

## Tests als Maßnahmen zur Förderung von Studierleistungen ?

**Abstract:** Während die Laborforschung die Lernwirksamkeit von Testen und Feedback hinreichend belegt hat, erweist sich eine direkte Umsetzung dieser Erkenntnis in die Schulpraxis insofern als schwierig, als dort etliche Varianten der Testung und ihrer Wirksamkeitsprüfungen zur Anwendung kommen, die nur sehr bedingt der Laborsituation entsprechen. Ziel der Arbeit war es, die Forschungsliteratur zu Testen und Feedback im Hinblick auf einige praktische Anwendungen im schulischen Umfeld zu durchleuchten. Zunächst werden Bedingungen aufgelistet, welche den Lern- oder Behaltenseffekt einer Testung beeinflussen könnten und in diesem Zusammenhang Faktoren wie Testfunktion, Testankündigung, Testkontrolle, Feedbacknutzung, und Testhäufigkeit sowie die Bedeutung der Testkonsequenzen behandelt. Anschließend werden theoretische Erwartungen expliziert, unter welchen Bedingungen Quiz bzw. Tests im Verlauf eines Seminars langfristiges Lernen überhaupt fördern könnten. Hierbei ist ein gewisser Transferanspruch an die Kriteriumsmaße einzufordern. Insgesamt hat sich Testen mit Feedback als eine praktikable lernwirksame Studiermethode erwiesen. Allerdings lassen sich keine Belege für Lernvorteile eines freiwilligen Online-Quizangebotes finden. Mit höheren Testwerten in zeitnahen Prüfungen ist vornehmlich dann zu rechnen, wenn auf der Bearbeitung der Quiz bestanden wird, was vorwiegend durch Bonusversprechen- oder Notenandrohung geschieht. Während Studierende die Verwendung von Quiz in der Regel subjektiv gutheißen und in gewisser Weise auch davon profitieren, erscheinen etliche Zweifel berechtigt, durch den Einsatz von Quiz deutliche Leistungseffekte im Abschlussexamen zu erwarten.

**Schlagerworte:** Testen, Quiz, Üben, Feedback, Notenanzreiz, Lernmotivierung, Lernförderung

### Einleitung

Der Einsatz pädagogischer Tests im Kontext wissenschaftlicher Studien umfasst mehrere Facetten, die hinreichend analysiert und eingeordnet werden müssen, um ihre Wirksamkeit in der Schulpraxis vernünftig abschätzen zu können. Die Wirkung eines Tests auf die Lernförderung hängt mitunter davon ab, ob der Test angekündigt wird, ob und wie er benotet wird, welche Lehrziele dort erfasst werden und welche Arten von Rückmeldungen wann im Anschluss an die Testung folgen. Zudem spielen das Zeitintervall zwischen Instruktionsphase und Test und das vom Test zum Kriterium sowie die näheren Bedingungen des Behaltenstests eine wesentliche Rolle. Insofern kann man eigentlich die Wirkung des Tests, z.B. eines Quiz, auf das schulische Lernergebnis gar nicht global beurteilen. Wenngleich entsprechende Informationen zur Interpretation bzw. dem Verständnis der Befunde häufig unerlässlich sind, werden diese in den Studien selten systematisch dargestellt und etliche wichtige Informationen sucht man vergeblich. Die aus der hoch kontrollierten Laborforschung resultierenden Erkenntnisse zur Wirkung von Tests und Feedback auf die Lernleistung lassen sich nicht bedenkenlos auf pädagogische Umwelten übertragen. Manche theoretisch unerwarteten Befunde der angewandten Forschung, etwa "Die Bearbeitung von Tests während eines Seminars hat keine Auswirkung auf das Ergebnis der Abschlussklausur", widerlegen nicht unbedingt die positive Wirkung des Testens, sondern basieren auf Faktoren, welche das Entdecken des erwarteten Unterschieds erschweren.

Im Folgenden werden die Begriffe Quiz und Test meist in synonyme Weise verwendet. Durch ein Quiz oder einen Test, bzw. durch mehrere Tests soll vornehmlich die Leistung in einem späteren Kriteriumstest gesteigert werden. Als Kriteriumstests fungieren meist

- Unittests: Der zu prüfende Lehrstoff bezieht sich auf einen thematischen Schwerpunkt und umfasst meist einige Kapitel in einem Lehrbuch,
- Midtermtests: eine Art Zwischenklausur, in der ca. die erste Hälfte des Seminarlehrstoffs überprüft wird, oder
- Abschlussexamen: Abschlussklausur, welche eine repräsentative Auswahl aus den Lehrzielen des Seminars erfassen soll.

Die verschiedenen Kriterien unterscheiden sich im Hinblick auf den Lehrziel relevanten Stoffumfang sowie den Zeitabstand zum Test/Quiz. Bei der Diskussion um die Lernwirksamkeit des Testens expliziert Tabelle 1 einige zentrale Unterscheidungen mit Ausprägungen etwas genauer. Anschließend werden mögliche Wechselwirkungen der Faktoren angedeutet und darauf hingewiesen, dass Lern- oder Leistungseffekte nur zu erwarten oder nachweisbar sind, wenn bestimmte Bedingungen erfüllt sind. Als empirische Stützen dienen vorwiegend Studien aus der angewandten Forschung im pädagogischen Feld, deren versuchsplantechnisches Niveau oftmals Einiges zu wünschen übrig lässt, was hier aber nicht über Gebühr problematisiert werden soll.

### **Tabelle 1: wichtige Bedingungsfaktoren bei Einschätzung von Leistungseffekten durch Tests**

#### Funktion des Tests

- Test als Vorbereitung auf eine Seminarsitzung, z.B. Quiz über einen zu lesenden Text, der dann im Seminar behandelt werden soll.
- Test als Übungsangebot zur Stabilisierung bzw. Konsolidierung des Wissens
- Test als Prüfung des durchgenommenen Seminarstoffes bzw. zur Erfassung der Lehrziele des Seminars.

#### Ankündigung eines Tests

- nicht angekündigt und völlig unerwartet.
- latent angekündigt bzw. angedroht: Quiz kann stattfinden, muss aber nicht: pop quiz, surprise quiz, random quiz.
- definitiv und rechtzeitig angekündigt: Test oder Prüfung finden statt am..... Zeit für Vorbereitung ist gegeben.

#### Testteilnahme

- verpflichtend: meist bei benoteten Tests
- freiwillig: häufig als optionales Übungsangebot

#### Bewertung des Tests

- unbenotet: Motivierungsansatz meist: "Test verbessert das Verständnis, gibt Leistungsrückmeldung, dient der Prüfungsvorbereitung"
- benotet:
  - a) mit relativ hohem Gewicht: z.B. Klassenarbeit, Klausur.
  - b) mit geringerem Gewicht: z.B. kleine Quiz.
  - c) extra credit - bonus - : Zugabe von leistungsabhängigen Punkten zum normalen Examen.  
Seminarnote = Examenspunktzahl + Bonuspunkte.
  - d) angekündigt und durchgeführt, aber nach Zufall bewertet: Beispiel: Es wird in jeder Sitzung ein Quiz geschrieben, aber nur 3 von 15 Quiz letztlich gewertet.

