ist davon auszugehen, dass ein Selbstselektionsprozess der Studierenden vorliegt. Um diesem Umstand Rechnung zu tragen, wird ein Heckman-Selektionsmodell geschätzt.

3 Beschreibung des Experiments

3.1 Personenkreis

Die Basis der Analyse bilden alle Studierende des Fachbereichs Wirtschaftswissenschaften, welche im Sommersemester 2002 an der Klausur zur Deskriptiven Statistik teilgenommen haben. Diese Klausur gehört zu einer Pflichtleistung für Studierende der folgenden Studiengänge:¹⁵

¹⁴Vgl. DIHOFF/BROSVIC/EPSTEIN (2003) und SLY (2000).

¹⁵Aus dieser Betrachtung ausgeschlossen wurden Lehramtsstudierende, da diese nur Teile des Pflichtprogramms für Diplom- und Bachelorstudenten lernen müssen. Darüber hin-

- Betriebswirtschaftslehre (BWL) mit dem Abschluss Diplom-Kaufmann/–frau
- Volkswirtschaftslehre (BA) mit dem Abschluss Bachelor of Arts
- Wirtschaftsinformatik (WIINF) mit dem Abschluss Diplom-Wirtschaftsinformatiker/in

3.2 Practice Test

Als Hilfestellung und Lernkontrolle besitzen die Studierenden die Möglichkeit, Übungsblätter zu bearbeiten. Auf jedem Übungsblatt werden den Studierenden mehrere Aufgaben zur Bearbeitung gestellt, welche auf den bis zu diesem Zeitpunkt behandelten Vorlesungs- und Übungsstoff abgestimmt sind. Es handelt sich dabei nicht um Klausuraufgaben, die Inhalte sind jedoch eindeutig lehrzielorientiert. Die Studierenden besitzen somit die Möglichkeit, sich bewusst mit den bereits behandelten Themengebieten zu befassen und einige Aufgaben selbständig zu bearbeiten. Gerade das selbständige Bearbeiten der Aufgaben verspricht einen hohen Lernerfolg, da nur so Methoden und Vorgehensweisen bei der Lösung der Aufgaben sinnvoll erlernt werden können und der Studierende danach realistisch einschätzen kann, welche Zeit die Lösung einer (unbekannten) Aufgabe in Anspruch nimmt.

Die bearbeiteten Übungsblätter können am Lehrstuhl eingereicht werden, wo sie korrigiert und gegebenenfalls kommentiert werden. Sie werden dann mit einer Musterlösung versehen und an die Studenten zurück gegeben.¹⁶

aus gab es zum Zeitpunkt der Klausur noch einige Studenten sogenannter *alter* Prüfungsordnungen. Diese befinden sich aber bereits weitgehend im Hauptstudium oder sind problemlos mit unter die drei genannten Studiengänge zu fassen.

¹⁶Musterlösungen werden dabei nicht pauschal verteilt, sondern lediglich an die Studierenden ausgeteilt, welche eigene Lösungsvorschläge eingereicht haben.

Da die Studenten nun nicht nur die Lösung, sondern auch den ausführlichen Lösungsweg kennen, kann man in diesem Zusammenhang vom *Knowledge of Correct Result* sprechen, was sich positiv auf das Ergebnis in der tatsächlichen Prüfungssituation auswirken sollte. Der Effekt einer bloßen Reproduktion der Lösungen ohne selbständigen Lösungsversuch wird in der Literatur nicht einheitlich beurteilt. Im Vergleich zur selbständigen Erarbeitung der Lösung ist der Lerneffekt nach Meinung der Verfasser jedoch größer einzuschätzen.

Auch wenn es sich hierbei nicht um eine Probeklausur im eigentlichen Sinn handelt, erfüllt die Ausgabe und Korrektur von Übungsblättern gemäß JACOBS (2000) doch die Anforderungen an eine Erfolg versprechende Maßnahme, da

- den Studenten eine realistische Einschätzung der eigenen Leistung ermöglicht wird,
- durch das Wissen der korrekten Lösung, die Lerndefizite aufgedeckt werden und
- gegebenenfalls die Motivation der Studenten angeregt wird, bis zur tatsächlichen Klausur diese Defizite zu beseitigen.

Insgesamt konnten im Sommersemester 2002 fünf Übungsblätter bearbeitet werden. Bei der Auswertung der Klausurergebnisse wird überprüft, ob Klausurteilnehmer, die die Übungsblätter bearbeitet haben, signifikant bessere Ergebnisse (gemessen anhand der erreichten Punktzahl und der Durchfallquote) erzielten als die Studenten, welche keine Übungsblätter bearbeitet haben.

3.3 Klausur

Die Klausur zur Deskriptiven Statistik besteht aus 6 zu bearbeitenden Aufgaben aus den Themengebieten Methodenlehre (4 Aufgaben á 20 Punkte) und Wirtschaftsstatistik (2 Aufgaben á 20 Punkte). Den Studierenden wird zweimal im Semester¹⁷ die Möglichkeit gegeben, an der Klausur teilzunehmen.¹⁸

Übersicht 2: Verwendete Variablen und deskriptive Statistik

Variable	Mittelw.	Stabw.	Min	Max	Beschreibung
ERFOLG	0,40	0,49	0	1	Klausur bestanden: (ja=1, nein=0)
PUNKTE	43,27	16,97	7	97	erreichte Punktzahl
NOTE	4,45	0,75	1,3	5	erreichte Note
UEB	0,24	0,43	0	1	Übungszettel abgegeben: (ja=1, nein=0)
NR-UEB	0,63	1,31	0	5	Anzahl der abgegebenen Übungszettel
BWL	0,70	0,46	0	1	studiert BWL (ja=1, nein=0)
BA	0,08	0,27	0	1	studiert BA (ja=1, nein=0)
WIINF	0,22	0,42	0	1	studiert WIINF (ja=1, nein=0)
TERMIN	0,59	0,49	0	1	Termin der Klausur (HT=1, NT=0)
VTERMIN	0,19	0,39	0	1	wiederholte Teilnahme (ja=1, nein=0)
ALTER	22,57	2,21	18	34	Alter des Klausurteilnehmers*)
SEM	2,38	1,13	1	10	Anzahl der Studiensemester
GENDER	0,68	0,47	0	1	Geschlecht (männlich=1, weiblich=0)
NAT	0,82	0,38	0	1	Nationalität (deutsch=1, andere=0)

*) Zum Zeitpunkt der Klausurteilnahme.

