

Autor: [Bernhard Jacobs](#), Fachrichtung 5.1 Erziehungswissenschaft der Universität des Saarlandes

created: 10.3. 2009

URL des Originals: http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/cash4quiz.html

Geld zur Steigerung der Testmotivation- und Leistung in einer unbenoteten Abschlussklausur ?

Abstract

In der letzten Seminarsitzung wurden Studierende des Lehramts nach Zufall auf 2 Bedingungen aufgeteilt und überraschend mit einer nicht benoteten Abschlussklausur konfrontiert. Um die Testmotivation möglichst hoch zu halten, versprach der Seminarleiter der EG eine leistungsabhängige Geldprämie von mindestens 10 Euro und vermittelte der KG Argumente zum Anstreben möglichst guter Leistungen. Beide Gruppen erzielten sehr ähnliche Werte in allen motivationsrelevanten Variablen (z.B. aktuelle Angst, subjektive Bedeutsamkeit guter Leistungen im Test, investierte Anstrengungsbereitschaft, Erfolgszuversicht) sowie vergleichbar gute Ergebnisse in der unbenoteten Abschlussklausur. Für den gegebenen Zweck erscheinen leistungsabhängige Geldanreize teuer, aber nicht wirksamer als einige pädagogisch vernünftige Argumente bzw. Versprechen, um ein hinreichend akzeptables Leistungsstreben anzuregen.

Schlagworte: extrinsische Motivierung, Geldanreiz, Testmotivation, monetary incentives, academic achievement

Ausgangspunkt der Studie

Vorliegende Studie entwickelte sich aus einer Problemstellung, die im Rahmen einer zuvor initiierten Untersuchung auftauchte. Dort hatten Studierende in der letzten Seminarsitzung für sie völlig überraschend eine unbenotete Abschlussklausur zu bearbeiten, welche den langfristigen Lernerfolg bestimmter Maßnahmen erfassen sollte. Hierbei stellte sich die Frage, wie man Studierende am Ende des Semesters zu einem hinreichenden Engagement für eine Klausurbearbeitung motivieren könne, die keinerlei Auswirkungen auf ihre Seminarnote ausüben konnte [low stake test]. Möglicherweise durch eine leistungsabhängige Geldprämie? Zugleich gab es das Bestreben, eine wirksame Alternative zum Geldanreiz zu finden, die zum einen kostengünstiger und zum anderen pädagogisch angemessener erschien.

Geld zur Steigerung der Testmotivation und Testleistung

Im Rahmen der Erwartungs-x-Werttheorie wird die aktuelle Motivation gesteigert, wenn das Individuum erwartet, durch seine Handlung könne es ein Ziel erreichen, welches eine positive Konsequenz nach sich zieht. (siehe dazu: Kap. [Geld als Lern und Leistungsanreiz](#) aus Jacobs 2007). Geld als generalisierter Verstärker ist eine derartige, ziemlich triviale, positive Konsequenz. Insofern überrascht es auch nicht sonderlich, dass etliche empirische Studien den Nachweis erbringen konnten, durch Geldversprechen nicht nur die aktuelle Motivation, sondern sogar kognitive Testleistungen gesteigert zu haben (siehe zusammenfassend Jacobs 2007). Allerdings fallen die Ergebnisse nicht einheitlich aus. Das übliche empirische Chaos divergierender Befunde erklärt sich teilweise aus der mangelnden theoretischen Fundierung. Dem Common Sense angelehnt, lautet die simple theoretische These: Der Geldanreiz erhöht die Testmotivation und mit steigender Motivation wächst auch die Leistung. Aber es müssen eine Reihe von Bedingungen zusammenkommen, damit ein Geldversprechen überhaupt eine deutliche motivierende Wirkung hinterlässt und weitere Bedingungen erfüllt sein, damit eine gesteigerte Motivation auch zu besserer Leistung führt. Kaum erreichbare oder ziemlich geringe Geldanreize können gar nicht motivieren oder gar demotivieren. Bei unzureichender Vorbereitung während des Seminars nützt erhöhte Testmotivation in einer Abschlussklausur recht wenig. Schwierige Aufgaben erfordern ein gewisses Maß an Intelligenz. Erhöhte Moti-

vation hilft dem wenig Begabten dann nicht weiter. Hoher Geldanreiz kann die Leistung sogar senken, insbesondere bei anspruchsvollen oder sehr komplexen Leistungen (siehe Ariely, Gneezy, Loewenstein & Mazar 2005). Für die Testdiagnostik besonders kritisch sind jedoch Probanden einzuschätzen, die bei der Testung nicht einmal ihre Normalform aktivieren und wegen fehlender Testmotivation letztlich Testergebnisse abliefern, die erkennbar unter ihrem Leistungspotential bleiben (z.B. Gneezy & Rustichini 2000). Dann geht es weniger darum, durch Geldanreize Höchstleistungen zu provozieren, sondern primär darum, ein angemessenes, akzeptables Testbemühen anzuregen. Um unzumutbare Trägheit, vermehrte Zufallsantworten oder gar Verweigerungsreaktionen zu unterbinden, werden deshalb in Erhebungen oder experimentellen Studien mitunter Geldanreize für gute Leistungen angeboten, um die Validität der abhängigen Variablen nicht zu gefährden. (z.B.: Carpenter, Pashler & Cepeda 2008).

