

**Autor:** [Bernhard Jacobs](#) created: 1.2.2007

Medienzentrum der Philosophischen Fakultäten der Universität des Saarlandes

URL des Originals: [http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische\\_motivierung/index.html](http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/index.html)

# Geld und Noten als extrinsische Motivatoren zur Verbesserung kognitiver Leistungen.

## Abstract

Im Mittelpunkt der Arbeit steht eine Analyse empirischer Untersuchungen zur Frage, ob und wie stark extrinsische Anreize kognitive Testleistungen sowie aktuelles Lernen beeinflussen. Zusammenfassende Analysen zur Wirkung von Geldanreizen aus dem Bereich der Wirtschaftswissenschaften deuten vornehmlich bei relativ einfachen Aufgabenstellungen auf gewisse Leistungssteigerungen durch leistungsabhängige Geldanreize hin. Ebenso gibt es gewisse Evidenzen, auch einige Lernprozesse durch Geldanreize zu beschleunigen oder zu verbessern. Kognitiv anspruchsvolle Testleistungen, wie z.B. Intelligenz oder hoch fähigkeitsabhängige Kompetenzen lassen sich durch Geldanreize hingegen kaum beeinflussen. Schwer einzuschätzen bleibt, ob sich Studienleistungen durch leistungsabhängige Stipendien verbessern. Neben Geld wird der Einfluss der Benotung auf die Testleistung abgeschätzt. Die Testmotivation bei benoteten Tests fällt überzeugend höher aus als bei unbenoteten Tests. Nur wenige Untersuchungen können jedoch eindeutige Aussagen dazu machen, wie stark die Benotung einer Klausur im Vergleich zu einer Nichtbenotung die Leistung in dieser Klausur erhöht, wenngleich alle angestregten Analysen in diesem Bereich bessere Leistungen durch die Androhung einer Benotung nahe legen. An verschiedenen Stellen wird auf Forschungsprobleme und am Rande auf Zielprobleme und potentielle Nebenwirkungen extrinsischer Motivierung hingewiesen.

## Einleitung

Extrinsische Konsequenzen als Anreiz oder Feedback ?

### **Geld als Lern- und Leistungsanreiz**

Zusammenfassungen aus empirischen Analysen

Ausgewählte Untersuchungen zur Wirkung von finanziellen Anreizen auf kognitive Leistungen

Bessere kognitive Testleistungen durch Geldanreize ?

Höhere Intelligenztestleistungen durch Geld ?

Geldanreize für bessere Testleistungen im Rahmen nationaler oder internationaler Studien

Bessere Lernleistungen durch Geldanreize ?

Besseres Einprägen durch Geldanreize

Geldanreize bei mehrfachen Lerndurchgängen

Geldanreize für schulische Leistungen

Schlussbetrachtung Geld als Lernanreiz für kognitive Leistungen.

### **Benotung als extrinsische Motivierung zu besseren Schulleistungen**

Testmotivation und Testleistung bei benoteten und unbenoteten Examen

Empirische Studien zur Wirkung einer angekündigten Benotung auf die Leistung

Anonyme, unbenotete Probeklausur gegen bewertete Abschlussklausur

Versuchsplans zur Überprüfung der Lernwirkung einer benoteten Klausur.

Gesamtresume

Literatur

## Einleitung

Gute Lernergebnisse sind nur dann zu erwarten, wenn der Lerner sich intensiv mit dem anstehenden Lernstoff auseinandersetzt und diesen aktiv verarbeitet. Zwar sollte, wo immer möglich, die Eigenbemühung des Lerners durch didaktisch angemessene Interessensweckung, das Lernen fördernde Lernmaterialien und pädagogische Hilfen unterstützt werden. Die erforderliche Eigenaktivität kann aber selbst durch die besten instruktionalen Maßnahmen nicht von außen aufgezwungen werden. Ein trivialer Ansatz besteht darin, von außen attraktive Anreize zu setzen, die den Lerner dazu bewegen sollten, die weiteren mit dem Lernen verbundenen Probleme möglichst selbstständig in Eigenverantwortung zu lösen. Letztlich geht es darum, den Lerner mit der Situation zu konfrontieren, die in ihm die Überzeugung stärkt, es lohne sich, Leistung zu erbringen. **In diesem Zusammenhang stellt sich hier die Frage, ob und wie stark extrinsische Anreize aktuelles Lernen und unmittelbare Leistung beeinflussen können.** Weitgehend ausgeblendet werden dabei Ausführungen zu pädagogischen Zielsetzungen, solche Anreize strikt zu vermeiden oder verstärkt einzusetzen sowie mögliche Folgewirkungen solcher Maßnahmen. Derartige Forschungsrichtungen legen ihren Forschungsfokus auf das zukünftige Verhalten und Erleben, das aus voraus gegangenen extrinsischen Konsequenzen für erbrachte Leistungen resultiert. Hierzu gehören etwa Fragen, welchen Einfluss Lob oder materielle Verstärker auf nachfolgende Leistungen oder Interessenseinschätzungen bzw. intrinsische Motivation ausüben. (z.B. Cameron et al. (2001))

Zu den wichtigsten extrinsischen Motivatoren zählen Geld und Noten. Beide Anreizqualitäten haben den Vorteil, mehrere, ganz unterschiedliche Bedürfnisse befriedigen zu können und versprechen somit eine ziemlich generelle interindividuelle Anwendbarkeit. Die Analyse soll auch nähere Erkenntnisse dafür liefern, in welchen Bereichen, unter welchen Bedingungen und in welchem Ausmaß diese motivationalen Faktoren zur Ausschöpfung des Leistungspotenzial überhaupt beitragen könnten.

## Extrinsische Konsequenzen als Anreiz oder Feedback ?

Geld als materieller Verstärker kann vordergründig betrachtet als eine motivationale Form des Feedbacks gedeutet werden (siehe [motivationales Feedback](#)). Diese Betrachtung trifft aber nur dann zu, wenn der Kandidat - analog der Taube in der Skinnerbox - nach jeder erfolgreich gelösten Aufgabe sein Geld als Verstärkung einsteckt. Im behavioristischen Sinne müsste dann im Verlauf der Testung die Leistung besser ausfallen als ohne Geld. In der Regel wird ein leistungsabhängiger Geldbetrag aber erst am Ende einer Aufgabenserie ausgehändigt. Der dann in Empfang genommene Geldbetrag kann dabei natürlich nicht rückwirkend die bereits erbrachte Leistung beeinflussen. Insofern trägt Geld als reine Rückmeldung höchstens in Form einer symbolisch vorweggenommenen Verstärkung etwas zum aktuellen Lernergebnis bei. Als direkte Verstärkung hat das ausgezahlte Geld nur für zukünftiges Lernen eine Bedeutung.

## Geld als Lern und Leistungsanreiz

Der generalisierte materielle Verstärker Geld wird hier jedoch in erster Linie als Anreiz betrachtet. Die aktuell zu erbringende Leistung sollte daher eher durch die Erwartung des Geldes beflügelt werden. Geld wirkt somit schon vor oder während des Leistungshandelns, was eine Subsumierung von Geld unter die Bezeichnung Feedback theoretisch durchaus in Frage stellt. Es klingt ziemlich plausibel anzunehmen, das glaubhafte Versprechen von Geld für bessere Leistung führe zu einer Erhöhung von Lern- bzw. Testleistungen genau dann, wenn

- A. der Geldanreiz zu höherem Leistungsbemühen in Form verbesserter Aufmerksamkeit, Informationsverarbeitung, Lernstrategie, Anstrengung oder Ausdauer führt **und zugleich**
- B. erhöhtes Leistungsbemühen bessere Lern- oder Testleistungen bewirkt.

Punkt A erfordert ein geeignetes Verfahren, die zu erbringende Leistung so mit versprochenen Geldbeträgen zu koppeln, dass der Kandidat sich optimal herausgefordert sieht, die notwendige Anstrengung oder Konzentration zu investieren. Punkt B schränkt die in Frage kommenden Lern- und Leistungsziele auf solche kognitiven Anforderungen ein, die durch ein aktuell erhöhtes Leistungsbemühen deutliche Leistungssteigerungen erwarten lassen. Soll eine bestimmte Leistung relativ kurzfristig erbracht werden, so fallen viele Lehrziele weg, die aufwändige, aufeinander aufbauende Lernprozesse voraussetzen. Auch Leistungen, die überwiegend von allgemeinen Fähigkeiten der Testanten abhängen, etwa das Lösen klassischer Intelligenztestaufgaben sind nur eingeschränkt durch verbesserte Motivation deutlich zu steigern. Ähnliches gilt für basale Fähigkeiten wie etwa die Verarbeitungskapazität des Arbeitsgedächtnisses, z.B. erfasst durch die Gedächtnisspanne oder für bestimmte Konzentrationsleistungen. Bei vielen anspruchsvollen Problemstellungen ist eine bestimmte Leistungsbemühung fraglos notwendig, aber bei weitem nicht hinreichend.

Nach der Erwartungs- x Werttheorie würde die Motivation für weitere Anstrengungen am besten angeregt, wenn starkes Bemühen mit sehr hoher Sicherheit zu einem Leistungsergebnis führte, das zwingend einen hinreichend attraktiven Geldbetrag nach sich zöge. Nur bei relativ einfachen Aufgaben, deren Lösungswahrscheinlichkeiten in hohem Maße vom aktuellen Leistungsbemühen abhängig sind, lässt sich durch Geldversprechen die aktuelle Leistung erkennbar erhöhen. Hierbei würde die mangelnde Attraktivität der kognitiven Tätigkeit selbst sowie sonstige Motivationskonstellationen wie etwa: gering ausgeprägtes Pflichtgefühl, schwache soziale Motivation, geringes Eigeninteresse an guten Leistungen im erfassten Aufgabenreich, hohe Motivation für alternative Tätigkeiten usw., den Effekt erst ermöglichen bzw. verstärken. Denn der Motivationseffekt bzw. die potentielle Leistungssteigerung durch einen Geldanreiz muss stets in Relation zu den ansonsten schon wirkenden Motivationsquellen betrachtet werden.

Wenn man nun etwa daran denkt, welche artifiziellen Aufgabenstellungen etwa die allgemeine Psychologie ihren Probanden mitunter zumutet, kann es schon verwundern, warum die Probanden auch ohne leistungsabhängige Anreize überhaupt verwertbare Ergebnisse abliefern. Versuchspersonen, die an psychologischen Experimenten teilnehmen, scheinen offensichtlich bereits in hohem Maße bereit, sich kognitiv anzustrengen. Die Ansicht, geistige Tätigkeit an sich sei aversiv und müsse deshalb mit Geld kompensiert werden, ist sicherlich nicht haltbar. Ideal für den Nachweis der Wirksamkeit von Geldanreizen wären jedoch Aufgaben, die man freiwillig nicht bzw. höchst ungern in Angriff nimmt, die aber bei starker Anstrengung mit sehr hoher Sicherheit bzw. an Sicherheit grenzender subjektiver Erfolgswahrscheinlichkeit eine anständige Belohnung nach sich zögen. Diese Erkenntnis erscheint jedoch ziemlich trivial. Man braucht offensichtlich kein Experiment durchzuführen, um herauszubekommen, dass man schon eine Menge Geld hinblättern muss, um etwa den Durchschnittsbürger überhaupt dazu bewegen zu können, die Lottozahlen des vergangenen Jahres auswendig zu lernen. Vermutlich stand im Mittelpunkt der Forschung die Frage, ob durch bestimmte Geldanreizsysteme die kognitive Leistung stärker ansteigt, als man dies normalerweise ohnehin erwarten konnte. Selbst interessante Aufgaben, die man anfangs vielleicht noch gerne bearbeitet, können auf die Dauer langweilig werden. Geistige Tätigkeiten verlangen mitunter große Disziplin, hohe Ausdauer und selbst bei kognitiv erträglicher Belastung immense Kraftanstrengungen, deren Bewältigung bei der Aussicht auf eine anständige Entlohnung viel-

leicht letzte Kraftreserven aktivieren. Andererseits könnten sehr attraktive Geldversprechen bei hoch anspruchsvollen Aufgaben die notwendige sachorientierte Aufmerksamkeit und Konzentrationsfähigkeit einschränken, die verfügbare Kapazität des Arbeitsgedächtnis durch Aufgaben irrelevante Gedanken limitieren und so eine Leistungsver schlechterung bewirken.

## Zusammenfassungen aus empirischen Analysen

Im Bereich der Wirtschaftswissenschaften wurden etliche Studien durchgeführt, deren Aufbau im Prinzip wie folgt beschrieben werden kann.

|     |   |  |
|-----|---|--|
| EG: | Instruktion hinsichtlich leistungsabhängiger Bezahlung:                                   | Lernen bzw. Aufgaben bearbeiten (= unmittelbarer Nachtest) |
| KG: | Aufforderung, sein Bestes zu geben bzw. Leistungsunabhängige Bezahlung für die Teilnahme. | Lernen bzw. Aufgaben bearbeiten (= unmittelbarer Nachtest) |

Manchmal geht der Aufgabenbearbeitung eine klassische Lern bzw. Lernaneignungsphase voraus. Dies trifft etwa für das Lernen aus Texten sowie sonstige Memorierungsanforderungen zu. Häufig fallen Lernen und Aufgabenbearbeitung aber auch zusammen, etwa wenn direkt Aufgaben vorgelegt werden und durch die Bezahlung die Testleistung gesteigert werden soll. Hierbei ist zwischen Aufgabenstellungen zu unterscheiden, die eher Fähigkeiten erfassen (z.B. Intelligenz oder bestimmte geistige Kompetenzen) und Lern bzw. Übungsaufgaben, die mit zunehmender Erfahrung noch erhebliche Lernsteigerungen zulassen. Außerdem spielt die zugestandene Lernzeit eine wichtige Rolle, da sich etwa Ausdauer effekte besser bei frei wählbarer Lernzeit auswirken können.

Als entsprechende Anreizverfahren werden häufig eingesetzt:

- fixer Geldbetrag für die Teilnahme am Experiment (=häufige Kontrollgruppe)
- Stücklohn (piece-rate): etwa fixer Geldbetrag pro korrekter Aufgabenlösung.
- variabler Stücklohn: z.B. im Durchschnitt folgt nach jeder 3. korrekten Lösung ein bestimmter Geldbetrag.
- das Erreichen eines festgelegten Leistungsstandards: (goal-contingent payment), eine Art Leistungskriterium, z.B. mindestens x richtige Lösungen oder
- Konkurrenzsysteme: z.B. die Ankündigung, nur die 5 % Besten bekämen attraktive Geldbeträge.

Gelegentlich werden die Verfahren auch miteinander kombiniert. So beginnt etwa der Stücklohn erst dann, wenn ein Mindestkriterium erreicht wird. Ein Sockelbeitrag kann ab einer bestimmten Leistungshöhe mit zusätzlichem Stücklohn aufgebessert werden oder mehrere Verfahren kommen gleichzeitig zum Einsatz. Manchmal werden auch Fehler mit Geldverlusten geahndet, die aber in der Regel geringer als die Geldeinnahmen ausfallen und aus der bisher erworbenen Geldmenge abgezogen werden, so dass echte Verluste praktisch nicht vorkommen können. Die Probanden der Kontrollgruppe erhalten meistens einen leistungsunabhängigen Geldbeitrag für die Teilnahme am Experiment. Dieser Teilnahmebeitrag kann der Experimentalgruppe ebenfalls zufallen, so dass die EG im Durchschnitt mehr Geld erhält. In anderen Experimenten erhält die KG annähernd gleich viel Geld wie die EG, braucht dieses aber nicht leistungsabhängig zu erwerben (= Kostenneutrale experimentelle Bedingungen).