Der Untersuchung liegen die Ergebnisse der Klausurteilnehmer zugrunde. Zusätzlich können die Informationen, ob und wie viele Übungsblätter die Studierenden während des Semesters bearbeitet haben, verwertet werden. Hinzu kommen weitere, individualisierte Daten, welche in *Übersicht 2*

¹⁷Dabei ist der Haupttermin (HT) zu Beginn und der Nachtermin (NT) zum Ende der vorlesungsfreien Zeit terminiert.

¹⁸Studierende, welche zum Haupttermin die Klausur nicht bestanden haben, besitzen somit die Möglichkeit, zum Nachtermin erneut an der Klausur teilzunehmen.

dargestellt sind.

Mit PUNKTE wird die Gesamtzahl der erzielten Punkte beschrieben. Insgesamt waren in der Klausur 120 Punkte zu erzielen. Als weiteres Merkmal zur Erfassung des Klausurerfolgs dient die Information, ob die Klausur bestanden wurde oder nicht (ERFOLG). Diese wird als Null–Eins–Variable verwendet, welche den Wert Eins bei einer erfolgreichen Teilnahme an der Klausur annimmt und Null, wenn die Klausur nicht bestanden wurde. Mit der Variable VTERMIN wird erfasst, ob ein Teilnehmer der Klausur zum Nachtermin bereits an der Klausur zum Haupttermin teilgenommen hat. Trifft dies zu, so wird die Variable auf den Wert Eins gesetzt. In diesem Fall kann die wiederholte Teilnahme¹⁹ auch als *repeatable Testing* oder *optionales Re-testing* bezeichnen,²⁰ wobei anzumerken ist, dass diese Möglichkeit nur den Studierenden geboten wird, welche zum Haupttermin nicht erfolgreich an der Klausur teilgenommen haben. Schließlich erfasst die Variable SEM die Anzahl der Studiensemester zum Zeitpunkt der Klausurteilnahme.

4 Ergebnisse

4.1 Deskriptive Analyse

Insgesamt haben im Sommersemester 2002 442 Personen an der Klausur teilgenommen. Davon entfielen 261 auf den Haupttermin und 181 auf den Nachtermin. Der überwiegende Teil der Teilnehmer gehört dem Studiengang BWL an (308 Teilnehmer), gefolgt von den Studierenden der Wirtschaftsinformatik (99) und den Bachelor–Studenten (35) (vgl. *Tabelle 1*). Von den 442 Teilnehmern haben 334 Personen kein und 108 haben (mindestens) ein Übungsblatt

¹⁹Der Teilnehmer hat ja mindestens ein zweites Mal an der Klausur teilgenommen.

²⁰Vgl. hierzu JACOBS (2001).

Tabelle 1: Teilnehmer der Klausur Deskriptive Statistik — Übersicht

	BWL	BA	WIINF	Insg.
Insg.	308	35	99	442
Termin				
HT	179	20	62	261
NT	129	15	37	181
Übungsblätter				
ja	81	1	26	108
nein	227	34	73	334
Anz. der. Übungsblätter				
1	25	0	9	34
2	17	1	7	25
3	14	0	5	19
4	11	0	3	14
5	14	0	2	16

bearbeitet.²¹ Besonders auffällig ist, dass ungefähr 26% aller Klausurteilnehmer der Studiengänge Betriebswirtschaftslehre und Wirtschaftsinformatik von dem Angebot Gebrauch gemacht haben, wogegen gerade einmal 3% der Bachelor-Studierenden, nämlich genau eine Person, sich dieser Lernkontrolle unterzogen.²²

4.1.1 Durchschnittliche Punktzahl

Ausgangspunkt der Analyse ist die Betrachtung der Klausurergebnisse. Die durchschnittlich erzielte Punktzahl beträgt für alle 442 Klausurteilnehmer

²¹Dabei wird optimistischerweise davon ausgegangen, dass die Personen, welche einen Zettel zur Korrektur einreichen, diesen auch selbständig bearbeitet haben, obgleich es den Studenten mittlerweile sehr wohl möglich ist, anderweitig (via Internet oder durch Kommilitonen) an eine Lösung der entsprechenden Übungszettel zu gelangen.

²²Dieses erstaunt umso mehr, da diese Teilnehmer mit Abstand die schlechtesten Ergebnisse in der Klausur erzielt haben, wie in *Tabelle 2* und *Tabelle 3* ersichtlich ist.