#### Testablauf

- streng kontrolliert: unter Aufsicht: Klassische Prüfungssituation
- weniger oder gar nicht kontrolliert: etwa Test als Hausaufgabe oder Test via Internet von beliebigen Orten.

### Testrückmeldung

- überhaupt keine: nur Test: reine Wirkung des Testens
- nur Gesamtergebnis: Punktzahl, Note
- Item-spezifischen Feedback (KCR, elaboriertes Feedback)
- unmittelbar nach Item oder Test bzw. verzögert.

### Transparenz der Testung

- sehr durchsichtig: Potenzielle Anforderungen in Quiz und Abschlussexamen gut antizipierbar
- Anforderungen eher intransparent

### Eigenschaften des Kriteriumstests

- Test ist zugleich Kriterium, etwa bei der Fragestellung: Welchen Einfluss hat die Benotung auf das Ergebnis im Test ?
- Transferdistanz des Kriteriums, memory: identische Items in Test und Kriterium (gilt für viele Laborstudien) lehrzielvalide, parallele, aber nicht identische Items in Test und Kriterium (selten)
- Zeitpunkt des Kriteriums: großes oder kleines Retentionsintervall
- Ankündigung und Relevanz des Kriteriums  
 angekündigt und benotet  
 angekündigt, aber nicht benotet  
 völlig unerwartet

Die einzelnen Faktoren sind natürlich nicht unabhängig voneinander, sondern treten meist in bestimmten Kombinationen auf. So sind Tests zur Überprüfung des Lehrstoffs in der Regel verpflichtend. Die prüfungsrelevanten Lehrziele können mehr oder weniger transparent ausfallen. Die Tests selbst werden angekündigt, unter strenger Kontrolle bearbeitet und mit Noten bewertet. Dem stehen etwa optionale Onlinetests gegenüber, die sehr dezidierte Rückmeldungen geben, zu Hause unkontrolliert bearbeitet werden und keinerlei Konsequenzen für die Studierenden nach sich ziehen.

### Reines Testen (mit Feedback) ohne Benotung

Die Studien zur Adjunct-question- Forschung (Hamaker 1986) sowie die meisten psychologischen Laborstudien (z.B. Roediger & Karpicke 2006) überprüften hauptsächlich die Wirkung reinen Testens im Anschluss an eine Instruktion (z.B. Lesen eines Textes) auf ein späteres Behalten. Hierbei werden weder der Übungstest noch der Kriteriumstest, welcher die Wirkung des Übungstests auf das Behalten messen soll, benotet. In der Regel ist zumindest der Kriteriumstest antizipierbar oder wird in irgendeiner Weise angekündigt. Das reine Testen hat sich hierbei als lernwirksam erwiesen. Als wesentliche Erklärung für den Testeffekt wird das Einüben des Abrufs aus dem Gedächtnis verantwortlich gemacht (retrieval effect). Dadurch verbessert sich der Zugriff auf das vorhandene Wissen. Der Testeffekt entfaltet deshalb seine Wirkung vornehmlich bei Aufgaben, die man mit hoher Wahrscheinlichkeit auch korrekt lösen kann.

Informatives Feedback im Anschluss an eine unbenotete Testung bewirkt in der Regel eine über die reine Testung hinaus gehende Behaltensstabilität, sofern bei der Testung eine nennenswerte Anzahl von Fehlern auftritt. (siehe zusammenfassend Jacobs 2008 c bzw. im Detail Jacobs 1998-2007). Das Feedback kann aber nur dann eine positive Wirkung versprechen, wenn eine hinreichende Motivation zur Korrektur von Fehlern vorliegt und der durch das Feedbackstudium erhoffte Lerngewinn so relativ einfach und sicher erreichbar erscheint, also eher nicht bei schwierigen und komplexen Aufgaben (Kluger & DeNisi 1996).

Deshalb ist die Lernwirksamkeit informativen, sachorientierten Feedbacks, insbesondere die Rückmeldung der korrekten Antwort (Knowledge of correct response (KCR)), auch relativ sicher und deutlich in Laborstudien an einfachen Lehrzielen (meist Faktenwissen) zu bestätigen. Am besten müsste dies gelingen, wenn man das Feedback den Lernern quasi in einer Form aufdrängt, dass diese eine bestimmte Zeit dem Feedback gar nicht ausweichen können. Dies geschieht in etlichen Laborstudien, welche die Feedbackzeit dadurch kontrollieren, auf dem Bildschirm einige Sekunden lang nur das Feedback zu belassen, bevor es weiter gehen kann. In der Praxis kann die erhoffte Fehlerkorrektur durch KCR-Feedback jedoch völlig versagen, wenn die Rückmeldung verzögert und umständlich gewährt wird, wodurch die Wahrscheinlichkeit sinkt, der Lerner nehme diese überhaupt zur Kenntnis.

Die überwiegend in Laborstudien erfassten Behaltensvorteile des Testens und Feedbacks können folglich nur dann auf schulische Umwelten übertragen werden, wenn die Bedingungen dort denen der Laborstudien entsprechen, wobei bereits stillschweigend vorausgesetzt wird, Studierende würden sich auch ohne Geld ähnlich anstrengen wie die Probanden, welche für ihre Teilnahme am Laborexperiment bezahlt werden. Das wäre etwa der Fall, wenn der Lehrer im Unterricht ein kurzes Lehrvideo einsetzt, dann ein Quiz zur Bearbeitung austeilte, anschließend die korrekten Lösungen nennt bzw. die Aufgaben mit den Schülern bespricht und diese auffordert, ihre Fehler durch die korrekten Antworten zu ersetzen. Dem Vorgehen im Labor weitgehend angelehnt, setzten z.B. Gier & Kreiner (2009) während der Powerpoint basierten Lehre im Seminar Quiz zur Bearbeitung an, die unmittelbar danach besprochen wurden. Sie berichten von deutlich besseren Posttest - und Examenresultaten im Vergleich zu einem Kurs ohne diese Quiz. Ebenso lassen sich einige empirische Belege für die Wirksamkeit von Personal Response Systemen (RPS) während einer Vorlesung finden. Bei elektronischen RPS werden im Verlauf einer Vorlesung MC-Fragen via Powerpoint zum gerade zuvor präsentierten Lehrinhalt gestellt, die dann von allen Studierenden mittels eines technischen Antwortgerätes beantwortet, unmittelbar vom System ausgewertet, teils sogar mit Noten bewertet werden und im Anschluss daran besprochen werden können (z.B. Mayer et. al. 2009, Shapiro 2009).

Testen in der Schule und Universität unterliegt aber häufig anderen Bedingungen als Testen im Labor, insbesondere dann, wenn sich der Lehrstoff bzw. Testumfang sowie das verlangte Lehrzielniveau als deutlich anspruchsvoller erweisen, der Abstand des Testens zur Instruktion wesentlich länger ausfällt, die Test- und Feedbacknutzung auf Freiwilligkeit beruht oder die Bewertungsfunktion des Testens stärker ins Blickfeld gerät.