Akzeptable Kontrolle zur Geldanreizbedingung

Der Vergleich Geld gegen gar nichts, erscheint ziemlich trivial und auch nicht sonderlich fair, da der potentielle Vorteil des Geldanreizes immer auch von den Anreizen der Kontrollbedingung abhängt. Beim Vergleich "Geld gegen kein Geld" drängt sich die Frage auf, wie die Probanden ansonsten zu kognitiven Leistungen bewegt werden (Jacobs 2007). Nützlicher erscheint mir der Ansatz, Argumente zu finden, um die Studierenden auch ohne bzw. durch eine mildere extrinsische Motivierung von der Sinnhaftigkeit einer engagierten Mitarbeit zu überzeugen. Aus diesem Grunde wurde als Vergleichsgruppe keine No-Treatment-Kontrollgruppe, sondern eine alternative Motivierungsmaßnahme gewählt, die hier aus Vereinfachungsgründen zwar Kontrollgruppe (KG) genannt wird, aber eher als zweite Experimentalgruppe fungiert. Damit verändert sich die Zielrichtung der Untersuchung von der Frage, ob Geldanreiz die Testmotivation oder Leistung erhöht, dazu, ob Geld für den angestrebten Zweck überhaupt notwendig erscheint und wie Geld möglicherweise durch pädagogische Argumente ersetzt werden könnte.

Optimale und tatsächlich realisierte Geldanreizbedingungen

Im Rahmen empirischer Forschung wurden etliche Geldanreizbedingungen ausprobiert, die sich im Wesentlichen wie folgt zusammenfassen lassen (siehe Jacobs 2007):

- fixer Geldbetrag für die Teilnahme am Experiment (=häufige Kontrollgruppe)
- Stücklohn (piece-rate): etwa fixer Geldbetrag pro korrekter Aufgabenlösung.
- das Erreichen eines festgelegten Leistungsstandards, z.B. mindestens x richtige Lösungen oder
- Konkurrenzsysteme: z.B. die Ankündigung, nur die 10 % Besten bekämen attraktive Geldbeträge.

Stücklohn und Leistungsstandards haben sich dabei relativ gut bewährt. Um möglichst alle Studierenden hinreichend durch Geld zu motivieren, wäre meiner Meinung nach ein Anreizsystem sinnvoll, welches zunächst einmal einen unverzichtbaren Minimalleistungsstandard fordert und darüber hinaus das individuelle Leistungsniveau berücksichtigt. Demnach müssten Zielkriterien gesetzt werden, die für jeden Einzelnen ein anspruchsvolles und zugleich erreichbares Leistungsniveau definierten, von dem die Geldprämie abhängig gemacht wird. Zugleich müsste die erwartete Geldbelohnung hinreichend hoch ausfallen, damit die Konsequenz überhaupt als erstrebenswert betrachtet werden kann.

Während man die entsprechenden Bedingungen in experimentellen Umgebungen noch relativ leicht herstellen kann, ergaben sich hier gewisse praktische Beschränkungen. Das individuelle Leistungsvermögen war zwar partiell durch die bisherigen Leistungen im vorangegangenen Quiz bekannt, aber die Zielkriterien schwierig zu bestimmen, da die Aufgabenschwierigkeit des zu bearbeitenden Abschlussquiz unbekannt war. Wegen der zu erwartenden Vergessensrate am Ende des Seminars konnte man nicht ernsthaft das Ziel setzen, ein Studierender müsse seine frühere Leistung verbessern. Es würde wohl ziemlich komisch klingen, den Studenten vorzuschlagen, man könne Geld bekommen, wenn man seine frühere Leistung nur um 5 Prozent verschlechterte.

Die tatsächliche Geldanreizbedingung [EG]

Aus diesem Grund wurde festgelegt, der Studierende müsse seinen Leistungsrangplatz gegenüber dem früheren Quiz verbessern. Damit kommt nun zwar eine soziale Bezugsnorm in die Zielsetzung hinein, weil man sich quasi nur auf Kosten Anderer verbessern kann. Gegenüber reiner sozialer Bezugsnorm haben aber alle Studierenden in etwa die gleichen Chancen. Schließlich musste die ursprüngliche Idee aufgegeben werden, sich lediglich um einen Rangplatz verbessern zu müssen, weil dadurch der verfügbare Geldrahmen gesprengt worden wäre. Somit wurde das Zielkriterium auf 3 Leistungsrangplätze [bei durchschnittlich ca. 30 Studierenden] verschärft. Diese strenge Forderung barg die Gefahr in sich, bei einigen Studierenden die Erfolgserwartung stark zu beeinträchtigen und infolgedessen die Zuversicht zu mindern, sein Ziel überhaupt erreichen zu können. Wer sein Ziel erreichte, sollte - bei ca. 35 bzw. 50 Minuten Arbeitszeit - mindestens 10 Euro Geldprämie erhalten, die sich aber noch deutlich erhöhen könne. Dabei gilt zu bedenken, dass die Studierenden das Quiz in einer verpflichtenden Seminarsitzung auf jeden Fall bearbeiten mussten, die 10 Euro also quasi ohne zusätzlichen Zeitbedarf erwerben konnten.

Die alternative Anreizbedingung [KG]

Die Studierenden wurden darauf hingewiesen, es lohne sich, in diesem Quiz gute Leistungen zu erbringen. Als Gründe wurden angeführt.