Die empirischen Studien zum Effekt des Geldanreizes auf kognitive Leistungen hinterließen erwartungsgemäß keine konsistenten Ergebnisse, was deshalb nicht verwundert, weil in die-

sen Experimenten eben selten theoretisch ideale Bedingungen vorherrschten. Geldanreize können die Leistungen verbessern, aber auch verschlechtern und schließlich keinen nennenswerten Effekt auslösen. **Camerer & Hogarth (1999)** analysierten die Ergebnisse aus 74 experimentellen Studien. Sie fanden vornehmlich bei eher einfachen Aufgaben häufiger Leistungsvorteile durch Geldanreize. Zu den einfachen Aufgaben gehören relativ triviale Wahrnehmungs- und Entscheidungsaufgaben (etwa Karteikarten ordnen), das Erlernen bzw. Behalten von Faktenwissen sowie relativ simple Bürötätigkeiten (etwa Daten aus Fragebögen in maschinenlesbare Form bringen, Texte korrigieren). Bei anspruchsvollen Problemlöseaufgaben bewirkten Geldanreize häufig keine Effekte bzw. sehr gemischte Ergebnisse, Leistungsvor- wie nachteile. Die Autoren führen eine Untersuchung an, die einsichtig aufzeigt, dass eine Leistungsverbesserung durch Geld nur bei der Probandengruppe gelang, welche über besondere Fähigkeiten verfügte, die Aufgaben auch lösen zu können.(=Interaktion zwischen Fähigkeit und Geldanreizwirkung)

Eine weitere umfangreiche Analyse von 85 Studien mit 131 Vergleichsmöglichkeiten stammt von **Bonner et al. (2000)**. Die prägnante Zusammenfassung der Ergebnisse erweist sich jedoch auch hier als schwierig, weil keine klassische Metaanalyse zugrunde liegt, die eine durchschnittliche Effektstärke vermitteln würde. Deren Sinn könnte man andererseits wegen der Heterogenität der Aufgaben und Anreizsysteme aber auch anzweifeln. Die Autoren fassen die Befunde unter einem worst-, sowie best-case-scenario zusammen und kommen zu der in Tabelle 1 dargestellten Einschätzung:

**Tabelle 1: Prozentsatz der durchgeführten Vergleiche (N=131)**

Die Wirkung des Geldanreizes auf die Leistung nach Bonner et al. (2000) ist...

|            | negativ | nicht<br>vorhanden | positiv |
|------------|---------|--------------------|---------|
| scenario   |         |                    |         |
| worst case | 10      | 60                 | 30      |
| best case  | 3       | 42                 | 55      |

Viele Vergleichstudien erbrachten keine signifikanten Unterschiede. Jedoch überwiegen in beiden Szenarien die Studien, welche eher positive als negative Effekte des leistungsabhängigen Geldanreizes auf die aktuelle Leistung vorweisen können. Am ehesten wurden günstigste Effekte finanzieller Anreize bei eher einfachen Anforderungen wie Büro-tätigkeiten, Vigilanz/Detection- und Memorierungsaufgaben gefunden, während Problemlöseaufgaben (z.B. mathematische Probleme, Schachprobleme) auch hier ein sehr gemischtes Bild vorweisen und nur im best case scenario überhaupt leichte Vorteile erkennen lassen.

Ein Vergleich der Anreizsysteme ergab mehr positive Befunde für Stücklohn und Quoten (Erreichen eines bestimmten Kriteriums) gegenüber einem Konkurrenzverfahren, welches nur für die Besten Geldbeträge vorsah. Das ist verständlich, weil die subjektive Wahrscheinlichkeit für Probanden mit mittleren oder geringer Fähigkeitseinschätzung sehr schwach ausfallen muss, die Geldbeträge überhaupt erreichen zu können. Motivationspsychologisch wesentlich effektiver erscheinen mir Zahlungen auf der Basis individueller Bezugsnormen (= Verbesserung gegenüber früherer Leistung bzw. das Erreichen oder leichtes Übertreffen eines individuellen Erwartungswertes). Bei verbalen Rückmeldungen, "etwa: du hast deine Leistung verbessert oder verschlechtert" erwies sich diese Form motivationalen Feedbacks häufig als sehr erfolgreich ( siehe Jacobs (2005) [Motivationales Feedback](#) ). Eine leistungsunabhängige Bezahlung (ausschließlich für die Teilnahme) führte nur in sehr seltenen Fällen zu besseren Leistungsergebnissen als gar keine Bezahlung. Um die Motivierung durch ein Geldanreizsystem

in Erfahrung zu bringen und so die Befunde besser interpretieren zu können, wäre es förderlich, stets auch subjektive Daten zur Anstrengungsbereitschaft zu erheben.

## Einige ausgewählte Untersuchungen zur Wirkung von finanziellen Anreizen auf kognitive Leistungen

Um einen besseren Eindruck zum Vorgehen und den Ergebnissen von Studien zum Geldanreiz zu vermitteln, werden im Folgenden einige Untersuchungen etwas genauer dargestellt. Zunächst werden Studien zur Wirkung des Geldanreizes auf die Ergebnisse in kognitiven Leistungstests dargestellt. Später werden Ansätze geschildert, Geldanreize bereits zu Beginn einer Lerneinheit einzuführen, um die Lernmotivation zu steigern und so den Lernprozess zu verbessern.

### Bessere kognitive Testleistungen durch Geldanreize ?

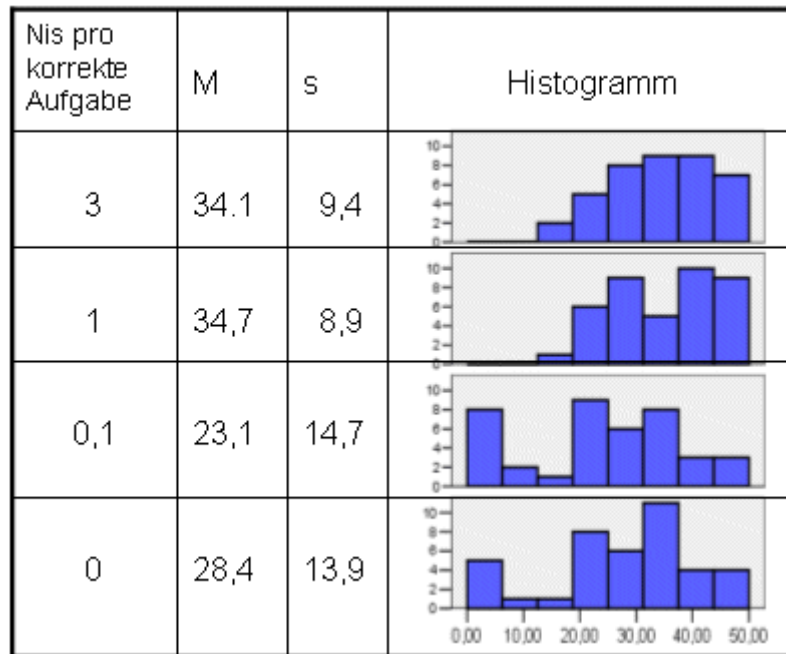
Werden extrinsische Lernanreize erst unmittelbar vor einer Testung angekündigt, so können diese lediglich die Testmotivation erhöhen. Bei der Erhebung kognitiver Leistungstests geht man normalerweise davon aus, alle Testkandidaten bemühten sich in vergleichbarer und hinreichender Weise, ihr Leistungspotential auch auszuschöpfen, da die Testwerte nur dann Fähigkeiten valide messen. Mitgeteilte Normen repräsentieren Unterschätzungen, wenn die Probanden der Normstichprobe den Test nur halbherzig beantworten würden. Da die Bearbeitungszeit in der Regel festliegt, kann eine stärkere Testmotivation nur zu einer erhöhten Konzentration oder größeren Anstrengung während der Testbearbeitung führen. Können Geldanreize die Intensität oder Qualität der Aufgabenbearbeitung über das normale Maß hinaus verbessern und so höhere Testleistungen bewirken?

### Höhere Intelligenztestleistungen durch Geld für jede korrekte Aufgabe in der Studie von Gneezy & Rustichini (2000) ?

Gneezy & Rustichini (2000) wollten herausfinden, ob finanzielle Anreize Intelligenztestleistungen verbessern und welcher Zusammenhang zwischen der Höhe der Bezahlung und den Testergebnissen besteht. An dem Experiment nahmen 160 undergraduate students teil. Alle Probanden kassierten bereits für die Teilnahme 60 NIS. Die Studenten wurden nach Zufall auf 4 Gruppen aufgeteilt und bearbeiteten einen Test unter 4 verschiedenen monetären Anreizbedingungen. Für jede korrekte Lösung wurden entweder 0, 0,1, 1 oder 3 NIS pro korrekt gelöster Aufgabe versprochen. (NIS = Neue israelischer Schekel: Zum Zeitpunkt des Experiments galt: 1\$ = 3.5 NIS.). Die experimentellen Sonderzahlungen beziehen sich somit auf potentielle zusätzliche Einnahmen, die nach der Testbearbeitung fällig waren: "The money will be paid to you, privately and in cash, at the end of the experiment." Die klassische Kontrollgruppe wird hier als 0 NIS-Gruppe bezeichnet. Ihr wurde natürlich nicht gesagt, sie bekäme gar kein zusätzliches Geld, sondern stattdessen: "Beantworten Sie so viele Aufgaben wie möglich!"

Die 50 Aufgaben des Tests sind kognitiv ziemlich anspruchsvoll, stammen aus einem IQ-Test und messen eher reasoning und computation als general knowledge. Der Test wurde den Probanden angekündigt als psychometrisches Verfahren und eine besondere Art von IQ-Test, der auch als Eingangstest der Universität genutzt werde. Die Autoren behaupten zwar: "The problems in the quiz were chosen to make the probability of a correct answer depend mostly on effort.", begründen ihre Itemauswahl bis auf den Ausschluss von general knowledge items aber nicht weiter. Die Behauptung der Autoren verwundert mich etwas, da die Lösung von Intelligenzaufgaben normalerweise hauptsächlich von der Intelligenz der Probanden abhängen sollte. Die Bearbeitungszeit war auf 45 Minuten festgelegt.

**Abbildung 1.: Anzahl korrekter Lösungen für alle Bedingungen (N pro Gruppe jeweils 40)**



Wie man aus den statistischen Kennwerten der Abbildung 1 abschätzen kann, entspricht der Leistungsvorteil der halbwegs anständig bezahlten Probanden (1 bzw. 3 NIS pro korrekter Lösung) gegenüber der Gruppe ohne monetären Anreiz einer Effektstärke von ca.  $d = 0.5$ . Das Experiment bestätigt die leistungssteigernde Wirkung positiver monetärer Anreize (schwach signifikant:  $p < 0.1$ ) auch für komplexere kognitive Aufgaben, was am ehesten mit der katastrophalen Testmotivation einiger nicht bzw. schlecht bezahlter Probanden erklärbar ist. Die Gruppe (zutreffender: Einige aus der Gruppe) ohne Geldanreize waren vermutlich gar nicht motiviert, ihr Leistungspotential überhaupt zu zeigen, geschweige denn voll auszuschöpfen. Vor dem Test wurde zwar an den guten Willen appelliert, das Testergebnis diene aber lediglich wissenschaftlichen Zwecken. Mögliche nicht monetäre Anreize (etwa: das Versprechen, über das erzielte Ergebnis des Test später zu informieren) fehlten und die Teilnahmevergütung von 60 NIS war nicht an bestimmte Mindestleistungsstandards gebunden.

Die Autoren teilen im Artikel die Originaldaten mit, welche als Grundlage für die in der Abbildung 1 dargestellten Histogramme dienten. Aus den Daten geht hervor, dass 20 % der mit 0,1 NIS bezahlten Gruppe und 10 % der nicht bezahlten Gruppe Testergebnisse mit 0 korrekten Lösungen abliefern. Derartige Testresultate sind **offensichtlich das Ergebnis reiner Testverweigerung**. Sie wurden aber bei der Berechnung der oben dargestellten Mittelwerte einbezogen. Entfernt man die Testverweigerer aus der Gruppe der nicht bezahlten Probanden, wie man dies in einer Normenstichprobe vernünftigerweise tun würde, so erhöht sich deren Mittelwert auf 31,6 und dann lässt sich kein signifikanter Vorteil gegenüber den mit 1 oder 3 NIS bezahlten Probanden nachweisen. Es erscheint mir aber nicht so einfach, diese Probanden aus statistischen Gründen einfach aus dem Datensatz zu entfernen, weil für diese eben "ideale theoretische Bedingungen einer Geldmotivierung" vorliegen. Zusätzlich zu den Testverweigerern findet man in der 0- bzw. 0,1- NIS-Gruppe noch einige sehr schwache Testergebnisse, die auf äußerst lasches Testverhalten schließen lassen. Die Probanden mit den sehr geringen Testwerten bei der 0 und 0,1 NIS Gruppe erklären auch die höhere Streuung bei diesen Gruppen. In der Incentive-Forschung wurde häufiger für die Gruppe mit finanziellem Anreizen die geringere Teststreuung festgestellt.

Zugleich deuten sich gewisse Grenzen monetäre Beeinflussung an. Schon 3 NIS erbrachten nicht mehr Testleistung als 1 NIS. Es sieht keineswegs so aus, dass eine Erhöhung auf 100 NIS pro korrekt gelöster Aufgabe nun noch mal einen Leistungszuwachs von  $d = .50$  bewirken könnte. Das Ergebnis entspricht einer Erfahrung von Camerer & Hogarth (1999): "Raising incentives from some modest level to a higher level is more likely to have no effect." Entscheidend ist offenbar ein finanzieller Mindeststücklohn, der nicht unterschrit-



ten werden darf, damit Geld überhaupt motivieren kann. "Pay Enough or Don't Pay At All." lautet ja auch der Titel von Gneezy & Rustichini (2000). Denn die Gruppe, die lediglich 0,1 NIS pro korrekter Aufgabe versprochen bekam, schnitt numerisch, aber nicht signifikant (bzw. nur  $p < 0.1$ ) schwächer ab als die Gruppe ohne jeglichen Geldanreiz, was auf die Gefahr verweist, mit symbolischen Geldbeträgen im Sinne von Peanuts eher Inaktivität oder Reaktanz als Leistungswillen zu beflügeln.

Insgesamt komme ich zu dem Ergebnis, es sei hier kein überzeugender Nachweis gelungen, durch finanzielle Anreize Intelligenztestleistungen von halbwegs motivierten Probanden nennenswert zu erhöhen. Durch Auswahl einer entsprechenden Probandenklientel ließe sich der experimentelle Effekt der Wirkung des Geldes auf die Intelligenztestleistung beliebig in die Höhe schaukeln. Eine valide Erhebung der Intelligenz setzt ein Mindestmaß an Testmotivation bei den Testkandidaten voraus. Das Versprechen attraktiver Geldbeträge für erfolgreiche Aufgabenlösungen kann ein Weg sein, diese erforderliche Testmotivation bei manchen Probanden abzurufen, die anderweitig nicht motivierbar sind. Kieffer & Goh (1981) konnten z.B. Intelligenztestleistungen von Unterschichtkindern deutlich steigern, in dem sie ihnen die von ihnen präferierten Rewards für gute Leistungen versprochen. Bei Mittelschichtkindern war diese Art der Motivierung aber nicht mehr notwendig, d.h. bei diesen führten Rewards zu keiner weiteren IQ-Verbesserung. Eine Studie von Manthei (2004), die teilweise auch als Replikation des Experimentes von Gneezy & Rustichini (2000) verstanden werden kann, aber zu anderen Ergebnissen führte, zeigt meiner Interpretation der Tabelle 2 auf Seite 23 zufolge ebenfalls auf, dass durch leistungsabhängige Geldanreize für korrekte Lösungen keine konsistenten, praktisch bedeutsamen Leistungsvorteile in einem Logiktest erzielt werden konnten, da die 4 möglichen Vergleiche der Geldanreizgruppen gegenüber der Kontrollbedingung Effektstärken von  $d = -.28$  bis  $d = .46$  ergaben. Desgleichen fanden sich keine Geldanreiz bezogenen Vorteile, sondern numerisch eher Nachteile hinsichtlich einer Aufgabenstellung, die an einen Konzentrationstest erinnert (siehe [Beispiel](#)).

## Geldanreize für bessere Testleistungen im Rahmen nationaler oder internationaler Studien

In breit angelegten wissenschaftlichen Untersuchungen zur Erfassung von Leistungsdaten auf nationaler oder internationaler Ebene (NEAP, TIMSS, PISA, IGLU) werden häufig Testverfahren angewandt, die keinerlei Konsequenzen für die Schüler nach sich ziehen (=low-stakes tests). Hierbei kommt immer wieder der Einwand, die Ergebnisse seien unbrauchbar, weil die Betroffenen keinen Sinn darin erblickten, ihr volles Leistungsvermögen unter Beweis zu stellen. Um diesen Einwand zu prüfen, hatten bereits O'Neil, Sugrue & Baker (1995/1996) eine Untersuchung im Rahmen der National assessment of educational process (NAEP)- Testung durchgeführt. 8- und 12 grader students wurden nach Zufall auf mehrere Bedingungen aufgeteilt. Unter der Bedingung "Geldanreiz" versprach man den Schülern unmittelbar vor der Testung 1\$ pro korrekt gelöster Aufgabe. Im Vergleich zur Bedingung "Standardinstruktion NAEP" erzielten die Schüler mit Geldanreiz insgesamt sehr vergleichbare Ergebnisse. Lediglich bei den 8 Graders ließen sich für die Geldanreizgruppe geringe Vorteile bei den leichtesten Items statistisch sichern. Dieser Vorteil verfehlte aber unter Hinzuziehung aller Items im Gesamttest die Signifikanz und schrumpfte auf eine ziemlich unbedeutende Effektstärke von  $d = 0,17$  zusammen.

Baumert & Demmrich (2001) untersuchten die Auswirkungen verschiedener motivationaler Anreize auf die Leistung einer Auswahl von 20 Pisa-Mathematikaufgaben sowie einige motivationale Variablen. Schüler aus mehreren Hauptschulen und Gymnasien wurden nach Zufall



auf 4 Bedingungen aufgeteilt: Die Schüler unter der Bedingung Geldanreiz erhielten folgende Instruktion:

"You will be given money if you do your best and perform well. We know how many questions a student with a particular grade in mathematics solves on average. Students who solve more items than would be expected on the basis of their math grade will be given 10 DM after the test. We therefore ask you to do your best."