## Testfunktionen im pädagogischen Bereich

In etlichen Studien wurden Quiz eingesetzt, um die Studierenden dazu anzuhalten, relevante Lehrtexte, etwa ein Kapitel im Lehrbuch durchzuarbeiten (=Preliminary quiz, z.B. Narloch et al. 2006, Ryan 2006, Johnson & Kiviniemi 2009). Eine ähnliche Funktion übernehmen Quiz, wenn die sorgfältige Bearbeitung von Arbeitsaufträgen angeregt werden soll (z.B. Hautau et al. 2006a). Diese Maßnahmen sollten die Anwesenheit im Seminar erhöhen - weil dort die Quiz anstehen -, eine bessere Vorbereitung auf die Seminarsitzung bewirken, eine aktivere und anspruchsvollere Mitarbeit im Seminar fördern und so insgesamt auch zu besseren Noten im Abschlussexamen führen. Der entsprechende Test wird zuvor angekündigt, das Ergebnis meist mit relativ marginalen, leistungsabhängigen Bonuspunkten oder Noten bewertet und zu den einzelnen Aufgaben mindestens KCR-Rückmeldungen gewährt. In der Regel führen derartige Testankündigungen zu deutlichen Leistungsvorteilen bei einer Lehrzielüberprüfung der diesen Tests zugrunde liegenden Lehrtexten oder sonstigen Informationen, aber nur teilweise zu besseren Abschlussexamensnoten. Zum anderen werden Tests eingesetzt, um den bisher behandelten Lehrstoff zeitnah zu testen und dabei insbesondere zu ergründen, ob eine häufigere Testung bessere Abschlussleistungen bewirkt (z.B. Bangert et al. 1991). Derartige angekündigte und benotete Testungen wollen den Studierenden dazu bewegen, dem Unterricht möglichst aufmerksam folgen, sich kontinuierlich

vorzubereiten und gegebenenfalls die dazu notwendigen Quellen (Seminarfolien, Texte, Notizen) heranzuziehen.

## Testen mit oder ohne Notenerwartung

Es kann kein Zweifel daran bestehen, dass die rechtzeitige Androhung einer Benotung in der Regel die Leistung im Test erhöht. Dieser Effekt wird z. B. sichtbar, wenn man die Ergebnisse von Tests mit und ohne Notenerwartung miteinander vergleicht (z.B.: Burns & Vinchur 1992, Jacobs 2007a Kap: Empirische Studien zur Wirkung einer angekündigten Benotung auf die Leistung). Der so erzielte massive Leistungsvorteil hat wenig mit dem Testeffekt zu tun, sondern ergibt sich im Wesentlichen durch vermehrte Lernaneignung als Folge deutlich gesteigerter Lernmotivation (Jacobs 2009). Der fast durchgängig nachgewiesene Leistungsvorteil einer angekündigt benoteten gegenüber einer angekündigt unbenoteten Testbedingung während des Seminars muss sich aber keineswegs zwingend in den Ergebnissen eines nachfolgenden, benoteten kumulativen Abschlussexamens manifestieren. Die Einführung benoteter Tests und benoteter Kriteriumstests erschwert die Interpretation, mögliche Leistungsgewinne kausal bestimmten Ursachen zuzuordnen, weil sich dann motivationale Faktoren, z.B. eigene Vorbereitung auf den Test, mit instruktionalen Maßnahmen, z.B. der Wirkung der Testung und des Feedbacks auf die Leistung, vermischen.

## Notengewichtung

Gegenüber klassischer Notengebung führt eine Bonusregelung niemals zu einer negativen Bewertung. Erzielte Bonuspunkte können sich nur positiv auf die Seminarendnote auswirken und versprechen so eine positiv getönte Motivaton sowie weniger aktuelle Leistungsangst bei der Testung. Wenngleich die Bedeutung der Gewichtung einzelner Tests auf die Examensnoten noch nicht endgültig geklärt erscheint, können bereits recht marginale Punktversprechen nachweisbare Motivationszuwächse bzw. Leistungssteigerungen nach sich ziehen. Bei Hautau et al. (2006a) bewirkten magere Bonuspunkte massive Leistungsvorteile in den sehr transparenten und recht einfach vorzubereitenden, geforderten Aufgaben ( $d = 1.6 - 2.3$ ), jedoch nicht mehr bei den üblichen Unittests, die für alle Studierende, also auch für die Kontrollgruppe eine Benotung vorsahen. Auch Hautau et al. (2006b) fanden sehr deutliche Leistungsvorteile leistungsabhängiger Bonuspunkte für die bonusrelevanten Aufgaben, bei streng kontingenter Bonuszuweisung im Gesamtumfang von 12% der Seminarendnote sogar etwas bessere Abschlussexamen. Einer Untersuchung von Kibble (2007) zufolge, bearbeiteten nur ca. 50% aller Studenten ein freiwilliges Quiz ohne jede extrinsische Konsequenz, das als Vorbereitungshilfe auf die Klausur gedacht war. Wurde jedoch allein die Quizteilnahme mit einem leistungsunabhängigen Creditpoint von 0.5% der Seminarleistung honoriert, so nahmen ca. 90% der Studierenden an der Testung teil. Der Anteil der Quizteilnehmer stieg auf ca. 98 Prozent, nachdem der Bonus nun abhängig von der Leistung auf maximal 2% erhöht wurde. Die Ergebnisse der Abschlussklausuren fielen jedoch unter allen Bedingungen annähernd gleich aus. Über 90% der Studierenden stimmten dem Statement zu: "I completed online quizzes because course credit was given for them." Die freiwillige Teilnahme an sowie die Leistung in einem Quiz lassen sich demnach relativ deutlich durch leistungsabhängige Bonuspunkte fördern. Eine Leistungssteigerung in den Quiz ist selbst dann zu erzielen, wenn lediglich eine Zufallstichprobe der bearbeiteten Quiz tatsächlich mit Kreditpunkten bewertet wird (Hautau et al. 2006b, Krohn et. al. 2009).

## Notenrelevante Testung und unbenoteter Kriteriumstest

Bei einigen Studien von Haynie (zusammenfassend Haynie 2007) rechneten die Studenten der EG mit einem benoteten Test und führten diesen durch, während die KG weder einen Test erwartete und durchführte. Beide Gruppen wurden zwei oder drei Wochen nach dem Test völlig unerwartet mit einem unbenoteten Kriteriumstest konfrontiert. Hier wird somit

die Wirkung benoteter Tests auf das längerfristige Behalten erfasst. Die benotete Testung führte hierbei ziemlich konsistent zu besseren Behaltensleistungen als keine Testung. Allerdings bleibt schwer abzuschätzen, wodurch der Effekt zustande kam. Haynie (z.B. Haynie 1997) prüfte dann in weiteren Untersuchungen, ob ein angekündigter und tatsächlich durchgeführter benoteter Test zu höheren langfristigen, nicht benoteten Behaltensleistungen führe als die alleinige Ankündigung eines benoteten Tests. Da die tatsächliche Testung meist mehr Behalten bewirkte als die alleinige Testankündigung, sprechen die Ergebnisse für die behaltensstabilisierende Wirkung der Testbearbeitung, die zumindest über einen reinen Lernmotivierungseffekt angekündigt benoteter Tests hinaus gehen kann. Allerdings fielen die Ergebnisse von Haynie nicht einheitlich aus und manche Befunde sind schwer zu verstehen.