1. Sie erhalten Rückmeldung über ihr Leistungsergebnis und können so einschätzen, wie gut sie wesentliche Lehrziele des Seminars noch beherrschen.
2. Sie erhalten die Chance sich zu verbessern. Hierbei wurde das dem Geldanreiz identische Bewertungssystem erklärt, als positive Konsequenz aber lediglich versprochen, der Seminarleiter würde eine Bestätigung zusenden, dass sie sich deutlich verbessert hätten. [Hinweis des Verfassers: Die Verbesserung hat natürlich keine Auswirkungen auf die Seminarnote]
3. Wissenschaftliche Untersuchungen haben nachgewiesen, intensive Quizbearbeitung stabilisiere das Wissen, fördere das Behalten und erleichtere nachfolgendes Lernen.

Die genauen Anweisungen für die beide Anreizbedingungen kann der Leser im [Anhang](#) einsehen.

Versuchsdurchführung

An der Untersuchung nahmen Studierende aus 4 Seminaren des Verfassers teil. Alle TeilnehmerInnen (73% Frauen; Durchschnittsalter: 21,5) sind Studierende des Lehramts und belegten im WS 08/09 ein Proseminar zum Themengebiet Pädagogische Diagnostik. Innerhalb jedes Seminars wurden die Studierenden zunächst nach den Ergebnissen ihrer vorangegangenen

nen, benoteten Quiz parallelisiert und die Paarlinge anschließend nach Zufall auf EG und KG aufgeteilt. Damit liegt ein strenger Randomisierungsplan zugrunde, der zudem hoch vergleichbare Vorquizzwerte garantiert.

Zwei Seminare, im folgenden *Diagnostik Aufgaben* genannt, bearbeiteten eine thematisch verkürzte Abschlussklausur, die sich auf die wichtigsten Inhalte des Seminars bezog, aus insgesamt 22 objektiven Aufgaben unterschiedlichen Aufgabenformates bestand, insgesamt 92 Antworten verlangte und eine Konsistenz von $\alpha=.69$ ($N=51$) sowie eine Korrelation zur benoteten Quiznote von $r=.42$ ($N=50$) erreichte. Die Abschlussklausur der beiden anderen Seminare, im folgenden *Seminare Diagnostik Grundlagen* genannt, deckte fast alle Themen der Veranstaltung ab, umfasste 14 Seiten mit 39 Aufgaben, erforderte insgesamt 181 Antworten und erzielte eine Konsistenzreliabilität von $\alpha=.87$ ($N=58$) sowie eine Korrelation zum Durchschnitt der benoteten Quiz während des Seminars von $r=.72$ ($N=56$). Die Klausurergebnisse wurden in die analoge Computerfassung übertragen und automatisch ausgewertet. Um die Wirkung von Testmotivationsunterschieden sensibler aufzufassen, würden sich allerdings Essaytestaufgaben eher eignen (Sundre & Kitsantas 2004), da die Beantwortung offener Fragen mehr aktives Engagement und eine höhere mentale Beanspruchung als die hier überwiegend verwendeten "Ankreuzaufgaben" erfordern.

Gegenüber den Studierenden wurde immer der Begriff Quiz verwandt, weil eine Klausur in diesem Seminar von der Prüfungsordnung her nicht zulässig war. Da der Umfang und die Breite des verlangten Lehrstoffs jedoch einer Klausur entsprechen, sind die Begriffe Klausur bzw. Abschlussklausur eher angebracht, die im folgenden synonym zu Quiz verwendet werden.

Unmittelbar vor der Klausurbearbeitung erhielten die Studierenden die genauen Anreizbedingungen schriftlich ausgehändigt und beantworteten danach ein paar Fragen zur aktuellen Angst, der intendierten Anstrengungsbereitschaft, der Erfolgszuversicht und der subjektiven Bedeutsamkeit des Quizergebnisses sowie der Zielerreichung. Der Seminarleiter kündigte sodann an, für die Klausur sei genügend Zeit vorgesehen. Wer früher fertig sei, könne die Klausur jedoch abgeben und sich still weiter beschäftigen. Während der Klausurbearbeitung kontrollierte der Seminarleiter die ordnungsgemäße Durchführung und achtete insbesondere darauf, wann einzelne Studierende ihre Klausur beendeten. Die Klausurunterlagen wurden unmittelbar nach individueller Fertigstellung eingezogen und nach dem Abgabetermin geordnet. Die Rangreihe der Quizabgabe sollte ein Maß für die Sorgfalt bzw. das Bemühen erfassen, ein möglichst gutes Ergebnis zu erzielen. In einer Studie zu Unterschieden zwischen benoteten und unbenoteten Quiz erbrachte dieses Rangmaß hypothesenkonform längere Bearbeitungszeiten für benotete Quiz (Jacobs in Vorbereitung).

Für die reine Klausurbearbeitung der Seminare Diagnostik Aufgaben wurden ca. 35 Minuten, für die Bearbeitung der Seminare Diagnostik Grundlagen ca. 50 Minuten Zeit gewährt. Nach der Klausurbearbeitung beantworteten die Studierenden weitere Fragen zur investierten Anstrengung, der Wichtigkeit eines guten Quizergebnisses und der Lernanregung durch das Quiz. Die Geldanreizgruppe (EG) musste zusätzlich ein paar Items zum Verständnis der näheren Bedingungen des Geldanreizes beantworten.