43,3 Punkte. Der Median beträgt 42 Punkte. Damit wurde die Klausur im Mittel nicht bestanden. Weiterhin sind folgende Ergebnisse erwähnenswert (vgl. *Tabelle 2*):

Tabelle 2: Erreichte Punkte — Deskriptive Statistik

	n	\bar{x}	σ
Insg.	442	43,3	16,97
BWL	308	42,7	16,34
BA	35	35,3	15,02
WIINF	99	48,0	18,30
Übungsblätter			
ja	108	49,9	17,57
nein	334	41,1	16,23
Anz. der. Übungsblätter			
1	34	44,5	14,82
2	25	52,9	15,50
3	19	49,5	20,43
4	14	50,1	19,93
5	16	57,3	18,7
Termin			
HT	261	41,2	16,82
NT	181	46,2	18,81
Geschlecht			
m	291	43,9	17,65
w	137	41,8	15,58
Nationalität			
deutsch	362	44,7	17,13
andere	79	37,1	14,73

- Die Studierenden der Wirtschaftsinformatik und BWL haben im Durchschnitt 13 bzw. 7 Punkte mehr erreicht als die Bachelor-Studierenden. Diese erzielten durchschnittlich gerade einmal 35 von 120 möglichen Punkten und waren somit weit von einer erfolgreichen Klausurteilnahme entfernt.

- Klausurteilnehmer, welche keine Übungszettel bearbeitet haben, haben in der Klausur durchschnittlich 9 Punkte weniger erzielt (41,1 Punkte zu 49,9 Punkte). Dieser Unterschied ist statistisch signifikant zum Niveau 1% und gewinnt besonders an Bedeutung, da in diesem Intervall die Grenze zwischen Bestehen und Nichtbestehen der Klausur liegt.²³
- Zum Haupttermin der Klausur wurden signifikant weniger Punkte erreicht als zum Nachtermin (Signifikanzniveau = 1%). Auch hier liegt die Grenze einer erfolgreichen Klausurteilnahme genau zwischen den beiden durchschnittlichen Punktzahlen.
- Ein ähnlich deutlicher Unterschied ist bei der Nationalität der Klausurteilnehmer zu beobachten. Zwar erreichen deutsche Teilnehmer signifikant mehr Punkte als ausländische Studierende, jedoch reicht die durchschnittlich erzielte Punktzahl nicht für eine erfolgreiche Klausurteilnahme aus.
- Weniger deutlich sind die Unterschiede der erreichten Punktzahl, wenn nach dem Merkmal Geschlecht unterschieden wird.

Bei der Analyse der Daten muss berücksichtigt werden, dass die Teilnahme an der lernfördernden Maßnahme dem Prozess der Selbstselektion unterliegt, da die Studenten selbst entscheiden können, ob sie an dem Practice Test teilnehmen oder nicht. Es ist zu vermuten, dass die Studenten eher dazu geneigt sind, ein Übungsblatt zur Korrektur einzureichen, welche a priori eine gewisse Affinität zur statistischen Materie aufweisen. Der positive Zusammenhang zwischen der Klausurleistung und der Anzahl der bearbeiteten Übungsblätter dürfte demnach von einer dritten – latenten – Variablen (wie z.B. generelles Interesse für Mathematik/Statistik) überlagert werden. Dass Personen, für die dieses zutreffend ist, ein besseres Klausurergebnis erzielen, ist dann nicht

²³Die Klausur wurde bestanden, wenn der Teilnehmer mindestens 45 Punkte erzielt hat.

verwunderlich. Dieser Tatsache wird jedoch im weiteren Verlauf der Analyse Beachtung geschenkt.

4.1.2 Klausurerfolg, ODDS und ODDS-Ratios

Von den 442 Teilnehmern haben 177 die Klausur bestanden. Das entspricht einer Quote von 40,05 Prozent. 265 Studenten (59,95%) haben die Klausur nicht bestanden (vgl. *Tabelle 3*). Die deskriptive Analyse des Klausurerfolgs zeigt zum Teil sehr deutliche Unterschiede für verschiedener Teilgruppen der Klausurteilnehmer. Diese Unterschiede werden mithilfe eines χ^2 -Tests statistisch untermauert.²⁴ Von den Studierenden der Wirtschaftsinformatik haben knapp 53% die Klausur bestanden. Bei den BWL- und Bachelor-Studierenden lag diese Quote bei 38% bzw. sogar nur 20%, also jeweils deutlich unter 50%. Bestätigt wird dieses Ergebnis — wie bereits erwähnt — durch eine χ^2 -Teststatistik, welche den Wert 12,67 annimmt und somit statistisch signifikant zum 1%-Niveau ist.

Eine andere statistische Kennzahl verdeutlicht den Unterschied zwischen einer erfolgreichen und nicht erfolgreichen Klausurteilnahme innerhalb einer bestimmten Personengruppe. Die *Chance*²⁵, definiert als das Verhältnis der Wahrscheinlichkeit, die Klausur zu bestehen und der Wahrscheinlichkeit, die Klausur nicht zu bestehen, beträgt für alle Klausurteilnehmer $0,4005/0,5995 = 0,6681$.²⁶ Die Wahrscheinlichkeit einer erfolgreichen Klausurteilnahme beträgt also nur ungefähr zwei Drittel der Wahrscheinlichkeit, die Klausur nicht zu bestehen. Es ist zu erkennen, dass die Chance einer

²⁴Vgl. ASSENMACHER (2000) oder VON DER LIPPE (1999).

²⁵Vgl. LONG (1997) und KOHLER/KREUTER (2001).

²⁶Chancen kleiner als Eins deuten demnach darauf hin, dass es (in der entsprechenden Gruppe) relativ unwahrscheinlicher ist, die Klausur zu bestehen, Werte größer als Eins deuten auf das Gegenteil hin.