Im normalen Schulsystem sind motivationale- und Testeffekte meist untrennbar miteinander verbunden und ihr jeweiliges Gewicht auf die Behaltensleistung bleibt ziemlich schwer abzuschätzen. Daten von Morrison et al. (1995, S. 39) deuten darauf hin, den Motivierungseffekt durch Noten selbst bei einem relativ überschaubaren Computerlernprogramm mindestens so hoch einzuschätzen ist wie die Lernwirksamkeit des Testens mit Feedback. Im Feldexperiment von Jacobs (2009) wurden die Auswirkungen der Benotung (benotetes Quiz gegen unbenotetes Quiz) auch auf die Leistung der entsprechenden Lehrziele in einer unangekündigt unbenoteten Abschlussklausur getestet. Dadurch sollte die reine Wirkung des Notendrucks auf das langfristige Behalten ermittelt werden. Die Ergebnisse belegen eindeutig, die Benotungserwartung bzw. die dadurch ausgelöste bessere Vorbereitung einschließlich der Testung und Rückmeldung bewirkten mehr Behaltensgewinn als eine verpflichtende unbenotete Testung mit demselben Feedback. Ich tendiere dazu, im Schulsystem [wie ich es hier vorfinde] dem Motivierungseffekt der Notenandrohung mehr Lernerfektivität zuzusprechen als dem Behaltenseffekt durch Testung und Feedback.

Der in Tabelle 2 angedeutete Untersuchungsvorschlag könnte mehr Klarheit hinsichtlich der Gewichtung von motivationalen und instruktionalen Wirkfaktoren des Testens auf das Behalten versprechen. Er greift das Vorgehen von Haynie auf und nimmt einige Modifikationen vor.

**Tabelle 2.** Vorschlag zur Abklärung der Wirkfaktoren erwartete Benotung (Motivierung) und Quizbearbeitung (Retrieval + Feedback) auf den langfristigen Lernerfolg

Exp. Gruppen	Quiz-Ankündigung	Quiz-bearbeitung	unangekündigter Kriteriumstest	
			Aufgaben A	Aufgaben B
1	benotet	findet statt	quizanalog	Neu
2	benotet	fällt aus	quizanalog	Neu
3	nicht benotet	findet statt	quizanalog	Neu

Bei allen 3 experimentellen Gruppen wird ein Test angekündigt, wobei der Gruppe 3 klar ist, dass dieser Test nicht benotet wird. Wie bei Haynie fällt die Testbearbeitung bei Gruppe 2 unter einem Vorwand aus. In den Kriteriumstest werden zum einen quizanaloge parallele Aufgaben aufgenommen, welche die Wirkung der Testung und des Feedbacks erfassen. Zum andern könnte der Kriteriumstest ganz neue, aber lehrzielrelevante Aufgaben enthalten, die durchaus auch im Quiz hätten vorkommen können. Bei diesen dürften dann nur motivationale Effekte zu Tage treten. Um den reinen Effekt der Testung mit Feedback zu erfassen wäre noch eine No-Treatment-Kontrollgruppe vonnöten.

## Testhäufigkeit

Bei vielen Studien steht die Frage im Raum, ob eine häufigere Testung während des Seminars auch die Leistung im benoteten Abschlussexamen verbessert (Bangert-Drowns, Kulik & Kulik 1991, Jacobs 2002, Kuo & Simon 2009). Durch eine Erhöhung der Testhäufigkeit sollten Studierende zu einem kontinuierlicheren Studium bewegt werden, was sich positiv auf ein kumulatives Abschlussexamen auswirken müsste. Die häufigen Tests selbst werden dabei benotet, gehen aber pro Test mit relativ geringem Gewicht in die Endnote ein. Abschlussexamen werden in der Regel angekündigt und tragen mit relativ hohem Gewicht zur Seminarendnote bei.

Der Leistungsvorteil häufiger benoteter Testung im benoteten Abschlussexamen wurde dann ziemlich eindeutig belegt, wenn viele benotete Tests mit gar keiner zwischenzeitlich benoteten Testung verglichen wurden. Werden statt einem Abschlussexamen, weniger umfangreiche und zeitnähere Zwischenexamen als Kriteriumstests verwendet, so lassen sich mitunter gewaltige Vorteile für den Einsatz benoteter Quiz erzielen (z.B. eine Effektstärke von  $d=1.44$  bei Tuckman 1996 sowie Effektstärken von 1,81 bzw. 2.26 bei Daniel und Broida 2004). Poljičanin et al. (2009) stellten durch die Einführung täglicher Miniquiz (insgesamt 35) sogar signifikant bessere Endnoten eines Anatomieseminars gegenüber dem Seminar des Vorjahres im Ausmaß einer Effektstärke von immerhin  $d = .57$  fest. In allen diesen Fällen bearbeiteten die Studierenden unter der Experimentalbedingung viele Testfragen vor dem Examen, während die Studierenden unter der Kontrollbedingung überhaupt keine Fragen beantworteten, was eher einer Variation der Testmenge gleichkommt. Letztlich wird so belegt, dass die Verwendung benoteter Kurzttests im Vergleich zu gar keiner Testung auch in der Schulpraxis eine deutliche Wirkung hinterlässt.

Ein anspruchsvolleres Verfahren verlangt jedoch den Nachweis einer Überlegenheit häufigerer benoteter Testung bei vergleichbarer Testmenge für EG und KG. Hierbei bearbeiten alle Studierenden dieselben Aufgaben. Die Items verteilen sich im Verlauf eines Seminars lediglich auf viele kurze oder wenige lange Tests. Die vielen kurzen Tests sollten dabei ein kontinuierliches Lernen fördern oder absichern. Dieser Ansatz führte in der Regel zu höheren Gesamtestwerten bei häufigerer Testung, vermutlich, weil der Testumfang bei kurzen Tests geringer ausfällt und man sich so effizienter auf kurze Tests vorbereiten kann. Im benoteten Abschlussexamen war der Leistungsvorteil dann aber meist nicht mehr zu entdecken (z.B. Deck 1998). Allerdings findet man auch Studien, die bei gleicher Testmenge, aber häufigerer Testung während des Seminars bessere Abschlussexamen fanden (z. B.: Khalaf & Hanna 1992 für das Fach Biologie, Myers & Myers 2007 im Fach Statistik). Möglicherweise hängt der Effekt auch von der Komplexität des Lehrstoffs bzw. den Lehrzielanforderungen ab.