Motivationale Variablen

Unmittelbar nach der Information zur Anreizbedingung wurde die [situative aktuelle Prüfungsangst \(SPA\)](#) und das Fearthermometer von Jacobs (1996-2005) erhoben. Dahinter steckte die Idee, vielleicht etwas höhere Werte bei der Geldanreizgruppe zu entdecken, insbesondere im

Emotionality-Bereich der Angst, der durch Adjektive wie aufgeregt, angespannt, nervös repräsentiert ist. Tabelle 1 zeigt die Items zur Erhebung einiger Variablen unmittelbar vor der Quizbearbeitung. Die gleichen bzw. sehr ähnliche Fragen wurden unmittelbar nach der Quizabgabe gestellt, wobei, wenn nötig, lediglich die Verbform angepasst wurde. So wird aus der zu Beginn erhobenen intendierten Anstrengungsbereitschaft (Ich werde alle meine Kräfte einsetzen...) nach dem Quiz die investierte Anstrengungsbereitschaft (Ich habe alle meine Kräfte eingesetzt...). Die aktuelle Testmotivation wird am ehesten durch die Anstrengungsbereitschaft erfasst. Die Testwerte der Variablen entsprechen dem Mittelwert der in richtige Richtung gepolten Items und liefern so auf deskriptiver Ebene eine gewisse grobe Orientierung, z.B. die Information dass die Anstrengungsbereitschaft insgesamt recht hoch ausfiel (siehe Tabelle 3).

Tabelle 1. Items der wichtigsten motivationalen Variablen

Variable	Items (stimmt ganz genau = 5 stimmt überhaupt nicht =1)
intendierte Anstrengungsbereitschaft	Ich werde alle meine Kräfte einsetzen, um das Quiz aufmerksam zu bearbeiten. Ich werde das Quiz voll konzentriert mit höchster Anstrengung bearbeiten. Ich sehe keine Veranlassung, mich im Quiz ganz besonders anzustrengen. (-)
Bedeutsamkeit guter Leistungen	Es ist für mich sehr wichtig, in diesem Quiz gut abzuschneiden. Das Ergebnis im Quiz hat für mich keine besondere Bedeutung. (-)
Erfolgsszuversicht	Ich bin zuversichtlich, meinen Rangplatz im Quiz genügend verbessern zu können. Ich rechne damit, keinen besseren Rangplatz als beim benoteten Quiz zu erreichen. (-).
Erfolgsschätzung	Wie hoch schätzen Sie die Wahrscheinlichkeit ein, Ihren Rang in diesem Quiz um 3 Plätze zu verbessern (1= mit Sicherheit...0.5= 50:50.....0 = unmöglich)

Trotz der geringen Itemanzahl belegten bereits frühere Untersuchungen sehr ansprechende Konsistenzen für die Bedeutsamkeit guter Leistungen und die Anstrengungsbereitschaft, sowie der weiter unten erwähnten Variablen Akzeptanz der Klausur (Jacobs 2008a,b). Die hier ermittelten Retest reliabilitäten (Zeitintervall: vor Klausur - nach Klausur) fallen ebenfalls sehr erfreulich aus [Bedeutsamkeit guter Leistung: $r = .76$ ($N=105$); Anstrengungsbereitschaft: $r = .67$ ($N=104$)]. Die Erfolgsszuversicht erreicht hingegen hier nur eine Konsistenz von $\alpha = .49$. Die Erhebung der Variablen verfolgte den Zweck, einige theoretische Voraussetzungen für die erwartete motivierende Wirkung des Geldanreizes auf die Testmotivation sowie ihrer Bedingungsfaktoren überprüfen. So kann der Geldanreiz die Testmotivation nur dann theoriekonform erhöhen, wenn wenigstens 3 Bedingungen erfüllt sind:

- Der Studierende versteht die Anreizbedingung in ausreichendem Maße
- Ein sehr gutes Leistungsergebnis bzw. genauer die Zielerreichung, welche letztlich das Geld nach sich zieht, stellt einen bedeutsamen Wert dar.
- Der Studierende hegt eine gewisse Zuversicht, sein Ziel auch tatsächlich erreichen zu können.

Sind diese Bedingungen in hinreichendem Maße erfüllt, so wäre eine höhere Anstrengungsbereitschaft zu erwarten. Lässt sich die höhere Anstrengungsbereitschaft unter der Geldanreizbedingung bestätigen, so wäre geklärt, dass Geld die Testmotivation im gegebenen Fall erhöht hat. Erst dadurch wäre die Voraussetzung geschaffen, eine höhere Klausurleistung theoretisch überhaupt erwarten zu dürfen.

Ergebnisse

Verständnis hinsichtlich der Geldanreizbedingung

Bei einigen Untersuchungen kannten oder verstanden etliche Probanden ihre Geldanreizbedingung nicht richtig, was die Treatmentfidelität solcher Studien in Frage stellt. Deshalb wurden unmittelbar nach der Klausur einige Wissens- und Anwendungsfragen zu den speziellen Anreizbedingungen gestellt, wobei die Befragung sich auf die Geldanreizgruppe beschränkte (siehe [Anhang](#)). Eine Ergebnisanalyse dieser Fragen kommt zu dem Gesamtergebnis, die meisten Studierenden hätten die wichtigsten Anreizbedingungen weitgehend durchschaut [im Durchschnitt 75% korrekte Antworten]. Allerdings lassen sich auch etliche Widersprüche sowie Unwissenheit aufdecken. 95% wissen zwar, dass Sie 3 Rangplätze besser abschneiden müssen, aber über 20% versagen bei konkreten Anwendungsfragen. 18% der Studierenden hatten nicht begriffen, dass auch sehr schwache Studierende eine echte Geldprämienschance hatten und 38% konnten nicht nachvollziehen, dass man sogar mehr als 10 Euro erhalten könne. Eine schriftliche Darbietung einer etwas komplizierten Anreizbedingung allein garantiert eben noch kein perfektes Verständnis derselben.