Tabelle 3: Klausurerfolg — Bedingte Häufigkeiten, Odds-Ratio und χ^2 -Test

	nicht bestanden	bestanden	Odds	Odds-Ratio	χ^2
Insg.	59,95%	40,05%	0,67		
BWL	61,69%	38,31%	0,62		12,67*
BA	80,00%	20,00%	0,25		
WIINF	47,47%	52,53%	1,11		
Übungsblätter					
ja	44,44%	55,56%	1,25	2,3188	14,32*
nein	64,97%	35,03%	0,54		
Anz. der. Übungsblätter					
1	58,82%	41,18%	0,70		23,97*
2	32,00%	68,00%	2,13		
3	52,63%	47,37%	0,90		
4	50,00%	50,00%	1		
5	18,75%	81,25%	4,33		
Termin					
HT	66,67%	33,33%	0,5	0,5056	11,96*
NT	50,28%	49,72%	0,99		
Geschlecht					
m	64,23%	35,77%	0,56	0,7716	1,47
w	58,08%	41,92%	0,72		
Nationalität					
deutsch	56,91%	43,09%	0,76	2,0914	7,36*
andere	73,42%	26,58%	0,36		

Anm.: * bedeutet signifikant zum 1%-Niveau.

erfolgreichen Klausurteilnahme ansteigt, wenn die Teilnehmer das Angebot der Lernerfolgskontrolle in Form von Übungsblättern angenommen haben. Die Durchfallquote der Teilnehmer, welche keine Übungsblätter bearbeitet haben, liegt bei ca. 65%. In der Gruppe der Studierenden, die Übungsblätter bearbeitet haben, beträgt die Durchfallquote rund 44%. Der Anstieg in der Wahrscheinlichkeit einer erfolgreichen Klausurteilnahme wird durch das *Chancenverhältnis* oder die *Odds-Ratio* ausgedrückt. Sie gibt an, um das Wievielfache die Chance einer erfolgreichen Klausurteilnahme ansteigt,

wenn die Teilnehmer Übungsblätter bearbeitet haben und beträgt hier

$$\text{Odds - Ratio} = \frac{1,2502}{0,5392} = 2,3188.$$

Das bedeutet, dass durch die Teilnahme am Practice Test die Chance einer erfolgreichen Klausurteilnahme um das 2,32-fache ansteigt.

Die Ergebnisse in *Tabelle 3* machen weiterhin deutlich, dass die Chance einer erfolgreichen Teilnahme an der Klausur mit der Anzahl der bearbeiteten Übungsblätter ansteigt. Der Anteil der Personen, welche die Klausur bestanden haben unter denjenigen, welche einen Übungszettel bearbeitet haben, beträgt ca. 41%, was einer Chance von 0,7 entspricht. Wenn diejenigen Teilnehmer betrachtet werden, welche alle fünf Übungsblätter bearbeitet haben, beträgt der Anteil der erfolgreichen Teilnehmer immerhin 81%, so dass für diese Teilnehmer die Chance einer erfolgreichen Klausurteilnahme um das 6,19-fache höher ist, als bei den Teilnehmern, welche lediglich einen Übungszettel bearbeitet haben. Auch hier wird der Zusammenhang zwischen der Bearbeitung der Übungsblätter und einer erfolgreichen Klausurteilnahme durch eine χ^2 -Teststatistik von 23,97 bestätigt, welche ebenfalls signifikant zum 1%-Niveau ist.

4.1.3 Nationalität der Teilnehmer

Aus *Tabelle 3* ist ersichtlich, dass der Klausurerfolg und die Nationalität der Klausurteilnehmer nicht unabhängig sind. Hierzu muss angemerkt werden, dass es sich bei den Schwierigkeiten ausländischer Studierender oftmals um Sprachprobleme handelt. Es ist zu vermuten, dass ausländische Studierende mit den Fragestellungen der Klausur erhebliche Probleme haben, diese inhaltlich zu verstehen. Gerade für diese Studierenden wäre es von Vorteil, wenn sie das Angebot der Lern- und Verständniskontrolle in Form der Übungsblätter annehmen würden. Bedauerlicherweise beträgt der Anteil der Studierenden,

welche von dieser Maßnahme Gebrauch machten auch unter den ausländischen Studierenden gerade einmal knapp 25%, nur hat das drastischere Auswirkungen, wie die Durchfallquote dieser Studenten belegt.

4.1.4 Zeitpunkt der Klausurteilnahme

Die Ergebnisse der Untersuchung deuten darauf hin, dass ein Zusammenhang zwischen der erfolgreichen Klausurteilnahme und dem Zeitpunkt, an dem an der Klausur teilgenommen wurde, zu erkennen ist, welcher statistisch signifikant zum 1%-Niveau ist. Zumindest in diesem Punkt kann ein immer wieder vorgebrachtes Vorurteil der Studierenden entkräftet werden, nämlich dass die Klausuren zum Nachtermin immer deutlich schwerer sind als die zum Haupttermin. Sowohl gemessen an der durchschnittlich erreichten Punktzahl als auch an der Quote der erfolgreichen Teilnahmen ist die Klausur zum Nachtermin deutlich besser ausgefallen. Anzumerken ist hierbei jedoch, dass im Sommersemester 2002 erstmalig eine Klausurbesprechung stattgefunden hat. Bei dieser wurden die Aufgaben der Klausur ausführlich besprochen. Für diejenigen Studierenden, welche zum Vortermine nicht erfolgreich waren, kann man bei erfolgreicher Teilnahme zum Nachtermin den ersten Versuch als *Pre-Test* auffassen, welcher durch die Kenntnis der korrekten Lösung sogar noch an Bedeutung gewinnt.