Auch Studierende ohne häufige benotete Quiznutzung sind an guten Leistungsergebnissen interessiert und bereiten sich entsprechend auf das Abschlussexamen vor. Narloch, Garbin, & Turnage (2006) fanden etwa heraus, dass Studierende, die während des Seminars keine Quiz erhielten, im Vergleich zu Studierenden unter mehrfach benoteter Quizbedingung deutlich mehr Vorbereitungszeit für das Abschlussexamen investierten. Der Notenanreiz des Abschlussexamens ist für die Leistung im Examen mitunter stärker als der Effekt häufigen Testens. Die Wirkung häufigerer benoteter Testung im Schulalltag würde sich deutlicher in unerwarteten, unbenoteten Abschlussklausuren bzw. Vergleichsarbeiten zeigen.

## Feedbacknutzung

Nach einem benoteten Examen fehlt jeglicher extrinsische Anreiz, den Fehler zu überwinden, da das Verständnis der Lösung "nichts mehr einbringt". Entsprechend nehmen viele Studierende das sachorientierte Feedback dann gar nicht zur Kenntnis. (siehe auch Jacobs 1999 Abschlussklausur). Brosvic & Epstein (2007) untersuchten bei mehreren classroom

examination die Wirkung des Feedbacks auf das langfristige Behalten. Ein deutlicher Feedbackeffekt zeigte sich nur bei unmittelbarem Feedback after item, also, einer gezielten Rückmeldung (in Form von Multiple Try Feedback bzw. KCR-Feedback) nach Bearbeitung jeder einzelnen Prüfungsaufgabe während des Exams. Wurde das KCR-Feedback unmittelbar nach der gesamten Testbearbeitung oder einen Tag später gewährt, so blieb von der Wirkung des Feedbacks auf das langfristige Behalten nicht mehr viel übrig. Sun & McDaniel 2008 ließen Tests bearbeiten und schickten später das Feedback via Email zu. Vergebens. Denn der Behaltenseffekt war hoch vergleichbar einer Testbedingung ohne jedes Feedback. Obwohl verzögertes Feedback bei adäquater Nutzung gemäß der Theorie des verteilten Lernens eigentlich mehr Behaltenserfolg verspricht als unmittelbares Feedback, erweist es sich in der Schulpraxis meist deshalb als unwirksamer, weil es nicht angenommen wird (siehe Jacobs 2008a).

Jacobs (2009) berichtet von offensichtlicher Desinteresse der Studierenden, nach einem benoteten Quiz im Seminar die korrekten Lösungen überhaupt erfahren zu wollen, möglicherweise, weil dort keine Abschlussklausur anstand, für welche man irgendwelche Lerngewinne hätte verwerten können. Deck (1998) gab den Studenten die Möglichkeit, auf Anfrage ihre benoteten Tests nach der Bewertung einzusehen und berichtet, sehr wenige Studenten hätten von dieser Möglichkeit Gebrauch gemacht, obwohl noch eine Abschlussklausur anstand. Auch in der Studie von Daniel & Broida (2004) folgte auf die benoteten Tests noch ein kumulatives Examen, was aber zu meiner Verwunderung die sachorientierte Feedbacknutzung der Quiz keineswegs besonders anregte. Einer Einschätzung von Daniel (persönliche Mitteilung) zufolge, interessierten sich die meisten Studenten nach den benoteten Quiz nur für ihr Leistungsergebnis und nahmen das informative Feedback, z.B. die korrekte Antworten, nicht weiter zur Kenntnis.

Es fehlen hinreichende Erfahrungen, wie intensiv Studierende bei Onlinetests das Feedback überhaupt nutzen. Im Gegensatz zu einem Laborexperiment will man dem Lerner das Feedback in natürlichen Lernumgebungen, etwa bei einem Onlinequiz, nicht einfach eine bestimmte Zeit lang aufzwingen, sondern ihm eine adaptive Nutzungszeit zugestehen. In der Studie von Jacobs 2007b hatten Studierende die Möglichkeit, nach der unbenoteten Testbearbeitung die korrekt gelösten Aufgaben einschließlich elaborierten Feedbacks einzusehen. Die Analysen zur Feedbackzeit ließen erkennen, dass ca. ein Viertel aller Studenten nach der Bearbeitung des Tests überhaupt kein Interesse an einem ernsthaften Studium der korrekten Lösungen zeigte. Ein einfacher Prompt: "Versuchen Sie, sich die Lösungen selbst verständlich zu machen" hatte für diese offenbar keinerlei Wirkung. Die Feedbackbearbeitungszeit betrug im Median 4,7 Minuten, ihre Angemessenheit ist aber schwer zu bewerten. Bei bestimmten Zwischentests, die auch prüfungsrelevante Lehrziele der Abschlussklausur tangieren, wäre ein aufmerksames Feedbackstudium eigentlich ratsam. Allerdings verzichten etliche Studenten dennoch darauf und begnügen sich damit, ihr Leistungsergebnis in Erfahrung zu bringen. Immerhin könnte auch die ausschließliche Rückmeldung des Leistungsergebnisses je nach Zielsetzung des Lerners motivationale Auswirkungen etwa auf die Intensität weiterer Prüfungsvorbereitung haben. Zudem wird dem sachorientierten Feedback dann mehr Aufmerksamkeit geschenkt, wenn in einer unbenoteten Übung ganz gezielt prüfungsrelevantes Wissen getestet wird (siehe Jacobs, 2000).

Jedoch selbst dann, wenn adäquat konstruierte Probeklausuren (siehe Jacobs 2003) mit elaboriertem Feedback effektiv genutzt werden würden, so garantierten sie keineswegs bessere Abschlussklausuren, weil die Notenrelevanz der Abschlussklausur alle Studenten stark motiviert, sich möglichst gut vorzubereiten und der Effekt der Person (Vorwissen, Fähigkeit, Motivation, Lernstrategie, Selbstkontrolle) den von Instruktionsmaßnahmen gänzlich aufreißt.

## Kontrolle der Testbearbeitung

Üblicherweise wird die Testbearbeitung bei benoteten Examen beaufsichtigt und streng kontrolliert. Stellenweise werden jedoch auch Noten für die relativ freie Bearbeitung von Onlinetests oder für Hausaufgaben vergeben, die deutlich weniger bzw. gar keine Kontrolle