Klausurbearbeitungszeiten von EG und KG

Wenn Geldversprechen deutlich mehr motivieren sollten als die Argumente der Kontrollbedingung, dann wäre zu vermuten gewesen, die Studierenden unter der Geldanreizbedingung hätten die verfügbare Zeit möglichst voll ausgeschöpft und ihr Quiz so spät wie möglich abgegeben. Denn bei Geldanreizen sind in der Regel längere Bearbeitungszeiten zu erwarten und auch schon nachgewiesen worden. (z.B.: Libby & Lipe. 1992, Borghans, Meijers, ter Weel, B. 2006, Konheim-Kalkstein & van den Broek 2008). Wie die Tabelle 2 aufzeigt, fallen die Unterschiede in den mittleren Rängen der Klausurabgabe für die einzelnen Seminare jedoch inkonsistent aus und lassen über alle Seminare hinweg keine klare Entscheidung für eine der Bedingungen zu. Lediglich Seminar 3 bestätigt die längeren Bearbeitungszeiten unter Geldanreiz signifikant in erwarteter Richtung. Seminar 2 erbrachte demgegenüber signifikant hypothesendiskonforme Resultate.

Tabelle 2: Mittlere Ränge bei der Klausurabgabe für die 4 Seminare

Seminar	Nr.	EG	KG	z-Wert nach U-Test
Diagnostik Grundlagen	1	11,3	15,4	-1,34
Diagnostik Grundlagen	2	9,5	17,5	-2,6*
Diagnostik Aufgaben	3	15,1	9,1	2,15*
Diagnostik Aufgaben	4	11,4	9,1	0,85

Im Mittel der Seminare wurde somit unter Geldanreiz nicht mehr Zeit für die Bearbeitung der Klausur eingesetzt als unter der KG-Bedingung. Quiznoten motivieren offensichtlich mehr als die Erwartung von potenziellen 10 Euro. Denn eine frühere Studie von Jacobs (in Vorbereitung) zum Quizabgabeunterschied zwischen benoteten und unbenoteten Quiz bezog sich auf die ersten beiden Seminare. Wurde dort das Quiz benotet, so fiel der mittlere Rang deutlich höher aus als beim identischen unbenoteten Quiz. Das hier untersuchte, unbenotete Abschlussquiz unter KG-Bedingungen wurde von den Seminaren 1 und 2 aber tendenziell später abgegeben als das Quiz unter Geldanreiz.

Ergebnisse zu den subjektiven Einschätzungen

Sollte der Geldanreiz stärker motivieren als die Kontrollbedingung, so wäre bei der Geldanreizgruppe mehr Angst, eine höhere Bedeutsamkeit guter Leistungen und eine größeren Anstrengungsbereitschaft zu erwarten. Die größere Anstrengungsbereitschaft sollte hierbei im Wesentlichen durch die höhere Bedeutsamkeit der Zielerreichung und nicht durch eine höhere Erfolgserwartung begründet sein. Die Ergebnisse in Tabelle 3 widerlegen die meisten Vermutungen sehr deutlich. Beide Gruppen weisen in nahezu allen subjektiven Variablen sehr hoch vergleichbare Werte auf.

Tabelle 3: Ergebnisse zu subjektiven Einschätzungen unter beiden Anreizbedingungen

	Geld		kein Geld		t	p	
	M	s	M	s			
aktuelle Prüfungsangst SPA	15.6	8.3	15.4	7.6	0.12	.90	ns
Fearthermometer FT	0.9	1.2	1.0	1.3	-0.47	.64	ns
Bedeutsamkeit vor Klausur	3.3	0.8	3.4	0.9	-0.5	.58	ns
Bedeutsamkeit nach Klausur	3.2	0.8	3.4	0.9	-1.5	.13	ns
intendierte Anstrengung	4.2	0.6	4.2	0.7	-0.14	.89	ns
investierte Anstrengung	4.0	0.8	4.2	0.7	-1.26	.21	ns
Erfolgsszuversicht vor Klausur	2.6	0.7	2.6	0.7	0.44	.66	ns
Erfolgswahrscheinlichkeit	0.38	0.2	0.39	0.2	-0.06	.95	ns
Akzeptanz der Klausur	3.4	0.7	3.4	0.7	-0.17	.87	ns

Anmerkung: Die Anzahl der Probanden schwankt bei der Gruppe Geld von 48 bis 50 und bei der KG zwischen 53 und 56. Das hier nicht in allen Einzelheiten erklärte Parallelisierungsverfahren erlaubte wegen ungerader Ausgangsgruppen keine vollständige Paarbildung. Die Probandenanzahl unter der Geldanreizbedingung ist auch zufallsbedingt geringer, da nicht alle Studierenden in der letzten Sitzung erschienen waren. Obwohl der sensitive t-test für abhängige Stichproben möglich und angemessen wäre, wird stets der konservativere t-Test für unabhängige Stichproben verwendet, da dieser alle Studierenden verrechnet. Hauchdünne signifikante Unterschiede sind für diese Arbeit sowieso ohne Belang.