4.2 Regressionsanalyse

Zusätzlich zu der bisher durchgeführten Analyse sollen die Einflussfaktoren einer erfolgreichen Klausurteilnahme bestimmt werden. Da der vorliegende Datensatz zwei Möglichkeiten der Erfolgsmessung erlaubt — die metrische Variable *Punktzahl* und die binäre Variable *Erfolg* — wird die Untersuchung zunächst mit einer OLS- bzw. einer Logit- und Probit-Schätzung begon-

nen.²⁷ Zur Erklärung des Klausurerfolgs werden die bereits diskutierten Variablen Studium, Übungsblätter, Termin, Geschlecht und Nationalität verwendet. Zusätzlich wird noch untersucht, ob das Alter und die Studiendauer einen Einfluss auf das Bestehen der Klausur hat.²⁸ Darüber hinaus wird der Effekt eines Pre-Tests in Verbindung mit *Knowledge of Correct Result* untersucht, indem die Studierenden betrachtet werden, welche zum Haupttermin die Klausur nicht bestanden und zum Nachtermin nochmals an der Klausur teilgenommen haben. Die zu schätzende Gleichung lautet somit:

$$\begin{aligned} \text{ERFOLG} = & \beta_0 + \beta_1 \text{STUDIUM} + \beta_2 \text{NR-UEB} + \beta_3 \text{SEM-Z} + \beta_4 \text{ALTER-Z} \\ & + \beta_5 \text{TERMIN} + \beta_6 \text{VTERMIN} + \beta_7 \text{GENDER} + \beta_8 \text{NAT} \quad (1) \end{aligned}$$

Die entsprechenden Ergebnisse sind in *Tabelle 4* abzulesen. Da die Interpretation der geschätzten Koeffizienten bei der Analyse kategorialer Daten nicht intuitiv ist, sind in der Tabelle neben den geschätzten Koeffizienten die entsprechenden Odds-Ratios angegeben. Diese geben dann an, um das Wievielfache sich die Chance der erfolgreichen Klausurteilnahme verändert, wenn die Variable x_i um eine Einheit steigt.²⁹

Zunächst ist zu erkennen, dass die bereits durch die deskriptive Analyse erzielten Ergebnisse durch den Regressionsansatz bestätigt werden. Sowohl der Studiengang, oder besser: die Studierenden des jeweiligen Studiengangs als auch die Anzahl der bearbeiteten Übungsblätter sind ein bedeutender Faktor für den Klausurerfolg. Beide Koeffizienten sind sowohl bei der Logit-

²⁷Die Ergebnisse der Logit- und Probit-Schätzung unterscheiden sich weder in den Vorzeichen der geschätzten Koeffizienten noch in deren Signifikanz. Im weiteren Verlauf werden somit lediglich die Ergebnisse der Logit-Schätzung dargestellt.

²⁸Dabei wird für das Merkmal Alter und Studiensemester die entsprechenden zentrierten Variablen ALTER-Z bzw. SEM-Z verwendet.

²⁹Dabei handelt es sich um einen multiplikativen Effekt, da die Chance bei einer Erhöhung von x_i um eine Einheit ein weiteres Mal um das e^{β_i} -fache ansteigt. Vgl. KOHLER/KREUTER (2001), S. 268.

als auch bei der OLS-Spezifikation statistisch signifikant. Zur Variable STUDIUM muss angemerkt werden, dass der Studiengang keinen originären Einfluss auf die Wahrscheinlichkeit einer erfolgreichen Klausurteilnahme hat.

Tabelle 4: Determinanten des Klausurerfolgs

Variable	Logit		OLS
	β	Odds-Ratio	β
KONST	-1,8575* (-4,57)		29,9364* (11,12)
STUDIUM	0,2785** (2,13)	1,3212	2,4960** (2,51)
NR-UEB	0,4282* (4,79)	1,5344	3,6812* (5,34)
SEM-Z	0,1259 (1,22)	1,1341	0,4967 (0,71)
ALTER-Z	-0,0467 (-0,81)	0,9544	-0,2429 (-0,59)
TERMIN	-0,6457** (-2,41)	0,5243	-3,7208*** (-1,88)
VTERMIN	0,8484** (2,57)	2,3359	7,8821* (3,31)
GENDER	0,3307 (1,35)	1,3920	2,3369 (1,41)
NAT	0,8790* (3,00)	2,4085	7,7749* (4,35)
Obs.	426		426
pseudo R^2	0,1101	R^2	0,1711
LR- χ^2	53,54	F	11,41
Prob	0,0000		0,0000

Anm.: Abhängige Variable ist ERFOLG sowie PUNKTE bei der OLS-Schätzung. Robuste Standardfehlerschätzung mittels Verfahren von Huber/White. Die t -Werte sind in Klammern angegeben. */**/** bedeutet signifikant zum 1%/5%/10%-Niveau.

Die Studierenden aller drei erwähnten Studiengänge besuchen die gleichen Veranstaltungen und haben exakt die gleiche Klausur zu lösen. Vielmehr ist zu vermuten, dass bei der Wahl des Studiengangs bestimmte Faktoren, welche das Ergebnis in Statistik positiv beeinflussen, eine mehr oder weniger

große Rolle spielen. Auch mag die Tatsache, dass der Studiengang Bachelor of Arts im Vergleich zu den beiden anderen Studiengängen BWL und Wirtschaftsinformatik keiner Zulassungsbeschränkung unterliegt, ein Indiz für das deutlich schlechtere Abschneiden der Studenten dieses Studiengangs sein.