der Aufgabenbearbeitung vorsehen. Die Wirkung eines benoteten Tests hängt mit davon ab, wie transparent die Anforderungen und wie stark die Eigenverantwortlichkeit bzw. das Eigenbemühen bei der Testvorbereitung und Testbearbeitung ausfällt. Kann der Lerner während der Testbearbeitung allerlei Hilfsmittel benutzen oder Antworten aus Texten kopieren, so leidet die Motivation, sich hinreichend vorzubereiten. Aus der Studie von Daniel & Broida (2004) ging etwa hervor, dass die leistungssteigernde Wirkung eines benoteten Tests auf das nachfolgende Examen ausblieb, wenn ein zu großer Freiheitsspielraum bei der Beantwortung eines benoteten, aber unkontrollierten Online-Tests gewährt wurde, der allerlei Missbrauch nach sich zog. Wie die Autoren dann nachwiesen, lässt sich auch die Lerneffektivität einer formal unkontrollierten Onlinetestung durch Einführung bestimmter Maßnahmen wie strenge Terminierung von Prüfungszeiten, eingegrenzte Aufgabenbearbeitungszeit sowie eine Zufallsauswahl von Testaufgaben steigern. Wellmann (2005) ermittelte etwas bessere Examensleistungen, wenn die Online-Testungen unter kontrollierten Bedingungen stattfanden. Tuckman (1996) fand höhere Leistungen durch benotete kontrollierte Kurztests zu Beginn eines Seminars als bei benoteten Hausaufgaben und führt dies darauf zurück, bei den dort aufgegebenen Hausaufgaben könne man einfacher Informationen kopieren. In einer Untersuchung von Kibble (2007) hatten Studierende die Möglichkeit, unkontrollierte Online-Tests mehrmals zu bearbeiten (Mastery quiz) und hierbei leistungsabhängige Creditpoints von maximal 2% der Seminarnote einzustreichen. Die so erreichten unrealistisch hohen Erfolgsquoten von im Mittel 98% korrekter Lösungen fielen im Abschlussexamen unter kontrollierten Bedingungen allerdings deutlich niedriger aus.

Einige Studien erbrachten allerdings den Nachweis, Hausaufgaben führten zu höheren Lernergebnissen, wenn diese zumindest nach der Abgabe kontrolliert und benotet oder mit Bonuspunkten bewertet wurden. (z.B. Hautau et al. 2006a,b, Ryan & Hemmes 2005). Die Arbeitsaufträge bei Hausaufgaben sollten vornehmlich so gestaltet werden, dass ihre Lösung eine aktive Auseinandersetzung mit dem Lehrstoff erfordert und ein Abschreiben aus einem Text so gar nicht möglich ist. Dann können benotete Hausaufgaben ähnlich gute Lernergebnisse erzielen wie kontrollierte, benotete MC-Quiz (Ryan 2006). Bei Ryan (2006) erzielten sehr dezidierte und transparente Arbeitsaufträge zu einem Lehrbuchtext bei den Hausaufgaben sogar einen höheren Lernerfolg als angekündigte und durchgeführte benotete Tests, denen lediglich die Information vorausging, der Lehrbuchtext werde getestet, ein Ergebnis, das im übrigen in einem gewissen Widerspruch zu den Befunden von Tuckman (1996) steht.

## Open book tests

Im Gegensatz zu streng oder weniger kontrollierten Onlinetests zielen open book tests geradezu darauf ab, Fragen unter Zuhilfenahme von Lehrtexten oder weiteren Lernmaterialien zu beantworten. Diese Testform erfordert weniger Memorierungsaufwand und könnte insbesondere in einem frühen Lernstadium sowie bei sehr umfangreichen Texten oder schwierigen anspruchsvollen Anforderungen durchaus sinnvoll sein. Wie aus zwei Laborexperimenten von Agarwal et al. (2008) zum Vergleich von open-book-test und konventionellem closed-book test mit Feedback hervorging, erzielten Studierende im open-book-test bessere unmittelbare und vergleichbare längerfristige Behaltenswerte. Für reine Lernzwecke, d.h. ohne den Effekt der Benotungserwartung sind open book tests echte Alternativen zu einer kontrollierten Testung. Sie erfordern vom Lerner eine gewisse Eigeninitiative, die verfügbaren Informationen bei Problemen selbst zu nutzen. Wenn es um notenirrelevantes Lernen geht, profitieren Studierende sogar vom Studium gelöster Aufgaben. (z.B. Jacobs 2008a). Bei echten Examen können open-book-tests jedoch Nachteile nach sich ziehen, die sich vornehmlich in einer weniger intensiven Vorbereitung äußern. (siehe Rakes (2008) zu einigen Ergebnissen der Open-book-testung).

Take home tests sind als eine besondere Variante von open book tests aufzufassen. Hierbei bearbeiten die Studierenden das Quiz als Hausaufgabe. Haynie (2003, 2007)

konnte die Behaltenswirksamkeit solcher (benoteter) Hausaufgaben teilweise bestätigen, fand insgesamt keine Leistungseinbußen gegenüber kontrollierten Quiz im Seminar, weist jedoch auf schwächere Ergebnisse bei solchen Fragen hin, die nicht im Test enthalten waren. Oliver & Williams (2005) untersuchten die Auswirkungen von Hausaufgabentests unter einem leistungsabhängigen und leistungsunabhängigen Bonusanreiz und fanden unter leistungsabhängiger Anreizbedingung deutlich bessere Leistungsergebnisse im Hausaufgabentest sowie mäßige, aber noch signifikante Vorteile in den Unittests.

## Erwartung der Lernförderung durch Quiz (Kurztests)

Die vermutete positive Lernförderung durch Tests basiert auf einigen idealisierten Annahmen. Um den möglichen Leistungseffekt eines Tests auf das Examen auch entdecken zu können, sind jedoch einige Voraussetzungen notwendig. Schließlich sollte der erwartete Nutzen durch eine Testung auch bestimmten Anforderungen genügen. Einige dieser Forderungen und Erwartungen werden kurz skizziert.

1. Die Quiz- und Abschlussklausur- bzw. Kriteriumsaufgaben messen im Idealfall dieselben Lehrziele mit nicht identischen Aufgaben. Denn es erscheint problematisch, den Quizteilnehmern bereits die Examensaufgaben vorzulegen. Lerneffekte sollten zumindest auch einen nahen Transfer erfassen.
2. Die Quiz erhöhen frühzeitig die Studiermotivation und bewirken ein kontinuierliches Studium während des Semesters [keeping students on track]. Dies ist eher bei angekündigten, benoteten Quiz zu erwarten. Dadurch erzielen die Studenten der benoteten Quizbedingung deutlich höhere Testwerte im Quiz als eine Gruppe, die ohne vorherige Ankündigung das Quiz ebenfalls bearbeiteten muss oder an einem angekündigten, aber unbenoteten Quiz teilnimmt.
3. Die Beherrschung der Lehrziele im Quiz fördert nachweislich das Erreichen weiterer Lehrziele im Examen. Das durch die Quiz angeregte verbesserte Studium und die Testung stabilisieren das Wissen, fördern die Kompetenz im Fachgebiet und erleichtern die Aufnahme neuen Wissens.
4. Die Quiz fördern eine effiziente Prüfungsvorbereitung auf die Abschlussklausur durch einen geringeren erforderlichen Lernaufwand, erhöhte Transparenz sowie gezielte Stoffwiederholung und verbessern so die Examensleistung und/oder die Stabilität des Wissens.