Der Wert des Geldanreizes wurde nur indirekt durch die subjektive Bedeutsamkeit guter Leistungen erfasst, wobei implizit unterstellt wird, gute Leistungen führten zur Zielerreichung. Einige zusätzliche Items sollten aber ganz präzise die Attraktivität der Zielerreichung oder die Enttäuschung einer möglichen Zielverfehlung akzentuieren [z.B. "Ich würde mich freuen, wenn ich meinen Rangplatz hinreichend verbessern könnte, oder "Ich wäre schon etwas enttäuscht, wenn ich meinen Rangplatz nicht genügend gesteigert hätte". Aber auch diese Items differenzierten nicht zwischen den beiden Anreizgruppen. Da die Zielerreichung selbst mehreren Zwecken dienen kann und die Freude oder Enttäuschung multifaktoriell bedingt sein könnte, bleibt die genaue Bedeutung unklar, wie viel die "mindestens 10 Euro" einem Studierenden wirklich wert sind. Jedoch lassen die Befunde den eindeutigen Schluss zu, **dass dieser Geldanreiz gegenüber der KG keinen Zuwachs an Attraktivität für die Zielerreichung bewirkt hat.**

Die Erfolgsszuversicht sowie die subjektive Erfolgswahrscheinlichkeit hängen im Wesentlichen von der Leistungsfähigkeit, dem Leistungsselbstkonzept und der Selbstwirksamkeit des Einzelnen ab. Da die Zufallszuweisung die Leistungsfähigkeit egalisierte und die experimentellen Maßnahmen keinen Einfluss auf Selbstkonzept -und wirksamkeit auszuüben beabsichtigten, wurden keine Unterschiede zwischen den Gruppen erwartet, was sich auch bestätigten

ließ. Die durchschnittliche subjektive Erfolgswahrscheinlichkeit von .385 entspricht im übrigen sehr gut der objektiven durchschnittlichen Zielerreichung von 40%. 62% aller Studierenden schätzen ihre Erfolgswahrscheinlichkeit auf .50, die nach dem Risikowahlmodell für Erfolgsmotivierte als optimale Zielwahl anzusehen ist. Da es hier aber um extrinsische Konsequenzen geht, würden meiner Meinung nach höhere Erfolgswahrscheinlichkeiten mehr motivieren. Auch der Mittelwert der Erfolgszuversicht stimmt pessimistisch, da er auf der 5 stufigen Likert-Skala leicht in die negative Richtung weist.

Von den hier untersuchten Studierenden liegen weitere Daten zu subjektiven Einschätzungen vor, die sich auf einige benotete wie unbenotete Quiz während des Seminars beziehen. (Jacobs in Vorbereitung). Die Motivierungsmaßnahmen bei unbenoteten Quiz im Verlauf des Seminars bestanden darin, Item spezifische, sachorientierte Rückmeldungen zu gewähren, den Prozentsatz der korrekten Lösungen mitzuteilen und die Vorteile einer Quizbearbeitung für die Stabilisierung des Wissens und die Förderung des Lernens zu beschwören. Auf das Anstreben eines bestimmten Rangplatzes wurde aber verzichtet. Die Mittelwerte aus mehreren unbenoteten Quiz während des Seminars bzgl. aktueller Angst (15.3) und der Bedeutsamkeit (3.1) gleichen in sehr hohem Maße den hier gefundenen Ergebnissen. Lediglich die investierte Anstrengung fiel hier mit ca. 4.1 gegenüber 3.7 höher aus als in den früheren unbenoteten Quiz. Benotete Quiz erbrachten hingegen bei all diesen Variablen deutlich höhere Werte als sie hier und bei den unbenoteten Quiz erreicht wurden. Eine gute Seminarnote ist sicher mehr Wert als 10 Euro, die dann im Schnitt nur mit einer Wahrscheinlichkeit von $p=.385$ eingestrichen werden können.

Die Akzeptanz der Klausur -erhoben nach der Klausurbearbeitung - bezieht sich auf Aspekte, wie gut es aus Sicht der Studierenden gelungen war, die Lernfreude anzuregen, das Interesse zu wecken oder die Lernwirksamkeit zu verbessern. Man hätte vermuten können, Geldprämien beeinträchtigten die Akzeptanz, weil die fachlichen Anreize der Klausur offenbar nicht ausreichten, echtes Engagement hervorzurufen und man quasi mit Geld eine Art Kompensation anbiete, sich doch richtig anzustrengen. Diese Erwartung konnte aber nicht bestätigt werden. Lediglich ein ad hoc aufgenommenes Item der Befragung lieferte gewisse Hinweise dafür, Geld könne den intrinsischen Wert des Kompetenzstrebens schwächen. Die Gruppe ohne Geldanreiz stimmte der Behauptung "Eine Leistungssteigerung an sich ist mir Belohnung genug." signifikant mehr zu als die Geldanreizgruppe [3,8 vs. 3.46: $t(104)= 2,036$; p zweiseitig=.044; Effektstärke $d =.40$]. So schön der Befund theoretisch auch ins Bild passen mag, so nährt er doch statistische Bedenken und verlangt nach einer Replikation. Denn bei den vielen angestregten statistischen Vergleichen, die vorsorglich auf eine Alphaadjustierung verzichteten, wäre ein signifikantes Ergebnis durch Zufall nun doch schlichtweg einmal fällig gewesen.