- Mit jedem zusätzlich bearbeiteten Übungsblatt steigt die Chance des Bestehens der Klausur im Mittel um das 1,5-fache an.
- Es ist weiterhin zu erkennen, dass sowohl der Zeitpunkt der Klausurteilnahme als auch eine wiederholte Teilnahme einen signifikanten Einfluss auf die Chance einer erfolgreichen Teilnahme haben. Diese ist bei den Studenten, welche an der Klausur mindestens zum zweiten Mal teilnehmen um das 2,33-fache größer als bei denjenigen, welche erstmalig an der Klausur teilnehmen. Somit sind bei der wiederholten Teilnahme deutliche Lerneffekte zu erkennen.
- Die Nationalität des Klausurteilnehmers wirkt sich ebenfalls statistisch signifikant (1%-Niveau) auf die Erfolgswahrscheinlichkeit aus. Deutsche Teilnehmer haben eine 2,4-fach höhere Chance, die Klausur zu bestehen, als ausländische Studierende.
- Sowohl das Alter als auch die Studiendauer der Teilnehmer zum Zeitpunkt der Klausurteilnahme üben keinen signifikanten Einfluss auf die Erfolgswahrscheinlichkeit aus.

Zur Analyse der Güte der Anpassung empfiehlt sich der Vergleich der tatsächlichen Beobachtungen und der auf der Basis des Regressionsansatzes prognostizierten Beobachtungen. Diese werden in einer Klassifikationstabelle gegenübergestellt (*Tabelle 5*), auf deren Hauptdiagonalen die Fälle stehen, welche vom Regressionsansatz korrekt spezifiziert worden sind, d.h.

Teilnehmer, welche die Klausur bestanden bzw. nicht bestanden haben, werden durch den Regressionsansatz als solche identifiziert.³⁰ Somit sind durch

Tabelle 5: Diagnostik der logistischen Regression

prognostiziert	wahrer Zustand		Insg.
	1	0	
1	83	43	126
0	88	212	300
Insg.	171	255	426
R_{Count}^2			0,6925
$R_{AdjCount}^2$			0,2339

den hier verwendeten Regressionsansatz $(83 + 212)/426 = 69,25\%$ aller Fälle korrekt klassifiziert worden. Da aber auch ohne Kenntnis des Modells mindestens 40% der Teilnehmer korrekt klassifiziert werden, können³¹ wird zur Analyse das *Adjusted Count r^2* verwendet.³² Es beträgt in diesem Fall 0,2339 und besagt, dass sich unter Kenntnis der unabhängigen Variablen der Vorhersagefehler um 23 Prozent im Vergleich zur Vorhersage allein auf der Basis der Randverteilung der abhängigen Variablen verringert, nämlich von 171 falsch kategorisierten Teilnehmern im Fall der Unkenntnis der unabhängigen Variablen auf $88+43 = 131$ in dem Fall, dass die unabhängigen Variablen bekannt sind.

³⁰Dabei wird üblicherweise allen geschätzten Wahrscheinlichkeiten größer als 0,5 ein geschätzter Beobachtungswert von Eins zugeordnet. Vgl. CRAMER (1991), S. 90 und KOHLER/KREUTER (2001), S. 273.

³¹Insgesamt genau 255, wenn alle Teilnehmer als nicht erfolgreich eingestuft werden, sowie 171, wenn alle Teilnehmer als erfolgreich klassifiziert werden.

³²Vgl. LONG (1997), S. 108.

4.2.1 Selektionsverzerrung

Wie bereits erwähnt, ist bei der hier beschriebenen Vorgehensweise zu vermuten, dass die erzielten Ergebnisse von Selbstselektion überlagert sind. Das hat zur Folge, dass bei Nichtberücksichtigung der Selektionsvariablen die Koeffizienten nicht erwartungstreu geschätzt werden.³³ Diesem Sachverhalt wird mittels eines von HECKMAN (1976) vorgeschlagenen Verfahrens Rechnung getragen, welches das Ausmaß der Selbstselektion bei der Bestimmung der Koeffizienten berücksichtigt.

In einer einfachen Ausgestaltung sieht diese Methode vor, zunächst die Faktoren zu ermitteln, welche zu einer Selektion führen. Im vorliegenden Fall bedeutet das, dass die Bestimmungsgründe für die Teilnahme an einem Practice Test ermittelt werden müssen. Dies ist notwendig, da die Teilnahme am Practice Test eben nicht zufällig ist, sondern die Studenten freiwillig über die Teilnahme entscheiden können.

Im ersten Schritt wird die Wahrscheinlichkeit Z_i einer Teilnahme am Practice Test gemäß folgender Gleichung geschätzt:

$$Z_i = \mathbf{w}_i \boldsymbol{\alpha} + \epsilon_i \quad (2)$$

Mit den geschätzten Koeffizienten lässt sich die inverse Mills-Ratio (IMR) ermitteln:

$$IMR = \lambda_i = \frac{\phi(v_i)}{\Phi(v_i)}, \quad \text{mit } v_i = \frac{\epsilon_i}{\sigma_{\epsilon_i}}. \quad (3)$$

Der zweite Schritt sieht vor, die inverse Mills-Ratio als zusätzlichen Regressor in die Bestimmungsgleichung des Klausurerfolgs aufzunehmen.

$$Y_i = \mathbf{x}_i \boldsymbol{\beta} + \gamma \lambda_i + u_i \quad (4)$$

³³Vgl. dazu GREENE (2003) und JOHNSTON/DI NARDO (1997).

Ist der geschätzte Koeffizient γ statistisch signifikant, deutet dies auf einen Selektionsprozess hin. Auf diese Weise lassen sich konsistente Schätzergebnisse ableiten.³⁴

Alternativ dazu können die beiden Gleichungen

$$Y_i = \mathbf{x}_i\boldsymbol{\beta} + u_i \quad (5)$$

$$Z_i = \mathbf{w}_i\boldsymbol{\alpha} + \epsilon_i \quad (6)$$

simultan geschätzt werden.³⁵ Dabei wird angenommen, dass die Störterme normalverteilt sind mit

$$\begin{pmatrix} u_i \\ \epsilon_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}; \begin{pmatrix} \sigma_u^2 & \rho\sigma_u \\ \rho\sigma_u & 1 \end{pmatrix} \right].^{36}$$

Zusätzlich zu den geschätzten Koeffizienten wird dann noch die Hypothese $H : \rho = 0$ getestet, was bei Unabhängigkeit der beiden Gleichungen der Fall sein müsste. Ist er Korrelationskoeffizient ρ statistisch gegen Null abgesichert, wird auch hier ein Selektionsprozess vermutet.