Das Anreizsystem basiert auf einer individuellen Bezugsnorm und gibt objektiv betrachtet, jedem Schüler eine relativ vergleichbare, reelle Chance von 50%, 10 DM zu verdienen. Vielleicht wäre es noch motivierender gewesen, die Anforderungen etwas zu entschärfen: etwa "mindestens ein Testergebnis zu erreichen, was deinen bisherigen Mathematikleistungen entspricht " Dennoch glaube ich, die meisten Schüler hätten für sich hohe Chancen erblickt, das Geld auch zu bekommen. Die Autoren teilen die in Tabelle 2 dargestellten Mittelwerte für die Bedingungen mit:

**Tabelle 2. Mittelwerte für verschiedene Anreizbedingungen in der Studie von Baumert & Demmrich (2001)**

|                              | Haupt-<br>schule | Gymna-<br>sium |
|------------------------------|------------------|----------------|
| <b>standard instruktion:</b> | <b>6,8</b>       | <b>13,1</b>    |
| <b>Geldanreiz:</b>           | <b>6,3</b>       | <b>12,8</b>    |
| Ergebnis wird benotet:       | 5,6              | 12,3           |
| Mitteilung Testergebnis:     | 5,6              | 13,3           |

Es konnten in beiden Schulformen keinerlei signifikante Leistungsunterschiede zwischen allen 4 Bedingungen festgestellt werden. Auch die subjektiven Einschätzungen fallen für alle Bedingungen sehr ähnlich aus. Insbesondere die vor der unmittelbaren Testung erhobene intendierte Anstrengung und die nach dem Test erhobene investierte Anstrengung erbrachten für alle Bedingungen sehr ähnliche Werte. Durch Geldanreiz konnte somit weder subjektiv das Leistungsbemühen gestärkt, noch die Testleistung gesteigert werden. Die in der Standardinstruktion vermittelte Begründung, die Untersuchung würde Politikern und Schulen helfen, die Leistungen von Schülern einzuschätzen und sei von größter Wichtigkeit für Schüler in der ganzen Welt, hat völlig ausgereicht.

O'Neil et al. (2001,2004,2005) überprüften die Wirkung des Geldanreizes auf Mathematikaufgaben der TIMSS (Third International Mathematics and Science Study) in mehreren Untersuchungen. 12th-graders aus insgesamt 23 kalifornischen Schulen wurden zufällig auf 2 Bedingungen aufgeteilt: entweder die Standardinstruktion von TIMSS oder die Bedingung finanzieller Anreiz. Als Geldanreiz wurden **10 \$ pro korrekt gelöster Aufgabe** versprochen. Die Glaubwürdigkeit dieser Zahlung wurde an 2 sehr einfachen Beispielen besonders herausgestellt. In der Hauptstudie kassierten die Schüler unter der Bedingung Geldanreiz im Durchschnitt insgesamt 100\$. Besonders begabte Mathematikschüler [Advanced Placement (AP) students in mathematics] einer gesonderten Studie [AP-Studie] verdienten durch die Testung im Durchschnitt 200 \$. Diese realisierten Einnahmen verdeutlichen ohne Zweifel, dass in diesen Untersuchungen sehr attraktive Geldsummen angeboten wurden, für die man doch schon besonderen Einsatz zeigen könnte. Die Schüler unter Geldanreiz und die Schüler unter der TIMSS-Standardinstruktion erzielten jedoch stets sehr vergleichbare Leistungen. Das Versprechen hoher Geldsummen führte folglich nicht zu signifikant besseren Mathematikleistungen. Selbst bei den leichten Items schnitten Schüler mit Geldanreiz nicht besser ab. Allerdings bekundeten die **bezahlten Schüler eine signifikant höhere Anstrengung** [Unterschied zur Kontrollgruppe ca.  $d = .5$ ]. Diese schlug sich jedoch nicht in höheren Testwerten nieder, da

subjektive Anstrengung und Testergebnis mit ca.  $r = .10$  nicht signifikant miteinander korrelierten.

Die im Ergebnis sehr konsistenten Studien zu finanziellen Anreizen bei Aufgaben aus internationalen Studien legen folgende Einschätzung nahe. Entscheidend für das Testergebnis dürfte die Mathematikfähigkeit bzw. die Intelligenz (Rindermann, 2006) sein. Um diese auch zu zeigen, bedarf es einer gewissen Bereitschaft, die Aufgaben in Angriff zu nehmen. Durch die Standardinstruktion gelingt es im Normalfall, die Schüler von der Nützlichkeit der Testung zu überzeugen und eine ausreichende Motivation zur Aufgabenbearbeitung zu erzeugen. Selbst wenn es durch das Versprechen hoher Geldbeträge noch gelingen sollte, die aktuelle Testmotivation zu toppen, so führt dieses erhöhte Bemühen zu keiner weiteren Leistungsverbesserung. In den durchgeführten Studien wurden keine repräsentativen Stichproben von Schülern aus den entsprechenden Ländern herangezogen. Insofern lassen sich die Ergebnisse natürlich nicht bedenkenlos, etwa auf ganz besonders schwierige Problemschulen in sozialen Brennpunktgebieten, generalisieren. Diese selbstverständliche Einschränkung der externen Validität gilt aber fast ausnahmslos für jede empirische Untersuchung.

### Bessere Lernleistungen durch Geldanreize ?

Während der Einfluss erhöhter Motivation auf die Erfassung stabiler Fähigkeiten wenig Spielraum für Veränderungen verspricht, sind deutlichere Leistungssteigerungen zu erwarten, wenn der Geldanreiz zu Beginn eines Lernprozesses eingeführt wird. Aber es müssen etliche Bedingungen hinzukommen, damit eine potentiell durch Geld bewirkte Erhöhung der Lernmotivation voll auf das Lernergebnis durchschlägt, da auch Lernentwicklungen in nicht unerheblichem Maße von vorgegebenen kognitiven Persönlichkeitseigenschaften abhängen.

### Besseres Einprägen durch Geldanreize

Libby und Lipe (1992) überprüften die Wirkung finanzieller Anreize auf die Ausdauer und die Behaltensleistung beim Einprägen von Faktenwissen. Als Lernmaterial dienten insgesamt 28 Regeln oder Aussagen zur Betriebs- bzw. Rechnungsprüfung [z.B. "Paid invoices are effectively cancelled"], die auf 4 Computerseiten verteilt waren. Die Studenten bekamen die Anweisung, diese Regeln sorgfältig zu lesen, da sie später getestet würden. Als Behaltensmaße waren zum einen ein Free recall- sowie ein Recognition-Test [true/false-Format] vorgesehen. Während der Lernphase konnten die Studenten frei bestimmen, wie lange sie auf welchen Bildschirmseiten verweilen wollten, um sich diese Regeln einzuprägen. Auch bei den Behaltenstests gab es keine Zeitbegrenzung. Alle Studenten erhielten 2 Dollar für die Teilnahme am Experiment. Die experimentellen Variationen beziehen sich auf die darüber hinaus gehenden Anreizbedingungen. Den Probanden der Incentive-Gruppen wurde mitgeteilt, für jede richtige freie Wiedergabe und jede richtige Antwort beim recognition test würden sie 10 Cent erhalten, sowie weitere 5 \$, wenn Sie zu den 5 besten Studenten gehören würden. Als experimentelle Gruppen fungierten:

- **Encoding group** incentives prior to encoding: : Die Probanden erfuhren die Anreizinformation vor der Lernphase.
- **Retrieval group:** incentives prior retrieval: Die Probanden erhielten die Anreizinformation erst unmittelbar vor den Behaltenstests.
- **Control group:** Die Studenten bekamen lediglich die 2 \$ für die Teilnahme am Experiment.

**Tabelle 3: Ergebnisse zum Free-recall-test in der Studie von Libby und Lipe (1992)**

|          | control    | retrieval   | encoding    |
|----------|------------|-------------|-------------|
| <b>M</b> | <b>9,8</b> | <b>11,6</b> | <b>12,3</b> |
| <b>s</b> | <b>4,2</b> | <b>5,5</b>  | <b>6,3</b>  |

Die Tabelle 3 beschränkt sich auf die Ergebnisse zum Free-recall-test, der die deutlichsten Behaltensunterschiede zwischen den Bedingungen aufwies. Wie man der Tabelle 3 entnehmen kann, erzielten die Anreizgruppen etwas höhere Behaltenswerte. Der unmittelbare Behaltensvorteil der Anreizgruppen von 6% bzw. 9 % im Vergleich zur Kontrollgruppe entspricht einem kleinen bis mittleren Effekt. Die Unterschiede beim einfacheren Recognitiontest fallen noch geringer aus. Die Autoren listen sehr detaillierte Angaben zu den Lern -und Testzeiten auf, die grob zusammengefasst im Durchschnitt für längere Zeiten der Anreizgruppen sprechen (ca. 3 Minuten mehr Zeit bei einer Gesamtzeit von ca. 18 Minuten für die Kontrollgruppe). Die Anreizgruppen investierten somit etwas mehr Lern- bzw. Testzeit und diese erhöhte Ausdauer könnte einen wesentlichen Teil des Behaltensvorteils mit bedingt haben. Die Autoren hatten hier bewusst freie Lern -und Testzeiten eingeführt, um zeitliche Ausdauerunterschiede überhaupt zu ermöglichen. Sie berichten von einigen früheren Studien, die bei starr festgelegten Lern- und Testzeiten für alle Probanden keine Anzeizeffekte beim Memorieren feststellten. Ich hätte einen höheren Lernerfolg für die Encodinggruppe erwartet, weil eine erhöhte Lernmotivation normalerweise mehr Lerneffekt verspricht als eine erhöhte Testmotivation. Aber der Sondereinsatz der Incentivegruppen von jeweils insgesamt 3 Minuten lässt ja auch keine deutlich höhere Ausdauer erkennen.

Geldversprechen führen selbst bei relativ einfachen Aufgaben wie das Erinnern an Fakten keineswegs immer zu höherer Leistung. Wie Camerer & Hogarth (1999, Table 2) bei ihrer Analyse empirischer Untersuchungen feststellten, konnten lediglich 3 von 7 Studien höhere Memorierungsleistungen [Item recognition/recall] durch Geldanreize nachweisen. Bei Bonner et al. (2000) überwiegen hingegen die positiven Effekte finanzieller Anreize bei Memorierungsaufgaben ziemlich deutlich. Das Ausmaß des Behaltensgewinns durch monetäre Anreize ist schwer einzuschätzen, sollte aber nicht überschätzt werden. In einigen Experimenten von Bauer & Peller-Porth (1990) hörten Kinder Serien von jeweils 10 Worten hintereinander und sollten dann die erinnerten Worte mitteilen (free recall). Somit wurde die verfügbare Lernzeit konstant gehalten. Die Behaltensleistungen unter Geldanreiz waren zwar auf dem Promillenniveau höher als die ohne Geldanreiz, die Behaltensvorteile durch Geldanreize bewegten sich aber durchschnittlich im Bereich von 2 bis 5 %. Durch reine Konzentrationsintensivierung bzw. einer aus dem Stehgreif versuchten Qualitätsverbesserung beim Enkodieren lässt sich vermutlich nicht viel mehr einprägen.

Mit deutlich höheren Effekten würde ich unter folgendem Szenario rechnen. Bei einer Schulklasse, deren Ergebnisse im unbenoteten Vokabeltest bisher durchschnittlich klar unter 50 % korrekter Zuordnungen ausfielen, werden 30 Vokabeln als Hausaufgabe mit dem Versprechen aufgegeben, bei mindestens 80 % korrekter Lösungen im unbenoteten Vokabeltest am nächsten Morgen erhalte jeder Schüler 30 Euro. Hier wäre zu vermuten, dass viele Schüler deutlich länger die Vokabeln einprägten als sie dies früher getan hatten. Allerdings würde ich mit wesentlich geringeren Unterschieden rechnen, wenn man diesen Geldanreiz mit einer zuvor angekündigten Benotung des Vokabeltests vergleichen würde.

### Geldversprechen vor der Lernaneignung lernwirksamer als unmittelbar vor der Testung

Im folgenden werden kurz zwei Studien referiert, welche den meiner Meinung nach theoretisch zu erwartenden Nachweis erbrachten, der Geldanreiz vor der Lernaneignung beeinflusse die Lernleistung stärker als ein Geldanreiz erst unmittelbar vor der Testung. Zudem werden in

beiden Studien Lehrziele erfasst, die zum Teil über ein reines Memorieren hinausgehen und ein gewisses Verständnis erfordern.

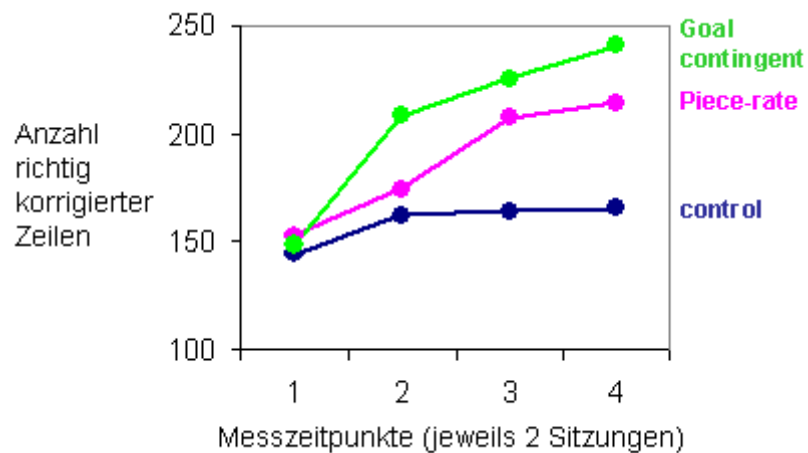
In der Untersuchung von Marsh (1984) erhielten alle Studenten zunächst einen Unterricht in Form eines Lehrervortrags via Video. Die Lehrziele im späteren Test erfassten 11 Items zu Faktenwissen sowie 15 Verständnisfragen. Als materiellen Anreiz bot man neben einem Sockelbetrag von einem Dollar fünf Cent pro korrekt gelöste Aufgabe an. Eine Experimentalgruppe (incentive-to-learn-and-perform) erhielt die Anreizinformation vor der Unterrichtslektion, eine zweite Experimentalgruppe (incentive-to-perform) erst unmittelbar vor der Testung. Die incentive-to-learn-and-perform-Gruppe erzielte signifikant bessere Testwerte als die incentive-to-perform-Gruppe und beide Experimentalgruppen waren einer Kontrollgruppe ohne Geldanreiz signifikant überlegen, womit der Nachweis gelang, materielle Anreize könnten sowohl die Lern- wie die Testmotivation in einem die Leistung steigernden Sinne beeinflussen.

Frase bestätigte bereits 1971 einen ähnlichen Effekt für das Erlernen bestimmter Texte, wobei in dieser Studie keinerlei Zeitbeschränkungen für das Lesen und Testen galten. Eine Geldanreizinformation (leistungsabhängige Einnahmen bis zu 4.2 \$ ) vor dem Lesen erbrachte mit 91,5% korrekter Antworten einen signifikant höheren Lernerfolg als dieselbe Information nach dem Lesen mit 82,4 Prozent. Geldanreize führten insgesamt auch zu höheren Lernergebnissen als kein Geld. Unerklärlich bleibt mir allerdings das Vorgehen, die Kontrollgruppe über die Geldanreize der anderen Experimentalgruppen zu informieren, ihnen dann aber mitzuteilen, sie gehörten nicht zu den Auserwählten. Dies könnte Reaktanz bei der Kontrollgruppe ausgelöst haben, in deren Gefolge sich deren Leistung auch hätte verschlechtern können.

### Geldanreize bei mehrfachen Lerndurchgängen

Nachfolgende Studien analysieren die Auswirkungen von Geldanreizen auf die Leistungsentwicklung bei einer speziellen Lernaufgabe. Eine komplexere Aufgabenstellung wird dabei mehrmals bearbeitet und das jeweilige Leistungsergebnis mit bestimmten Geldversprechen gekoppelt. Dann kann geprüft werden, wie sich der Leistungsverlauf bzw. die Gesamtleistung in Abhängigkeit von der monetären Bedingung gestaltet.