### **Parallele Aufgaben im Quiz und Kriterium**

Punkt 1 bezieht sich auf die Diagnostik der Testungen im Quiz und Kriterium und kann prinzipiell durch die Konstruktion paralleler Aufgaben gelöst werden. Jedoch liefert die Forschung eindeutige Belege dafür, mit wachsender Transferdistanz nähmen Test- und Feedbackeffekte ab, was im Übrigen für die meisten Instruktionsmaßnahmen zutreffen dürfte. Schon kleine Modifikationen im Satzbau oder die Verwendung von Synonymen in einer Aufgabe können den Behaltenserfolg merklich beeinträchtigen (Jacobs 1999). Wichtiger als Faktenwissen im Kriteriumstest sprachlichen Modifikationen zu unterwerfen, wäre es freilich, vornehmlich echtes transferfähiges Wissen zu vermitteln und zu untersuchen. Man darf aber kaum erwarten, dass auch solche Lehrziele im Abschlussexamen eher erreicht werden, die in den Quiz gar nicht eingeübt wurden, es sei denn, Notendruck bewirke eine umfassende Vorbereitung auf die Quiz. Ein möglicher Grund für fehlenden oder schwachen Lernvorteil einer Quizgruppe im Abschlussexamen mag mit auf eine unzureichende Überlappung zwischen Quiz- und Kriteriumsaufgaben zurückgehen. (Burns & Vin-

chur 1992, Kling et al. 2005, Brown & Sternberg 2009). Bei der Itemkonstruktion steht die ziemlich knifflige Forderung an, für den Kriteriumstest akzeptable parallele Aufgaben zu finden, die den Transferanspruch gewährleisten, aber nicht überstrapazieren.

### **Förderung des Verständnisses im Verlauf des Seminars ?**

Während die in Punkt 2 angesprochene höhere Lernmotivierung und intensivere Vorbereitung durch die Erwartung benoteter Quiz im Verlauf des Seminars relativ unstrittig sein dürfte, sehe ich die größte Problematik bei Punkt 3. Denn die Lehrziele vieler - untersuchter - Seminare basieren meist nicht auf einer gültigen Lehrzielhierarchie im Sinne "vorheriges Wissen ist notwendig oder förderlich für nachfolgendes Wissen", sondern lassen sich beliebig additiv aneinander reihen. So erweist sich etwa die Sequenzierung des Lehrstoffs in einem Seminar wie "Lehren und Lernen" keineswegs als didaktisch zwingend, weil Prinzipien der klassischen und operanten Konditionierung in keinem direkt nachvollziehbar lernförderlichen Zusammenhang mit Gedächtnismodellen, Begriffslernen oder Problemlösung stehen. Andererseits benötigt ein Studierender in Lerngebieten mit streng aufeinander aufbauenden Wissensbeständen kaum kontinuierliche Testungen, um einzusehen, ohne kontinuierliche Arbeit könne er dem Seminar überhaupt nicht mehr folgen. In solch anspruchsvollen Lehrgebieten benötigen manche Studierende bei unzureichenden Quizleistungen vermutlich zusätzliche hilfreiche Unterstützung, die deutlich über das im Quiz gewährte Feedback hinausgehen müsste. Hier wäre quasi ein Mastery learning mit gezielter remedialer Nachbearbeitung vonnöten.

### **Effizientere Vorbereitung auf das Abschlussexamen ?**

Wird derselbe Lehrstoff vor dem Quiz und dem Abschlussexamen eingeübt, so profitiert die Quizgruppe vom verteilten Lernen, während etliche Probanden der Kontrollgruppe Gefahr laufen, die gesamte Prüfungsvorbereitung massiert gegen Ende des Prüfungstermins zu platzieren [=cramming]. Massiertes Lernen erzielt jedoch kurzfristig einen mindestens ähnlich hohen Behaltenserfolg wie verteiltes Lernen, was bis zur Abschlussklausur durchaus reichen kann. Der Vorteil verteilten gegenüber massiertem Lernen manifestiert sich erst bei längeren Behaltensintervallen. Die Anforderungen der Klausuren lassen sich relativ transparent gestalten und können auch von Studenten ohne Quiznutzung gut antizipiert werden. In etlichen Studienfächern lässt sich so der Prüfungsstoff auch ein bis zwei Wochen vor der Abschlussklausur noch hinreichend vorzubereiten. (siehe z.B. die deutlichen Leistungssteigerungen von der Probeklausur zu der eine Woche später folgenden echten Klausur in Didaktik: Lehren und Lernen von fast 20%). Der Erfolg im Examen wird somit in erster Linie von der Fähigkeit und dem Lernmanagement des Studierenden bestimmt, der sich den jeweiligen Bedingungen jeweils adäquat anpassen muss. Durch den Prüfungswahn der Bachelorstudiengänge könnte eine ausreichende Prüfungsvorbereitung auf die Abschlussklausuren in einem Semester allerdings auf eine harte Probe gestellt werden.

Deutlich bessere, durch vorausgehende Quiz bewirkte Ergebnisse im benoteten Abschlussexamen wären nur dann zu erwarten, wenn ihre Bearbeitung durch keinerlei selbst kontrolliertes Eigenstudium kompensierbar wäre und/oder die notwendige Vorbereitung für die Klausur einfach nicht mehr reicht oder aufgebracht werden kann. Hierbei müssten unfaire Bedingungen ausgeschlossen sein, die sich z.B. daraus ergäben, wenn nur aus den Quiz erkennbar wäre, wie die Aufgaben im Examen in etwa aussehen könnten. Einige Studien belegen in der Tat, dass Studierende ohne Quiz während des Semesters im späteren Mitterm- oder Unitexamen ziemlich chancenlos waren (z.B. Tuckman 1998, Daniel & Broida 2004). Die meisten Studien finden jedoch eher mäßige Leistungsunterschiede im benoteten Abschlussexamen oder signifikante Vorteile der Quiz nur für ausgewählte Studierende (meist leistungsschwache oder durchschnittliche leistungsfähige Studierende, z.B. Graham 1999).

## **Bessere Erfassung der reinen Quizwirkung auch in der Schulpraxis**

Die reine Wirkung des Testens sowie die des Feedbacks im Vergleich zu einer definierten Kontrollbedingung kann in der Regel nur bei unbenoteten Übungs- und Kriteriumstests, am besten unter kontrollierten Bedingungen im Labor geprüft werden. In der Praxis scheint ein überraschender, unbenoteter Kriteriumstest dann gut anwendbar, wenn zuvor keinerlei Anlass vorhanden war, die quizrelevanten Lehrziele irgendwie nachzuarbeiten. Jedenfalls sollten herbeigesehnte Kompensationsmöglichkeiten für frei gewählte, weitere Lernaktivitäten zwischen Quiz und Kriterium ausgeschlossen sein. Ob häufigere Testung zu mehr Behaltensstabilität führt, ließe sich deshalb besser in schulischen Umwelten ermitteln, die keine Abschlussklausuren vorsehen, etwa beim Vergleich von 4 gegenüber 2 Klassenarbeiten pro Schulhalbjahr. Als Kriteriumstest müsste dann gänzlich unerwartet eine unbenotete Abschlussarbeit zur Bearbeitung vorgelegt werden, welche die Lehrziele des gesamten Schulhalbjahres erfasst. Ein möglicher Vorteil häufigen Testens könnte sich möglicherweise trotz benoteter Abschlussklausur ermitteln lassen, wenn man einen weiteren unbenoteten Test einige Wochen später als Behaltenskriterium heranzieht. Bei häufiger Testung und Abschlussklausur müsste das verteilte Üben (derselbe Lehrstoff vor dem Quiz und vor dem Examen) eine höhere Behaltensstabilität bewirken.