Ergebnisse Leistungsdaten

Nach den bisher bekannten Befunden der vergleichbaren Motivationswerte unter beiden Anreizbedingungen ist nicht mehr mit einem Leistungsunterschied der Gruppen zu rechnen. Da die Klausurmittelwertsergebnisse der Seminare Diagnostik Aufgaben und Diagnostik Grundlagen mit 65,5 und 64,6 Prozent korrekter Lösungen fast identisch ausgefallen waren, wurden die Klausurergebnisse aller Studierenden zusammengefasst und die Auswirkungen der beiden Anreizbedingungen an der gesamten Untersuchungss Stichprobe getestet. Wie man aus der Tabelle 4 entnimmt, fallen die Klausurergebnisse unter beiden Bedingungen vergleichbar aus. Die Geldprämie bewirkte eindeutig keine höhere Klausurleistung.

Tabelle 4: Prozentsatz korrekter Lösungen in der Abschlussklausur unter EG (Geld) und KG-Bedingung (kein Geld)

	M	s	N	
Geld	63.6	14.7	50	$t(107) = -1,06 \quad p = 0.29$
kein Geld	66,2	10.6	59	

37 % der Studierenden unter der Geldanreizbedingung und 43 % unter der KG- Bedingung, erreichten schließlich das geforderte Zielkriterium, welches die Geldprämie [tatsächlich 16.5 €] oder die Mitteilung einer Verbesserung (siehe [Anhang](#)) nach sich zog. Es kann somit auch keine Rede davon sein, dass die in Aussicht gestellte Geldprämie die Zielerreichung mehr als die KG-Bedingung gefördert hätte.

Zusammenfassung und Diskussion

Zwei verschiedene Anreizbedingungen bewirkten eine vergleichbar hohe Testmotivation in einer unbenoteten Abschlussklausur. Beiden Anreizbedingungen lag dasselbe, vorwiegend auf individueller Bezugsnorm basierende Leistungsziel zugrunde. Die Bedingungen unterschieden sich lediglich durch die fokussierten Konsequenzen im Falle der Zielerreichung. Die EG konnte mit Geld rechnen, die KG lediglich mit einer Mitteilung, sich verbessert zu haben. Zudem wurden der KG Leistungsrückmeldungen versprochen und auf die das Lernen fördernde Bearbeitung von Tests hingewiesen. Diese pädagogischen Argumente im Verbund mit einer die Leistungsmotivation fördernden Zielsetzung in der KG führten zu einer Leistungsbeurteilung, die der dem Versprechen einer Geldprämie in der EG entsprach. Folgerichtig ergaben beide Anreizbedingungen auch vergleichbare Klausurleistungen. Die praktische Konsequenz daraus lautet: Für den anstehenden Zweck benötigt man kein Geld.

Mehrere Studien zur Wirksamkeit extrinsischer Anreize auf die Leistung bei anspruchsvollen Tests (z.B. TIMMS, PISA) ergaben in der Regel keine Vorteile gegenüber der Standardinstruktion (z.B. Baumert & Demmrich (2001), O'Neil et al. (2005)). Dies lässt vermuten, die Testanten seien durch die Instruktion, welche den Sinn der nachfolgenden Testung verständlich machte, bereits ausreichend motiviert worden. O'Neil et al. (2005) gelang es zwar, Testanten durch sehr viel Geld [10\$ pro korrekter Aufgabe] mehr zu motivieren, aber dieser signifikante Motivationsschub lief ins Leere und führte zu keiner Leistungsverbesserung, was angesichts der Intelligenzlastigkeit solcher Testverfahren auch nicht besonders verwundert. Das hier eingesetzte Anreizsystem entspricht weitgehend dem Vorgehen zur Festlegung des finanziellen Anreizes bei Baumert & Demmrich (2001) und führte auch zum selben Ergebnis. In der Studie von Baumert & Demmrich (2001) musste der Schüler im Mathematik-Pisa-Test eine Leistung erbringen, die etwas über seiner bisherigen Leistung in Mathematik lag, um im Erfolgsfall dann 10 DM zu kassieren. Die finanzielle Anreizvariante erzielte gegenüber der Standardinstruktion von PISA weder Leistungsvorteile, noch ließen sich Unterschiede bei der intendierten und investierten Anstrengung aufdecken.

In der vorliegenden Studie waren die Bedingungen für die Zielerreichung ziemlich kompliziert und hatten bei einigen Studierenden ungerechtfertigt pessimistische sowie objektiv falsche Schlussfolgerungen nach sich gezogen, etwa die, bei schwachen Vorquizzwerten könne man keine Geldprämie bekommen. Die Komplexität des Anreizsystems hat möglicherweise mit dazu beigetragen, die eingeschätzte Erfolgswahrscheinlichkeit im Mittel niedriger als 50% einzuschätzen, was aber der objektiven Sachlage durchaus gerecht wird. Ich schätze als optimal motivierend ein: **"90% Erfolgswahrscheinlichkeit unter der Bedingung höchst möglicher Anstrengung."** Es wäre vermutlich besser gewesen, das Zielkriterium für die Geldprämie wesentlich zu vereinfachen

und die genaueren Details offen zu lassen, etwa in der Form: "Geld wird ausgezahlt, wenn Sie sich gegenüber früher verbessern."