Huber (1985) untersuchte die Leistungsentwicklung beim Erlernen des Korrekturlesens unter verschiedenen Geldanreizbedingungen. Beim Korrekturlesen müssen bestimmte Fehler in besonderer Weise gekennzeichnet werden und dabei verschiedene Symbole verwendet werden. Das Randomisierungsexperiment umfasste insgesamt 10 Termine, die sich über 5 Tage verteilten. In den ersten beiden Sitzungen wurden die Probanden im Korrekturlesen eingewiesen und Anfangsleistungen erhoben. In den folgenden 8 Sitzungen bearbeiteten sie die Texte unter verschiedenen experimentellen Bedingungen. Als abhängige Variable diente die Anzahl der korrekt bearbeiteten Zeilen innerhalb einer fest gelegten Zeit. Alle Probanden erhielten bereits für die Teilnahme 30 \$. Die Kontrollgruppe musste sich damit begnügen und bekam lediglich die Anweisung, ihr Bestes zu geben. Unter der Stücklohnbedingung (piece-rate) war zunächst eine Mindestleistung (Durchschnittswert) zu erzielen. Erst danach konnte der Proband für jeweils 5 weitere korrekte Korrekturzeilen 4 Cent hinzu verdienen. Weitere Geldanreizbedingungen wurden vom Erreichen selbst gesetzter bzw. fremdbestimmter Zielsetzungen abhängig gemacht. Die Abbildung 2 beinhaltet nur 3 der insgesamt 8 experimentellen Bedingungen und stellt die für jeweils 2 Termine zusammen gefassten Ergebnisse der Kontrollgruppe den günstigsten Varianten bestimmter Geldanreizbedingungen gegenüber.

**Abbildung 2: Leistungsergebnisse in Abhängigkeit vom Geldanreiz nach Huber (1985)**

Im Gegensatz zur Kontrollgruppe konnte bei allen Geldanreizbedingungen, auch bei den in der Abbildung 2 nicht dargestellten Varianten, ein signifikanter positiver linearer Trend gesichert werden. In der hier gezeigten Goal-contingent-Bedingung setzte der Versuchsleiter bestimmte Ziele, die im übrigen auch relativ sicher erreicht werden konnten. Nur beim Erreichen oder Übertreffen dieser Ziele war eine Geldzahlung fällig. Leider teilt Huber keine Streuungen mit. Wegen der geringen Vpn-Anzahl von ca. 10 pro Gruppe sind statistische Effekte natürlich schwer zu sichern. Im Gesamtmittelwert erzielte nur die goal-contingent-Gruppe ( $M=206$ ) signifikant bessere Leistungen als die Fix-Payment-Kontrollgruppe ( $M=159$ ). Der mitgeteilte F-Wert von 9,86 erlaubt eine Effektstärkeschätzung, die mit  $d = 1.4$  sehr bedeutsam ausfällt. Die piece-rate-Gruppe erreichte immerhin einen signifikant anderen Trend als die Kontrollgruppe. Die Ergebnisse lassen insgesamt keinen Zweifel daran aufkommen, durch leistungsabhängige Geldanreize eine Effizienzsteigerung beim Korrekturlesen erreichen zu können.

Wegen der relativ geringen kognitiven Beanspruchung sei hier nur am Rande verwiesen auf eine neuere Studie von Goette & Lienhard (2006) zur Wirkung des Geldanreizes auf die quantitative Leistung bei der Dateneingabe Hand geschriebener Lieferscheine in den Computer. Eine Piece-rate-Gruppe erhielt letztendlich mit 0,2 CHF pro Lieferschein im Mittel genauso viel Geld wie die Gruppe, welche leistungsunabhängig einen festen Geldbetrag von 50 CHF kassierte. Die Autoren ermittelten eine signifikant höhere Anzahl von Dateneingaben für die piece-rate-Gruppe und dies bei vergleichbarer Fehlerrate. Der Leistungsvorteil kristallisierte sich nach den ersten Lernperioden heraus und blieb dann weitgehend konstant erhalten. Die zwischenzeitlich mit einer Adjektivliste mehrmals erhobenen Daten zur emotionalen Befindlichkeit ergaben signifikant ungünstigere Ergebnisse für die Stücklohngruppe. Die Stücklohngruppe fühlte sich **unwohler, unruhiger, weniger entspannt und weniger gelassen**. Sie konzentrierte sich jedoch stärker auf die Arbeit ("The activity required my entire concentration.") und hatte mehr Spaß an der Arbeit ("I found the activity to be fun.").

Im Experiment von Bailey, Brown und Cocco (1998) war eine neue Lernaufgabe, das Zusammensetzen eines Krans nach vorgelegtem Plan, mehrmals vorzunehmen. Die gesamte Prozedur erstreckte sich über ca. 4 Stunden. Während dieser Zeit können die Probanden deutlich hinzulernen und die anstehende Montagearbeit somit zunehmend verbessern. Es wird demnach auch hier eher der Erwerb einer Fertigkeit als ein einmaliger Lernprozess analysiert oder lediglich vorhandene Fähigkeiten getestet. An dem Experiment nahmen 72 Probanden teil, die nicht nach Zufall auf 3 Gruppen aufgeteilt wurden, aber hinsichtlich wichtiger Variablen vergleichbar waren. Die Kontrollgruppe (fixed payment) erhielt 20 \$ für die Teilnahme. Die Stücklohngruppe (Piece-rate) bekam für jede Montagefertigstellung 1.8 \$. Die zweite Geldanreizgruppe (Goal-contingent) erhielt einen festen Beitrag sowie eine Bonuszahlung beim Erreichen bestimmter Ziele. Die Bedingungen waren so gewählt worden, dass letztlich

alle Bedingungen am Ende des Experimentes in etwa vergleichbare Kosten verursachten.(=kostenneutrale Bedingungen). Als abhängige Variable diente die Anzahl der Fertigstellungen (Wiederholungen) in einer Zeiteinheit:

**Tabelle 4: Ergebnisse des Experimentes von Bailey Brown und Cocco (1998)**

|                 | Anzahl der Wiederholungen / Zeiteinheit |     |
|-----------------|---|-----|
|                 | M                                       | s   |
| Kontrollgruppe  | 4,8                                     | 1,2 |
| Stücklohn       | 5,5                                     | 1,3 |
| goal-contingent | 6,0                                     | 1,3 |

Wie Tabelle 4 nahelegt, erzielten die Geldanreizgruppen signifikant höhere Lernleistungen als die Kontrollgruppe. Weitere Analysen ergaben, dass dieser Vorteil im Wesentlichen auf die erste Montage zurückzuführen war. Die Geldanreizgruppen gingen die Arbeit somit zügiger an, ihr Lernzuwachs nahm im Verlauf aber nicht deutlicher zu als derjenige der Kontrollgruppe. Um Lernzuwächse zu fördern, müsste sich das monetäre Schema der Vergütung an Leistungsverbesserungen ausrichten.

Sprinkle (2000) analysierte ebenfalls die Auswirkung leistungsabhängiger monetärer Anreize über mehrere Lerndurchgänge. Die Probanden mussten dabei jeweils ein neues kompliziertes mehrstufiges Entscheidungsproblem (von offenbar gleicher Tiefenstruktur) lösen. Dabei erhielten alle Probanden stets ein gewisses sachorientiertes Feedback, dessen Bedeutung allerdings selbst erschlossen werden musste. Die Kontrollgruppe bekam einen Fixbetrag für die Teilnahme am Experiment versprochen. Die Probanden der Experimentalgruppe bearbeiteten das Entscheidungsproblem unter der Bedingung einer komplizierten, leistungsabhängigen Bezahlung mit Lotteriecharakter. Die Zahlung war fällig, wenn eine bis zu einer gewissen Höhe mögliche Zufallszahl kleiner als die erzielte Punktzahl bei der Aufgabe ausfiel. Dieses Vergütungsschema führte im Mittel zu vergleichbar hohen Geldeinnahmen wie die Fixpayment-Bedingung. Die Studenten unter leistungsabhängiger Bezahlung zeigten bei allen Aufgaben durchgängig eine höhere Ausdauer. Sie erzielten aber erst in den letzten Lerndurchgängen bessere Leistungsergebnisse. Sprinkle vermutet, dieser größere Leistungszuwachs gehe nicht nur auf die umfangreichere Bearbeitungszeit, sondern auch auf die verbesserte Qualität der Aufgabenbearbeitung zurück.

**Leider habe ich keinerlei weitere empirische Belege dafür finden können, dass eine durch Geldversprechen kontrollierte, erhöhte Lernmotivation bei einer umfangreichen Lernserie zu einer qualitativ besseren Problemlösung beigetragen hätte.**

Eisenberger, Armeli, & Pretz (1998) prüften in einem Experiment, ob Geldversprechen kreative Leistungen erhöht. Als Trainings- und zugleich Kriteriumsaufgaben dienten Zeichenvorlagen, die zu kreativen Bildern ergänzt werden sollten. Insgesamt waren 24 solcher Zeichenvorlagen zu bearbeiten. Der Leistungsunterschied zwischen der Gruppe mit Geldanreiz (offenbar leistungsunabhängig insgesamt 75 Cent) und einer Gruppe ohne jeden Geldanreiz trat deutlich ( $p < .005$ ) in Erscheinung, wenn zuvor ganz klar expliziert wurde, dass eine kreative Leistung verlangt war und welche Kriterien eine kreative Leistung ausmachten. Hierbei trat der Kreativitätsvorteil der Geldanreizgruppe erst im Verlauf der Itembearbeitungen zu Tage und zeigte sich ganz klar beim letzten Drittel der Aufgaben. Dieser Entwicklungsverlauf könnte auf anfänglich ganz interessante Aufgabenstellungen hinweisen, die mit der Zeit ihre intrinsische Attraktivität verlieren, die Geldanreizgruppe aber zu einer größeren Ausdauer bewegten, das kreative Verhalten länger durchzuhalten. Das Experiment beweist keineswegs, Geld verbessere die Kreativität, sondern verdeutlicht lediglich, durch Geldanreize ließen sich auch kreative Leistungen analog sonstigem Verhalten, welches im Leistungsvermögen der Probanden angelegt ist, eher hervorlocken.

## Geldanreize für schulische Leistungen

Wie oben aufgezeigt, führt eine durch Geldversprechen intendierte höhere Testmotivation kaum zu erwähnenswerten Leistungsvorteilen. Die experimentellen Studien zu Lernvorteilen durch Geldanreize liefern zwar gewisse Evidenz für ihre potentielle Wirksamkeit, engen den Spielraum des einzelnen vor allem hinsichtlich der möglichen Arbeitszeit aber immer noch relativ stark ein. Das größte Lern- bzw. Leistungspotential extrinsischer Anreize sehe ich in einer entsprechend verbesserten Arbeitsintensivierung während des Semesters bzw. der Prüfungsvorbereitung. Geldversprechen müssten somit rechtzeitig und nicht unmittelbar vor der Prüfung bzw. einer Testung ansetzen. Allerdings motivieren Noten bereits in sehr hohem Maße. Somit stellt sich die Frage, ob bzw. unter welchen Bedingungen zusätzliches Geldversprechen noch mehr Leistung zu evozieren vermag. Das scheint zumindest auch gelingen zu können. Goetz, Tucker und Weber (1973) erreichten bei college freshmens eine Notenverbesserung durch Geldanreize für ein Übertreffen ihres normalen Leistungsniveaus und kommen zu dem Ergebnis .." the payment rates appeared to produce differences of from roughly a fourth to a third of a letter grade in the quarterly grade averages of some large segments of the test groups."

## Monatliches Stipendium in Abhängigkeit von der Schulleistung ?

Meistens werden finanzielle Unterstützungen zur Verbesserung des Bildungswesens in pädagogische Einrichtungen wie Lehrer, Schulgebäude oder Lehrmaterialien investiert (top-down approach) und kommen nicht direkt den Schülern (bottom-up approach) zu Gute. Eine private Organisation in Amerika stellte nun Mittel zur Verfügung, um die Schulleistungen von high-school-Schüler durch direkte monetäre Anreize zu verbessern. Als förderungswürdig galten allerdings nur high-achieving, low-resource public high school students, also solche Schüler, die bereits relativ gute Schulleistungen erzielten, aber in ziemlich ungünstigen Umwelten lebten.

Die Untersuchung von Spencer, Noll. & Cassidy (2005) basiert auf einem für Experimentieren in der realen Umfeld durchaus ansprechenden versuchsplantechnischen Niveau. Die förderungsberechtigten Highschool-Schüler wurden nach bestimmten Variablen stratifiziert nach Zufall auf 2 Gruppen zugeteilt:

1. **Leistungsabhängiges Stipendium:** Es wurde abhängig vom Schuljahr ein monatliches Stipendium zwischen 50 und 65 \$ gewährt, wobei die Auszahlung von der Erfüllung bestimmter Zulassungskriterien (Mindestnoten in bestimmten Fächern) abhing, die monatlich überprüft wurden.
2. **Kontrollgruppe:** Kein Stipendium, sondern Wartegruppe für ein Stipendium im nächsten Jahr.

Als abhängige Variable diente die Anzahl der Schüler, die am Ende des Schuljahres die gesetzten Schulleistungskriterien erfüllten. Der entsprechende Prozentsatz war unter der Bedingung leistungsabhängiges Stipendium mit 61% signifikant höher als unter der Bedingung "kein Stipendium" (51%) und bestätigt somit im Mittel eine Leistungssteigerung durch materielle Anreize. Allerdings war der Erfolg der Maßnahme auch abhängig vom Geschlecht, dem Schuljahr sowie der ethnischen Gruppe. Zudem verbietet sich eine vorschnelle Generalisierung der Befunde auf alle Schüler, weil die Untersuchung nur high achieving pupils einbezog. Auch wenn monetäre Anreize die Leistungsbereitschaft und den Arbeitseinsatz erhöhen sollten, so folgt daraus eben keineswegs zwingend eine höhere Schulleistung für alle Schüler.



Bereits Reid & Bailey-Dempsey, C. (1995) führten bei gefährdeten Teenagern unter anderem auch eine Payment-Bedingung ein, welche über 2 Jahre hinweg Geldzahlungen an bestimmte Leistungsverbesserungen in der Schule koppelte. Die Schülerinnen wurden nach Zufall auf die Bedingungen zugeteilt. Die Payment-Gruppe erzielte insgesamt bessere Schulleistungen als die nicht bezahlte Kontrollgruppe. Allerdings fielen die Leistungen im Folgejahr wieder ab, als die experimentellen Programme beendet wurden, was eigentlich für die unmittelbare Wirksamkeit der Geldzahlungen spricht.

Einige hier nicht weiter verfolgte empirische Studien an Schülern aus Israel, Kolumbien oder Kenia deuten ebenfalls auf gewisse Leistungseffekte durch Geldzahlungen hin, sind aber in ihrem methodischen Status schwer durchschaubar und erlauben meiner Meinung nach keine eindeutigen Schlussfolgerungen. Häufig bieten solche Geldzahlungen in Entwicklungsländer für die Betroffenen erst die Möglichkeit, sich angemessen vorbereiten zu können.

Geldanreize für ein zügiges Studium in der Studie von Leuven, Oosterbeek und Klaauw (2003) ?

Leuven, Oosterbeek und Klaauw (2003) untersuchten in einem klassischen Experiment die Auswirkungen von Geldanreizen auf Studienleistungen in den Wirtschaftswissenschaften. Voraussetzung für die Geldzahlung am Ende eines Studienjahres war das Erreichen aller möglichen Creditpoints im untersuchten Jahr [passing all first year exams before the start of their second academic year.] Ca 250 Studenten wurden nach Zufall auf 3 Bedingungen aufgeteilt:

- kein Geld (=Kontrollgruppe)
- 227 Euro
- 681 Euro

Es ergaben sich geringfügig, aber nicht signifikant mehr erfolgreiche Abschlüsse des akademischen Jahres für die Studenten unter der 681 Euro-Bedingung im Vergleich zur Kontrollgruppe. (first year pass rate. 24 % gegenüber 20%) , "**the rewards have small and non-significant effects**"). Die Leistungsschwachen erzielten unter der 681 Euro –Payment-bedingung eher schlechtere Ergebnisse. Die im Durchschnitt erzielten numerischen Vorteile unter der 681 Euro-Bedingung basierten vornehmlich auf den besseren Leistungen der leistungsfähigsten Studenten dieser Incentive-Bedingung. Diese hatten auch objektiv höhere Chancen, das strenge Paymentkriterium überhaupt zu erfüllen. Insgesamt waren die Anforderungen zum Erreichen des Geldbetrages sehr hoch, für die meisten völlig unrealistisch und letztlich ziemlich zeitintensiv. Viele Studenten in den Niederlanden nehmen Arbeit an, um ihr verfügbares Einkommen aufzubessern. Wenn man bedenkt, wie viel Lerneinsatz es erforderte, das versprochene Geld unter hohem Risiko überhaupt zu erhalten, dann hätten die meisten Studenten bei vergleichbarem Zeiteinsatz durch konventionelle Arbeit deutlich höhere Erfolgchancen gehabt, insgesamt mehr Geld zu verdienen. Die Studenten unter finanziellen Anreizen behaupteten zwar, härter für das Studium gearbeitet zu haben, was sich jedoch nicht durch entsprechend höhere Studierzeiten im Vergleich zur Kontrollgruppe belegen ließ. Die empirischen Befunde entsprechen eigentlich theoretischen Erwartungen. Geldversprechen für sehr schwer erreichbare Kriterien können Studenten nicht zu einem zügigeren Studium bewegen. An der Universität des Saarlandes versucht man es denn auch mit Strafgebühren für Langzeitstudierende, um der Studienbummelei Einhalt zu gebieten.