Ein Problem bei der Anwendung dieses Verfahrens ist u.a. die Identifikation von Faktoren, welche für die Abgabe von Übungsblättern maßgeblich sind, aber *nicht* die Erfolgswahrscheinlichkeit beeinflussen.³⁷

In dieser Untersuchung wurden beide Methoden verwendet. Die Ergebnisse sind in *Tabelle 6* abgebildet. Es ist zu erkennen, dass die Wahrscheinlichkeit, an dem Practice Test teilzunehmen, vom gewählten Studiengang beeinflusst wird.³⁸ Weiterhin ist ersichtlich, dass Selbstselektion vorliegt. Der geschätzte Koeffizient der IMR ist hoch signifikant und die geschätzte Korrelation

³⁴Vgl. JOHNSTON/DINARDO (1997), S. 447 ff.

³⁵Vgl. LONG 1997, S. 215f.

³⁶Die Varianz von ϵ wird aus Gründen der Identifizierbarkeit auf Eins normiert.

³⁷Vgl. DOUGHERTY (2002), S. 450.

³⁸Es sei nochmals darauf hingewiesen, dass die Variable STUDIUM lediglich als Proxy-Variable gelten kann für eine latente Variable, welche möglicherweise die Motivation, das Verantwortungsbewusstsein o.ä. widerspiegelt.

Tabelle 6: Schätzergebnisse des Selektionsmodells

	2SLS	ML
	β	β
UEB		
STUDIUM	-0,3735*	-0,3516*
	(-9,49)	(-6,77)
ALTER-Z	0,0707**	0,0394
	(2,22)	(1,36)
SEM-Z	-0,1154***	-0,0577
	(-1,64)	(-0,96)
PUNKTE		
NR-UEB	3,5236*	3,3279**
	(2,60)	(2,57)
TERMIN	4,2317	3,7459
	(0,79)	(0,65)
VTERMIN	14,2242*	15,6339*
	(2,22)	(3,32)
NAT	12,5953*	12,4105*
	(2,76)	(3,04)
GENDER	1,6700	0,2373
	(0,45)	(0,05)
λ	19,802*	—
	(4,29)	
Obs.	439	439
Wald- χ^2	33,22	51,55
Prob.	0,0000	0,0000
ρ		0,9580
Wald-Test		6,56
Prob.		0,0104

Anm.: Robuste Standardfehlerschätzung mittels Verfahren von Huber/White. Die t -Werte sind in Klammern angegeben. */**/** bedeutet signifikant zum 1%/5%/10%-Niveau.

ρ von 0,9580 ist zum Signifikanzniveau von 1% gegen Null abgesichert. Ein Vergleich mit den Ergebnissen der einfachen Regression in *Tabelle 4* macht deutlich, dass die Nichtberücksichtigung der Selektionsvariablen z.T. deutlich unterschiedliche Ergebnisse hervorbringt. Am deutlichsten wird dies bei Betrachtung der Variablen VTERMIN und NAT, deren Einfluss in der einfachen Regressionsgleichung deutlich unterschätzt wurde. Der Effekt der Anzahl der bearbeiteten Übungsblätter wurde in der ursprünglichen Schätzung jedoch leicht überschätzt.

5 Fazit

In dieser Arbeit ist untersucht worden, ob sich die Teilnahme an einer Lernwirksamkeitskontrolle in Form von freiwillig zu bearbeitenden Übungsblätter positiv auf den Klausurerfolg auswirkt. In Übereinkunft mit den Ergebnissen von GRETES/GREEN (2000) sowie JACOBS (2003) in Bezug auf die Untersuchung des Statistikseminars ist in der vorliegenden Untersuchung eine Bestätigung für den positiven Einfluss von Practice Tests auf die reale Prüfungssituation gefunden worden. Diejenigen Teilnehmer an der Statistik-Klausur im Sommersemester 2002, welche freiwillig Übungszettel bearbeitet haben, erzielten deutlich bessere Ergebnisse in der Klausur — sowohl bei Betrachtung der durchschnittlich erreichten Punktzahl als auch bei Betrachtung der Erfolgsquote. Außerdem war eine deutliche Verbesserung der Ergebnisse mit zunehmender Anzahl der bearbeiteten Übungszetteln zu verzeichnen.

Darüber hinaus konnte festgestellt werden, dass der Klausurerfolg durch weitere Faktoren beeinflusst wird, nämlich durch die Nationalität der Teilnehmer, der Art des Studiengangs und der Häufigkeit, mit der an der Klausur teilgenommen wurde. Auch deutet die im Anschluss an die Regressionsrechnung durchgeführte Diagnose darauf hin, dass der hier verwendete

Erklärungsansatz einen hohen Erklärungsgehalt besitzt, da durch ihn der Prognosefehler immerhin um 23 Prozent gesenkt werden konnte.

Aufgrund der freiwilligen Bearbeitung der Übungszettel konnte das Vorliegen von Selbstselektion vermutet und durch die erweiterte Analyse auch bestätigt werden. Dabei stellte sich heraus, dass der Studiengang, das Alter und die Semesterzahl einen Einfluss auf die Bearbeitungsbereitschaft haben. Die Berücksichtigung der Selbstselektion änderte nichts an den zuvor ermittelten Ergebnissen. Lediglich die Effektstärke wurde zuvor nicht korrekt ermittelt. So wurde der Einfluss der wiederholten Teilnahme und der Nationalität deutlich unterschätzt und der Effekt der Anzahl der bearbeiteten Übungszettel leicht überschätzt.