Die hier verfolgte Untersuchungssituation unterscheidet sich deutlich von sonstigen Erhebungen oder Experimenten, die auf die Freiwilligkeit der Teilnehmer angewiesen sind und Ziele verfolgen, die gänzlich losgelöst von den Lebensinteressen der Beteiligten sind (z.B. ein Experiment zur allgemeinen Psychologie). Die gestellten Fragen bezogen sich auf Inhalte des Seminars, im Seminar 1 sogar auf Inhaltsbereiche, die 2 Stunden später noch in einer Klausur überprüft wurden. Der Seminarleiter legte in der Sitzung fest, dass nun ein Quiz zu bearbeiten sei und machte damit die Klausurbearbeitung zur Pflicht. Diese Situation erschwert es, der Aufforderung nicht ordentlich nachzukommen. Der Student fühlt sich vermutlich verpflichtet, zumindest eine mäßige Leistung anzustreben, zumal die Studierenden ihre Namen angeben mussten. Möglicherweise löst eine derartige schulische Leistungssituation teilweise "reflexhaft" eine Leistungsbereitschaft aus, nun ein möglichst gutes Ergebnis anzustreben. Insofern vermute ich, die Ergebnisse wären nicht wesentlich anders ausgefallen, wenn überhaupt keine besondere Maßnahme zum Einsatz gekommen wäre.

In einer solch starken Testsituation reicht vermutlich eine gewisse Bereitschaft aus, die Aufgaben konstruktiv in Angriff zu nehmen, um eine Leistung in der Nähe des aktuellen Leistungspotenzials zu realisieren. Eine darüber hinausgehende Anstrengung vermag, wenn überhaupt, nur noch marginale Leistungsvorteile zu bringen, da mehr Anstrengung ja nicht ohne weiteres mehr Wissen produzieren kann. Verminderte Anstrengung könnte höchstens das wahre Leistungspotenzial unterschätzen lassen. Die erhöhte Anstrengung in einer Klausur kommt natürlich definitiv zu spät. Eine Erhöhung der Lernmotivation in der Vorbereitungszeit bewirkt eine wesentlich größere Leistungseffizienz als eine gesteigerte Testmotivation, da sich Klausurleistungen durch vermehrte Vorbereitung eher steigern lassen. Auch Geldanreize versprechen theoretisch eher Leistungsvorteile gegenüber einfachen pädagogischen Ratschlägen, wenn der Lerner die Lernzeit frei bestimmen kann. Allerdings sollte man sich keine allzu große Hoffnungen machen, durch bestimmte Geldanreize schulische Leistungen wirksam verbessern zu können (Jacobs 2008b). Selbst beträchtliche Leistungsprämien von etlichen 1000 \$ garantieren keine besseren Studienleistungen.

Literatur

- Ariely, D. Gneezy, U., Loewenstein, G. & Mazar, N. (2005). LARGE STAKES AND BIG MISTAKES Working Paper 05-11
<http://econ.princeton.edu/seminars/BEHAVIORALECO/BEHAVIORAL%20ECO/BE%20Fall%202005/Ariely.pdf>
 [27.6.2006]
- Baumert, J., Demmrich, A. (2001). Test motivation in the assessment of student skills: The effects of incentives on motivation and performance.
 European Journal of Psychology of Education, Vol. XVI, No. 3, 441-462.
- Borghans, L., Meijers, H., ter Weel, B. (2006). The Role of Noncognitive Skills in Explaining Cognitive Test Scores. IZA Discussion Paper Series No. 2429
 URN: urn:nbn:de:bsz:291-psydok-16469
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2008/1646/>
- Carpenter, S. K., Pashler, H., & Cepeda, N. J. (2008). Using tests to enhance 8th grade students' retention of U. S. history facts.
Applied Cognitive Psychology. Published Online: 26 Aug 2008.
- Gneezy, U. & Rustichini, A. (2000). Pay enough or don't pay at all.
The Quarterly Journal of Economics. Vol. 115, Issue 3, 791-810.
- Jacobs, B. (1996-2005). Situative, aktuelle Prüfungsangst (SPA).
<http://www.phil.uni-sb.de/FR/Medienzentrum/verweise/psych/wwwfrage/spajava2.html>
- Jacobs, B. (2007). Geld und Noten als extrinsische Motivatoren zur Verbesserung kognitiver Leistungen.
 URN: <urn:nbn:de:bsz:291-psydok-9644>
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2007/964/>

- Jacobs, B. (2008a) Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback,
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/testread/testread2.html>
 Kurzfassung: <http://psydok.sulb.uni-saarland.de/volltexte/2008/1559/>
- Jacobs, B. (2008b). Geld für bessere schulische Leistungen.
 Internes Diskussionspapier
- Jacobs, B. (in Vorbereitung). Leistungssteigerung durch Notendruck ?
 Die Wirkung der Benotung auf die Lernleistung in einem Seminar.
- Konheim-Kalkstein, Y. & van den Broek, P. (2008). The effect of incentives
 on cognitive processing of text. *Discourse Processes*, 45, 180-194.
- Libby, R. & Lipe, M.G. (1992). Incentives, Effort, and the Cognitive Processes
 Involved in Accounting-Related Judgments. *Journal of Accounting Research* 30, 249-273.
- O'Neil, H. F., Abedi, J., Miyoshi, J. & Mastergeorge, A. 2005). Monetary Incentives for
 Low-Stakes Tests. *EDUCATIONAL ASSESSMENT*, 10(3), 185-208.
- Sundre, D. L. & Kitsantas, A. (2004). An exploration of the psychology of the examinee:
 Can examinee self-regulation and test-taking motivation predict consequential and
 non-consequential test performance? *Contemporary Educational Psychology* 29, 1, 6-26.

Die Forschung wurde von der Fachrichtung 5.1 Erziehungswissenschaft der Universität des Saarlandes finanziell unterstützt.

Anhang siehe:

http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/cash4quiz.html