## Schlussbetrachtung Geld als Lernanreiz für kognitive Leistungen.

Da Geld nur über eine Motivationserhöhung kognitive Leistungen beeinflussen kann, hängt der potentielle Leistungsvorteil des Geldanreizes immer auch von den Anreizen der Kontrollbedingung ab. Beim Vergleich "Geld gegen kein Geld" drängt sich die Frage auf, wie die Probanden ansonsten zu kognitiven Leistungen bewegt wurden. Geld fungiert hierbei entweder als zusätzlicher Anreiz gegenüber irgendeiner Normalbedingung, etwa: "Der Versuchsleiter gibt einfach nur die Anweisung, Aufgaben zu bearbeiten bzw. sein Bestes zu geben" oder als echte Alternative gegenüber sonstigen, als ähnlich wirksam angesehenen Motivatoren (z.B. Noten, Rückmeldung der Ergebnisse, angepriesener Nutzen der Testung für andere usw.). Eine generalisierbare Normalbedingung ist allerdings schwer vorstellbar, weswegen diese Art der Kontrolle theoretisch nicht viel hergibt, da ihr Spektrum von gänzlich unzureichender bis zu sehr guter Motivierung reicht.

In der Regel erhalten die Probanden der Kontrollgruppe einen bestimmten Geldbetrag für die Teilnahme an der Untersuchung. Damit gehen viele die Verpflichtung ein, wenigstens eine mäßige Leistung abzuliefern. Das Motivierungspotential von Geld für kognitive Leistungen läge höchstwahrscheinlich wesentlich höher, wenn man attraktive Geldversprechen gegen völlig freigestellte Lernbemühung testen würde. Dies entspräche etwa einer schulischen Situation, bei der allen Schülern die freiwillige Bearbeitung eines bestimmten, nicht notenrelevanten Lehrtextes empfohlen wird, die EG-Schüler im Gegensatz zu den KG-Schülern dann aber zusätzlich sehr attraktive, von der Lehrzielerreichung abhängige, Geldbeträge erwarten könnten. Der pädagogische Trivialfall läge schließlich vor, wenn ein Kandidat überhaupt nur noch bei Geldversprechen dazu bereit ist, irgendeine Leistung zu erbringen.

Von einem sehr strengen Versuch kann man sprechen, wenn leistungsabhängige Geldanreize mit einem vergleichbar hohen, leistungsunabhängigen Fixbetrag (fixed-wage oder fix payment -treatment,) verglichen werden. Hier wäre nun motivationstheoretisches Wissen verlangt, die optimale Geldanreizbedingung zu begründen. Die Arbeiten aus den Wirtschaftswissenschaften orientieren sich meist an praktischen Modellen in der Arbeitswelt. Stücklohn im Sinne eines fixen Geldbetrags pro korrekt gelöste Aufgabe hat sich häufig als wirksam erwiesen. Ein anstrengungssensitives, am individuell erreichbaren Leistungsniveau orientiertes Geldversprechen bewirkt möglicherweise einen noch stärkeren motivationalen Schub. Zumindest einigen handwerklichen Fehlern, die bei guten Untersuchungen nicht vorkommen, könnte vorgebeugt werden. So sollte man den leistungsunabhängigen Geldbetrag an bestimmte untere Mindestleistungen koppeln, um Arbeitsverweigerungen oder unzumutbar schwaches Leistungsverhalten zu vermeiden. Die jeweilige Geldanreizbedingung sollte durchsichtig gestaltet und den Probanden stets bewusst sein. In manchen Experimenten ergab eine nachträgliche Befragung bei etlichen Probanden eine offensichtliche Unkenntnis ihrer Anreizbedingung.

Zunächst müsste stets der Nachweis erfolgen, dass die Geldbedingung eine deutlich höhere Anstrengungsbereitschaft oder Ausdauer nach zieht, was man zumindest auf subjektiv verbaler Ebene erfassen, aber nach Möglichkeit auch durch objektive Daten [Beobachtung, Lernzeiten] belegen könnte. Im positiven Falle stellt sich die Frage des kausalen Zusammenhangs zwischen Motivationsveränderung und Lern- bzw. Testerfolg. Das verfügbare Wissen erlaubt vermutlich keine exakten Prognosen für eine Leistungssteigerung durch Motivationserhöhung, da hier ein Bündel von Faktoren eine wesentliche Rolle spielt, etwa Vorwissen, Experteniveau, Intelligenz, Aufgabenkomplexität, Aufgabenschwierigkeit, verfügbare Zeit usw. Die experimentelle Motivationsforschung hat mit Geldanreizen immerhin die Möglichkeit, Lernmotivation auf experimentellem Wege zu variieren und zu aussagekräftigeren Resultaten

zu gelangen als dies die üblichen Korrelationen zwischen einem Lernmotivationstest und der Lernleistung oder pfadanalytische Rechenispiele suggerieren. Einige Experimente zu den Auswirkungen leistungsabhängiger Geldanreize auf das Ergebnis in internationalen Tests steuerten durch die Hinzunahme der Motivationsvariablen verstehbare Befunde für die Unwirksamkeit des Geldanreizes bei der Erfassung bestimmter mathematischer Kompetenzen bzw. Intelligenz, weil dieser die bereits recht hohe Anstrengungsbereitschaft meistens nicht weiter erhöhte. Dort, wo das Geldversprechen dennoch höhere Anstrengungsinvestitionen bewirkte, gab es aber keinen Zusammenhang zwischen Anstrengung und Testleistung.

Der Geldanreiz selbst sollte ein angemessenes Niveau umfassen. Sehr geringe Geldbeträge zeigen keine oder sogar nachteilige Wirkungen. Auch extrem hohe Geldversprechen können unter bestimmten Bedingungen, etwa hohem Zeitdruck und starker kognitiver Beanspruchung, ungünstige Leistungsergebnisse nach sich ziehen. Ariely et al. (2005) fanden bei etlichen kognitiven Anforderungen, u. a. Spiele, die Kreativität und Konzentration verlangten, schwächere Leistungen für sehr hohe gegenüber mittleren und geringeren Geldanreizen. So mussten die Probanden etwa aus einer Matrix von 12 Zahlen [jeweils 3 Ziffern, davon 2 nach dem Dezimalpunkt] diejenigen beiden Zahlen finden, deren Summe genau 10 ergab. Diese Aufgabe ist zwar relativ sicher lösbar, aber unter Zeitdruck kognitiv sehr anstrengend. (siehe: [Beispiel](#)). Von insgesamt 20 dieser Matrixen waren innerhalb einer Zeit von 4 Minuten mindestens 10 korrekt zu lösen, bevor überhaupt ein Geldbetrag erzielt werden konnte. Jede weitere korrekte Lösung ergab dann einen zusätzlichen Stücklohn. Unter der sehr extremen Geldanreizbedingung von bis zu 300\$, erreichten die Probanden signifikant schlechtere Leistungen als unter einem geringeren Geldanreiz von bis zu höchstens 30\$, was möglicherweise -analog dem Yerkes Dodson Gesetz- auf den erhöhten Stress unter dem extrem hohen Geldanreiz zurück geht.

Schwer einzuschätzen bleibt, ob durch leistungsabhängige Geldstipendien Schul- bzw. Studienleistungen deutlich verbessert werden, da meist leistungsfähige Extremgruppen aus einkommensschwachen Verhältnissen untersucht wurden. Condly et al. (2003) kommen in einer neueren Metaanalyse zur Wirksamkeit von Geldanreizen in der Arbeitswelt zu dem Ergebnis, auch für Tätigkeiten, die überwiegend kognitive Leistungen erfordern, würden leistungsabhängige Geldanreize die Leistungen verbessern. Die Autoren teilen für mental bzw. cognitive work eine auf 38 Studien basierende durchschnittliche Effektstärke von  $d = .6$  zugunsten leistungsabhängiger Bezahlung mit.

Rein pragmatisch lässt sich aus den durchgeführten Untersuchungen ableiten, durch Geldanreize könnten auch kognitive Lernleistungen verbessert werden. Dies wurde ziemlich gut für relativ einfache geistige Aufgabenanforderungen nachgewiesen, ist aber auch bei komplexen Aufgaben nicht unmöglich. Selbst bei vergleichbarer Lernzeit und somit nicht allein durch vermehrte Ausdauer ließen sich Vorteile für leistungsabhängige Geldanreize finden, was dann eine höhere Lerneffizienz belegt. Hoch fähigkeitsabhängige Testleistungen sind hingegen nicht zu beeinflussen. Mir ist lediglich die Untersuchung von Goette & Lienhard (2006) bekannt, welche neben aktuellen Leistungen auch Daten zu aktuellen emotionalen Befindlichkeiten beisteuert. Die Ergebnisse dieser Studie werfen die Frage auf, ob die durch Geldanreize bewirkte Leistungssteigerung mit höherem Unwohlsein, verstärkter Anspannung und größerer Nervosität bei der Leistungsausführung erkauft werden muss. Ab einer bestimmten kognitiven Beanspruchung wirken sich derartige Emotionen eher negativ auf die Leistung aus.

## Benotung als extrinsische Motivierung zu besseren Schulleistungen

Neben Geld fungiert die Benotung als klassische extrinsische Motivationsquelle. Das Erreichen guter Noten gehört ohne Frage zu den wichtigsten Zielen eines Schülers. Die Zertifizierung guter Noten ist den Betroffenen häufig wichtiger als der entsprechende Wissenszuwachs bzw. die gesteigerte Kompetenz, welche die Noten eigentlich diagnostizieren sollten ([Entwurf für eine Aufgabe zur empirischen Überprüfung dieser Behauptung](#)). Da von der Benotung vielfältige weitere intrinsische wie extrinsische Konsequenzen abhängig sind bzw. genauer, vom Individuum abhängig gemacht werden, wirken Noten als ziemlich generalisierte Anreize und darüber hinaus als Bedrohungen. Während bei klassischen Geldanreizexperimenten im ungünstigsten Fall nichts gewonnen werden kann, wirkt die Benotung insgesamt noch breiter durch den zusätzlichen Anreiz, drohende aversive Konsequenzen abzuwenden. Culler et. al (1975) konnten z.B. nachweisen, dass die Erledigung eines Arbeitsauftrags (a one page library assignment) durch das Versprechen zusätzlicher Punkte bei der Notevergabe anstieg. Der Prozentsatz der Schüler, welche den Arbeitsauftrag erfüllten, war aber noch etwas höher, wenn die Nichterfüllung des Arbeitsauftrags mit Punktabzügen bei der Note geahndet wurde.<sup>1</sup>

Es bedarf keiner großen Phantasie anzunehmen, benotete Klausuren bzw. Klassenarbeiten führten im Vergleich zu unbenoteten Klausuren zu besseren Leistungsergebnissen. Möglicherweise ist die hohe subjektive Evidenz dieser Hypothese ein Grund für die Kargheit der empirischen Untersuchungen zu diesem Thema. Die Erwartung einer benoteten Leistungsprüfung dürfte sich vornehmlich in einer verbesserten Vorbereitung niederschlagen, die dann zu höheren Klausurergebnissen führen könnte bzw. sollte. Der Einfluss einer potentiell gesteigerten Testmotivation durch Noten ist demgegenüber deutlich geringer zu veranschlagen, da diese bei unzureichender Vorbereitung keine deutliche Erhöhung der Testleistung bewirken kann. Eine mögliche Überlegenheit der Bedingung "Benotung bei der Testung" könnte aber durch mangelndes Interesse bei Nichtbenotung zustande kommen. Im Übrigen hängt die Größe der überhaupt erreichbaren Effekte ganz entscheidend davon ab, wie gut die zu prüfenden Lehrziele durch vermehrte Vorbereitungsintensivierung bzw. Anstrengung auch erreicht werden können. In diesem Zusammenhang fungieren erforderliches Vorwissen, Lehrzielniveau, Transparenz der Prüfungsanforderungen, didaktische Aufbereitung des Lehrstoffs sowie verfügbare Zeit als wichtige Einflussgrößen. Je mehr die Leistungsprüfungen den Charakter von Intelligenztests oder hoch fähigkeitsabhängiger Kompetenzerfassung annehmen (z.B. Pisa), um so weniger positiver Lerneffekt können eine Benotung sowie sonstige extrinsische Konsequenzen ausrichten. (Siehe dazu: [die Ergebnisse von Baumert und Demmrich 2001](#), sowie Kiplinger and Linn (1992) ). In derartigen Fällen ist zumindest bei leistungsängstlichen Personen sogar eher eine Leistungsverschlechterung zu erwarten.

Die realistische Einschätzung des Leistungsvorteils benoteter gegenüber unbenoteten Arbeiten steuert auch wertvolle Informationen zur Bewertung freiwilliger Lernangebote, etwa unbenoteter Online-Tests bei. Müssen Zwischentests während des Seminars z.B. benotet werden, damit sie die Leistungen im Abschlussexamen auch erhöhen oder genügen hierfür unbenotete Online-Tests? Lernt der Student vorwiegend durch die Testbearbeitung oder doch wesentlich mehr durch eine vernünftige Testvorbereitung, welche durch Notenrelevanz möglicherweise stärker angeregt wird?

## Testmotivation und Testleistung bei benoteten und unbenoteten Examen

Im Mittelpunkt einiger Studien steht die Frage, wie stark das Wissen um eine Benotung das Leistungsergebnis und die Testmotivation in einem Examen beeinflusst. Offensichtlich sind dabei aber nur die Auswirkungen der Benotung auf die Bearbeitung während des Exams von Interesse. Die Untersuchung beginnt quasi erst unmittelbar vor dem Austeilen der Klausurunterlagen. Hier erfahren z.B. die Studenten in der Kontrollbedingung, dass sie zwar am Examen teilnehmen sollen, das Ergebnis aber nicht benotet wird. Derartige Studien erfassen somit nicht die mögliche verbesserte Prüfungsvorbereitung angesichts einer benoteten Klausur.

Wolf & Smith (1995) überprüften die Frage, ob ein benotetes Examen höhere Leistungen nach sich zieht als ein unbenotetes Examen. Ich gehe davon aus, dass alle Studenten ein benotetes Examen antizipierten. Denn, die ca. 150 Studenten wurden erst am Tag des Exams nach Zufall zu 2 Bedingungen aufgeteilt: Auf der ersten Seite der Examensunterlagen stand entweder "Das Examensergebnis wird gewertet" oder "Das Examensergebnis hat keinerlei notenrelevante Konsequenzen". Am nächsten Tag fand ein Treatmentreversal statt, so dass alle Studenten einmal ein benotetes und einmal ein unbenotetes Examen bearbeiteten. Die Studenten erzielten unter der Bedingung "Testergebnis wird gewertet" mit  $d = .26$  etwas, wenngleich hochsignifikant, höhere Testwerte. Der Unterschied hinsichtlich der subjektiven Motivation (z.B. "I was highly motivated to do well on this test" oder "I gave my very best effort on this test") fiel mit  $d = 1.45$  allerdings sehr überzeugend zu Gunsten des bewerteten Exams aus. Die Benotung des Exams hatte folglich dramatische Auswirkungen auf die subjektive Bedeutung und Anstrengungsbereitschaft während der Testbearbeitung, wirkte sich aber nur relativ marginal auf die Testleistung aus.

Wolf, Smith, and DiPaulo (1996) fanden ähnliche Ergebnisse bei mehreren Examen. Die Leistungsunterschiede fielen im Schnitt noch etwas höher aus. Die Effektstärke bzgl. der höheren Testmotivation erreichte mit  $d = 1.53$  wiederum riesige Ausmaße. Meine Ergebnisse zur [Angst in unterschiedlichen Leistungssituationen](#) bestätigen im übrigen die höhere Motivierung durch benotete Klausuren. Die Daten deuten darauf hin, dass die aktuelle Angst in benoteten Klausuren deutlich höher ausfällt als in anonymen, unbenoteten Klausuren. Eine Ausnahme bilden hier jedoch anonyme, unbenotete Probeklausuren im Sinne von Prüfungssimulationen, die ähnlich hohe aktuelle Angst auslösen wie echte Klausuren..