Um sowohl die Wirksamkeit der Practice Tests als auch die Einflussfaktoren des Klausurerfolgs genauer identifizieren zu können, bedarf es umfassenderer Informationen über die Klausurteilnehmer, z.B. über deren Lernaufwand und ob sie weitere Lernhilfen, wie Nachhilfe oder Repetitorien, in Anspruch genommen haben.

Literaturverzeichnis

ASSENMACHER, W. (2000): *Induktive Statistik*; München, Springer.

BEESON, R.O. (1973): *Immediate Knowledge of Results and Test Performance*; In: The Journal of Educational Research, 66 (5), 224 – 226.

BOL, L./HACKER, D.J. (2001): *A Comparison of the Effects of Practice Tests and Traditional Review on Performance and Calibration*; In: Journal of Experimental Education, 69 (2), 133 – 152.

BROSVIC, G.M./COHEN, B.D. (1988): *The Horizontal–Vertical Illusion and*

- Knowledge of Results*; In: *Perceptual and Motor Skills*, 67, 463 – 469.
- CRAMER, J.S. (1991): *The logit model*; In: New York, Arnold.
- DEMPSEY, J.V./LITCHFIELD, B.C./DRISCOLL, M.P. (1993): *Feedback, Retention, Discrimination Error, and Feedback Study Time*; In: *Journal of Research on Computing in Education*, 25 (3), 303 – 326.
- DIHOFF, R.E./BROSVIC, G.M./EPSTEIN, M.L. (2003): *The Role of Feedback During Academic Testing: The Delay Retention Effect Revisited*; In: *The Psychological Record*, 53 (4), 533 – 549.
- DOUGHERTY, CH. (2002): *Introduction to Econometrics*, 2nd ed.; New York, Oxford University Press.
- EPSTEIN, M.L./EPSTEIN, B.E./BROSVIC, G.M. (2001): *Immediate Feedback During Academic Testing*; In: *Psychological Reports*, 88, 889 – 894.
- EPSTEIN, M.L./LAZARUS, A.D./CALVANO, T.B./MATTHEWS, K. (2002): *Immediate Feedback Assessment Technique Promotes Learning and Corrects Inaccurate First Responses*; In: *The Psychological Record*, 52, 187 – 201.
- GREENE, W.H. (2003): *Econometric Analysis*, 5th ed.; Upper Saddle River, Prentice–Hall.
- GRETES, J.A./GREEN, M. (2000): *Improving Undergraduate Learning with Computer–Assisted Assessment*; In: *Journal of Research on Computing in Education*, 33 (1), 46 – 57.
- HACKER, D.J./BOL, L./HORGAN, D./RAKOW, E. (2000): *Test Prediction and Performance in a Classroom Context*; In: *The Journal of Educational Psychology*, 92, 160 – 170.

HANNA, G.S./LONG, C.A. (1979): *Effect of Answer Until Correct Testing on Reliability*; In: *Perceptual and Motor Skills*, 9, 464 – 466.

HECKMAN, J. (1976): *The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models*; In: *Annals of Economic and Social Measurement*, 5, 475 – 492.

JACOBS, B. (1998): *Aufgaben stellen und Feedback geben*; www.phil.uni-sb.de/jakobs/wwwartikel/feedback/index.htm [23.01.2004].

JACOBS, B. (2000): *Feedback nach Bearbeitung einer Probeklausur — Wird es genutzt?*; www.phil.uni-sb.de/jakobs/wwwartikel/feedback/probfeedback.htm [23.01.2004].

JACOBS, B. (2001): *Die Wirkung von Probeklausuren bzw. Prüfungssimulationen auf das Ergebnis in der Klausur*; www.phil.uni-sb.de/jakobs/wwwartikel/feedback/probeklausur.htm [23.01.2004].

JACOBS, B. (2003): *Die Wirkung einer Probeklausur auf Lernleistung und aktuelle Angst in einer echten Klausur*; <http://www.phil.uni-sb.de/jakobs/wwwartikel/probeklausur/index.htm> [23.01.2004].

JOHNSTON, J./DINARDO, J. (1997): *Econometric Methods*, 4th ed.; New York, McGraw–Hill.

KLADROBA, A. (2002): *Eine Analyse der Statistik I Hauptklausur im Sommersemester 2002*; www.vwl.uni-essen.de/dt/stat/dokumente/klausuranalyse_stat1_ht_ss2002.pdf [09.01.2004].

KOHLER, U./KREUTER, F. (2001): *Datenanalyse mit Stata: Allgemeine Konzepte der Datenanalyse und ihre praktische Anwendung*; München, Oldenbourg.

LEE-SAMMONS, W.H./WOLLEN, K.A. (1989): *Computerized practice tests and effects on in-class exams*; In: Behavior Research Methods, Instruments & Computers, 21 (2), 189 – 194.

LONG, J.S. (1997): *Regression Models for Categorical and Limited Dependent Variables*; London, Sage.

LIPPE, P. VON DER (1999): *Induktive Statistik*; München, Oldenbourg.

PRESSEY, S.L. (1926): *A simple device which gives tests and scores and teaches*; In: School and Society, 23, 373 – 376.

SLY, L. (2000): *Computer Managed Learning Assessment in Higher Education: The Effect of a Practice Test*; PhD thesis, Science and Mathematics Education Centre, Curtin University of Technology, Perth Western Australia.