In der Untersuchung von Sundre & Kitsantas (2004) war den Studenten zu Beginn des Seminars eindeutig klar, dass zu einem bestimmten Termin ein echtes Examen anstand, wodurch vergleichbare Prüfungsvorbereitungen für alle Studenten zu erwarten waren. Ein Teil der ausgeteilten Prüfungsunterlagen war eindeutig als notenrelevant, der andere parallele Teil als "ohne jede Notenrelevanz" gekennzeichnet. Als Testformate dienten MC-Aufgaben (60% Gewicht für die Note), sowie eine [Essaytest-Aufgabe](#) (40% Gewicht für die Note). Tabelle 5 zeigt die wesentlichen Leistungsergebnisse:

**Tabelle 5: MC- und Essay-Testergebnisse für benotete und unbenotete Klausuren in der Studie von Sundre & Kitsantas (2004)**

|                        | MC          |     | ESSAY       |     |
|------------------------|-------------|-----|-------------|-----|
| Leistungsergebnis wird | M           | s   | M           | s   |
| benotet:               | <b>49.9</b> | 4.9 | <b>37.0</b> | 2.7 |
| nicht benotet:         | <b>46.1</b> | 7.5 | <b>28.6</b> | 9.4 |

Die Leistungsvorteile unter Benotung lagen beim MC-Test im mittleren Bereich. Ganz deutliche Leistungsunterschiede zeigten sich aber beim Essay-Test. Das ist verständlich, weil die

freie Beantwortung in Form eines kleinen Aufsatzes recht viel Mühe macht, die man sich bei Notenirrelevanz nicht unbedingt zumuten muss. Das Ergebnis bestärkt mich in der Ansicht, es sei wenig nutzbringend, in einem unbenoteten E-learning-Angebot Essay-Test-Aufgaben als Übungsmöglichkeiten anzubieten. Die ernsthafte Bearbeitung solcher Aufgaben hängt sehr stark von der Bereitschaft des Lernalers ab. Die Autorinnen stellten für den unbenoteten Essaytest im Gegensatz zum benoteten Essaytest auch einen signifikanten Zusammenhang zwischen [Testmotivation](#) und dem Testergebnis fest. Die subjektiv erhobene Testmotivation war bei allen möglichen Vergleichen zwischen den benoteten und unbenoteten Testbedingungen stets deutlich höher zugunsten der notenrelevanten Testung. Weil alle Probanden unter Notenrelevanz ziemlich stark motiviert waren, konnten dort auch keine klaren Zusammenhänge mehr zwischen Testmotivation und Leistungsergebnissen gefunden werden.

Wise & DeMars (2005) listen weitere Studien auf, die einen Leistungsvergleich zwischen unbenoteten Tests und Tests mit extrinsischen Konsequenzen durchführten. Bei allen diesen Studien sind die potenzielle Leistungsunterschiede ausschließlich der Testmotivation während der Testbearbeitung anzulasten und sagen nichts über die potentiellen Unterschiede des Arbeitseinsatzes während des Seminars oder der Klausurvorbereitung aus. Zum Vergleich "benotetes vs. unbenotetes Examen" teilen die Autoren in ihrer Tabelle 1 insgesamt 7 Effektstärken aus 3 Studien mit, die alle für höhere Leistungen in benoteten Examen sprechen und zwischen 0,26 und 1,38 schwanken.

Man muss sich bei diesen Experimenten allerdings die Frage stellen, warum nicht nahe liegenden Möglichkeiten in Erwägung gezogen wurden, auch der Bearbeitung unbenoteter Examen etwas mehr Sinn zu verleihen als der Maxime des Versuchsleiters zu folgen: "Tut mir den Gefallen und beantwortet die Aufgaben!" Hierbei böte sich unter anderem das Versprechen an, das Leistungsergebnis als Prozentsatz der korrekten Lösungen rückzumelden oder Vergleiche von eigenen Antworten und Musterlösungen anzubieten. Aus diesem Grunde bleibt schwer einzuschätzen, ob der gefundene Leistungs- und Testmotivationsvorteil benoteter Klausuren auch gegenüber unbenoteten Computertests mit vielfältigen Rückmeldemöglichkeiten noch gilt. Zudem könnte einer unbenoteten Klausur eine bedeutsamere didaktische Funktion verliehen werden? Ich schätze z.B. die Testmotivation höher ein, wenn die unbenotete Klausur in Form einer Probeklausur 2 Wochen vor der echten Klausur stattfindet, weil der Student in einer derartig unbenoteten Klausur ein echtes Interesse daran haben müsste, seine Leistung möglichst valide einzuschätzen.

**Hinweis:** Die schwächeren Ergebnisse der unbenoteten Examen hinsichtlich Testmotivation und Leistungsergebnis sind nicht einfach auf eine Testung von Kompetenzen im Rahmen nationaler und internationaler Studien zu extrapolieren (siehe oben). Bei den internationalen Studien zeigte sich konsistent und ganz eindeutig, dass extrinsische Anreize, u. a. auch echte Benotung [Test ersetzt eine Mathematikklassenarbeit], die Testergebnisse keineswegs verbesserten und auch die Testmotivation meistens gar nicht und wenn, dann deutlich geringer beeinflussten als bei üblichen Examen. "Tendenziell erbrachten Schülerinnen und Schüler, die glaubten, an der PISA-Studie teilzunehmen, sogar bessere Leistungen als diejenigen, deren Testergebnisse benotet werden sollten." [Kurzfassung \(9.1.2002\)](#). Hierfür ist möglicherweise die Standardinstruktion bei derartigen Tests mitverantwortlich. Es gelingt dabei offensichtlich, die Schüler vom Sinn der Untersuchung zu überzeugen und zu hinreichender Anstrengung zu motivieren.

## Empirische Studien zur Wirkung einer angekündigten Benotung auf die Leistung

In dem Beitrag: "Der Einfluss von Tests auf die Lernleistung" werden unter dem Kapitel: [Tests als Motivierungsnahmen zu einem effizienteren Durcharbeiten des Lehrstoffs](#) die Studien von Morrison et al (1995), Tuckman und Mitarbeiter, Burns & Vinchur (1992) und Elkai, & Baker (1988) ausführlich beschrieben. Diese Studien bestätigen weitgehend bessere Leistungsergebnisse bei benoteten Prüfungsarbeiten, weisen aber häufig kein hinreichend



befriedigendes experimentelles Niveau auf. Den Studien lagen meist keine umfassenden Abschlussklausuren zugrunde, sondern Tests bzw. Quiz, die sich auf kleinere Lerneinheiten beziehen. Obgleich in diesen Studien die Leistungsvorteile bei einer Benotung teilweise auch auf die höhere Testmotivation zurückgehen könnten, ist der Hauptanteil des Leistungsunterschiedes meiner Meinung nach dem Lernengagement vor dem Examen anzulasten.

In der Studie von Morrison et al (1995) bearbeiteten die Teilnehmer ein Computerlernprogramm unter verschiedenen Bedingungen. Diejenigen Gruppen, die wussten, ihr Ergebnis gehe in die Seminarnote ein, erzielten deutlich bessere Ergebnisse im unmittelbaren Lernerfolgstest. Die benoteten Studenten arbeiteten offenbar das Tutorium aufmerksamer durch. [Burns & Vinchur \(1992\)](#) wiesen bessere Prüfungsergebnisse bei der Erwartung einer anstehenden Benotung nach. Die Autoren verglichen während eines Seminars in einem Wiederholungsdesign angekündigte benotete mit völlig unerwarteten, unbenoteten Tests und fanden eindeutige Belege für die Lernwirksamkeit der angekündigt benoteten Prüfungen. Dieser Effekt geht vermutlich im Wesentlichen auf die bessere Vorbereitung der angekündigten Prüfungen zurück, da die Quiz aus wenigen MC-Aufgaben bestanden, bei welchen kaum Leistungsunterschiede durch eine höhere Testmotivation zu erwarten sind. Ryan & Hemmes (2005) überprüften in zwei Experimenten die Wirksamkeit benoteter gegenüber fakultativen, unbenoteten Hausaufgaben in einem within-subject design mit etlichen Treatmentreversal-Bedingungen. Die Studenten waren stets darüber informiert, zu welchen Zeitpunkten ihre benoteten oder unbenoteten Hausaufgaben anstanden. Ryan & Hemmes fanden in den jeweils nachfolgenden, den Lehrzielen der Hausaufgaben analogen Kurztests signifikante Leistungsvorteile unter der Bedingung "benotete Hausaufgabe", die im Mittel ca. 10 % ausmachten, was einer ganzen Notenstufe entspricht. Unter der Bedingung "Hausaufgabe mit Benotung" gaben alle Studenten ihre Hausaufgaben stets zur Bewertung ab und verbesserten gegebenenfalls ihre Fehler, unter der Bedingung "unbenotete Hausaufgabe" wurde die Hausaufgabe in den meisten Fällen gar nicht erst abgeliefert, obwohl der Experimentator entsprechende Rückmeldungen versprach und eigens darauf hinwies, die Bearbeitung der Hausaufgabe würde die nachfolgenden Quizleistungen verbessern. Der Notendruck hat hier offenbar bewirkt, die Hausaufgabe zuverlässig zu bearbeiten und so zu einer Verbesserung der nachfolgenden Quizleistungen beigetragen.

Von besonderer Relevanz sind Studien, die auch längerfristige Auswirkungen von benoteten Tests erfassen und so überprüfen können, ob eine Benotung von Leistungen während des Seminars z.B. die Ergebnisse im Abschlussexamen positiv beeinflusst. In der Studie von Elikai, & Baker (1988) bearbeiteten Studenten während des Seminars insgesamt 10 Kurztests. In einem Seminar wurden die Kurztests benotet, im anderen ohne Benotung lediglich bearbeitet. Die Benotung der Kurztests führte nicht nur zu besseren Leistungen in den Kurztests und belegt so die unmittelbare Wirkung der Benotungserwartung auf aktuell anstehende Leistungen, sondern auch in den nachfolgenden Examen. Dieses Ergebnis legt die Vermutung nahe, kontinuierliche Benotung könne auch zu einer höheren Kompetenz bzw. stabileren Lernleistung beitragen. Ein "natürliches Experiment" von Grove & Wasserman (2006) scheint diese These zu belegen. Die Autoren analysierten die Auswirkungen benoteter (Gewicht 15 %) gegenüber unbenoteter Hausaufgaben auf nachfolgende Examensleistungen. Studenten unter der Bedingung "benotete Hausaufgaben" erzielten im Ausmaß zwar geringe, aber signifikant bessere Examensleistungen. Den Ergebnissen von Tuckman und Mitarbeiter zufolge, erbrachten hauptsächlich Personen mit hohem Procrastination ["ist eine Bezeichnung für das Verhalten von Menschen, welche regelmäßig die Erledigung ihnen wichtiger Dinge immer wieder in die Zukunft hinaus verschieben" nach Wikipedia <http://de.wikipedia.org/wiki/Procrastination>] bei häufigerem Notendruck signifikant bessere Leistungen. Studenten mit mittlerem oder geringem Procrastination profitierten



hingegen nicht von den wöchentlich benoteten Tests, da die Leistungen auch ohne ständig benotete Tests am Ende des Seminars vergleichbar gut ausfielen.

Deci und Ryan (2005) referieren kurz eine schwer zugängliche Studie von Kage (1991), die im Widerspruch zu den mir bisher bekannten Ergebnissen steht. Japanische Mittelschüler nahmen an wöchentlichen Quiz im Fach Geschichte teil. In manchen Klassen wurden diese Quiz benotet. In den unbenoteten Klassen bewerteten die Schüler ihr Quiz selbst, was möglicherweise als zusätzliche Lerngelegenheit interpretierbar wäre. Die unbenoteten Klassen erzielten im Abschlussexamen bessere Leistungen.

### **Die umfangreiche Studie von Morgan (1978) zur Lernwirksamkeit echter Abschlussklausuren**

An der Untersuchung von Morgan (1978) waren insgesamt 23 Kurse beteiligt. Diese wurden in einem Zeitraum von 3 Jahren vom selben Dozenten unterrichtet. Im Verlauf eines Kurses absolvierten die Studenten bereits mehrere Tests (teils Mastery-Tests). Das Ergebnis dieser Tests bildet zusammenfassend eine Pre-final-examination-note, im Folgenden auch Vornote genannt, welche einen Großteil der Seminarendnote bestimmt. Die 23 Kurse wurden nach Zufall auf 2 Bedingungen aufgeteilt. Obgleich dieses Verfahren keiner klassischen Randomisierung der Vpn zu den Bedingungen entspricht, erlaubt es wegen der relativ hohen Anzahl der Kurse eine recht wirkungsvolle Kontrolle im natürlichen Umfeld und schließt somit viele Störfaktoren, z.B. Selbstselektion der Probanden, aus. Die experimentelle Variation bezieht sich auf die Bedingung der Abschlussklausur (final examination), welche den Studenten zu Beginn des Seminars angekündigt wurde.

- **EG: course grade contingent condition:** Die Note der Abschlussklausur geht in jedem Fall zu 33% in die Seminarendnote ein.
- **KG: non course grade contingent condition:** Es wird eine Abschlussklausur geschrieben, deren Ergebnis aber nur dann (ebenfalls zu 33%) gewertet wird, wenn es besser als die Pre-final-examination-note ausfällt. Schwache Abschlussklausurleistungen haben somit keine Auswirkungen auf die Seminarendnote.

Leider teilt der Autor keine Abschlussklausurmittelwerte für beide Bedingungen mit. Statt dessen stellt er in seiner Tabelle 1 Häufigkeiten dar, welche Veränderungen für etliche Subgruppen zeigen und die von mir wie in Tabelle 6 aufgeführt, zusammengefasst wurden.

**Tabelle 6: Anzahl sowie Prozentsatz der Studenten für die Abschlussexamensergebnisse im Vergleich zu ihren Vornoten nach Morgan (1978, Basis Tabelle 1)**

| Leistung in Abschlussklausur<br>gegenüber Vornote | Anzahl |     | Prozentsatz |    |
|---|--------|-----|-------------|----|
|   | EG     | KG  | EG          | KG |
| verbessert  | 113    | 0   | 43          | 0  |
| gleich  | 131    | 94  | 49          | 47 |
| verschlechtert                                    | 21     | 106 | 8           | 53 |

EG = die Abschlussexamensnote geht zu 33 Prozent in die Kursendnote ein:

KG = das Abschlussexamen wird bei schwachen Leistungen nicht gewertet,  
nur wenn der Student besser abschneidet als den bisherigen Leistungen entspricht.

Die Tabelle 6 vermittelt in erschreckender Deutlichkeit die Leistung steigernde Wirkung der Erwartung einer zwingenden Benotung. Diese Interpretation wäre völlig schlüssig, wenn die Vornoten beider Gruppen vergleichbar ausgefallen wären, was zwar vermutet werden kann, aber leider unbekannt bleibt. Die Studenten der KG sahen möglicherweise keine Veranlassung, sich vor bzw. in der Abschlussklausur noch richtig ins Zeug zu legen, vermutlich auch

deshalb, weil die Vornoten in vielen Fällen - teilweise bedingt durch Mastery-Tests - bereits recht gut ausfielen. Vielleicht deuten auch diese Ergebnisse auf die Bedeutung der Bedrohung als wirksame motivierende Funktion von Prüfungen hin. Eine drohende Notenabwertung motiviert offenbar mehr als eine mögliche Notenverbesserung. Bei hoher Bedeutsamkeit gilt vielleicht ganz allgemein: "Die Furcht, etwas verlieren zu können treibt mehr an als die Hoffnung, etwas gewinnen zu können". Morgan (1978) ist es jedenfalls gelungen, durch eine zwingend benotete Abschlussklausur die Lernleistungen der Studenten deutlich zu steigern. Er fand zudem keine Hinweise für die unterschiedlichen Bedingungen im Hinblick auf die Bewertung der Kurse und die intrinsische Motivation. Die zu Beginn des Artikels zusammengefasste Essens seiner Arbeit lautet: "The extrinsic manipulation did modify final examination performance without any detriment to intrinsic interests."

Befunde von Murray (1982) weisen in die gleiche Richtung. Den Schülern einer fear elimination group war eine mindestens durchschnittliche Note garantiert worden. Die Leistungen dieser Gruppe fielen im schriftlichen Examen sowie bei den Abschlussnoten signifikant schlechter aus als bei Schülern unter konventioneller Notengebung. Zu milde Benotung (Noteninflation) lässt schwächere Leistungen erwarten. Etliche korrelations- bzw. regressionsanalytische Studien ergaben signifikante Zusammenhänge zwischen der Benotungspraxis des Lehrers oder den Bewertungsstandards von Schulen und den Leistungen der entsprechenden Schüler. Diese Zusammenhänge werden meist in der Weise interpretiert, dass strengere Bewertungen bessere Leistungen -und Leistungsentwicklungen (vornehmlich bei guten Schülern) nach sich ziehen. Obgleich korrelative Versuchspläne in ihrer Wirkungsweise schwer durchschaubar sind und gar keine kausalen Schlussfolgerungen erlauben, deuten die Ergebnisse in ähnliche Richtung wie die experimentellen Befunde.

### Anonyme, unbenotete Probeklausur gegen bewertete Abschlussklausur

Wegen der Kargheit empirischer Studien zur Wirksamkeit der Benotung auf die Lernleistung werden im Folgenden auch solche Untersuchungen herangezogen, die einen Vergleich von unbenoteten Probeklausuren mit echten Klausuren ermöglichen. Die bewertete, aber unbenotete Probeklausur findet in der Regel ein oder zwei Wochen vor der echten notenrelevanten Klausur statt. Nach der Probeklausur ist noch Zeit, sich ordentlich vorzubereiten. Auch wenn diese Studien wegen des Versuchsplans keine zweifelsfreien Schlussfolgerungen zur Wirkung der Benotung auf die Klausurleistung erlauben, so liefern sie Daten mit ziemlich hoher Evidenz für die Bestätigung dieser Hypothese.

Jacobs, Bernd und Fey (2004) überprüften die Wirkung von Probeklausuren auf das Ergebnis des Abschlussexamens. Probeklausur und Abschlussklausur beinhalteten parallele Aufgaben und messen somit dieselben Lehrziele. Die Studenten wurden nach Abiturnotendurchschnitt parallelisiert und dann zufällig auf 2 Bedingungen aufgeteilt. Teilnahme an der Probeklausur (EG) oder Teilnahme an einer alternativen Prüfungsvorbereitung (KG). Die Studenten waren zuvor nicht darüber informiert, dass eine Probeklausur anstand. Die anonymen und somit unbenoteten Probeklausuren fanden eine Woche vor der benoteten Abschlussklausur statt. Die Tabelle 7 verdeutlicht den Versuchsplan.

**Tabelle 7: Versuchsplan der Studie von Jacobs et. al. 2004**

|   |    | Probe<br>Klausur<br>unbenotet | echte<br>Klausur<br>benotet |
|---|----|-------------------------------|-----------------------------|
| R | EG | ○                             | ○                           |
| R | KG |                               | ○                           |

Nachfolgende Ergebnisse erlauben mehrfache Vergleiche zwischen benoteten und unbenoteten Klausuren.

**Tabelle 8: Prozentsatz korrekter Lösungen (Mittelwert und Standardabweichung) in Probeklausur und echter Klausur für die Seminare Statistik und Didaktik nach Jacobs 2003**

|           | Probe<br>klausur<br>unbenotet | echte<br>Klausur<br>benotet         | t-test | df | p     |
|-----------|-------------------------------|-------------------------------------|--------|----|-------|
| Statistik | <b>56.4</b><br>12.2           | 66.0<br>11.7<br><b>60.1</b><br>15.2 | 5.1    | 24 | <.001 |
| Didaktik  | <b>52.2</b><br>9.6            | 70.0<br>14.1<br><b>69.1</b><br>13.2 | 5.5    | 21 | <.001 |

Wie die Tabelle 8 aufzeigt, fallen die Ergebnisse der Probeklausurteilnehmer in der benoteten Klausur deutlich und signifikant höher aus als in ihrer unbenoteten Probeklausur ( $p$  jeweils  $<0.001$ ). Leider basiert dieser Vergleichstests auf dem [vorexperimentellen](#) Vortest-Nachtest-Eingruppenplan Design. Da die Leistungssteigerung zumindest teilweise durch die Bearbeitung der Probeklausur zustande kommen sollte und zum Teil auch gekommen sein könnte, ist ein Vergleich zwischen Probeklausurergebnis und benotetem Abschlussklausurergebnis von Studenten ohne Probeklausur (fett gedruckte Mittelwerte) möglicherweise aussagekräftiger. Dieser Vergleich basiert auf zufälliger Zuteilung der Probanden, wobei lediglich die Messzeitpunkte eine Woche auseinander liegen.

Zumindest in Didaktik ist der Unterschied zwischen unbenoteter und benoteter Klausur unübersehbar. Die Effektstärke des Unterschieds zwischen den fett gedruckten Mittelwerten in Didaktik beträgt  $d = 1.46$ . In Didaktik kann kaum ein Zweifel bestehen, dass eine benotete Klausur bessere Lernergebnisse bewirkte als eine nicht benotete Klausur, was höchstwahrscheinlich den Lehrzielen des Seminars anzulasten ist, die im Gegensatz zur Statistik einfacher bzw. in kürzerer Vorbereitungszeit zu erreichen sind.

### Practice-Tests und Examen bei Sly

Umfangreiche Studien von Sly (2000) zur Wirkung von Practice-Tests weisen alle in die gleiche Richtung. Sly (2000) unternahm insgesamt 9 Untersuchungen zur Wirkung von Practice-Tests auf die Leistung im Examen. Sie verwandte hierbei allerdings keinen experimentellen Versuchsplan. Die Teilnahme an einem unbenoteten Practice-Test (Probeklausur) war meistens freiwillig. Der Versuchsplan kann am ehesten wie folgt formalisiert werden:

**Tabelle 9: Versuchsplanformalisierung der Studien von Sly (2000)**

|   |    | Practice<br>unbenotet | Examen<br>benotet |
|---|----|-----------------------|-------------------|
| N | EG | ○                     | ○                 |
| N | KG |                       | ○                 |

Sly (2000) findet in allen ihren Studien deutliche höhere Leistungsergebnisse im realen Test als im jeweils vorausgehenden Practice-Test. Die in Tabelle 4.16 aufgelisteten 9 Effektstärken dieses Unterschieds schwanken zwischen 0.45 und 2.06. Die Teilnehmer am nicht bewerteten Practice-Test (EG) verbesserten demnach ihre Examensleistung in praktisch bedeutsamem Ausmaß (ungewichteter Durchschnitt  $d = 1,14$ ). Da dieser Vergleich aber genauso wie bei Jacobs et al. (2003) auf einem Vortest-Nachtest-Eingruppenplan basiert, kann der Leistungsgewinn nicht einfach der Testbearbeitung des Practice-Tests oder der Benotung zugeschrieben werden.

Denn auch die Kontrollgruppen, welche keine Practice-Tests durchführten, erzielten im echten Examen stets höhere Leistungen als die Experimentalgruppen im ihrem Practice-Test. Dieser Unterschied geht vermutlich hauptsächlich zu Lasten der Benotung, wird aber wegen der Selbstselektion der Probanden möglicherweise noch unterschätzt, da die besonders motivierten und möglicherweise etwas leistungsfähigeren Studenten höchstwahrscheinlich das Angebot einer Probeklausur eher in Anspruch genommen haben. Wie eine eigene Berechnung auf der Basis der Tabellen bei Sly (2000) ergab, erzielte die Kontrollgruppe im Examen im ungewichteten Durchschnitt über alle 9 Studien 10,6 Prozent höhere Leistungen als die EG im Practice test. Die Effektstärken dieses Vergleichs schwanken über die 9 Studien zwischen  $d = 0,1$  bis  $d = 1,87$ . Der ungewichtete durchschnittliche Effektstärkemittelwert beträgt  $d = 0,74$ . Tabelle 10 fasst die wesentlichen Vergleiche zusammen.

**Tabelle: 10: Wichtige Effektstärken aus den Studien von Sly (2000) und vermutete Ursachen**

| Mittelwertsvergleich         | Effektstärke $d$ |               | Vermutete Ursachen                        |
|------------------------------|------------------|---------------|---|
|                              | Mittel           | Spannweite    |   |
| EG Examen - EG Practicetest: | 1.14:            | (0,45 - 2.06) | Practicetest+Benotung                     |
| KG Examen - EG Practicetest: | 0.74:            | (0,10 - 1,87) | Benotung, konfundiert mit Selbstselektion |

**Fazit:** Alle Ergebnisse der Studien zur Wirkung von Probeklausuren auf das Ergebnis im Examen liegen in einer Richtung und sind mit der Hypothese "Benotete Klausuren führen zu besseren Klausurleistungen als unbenotete Klausuren" vereinbar, obgleich sie alternative Erklärungen nicht völlig ausschließen. Das Ausmaß des Effektes ist recht massiv und ohne Zweifel von hoher praktischer Bedeutsamkeit. Eine eindeutige quantitative Abschätzung des Effektes erweist sich wegen der schwachen Versuchsplan als schwierig. Höchstwahrscheinlich wurde die Wirkung der Benotung sogar noch deutlich unterschätzt, da während des Seminars ja bekannt war, dass langfristig eine Benotung ansteht und ein Teil der Vorbereitung, die vermutlich ohne anstehende echte Klausur gar nicht in Angriff genommen worden wäre, schon vor der Probeklausur stattfand. Der Leistungseffekt durch Benotung hängt vermutlich von weiteren Faktoren ab, z.B. den Lernzielen, der Schwierigkeit, der Vorbereitbarkeitseffizienz auf die Klausur. Diese Studien sagen aber selbstverständlich gar nichts über den langfristigen Behaltenserfolg benoteter Klausuren aus.

## Vorschlag eines Versuchsplans zur Überprüfung der Lernwirkung einer benoteten Klausur.

Im folgenden wird ein hypothetischer Vorschlag entwickelt, die Lernwirksamkeit einer benoteten Klausurankündigung zu testen, um besser einschätzen zu können, ob bzw. in welchem Ausmaß der Notenanreiz letztlich wesentlich das gesamte Lernengagement des Studenten während des Semesters bestimmt. Hierzu zählen die Mitarbeit während des Unterrichts, die Nachbereitung zu Hause sowie die gesamte Prüfungsvorbereitung.

Die Studenten einer hinreichend gut besuchten Vorlesung ( $N = 500$ ) werden nach Zufall auf 5 Gruppen zu je 100 aufgeteilt. Im Verlauf der Vorlesung bearbeiten manche Probanden einen Test (Zwischenklausur), der entweder rechtzeitig angekündigt wird oder gar nicht und zum Teil mit einem Gewicht von ca. 20 Prozent in die Seminarendnote eingeht oder auch überhaupt nicht. Am Ende der Vorlesung bearbeiten alle Studenten eine Abschlussklausur, die unter anderem parallele Items der Zwischenklausur beinhaltet, um auch die langfristigen Auswirkungen einer Testung mit sowie ohne Benotung zu prüfen. Tabelle 11 verdeutlicht die näheren Bedingungen.

**Tabelle 11: vereinfachte Versuchsplanformalisierung zur Testung der Lernwirksamkeit benoteter Klausuren.**

|   | Test     | Abschluss-<br>klausur |
|---|----------|-----------------------|
| A | -        | <b>B</b>              |
| B | -        | <i>N</i>              |
| C | <b>B</b> | <b>B</b>              |
| D | <b>N</b> | <b>B</b>              |
| E | <i>N</i> | <b>B</b>              |

---

B = Leistung wird benotet: N = Leistung wird nicht benotet:

**Fettdruck= Test bzw. Klausur angekündigt; kursiv= Test bzw.**

*Klausur nicht angekündigt*

Gruppe A weiß zu Beginn der Vorlesung, dass am Ende eine Abschlussklausur ansteht. Gruppe B erhält zu Beginn der Vorlesung die Information, eine Abschlussklausur fände für sie gar nicht statt. Im eigenen Interesse sollten sich die Studenten um ein hinreichendes Verständnis bemühen. Am letzten Semestertermin wird Gruppe B aber völlig unerwartet dazu aufgefordert, anonym an der Abschlussklausur teilzunehmen, wobei noch zu überlegen bleibt, welche Anreize Gruppe B zu einer hinreichenden Testmotivation bewegen sollten. Durch den Vergleich der Gruppen A und B lässt sich die durch vorausgehendes Lernengagement bedingte Auswirkung der Benotung auf klassische Weise testen. Die Hinzuziehung der Gruppen C, D, E erlaubt bei der Analyse der Abschlussklausur die Überprüfung der stabilisierenden Funktion vorausgehender Testung (mit oder ohne Benotung). Ein Vergleich der Gruppen C bis E beim ersten Test gibt Auskunft über die unmittelbare Wirkung eines benoteten bzw. angekündigt unbenoteten Tests auf die Testleistung sowie den Einfluss der vorherigen Information über diese Testung. Der potentielle Unterschied zwischen D und E beim Test lässt etwa Rückschlüsse darauf zu, ob das angekündigte Angebot einer unbenoteten Testung eine Wirkung auf die Vorbereitung ausübt, was wesentliche Erkenntnisse zur Lernwirksamkeit entsprechender E-learning-Angebote beisteuern würde. Besonders interessant wären Unterschiede der Gruppen C bis E in der Abschlussklausur.

Ich erwarte angesichts der Trivialität der Fragestellung sehr deutliche Leistungsvorteile für angekündigt benotete Klausuren. Deutlich anspruchsvoller wäre freilich die Konzipierung

einer Kontrollbedingung, welche ohne Notenandrohung ein ernsthaftes Durcharbeiten des Lehrstoffs anregen würde.

## Gesamtresumé

Die Analyse der empirischen Befunde lässt keinen Zweifel daran aufkommen, Geld und Benotung könnten die aktuelle Leistung steigern. Allerdings müssen dabei bestimmte wichtige Voraussetzungen und Bedingungen erfüllt sein. Zum einen setzt die Fähigkeit der Probanden, die gesetzten Lehrziele bzw. Kriterien überhaupt erreichen zu können, definitive Grenzen jedweder Leistungssteigerung durch höhere Lernmotivierung. Deshalb werden extrinsische Anreize wie etwa Geld die Leistungen bei vielen anspruchsvollen schulischen Lehrzielen keineswegs automatisch verbessern. Zum andern muss das Anreizsystem optimal auf das Motivationsgefüge passen, um hinreichende Motivation zu beflügeln. Dabei stehen die extrinsischen Anreize häufig in Konkurrenz zu sonstigen Motivatoren, die zum einen eine hinreichend hohe Lernmotivation bewirken und keiner Übermotivierung bedürfen, zum andern aber mehr Erfolg oder Wert versprechen als die von außen gesetzten extrinsischen Anreize.

Das extrinsische Motivationsmodell passt am besten auf ein Menschenbild, was davon ausgeht, der Geist würde sich nur dann in Bewegung setzen, wenn Geld fließe oder der Notendruck greife. Das ist natürlich eine ziemlich absurde Vorstellung, auch wenn man in der Praxis Sonderfälle finden mag, die an dieses Bild heranreichen. Dennoch mobilisieren gut organisierte und kontrollierte extrinsische Anreize bei etlichen Schülern und Studenten zusätzliche Motivationsreserven. (siehe etwa: [Wann erhöhen Belohnungen die Lernleistung ?](#) ). Insbesondere Notendruck trägt mit dazu bei, erhebliche Arbeitsmotivation auf die erwünschten prüfungsrelevanten Lerninhalte zu konzentrieren. Diese sollten anspruchsvolle Lehrziele beinhalten, um prüfungstaktische Ausweichmanöver bzw. unproduktives Lernen zu unterbinden, dürfen andererseits aber auch nicht zu hohe Kompetenzen einfordern, da diese durch erhöhten Lerneinsatz im Mittel nicht erreichbar sind. Weil Notendruck keine Personen abhängige, freie Gestaltung optimaler Motivationsbedingungen zulässt, verfehlt er bei etlichen Betroffenen seine erhoffte Wirkung höherer Leistungsbereitschaft und fördert stattdessen eher Selbstzweifel oder Resignation. Da viele Schüler keine guten Noten erwarten können, wirkt die Ankündigung einer benoteten Prüfung weniger herausfordernd, sondern ziemlich stressig in Form einer aversiven Motivierung, die anstehende Bedrohung zu vermeiden. Potenzielle Vorteile leistungssteigernder Maßnahmen sollten nicht den Blick auf die möglichen unerwünschten Nebenwirkungen verstellen! Manche Problemschüler haben gegenüber Noten bereits eine Gleichgültigkeit entwickelt, die den Anreizwert der Noten auf Null reduziert. Bei vielen gewissenhaft arbeitenden Studenten ist Notendruck schlichtweg unnötig und verhindert eher eine verstärkte Beschäftigung mit innovativen, anspruchsvolleren geistigen Tätigkeiten. Für etliche Prüfungszwecke könnte vielleicht ein Bonussystem hinreichend motivieren. Hierbei erwirbt der Student in einem Test leistungsabhängig relative marginale Punkte, die dann den Punkten einer echten Klausur hinzugefügt werden können. (siehe [Sansgiry et al. \(2006\)](#)) So wird der Fokus nicht so sehr auf eine potentielle Bedrohung, sondern eher auf mögliche Chance verlagert.

Extrinsische Lernanreize für bestimmte Prüfungen erhöhen höchstwahrscheinlich die Vorbereitungsintensität für diese Prüfungen, sorgen möglicherweise dafür, dass bei einigen überhaupt ein ernsthaftes Lernengagement zustande kommt, könnten andererseits aber auch zwanglos investierte Lernzeit für weniger Anreiz orientierte Lerngebiete senken. Durchaus sinnvolle, anspruchsvolle Lernangebote werden aus prüfungstaktischen Erwägungen nicht wahrgenommen. Interessante Seminare ohne Noten relevante Scheine lässt man notgedrungen schleifen. Bei freier Auswahl entscheidet man sich vornehmlich für solche Kurse, die mit re-

lativ wenig Aufwand gute Noten erwarten lassen. Der Student lernt, rein Interesse orientiertes Lernen notwendigen Sachzwängen angemessen unterzuordnen und sich den formalen gesellschaftlichen Anforderungen möglichst Kräfte schonend anzupassen. Das ist in gewisser Weise und in gewissem Ausmaß durchaus vernünftig und bereitet den Betroffenen auf die Anforderungen der Arbeitswelt vor. Die eigentliche Kunst liegt denn auch darin, dem, was verlangt wird, ein gewisses Interesse abzurufen.

Die höhere, eher aversive emotionale Aktivierung vor bzw. während einer notenrelevanten Testung müsste die aktuelle Leistung der Hochhängstlichen etwas beeinträchtigen. Jedoch dürften auch Prüfungsängstliche insgesamt in rechtzeitig angekündigten, benoteten Prüfungen höhere Leistungen erzielen als in unbenoteten Prüfungen. Lernen unter Notendruck mag mit dazu beitragen, intrinsische Motivation zu schwächen. Aber ein Student, der die gesamte schulische Sozialisation mit allen benoteten Leistungsprüfungen hinter sich gebracht hat, wird am Ende seines Studiums nicht jedes Interesse an seinem Studienfach verloren haben. Die in der Zwischenzeit gewachsene Kompetenz könnte das Interesse später unter günstigeren Umweltbedingungen vielleicht sogar fördern.

Die vielfältigen Forschungsbemühungen zu den Auswirkungen extrinsischer Anreize auf die intrinsische Motivation beschränken sich in der Regel auf Laborstudien und verwenden Aufgaben, die man meist mit Interesse auch ohne extrinsische Anreize hinreichend bearbeitet. Die dort gefundenen Effekte sind im Übrigen ziemlich gering. (z.B. Cameron et al. 2001). Selbst wenn das intrinsische Interesse sinken sollte, so lässt sich aus Ergebnissen solcher Studien allerdings nicht einfach ableiten, logische Folge dieser gesunkenen intrinsischen Motivation sei eine eher negative Leistungsprognose. Denn im Schulsystem werden ja kontinuierlich extrinsische Anreize gesetzt und es gibt wenig Anlass anzunehmen, bei geringerem Interesse würden diese ihre Wirkung einbüßen, die aktuelle Lernleistung zu steigern.

In unserem Bildungssystem sind Noten fest verankert. Ein Vergleich zum selben System einer Testung ohne Benotung ist somit unmöglich. Jedoch weisen Schulleistungsvergleiche darauf hin, dass auch ohne Noten hohe Kompetenzen erzielt werden können, etwa in skandinavischen Ländern. Die wenigen Experimente zur aktuellen Leistungssteigerung durch Benotung erfassen meist nur einen kleinen Ausschnitt der aktuellen Wirkung, ziehen mögliche ernsthafte Alternativen nicht ins Kalkül und erlauben keine Aussagen zu den langfristigen Folgen. Ungeklärt bleibt etwa die Frage, ob die gesamte effektiv genutzte Lernzeit durch die Benotung zunimmt oder ob sich die Lernzeit spätestens ab einer bestimmten Belastungshöhe lediglich anders verteilt, nämlich zu Gunsten fremdbestimmten, notenrelevanten Lehrstoffs. Ebenso schwer einschätzbar ist, ob und wie sich die durch extrinsische Faktoren erreichbare, höhere aktuelle Leistungsperformanz langfristig auch in verbesserter Kompetenz niederschlägt.

Mehr benotete Tests im Schulsystem lassen im Mittel eine leicht verbesserte Leistung erwarten, weil damit einige Schüler zu einem kontinuierlicheren Arbeiten angeregt werden könnten. Mehr benotete Tests werden sogar von Schülern präferiert, vermutlich, weil diese das Gewicht der Testung für die Abschlussnote senken und den Lernumfang pro Prüfung reduzieren. Sie kosten aber auch mehr Geld. Ich sehe für die meisten Studenten keine wirksame Leistungsverbesserung, wenn über den extrinsischen Anreiz von Noten hinaus nun noch leistungsabhängige Stipendien hinzukommen würden. Leistungsabhängige Stipendien könnten sich für solche leistungsfähige Schüler und Studenten auszahlen, die ohne entsprechendes Einkommen nur wenig Chancen haben, sich hinreichend auf ein Studium zu konzentrieren, wie durch Einführung von Studiengebühren in Deutschland zunehmend zu erwarten ist, aber in weit stärkerem Ausmaß in etlichen armen Ländern dieser Erde der Fall ist. Noch sinnvoller wäre es freilich, die finanziellen Ressourcen verstärkt in Bildung zu investieren, um möglichst



allen an Bildung wirklich Interessierten die Chance zu geben, sich innerhalb ihrer Lernmöglichkeiten voll entfalten zu können.

Die offenbar simple Implementation und der schnell erhoffte, aber keineswegs sichere Erfolg extrinsischer Anreize darf selbstverständlich nicht den Blick darauf verstellen, durch geeignete Unterrichtssituationen bzw. Lernmaterialien eine am Interesse orientierte Lernmotivierung anzubieten sowie weitere subtilere Motivationsquellen zu nutzen. **Auch wenn extrinsische Anreize wie Geld oder Noten gegenüber "gar nichts" wirken, so folgt daraus eben nicht, andere Motivatoren könnten keine echten oder die besseren Alternativen sowie darüberhinaus unverzichtbare Elemente guten Unterrichts sein.** Es gehört schließlich zu den ganz besonderen Herausforderungen der Pädagogik, Lernen ohne Rückgriff auf diese trivialen Lernanreize wirksam und zuverlässig anzuregen. Die besten Beweise dafür würden freilich schlüssige Experimente eines entsprechenden Leistungsnachweises unter realen schulischen Bedingungen liefern. Diese aber sind sehr schwer zu erbringen, wenn sie gegen eine Übermacht von benoteten Leistungsnachweisen antreten müssen. Eine mögliche Frage kann daher nicht etwa ernsthaft lauten: "Führt Weckung von Neugier zu mehr Lernerfolg als die Androhung einer Benotung?", sondern eher "Macht Lernen aus Neugier mehr Freude. Gelingt es dadurch, Lernen wirksam anzuregen und das rein Anreiz orientierte Lernen etwas in den Hintergrund zu drängen? Letztlich stellt sich für engagierte Lehrer das Problem, den pädagogischen Spielraum möglichst gut auszunutzen, den Lerner zu befähigen sowie im positiven Sinne zu motivieren, unter erträglichen Bedingungen den geforderten Leistungsnachweisen mit extrinsischen Konsequenzen besser nachkommen zu können.

---

## Fußnoten

<sup>1</sup>Van de Weghe, P. & Bruggeman (2006) gelang der Nachweis, unter objektiv vergleichbaren Gewinnchancen bewirke ein Penaltysystem mehr Leistung in einer Codier- und Sortieraufgabe als ein Bonussystem. Unter einem Penaltysystem versprach man den Probanden standardmäßig 12 Euro. Sie würden jedoch 6 dieser 12 Euro verlieren, wenn sie das gesetzte Ziel nicht erreichten. Probanden unter Bonussystem erwarteten standardmäßig 6 Euro, konnten aber bei Zielerreichung weitere 6 Euro hinzu gewinnen.

---

## 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%2520ECO/BE%2520Fall%25202005/Ariely.pdf> [27.6.2006]
- Bailey, C. D.; Brown, L. D.; Cocco, A. F. (1998). The Effects of Monetary Incentives on Worker Learning and Performance in an Assembly Task. Journal of Management Accounting Research, Vol. 10, p119, 13p
- Bauer, R. H.; Peller-Porth, V. (1990). The effect of increased incentive on free recall by learning-disabled and nondisabled children. Journal of General Psychology, Vol 117(4), pp. 447-461.
- 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.
- Bonner, S. E., R. Hastie, G. B. Sprinkle and S. M. Young (2000). A Review of the Effects of Financial Incentives on Performance in Laboratory Tasks: Implications for Management Accounting. Journal of Management Accounting Research, 19-64.
- Camerer, C. & Hogarth, R. (1999) The effects of financial incentives in experiments: a review and capital-labor-production framework. Journal of Risk and Uncertainty 19, 7-42.

- Cameron, J., Banko, K. M. & Pierce, W. D. (2001). Pervasive Negative Effects of Rewards on Intrinsic Motivation: The Myth Continues. *The Behavior Analyst*. 24, 1–44.
- Condly, S.J., Clark, R.E. & Stolovitch, H.D. (1993). The effects of incentives on workplace performance: A meta-analytic review of research studies. *Performance Improvement Quarterly*, 16(3) pp. 46-63
- Cullen J. F., Cullen, J. B., Hayhow, Van L. & Plouffe, J. T. (1975). The Effects of the Use of Grades as an Incentive. *Journal of Educational Research*; Vol. 68, Issue 7.
- Deci, E.L. & Ryan, R.M. (2005) Test-Based Accountability, Incentives, and Motivation for Learning. [http://www7.nationalacademies.org/bota/10Bota\\_Incentives\\_Papers%20and%20PPTs.pdf](http://www7.nationalacademies.org/bota/10Bota_Incentives_Papers%20and%20PPTs.pdf) [29.6.2006]
- Eisenberger, R. & Armeli, S. (1997). Can salient reward increase creative performance without reducing intrinsic creative interest. *Journal of Personality and Social Psychology*, 72,3,652-663.
- Eisenberger, R. & Armeli, S. & Pretz, J. (1998). Can the promise of reward increase creativity. *Journal of Personality and Social Psychology*, 1998, 74,3, 704-714.
- Goette, L. & Lienhard, M. (2006). Productivity and Well-being under Fixed Wages and Piece Rates. University of Zurich. Preliminary Draft. February 14, 2006  
[www.isu.unizh.ch/emap/docs/2006/poek/LienhardGoette.pdf](http://www.isu.unizh.ch/emap/docs/2006/poek/LienhardGoette.pdf) [6.7.2006]
- Goetz, C.J., Tucker, J.F. & Weber, W.E. (1973). STUDENT INCENTIVE PAYMENTS AND ACADEMIC ACHIEVEMENT: AN EMPIRICAL TEST. *Social Science Quarterly* (Southwestern Social Sciences Association Vol. 54 Issue 1, p159-167. - Nur das Abstract der Studie war mir verfügbar.
- Gneezy, U & Rustichini (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, Vol. 115, Issue 3, 791-810.
- Grove, W. A.; Wasserman, T. (2006). Incentives and Student Learning: A Natural Experiment with Economics Problem Sets. *American Economic Review*, Vol. 96, Issue 2, p447-452.
- Huber, V. L. (1985) Comparison of monetary reinforcers and goal setting as learning incentives. *Psychological Reports*, Vol 56(1). pp. 223-235.
- Jacobs, B. (2003) Die Wirkung einer Probeklausur auf Lernleistung und aktuelle Angst in einer echten Klausur. URL: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/probeklausur/index.htm>
- Jacobs, B., Bernd, H., Fey, A. (2004). [Die Wirkung einer Probeklausur auf Klausurleistung und Angst in einer Statistiklausur.](http://www.phil.uni-sb.de/~jakobs/wwwartikel/probeklausur/index.htm)  
URN: urn:nbn:de:bsz:291-psydok-2720  
URL: <http://psydok.sulb.uni-saarland.de/volltexte/2004/272/>
- Jacobs, B. (2005). Motivationales Feedback und Lernleistung. URL: <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/motivation.htm> [24.8.2006]
- Kage, M. (1991, September). The effects of evaluation on intrinsic motivation. Paper presented at the Japan Association of Educational Psychology, Joetsu, Japan. zitiert nach Deci & Ryan (2005.)
- Kieffer, D. A.; Goh, D S. (1981) The effect of individually contracted incentives on intelligence test performance of middle- and low-SES children. *Journal of Clinical Psychology*, Vol 37(1), Jan 1981. pp. 175-179.
- Kiplinger, V. L., & Linn, R. L. (1992, April). Raising the stakes of test administration: The impact on student performance on NAEP. Paper presented at the annual meeting of the American Educational Research Association, Los Angeles. (ERIC Document Reproduction Service No. ED378221). zitiert nach Wise & DeMars (2005).
- Leuven, E. Oosterbeek, H. van der Klaauw, B. (2003) The effect of financial rewards on students' achievement: Evidence from a randomized experiment. No 3921, CEPR Discussion Papers from C.E.P.R. Discussion Papers  
<http://ecodip.net/pdf/bonus2.pdf> [27.4.2006]
- Libby, R., & Lipe, M.G. (1992). "Incentives, Effort, and the Cognitive Processes Involved in Accounting-Related Judgments," *Journal of Accounting Research* 30, 249-273.
- Manthei, K. (2004). Pay - But Don't Pay Too Much: An Experimental Study on the Impact of Incentives, Department of Personnel Economics, University of Cologne, DP 04-27 K.  
<http://www.whu.edu/orga/geaba/Papers/2004/Manthei.pdf> [7.7.2006]
- Marsh, H.W. (1984). Experimental manipulations of university student motivation and their effects on examination performance. *British Journal of educational Psychology*. 54,206-213.
- Morgan, R. R. (1978) Utilization of Extrinsically Manipulated Grade Contingencies as Motivators of School Learning. *Teaching of Psychology*, Vol. 5, Issue 4, p191, 3p.
- Murray, J. F. (1982). The elimination of fear of failure as an incentive for learning.

- International Journal of Sport Psychology, Vol 13(1), pp. 13-20.  
- Artikel war mir nur im Abstract zugänglich -
- O'Neil, H. F.; Sugrue, B.; Baker, E. L. (1995/1996). Effects of Motivational Interventions on the National Assessment of Educational Progress Mathematics Performance; Educational Assessment; Vol. 3 Issue 2, p135-157, 24p
- O'Neil, H. F., Abedi, J., Lee, Ch., Miyoshi, J. and Mastergeorge, A. (2001). Monetary Incentives for Low-Stakes Tests. Education Statistics Quarterly, Vol 3, Issue 2, URL: [http://nces.ed.gov/programs/quarterly/vol\\_3/3\\_2/q6-1.asp](http://nces.ed.gov/programs/quarterly/vol_3/3_2/q6-1.asp) [5.4.2006]
- O'Neil, H. F., Abedi, J., Lee, Ch., Miyoshi, J. and Mastergeorge, A. (2004). Monetary Incentives for Low-Stakes Tests. CSE Report 625 [www.cse.ucla.edu/reports/R625.pdf](http://www.cse.ucla.edu/reports/R625.pdf) [5.4.2006]
- O'Neil, H. F., Abedi, J., Miyoshi, J. and Mastergeorge, A. (2005). Monetary Incentives for Low-Stakes Tests. EDUCATIONAL ASSESSMENT, 10(3), 185–208
- Reid, W. J. & Bailey-Dempsey, C. (1995). The Effects of Monetary Incentives on School Performance. Families in Society; Jun95, Vol. 76, Issue 6, p331-340.
- Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? Psychologische Rundschau, 57(2), 69-86
- Ryan, C. S. & Hemmes, N. S. (2006). Effects Of The Contingency For Homework Submission On Homework Submission And Quiz Performance In A College Course. Journal of Applied Behavior Analysis. 38(1): 79–88.
- Spencer, M. B., Noll, E. & Cassidy, E. (2005). Monetary Incentives in Support of Academic Achievement. Results of a Randomized Field Trial Involving High-Achieving, Low-Resource, Ethnically Diverse Urban Adolescents. Evaluation Review, Vol. 29, No. 3, 199-222
- 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
- Sly, J. L. (2000). Computer Managed Learning Assessment in Higher Education: The Effect of a Practice Test. PhD Thesis. Curtin University of Technology. <http://adt.curtin.edu.au/theses/available/adt-WCU20030115.113425/> [30.5.2003]
- Sprinkle, G. B. (2000). The Effect of Incentive Contracts on Learning and Performance. The Accounting Review. Vol. 75, Issue 3, 299-326.
- Sansgiry, S.S., Chanda, S., Lemke, Th. & Szilagyi, J.E (2006). Effect of Incentives on Student Performance on Milemarker Examinations. American Journal of Pharmaceutical Education 2006; 70 (5) Article 103. <http://www.ajpe.org/view.asp?art=aj7005103&pdf=yes>
- Wise, S. L. & DeMars, C.E. (2005). Low Examinee Effort in Low-Stakes Assessment: Problems and Potential Solutions. EDUCATIONAL ASSESSMENT, 10(1), 1–17.
- Wolf, L.F. & Smith J. K. (1995) The consequence of consequence: Motivation, anxiety, and test performance, APPLIED MEASUREMENT IN EDUCATION, 8(3), 227-242
- Wolf, L. F., Smith, J. K., & DiPaolo, T. (1996, April). The effects of test specific motivation and anxiety on test performance. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.  
zitiert nach Wise & DeMars (2005)
- Van de Weghe, P. & Bruggeman, W. (2006). The Impact of the Number of Performance Measures and Incentive Framing on Performance in a Multidimensional Task Environment Working Paper. Universiteit Gent. [http://www.feb.ugent.be/fac/research/WP/Papers/wp\\_06\\_404.pdf](http://www.feb.ugent.be/fac/research/WP/Papers/wp_06_404.pdf) [27.11.2006]