## **Alternativen zu Tests**

Testen mit Feedback führt zwar nachweislich zu besseren Behaltensleistungen. Aber Testen erscheint nicht zwingend lernnotwendig, weil sonstige Maßnahmen ebenfalls bessere Leistungen erwarten lassen. (z. B. wirksame Lerntechniken zum besseren Verständnis von Lehrtexten anwenden, gepromptete Lerntagebücher schreiben, Zusammenfassungen anfertigen oder Concept maps herstellen oder durcharbeiten, usw. siehe etwa Abraham & Sherman 2009). Wenn viele Studierende trotz Testangebot auf die Nutzung freiwilliger Quiz verzichten, so mögen manche von ihnen nicht zu Unrecht der Auffassung sein, es gäbe vergleichbar gute oder gar effizientere Maßnahmen der Vorbereitung auf ein Examen. Auf die von den Studierenden hierbei selbst initiierten Lernstrategien soll an dieser Stelle nicht weiter eingegangen werden. Als weitere mögliche Ergänzungen oder echte instruktionale Alternativen mit entsprechender Hilfestellung bzw. Kontrolle durch den Lehrenden betrachte ich,

- die Studierenden aufzufordern, eigene Fragen an den Lehrstoff zu richten, diese zu beantworten und als Kontrolle einzureichen,
- die Bearbeitung bereits beantworteter Aufgaben (Read Statements) zu ermöglichen und anzuregen,
- die prüfungsrelevanten Lehrziele transparent zu machen.

## **Selbst Fragen generieren**

Nach bisherigem Stand der Laborforschung führt das Generieren von Fragen an einen Text zu vergleichbar guten Lernergebnissen wie das Beantworten vorgegebener Fragen, was letztlich mit Testen gleichzusetzen ist. (siehe Jacobs 2005). Allerdings lassen sich Tests einfacher implementieren, informative Rückmeldungen leichter automatisieren und das Testen insgesamt besser kontrollieren. Es dürfte weitaus schwieriger sein, Studierende langfristig zum Generieren und Beantworten ihrer eigenen Fragen zu bewegen, diese Leistungen konsequent zu überprüfen und gegebenenfalls Fehler zu korrigieren. Zudem sind mir keine Untersuchungen bekannt, welche die Auswirkungen eigenen Fragestellens auf ein Examensergebnis überprüft hätten. Andererseits handelt es sich beim Fragenstellen an einen Text um eine basale Lerntechnik des Textverständnisses, deren Anwendung von außen zumindest gelegentlich kontrolliert eingefordert werden könnte

## **Sich korrekt beantwortete Quizaufgaben verständlich machen**

In einem frühen Lernstadium ist das Bearbeiten vorgegebener Aufgabenlösungen bzw. Lösungsbeispiele eigenen Problemlösebemühungen deutlich überlegen. Bei etwas anspruchsvolleren Fragestellungen dürfte das gezielte Studieren gelöster Aufgaben im Mittel der eigenen Bearbeitung von Testaufgaben zumindest ebenbürtig sein. Auch bei Aufgaben eines niedrigeren Lehrzielniveaus, etwa Verstehen, führt Testen nicht zwingend zu besseren Lernleistungen als das aufmerksame Lesen der bereits korrekt beantworteten Aufgaben (=Read Statement), obwohl die meisten Labor- und Feldstudien Vorteile für bestimmte Testformen, insbesondere freie Beantwortung, nachweisen konnten. Nur bei Aufgaben, die man mit relativ hoher Wahrscheinlichkeit richtig löst, dürften Tests wirkungsvoller ausfallen als das Bemühen, sich die korrekte Aufgabenlösung verständlich zu machen. Im Feldexperiment von Jacobs (2008) bearbeiteten Studierende Testaufgaben zum Thema klassische und operante Konditionierung und erhielten im unmittelbaren Anschluss an ihre Antwort sehr informatives Feedback. Unter der Bedingung gezieltes Studieren (=Read Statement) lagen den Studierenden diese Aufgaben korrekt beantwortet und mit dem elaborierten Feedback versehen direkt als zu erarbeitender Text vor Augen. Diese Art der Instruktion verdichtet wesentliche Aspekte eines Lehrthemas und konkretisiert die wichtigsten Inhalte. Gezieltes Studieren gelöster Aufgaben hat sich dabei im Vergleich zur Testung mit sehr informativem Feedback als gleich lerneffektiv und zudem lerneffizienter erwiesen. Am Feldexperiment nahmen allerdings nur solche Studierende teil, die freiwillig eine Übung bearbeiten wollten und die offensichtlich eine hinreichende Lernmotivation aufwiesen. Im Gegensatz zu Tests lässt sich die adäquate Nutzung von Read Statements nicht ohne weiteres kontrollieren oder erzwingen. In der normalen Schulpraxis werden Read Statements möglicherweise nur dann gut durchgearbeitet, wenn man ein Bonusquiz oder eine Prüfung ankündigt, die später ähnliche Aufgaben testen.

## **Explizierung prüfungsrelevanter Lehrziele**

Die Mitteilung der prüfungsrelevanten Lehrziele in sehr transparenter Form mag Studierende dazu motivieren, die entsprechenden Lehrmaterialien gezielt selbst zu erarbeiten. Auf diese Weise lässt sich die Prüfungsvorbereitung auf die wichtigen Lehrziele eingrenzen und für den Studierenden ziemlich effizient gestalten, da so der Einstellung entgegengewirkt wird, vieles "umsonst" zu lernen und von der Stofffülle überwältigt zu werden. Die Explizierung der prüfungsrelevanten Lehrziele könnte ihn dazu bewegen, selbst potentielle Prüfungsfragen zu generieren. Unter günstigen Bedingungen sind dann deutlich bessere Examenleistungen zu erzielen (Jacobs, Bedersdorfer & Bohse 1983). Bei Jacobs 2003 wurde die Wirksamkeit einer Probeklausur [Testen mit Feedback] mit einer Vorbereitungssitzung verglichen, die in sehr transparenter Form die Klausuranforderungen vermittelte. Unter beiden Bedingungen erzielten die Studierenden teilweise annähernd vergleichbar gute Klausurergebnisse. Nur in Statistik profitierten vor allem sehr schwache Studierende mehr von der Probeklausur. Bei sehr komplexen Anforderungen hinterlassen allerdings auch sehr exakte Lehrzielangaben keine Leistungssteigerung, was jedoch in gleicher Weise auch für Tests mit Rückmeldungen gilt. (Kluger & DeNisi 1998)

Eine ähnliche Funktion könnten Studierfragen an den richtigen Stellen im Lehrmaterial übernehmen. So konnte Haynie (2003) beispielsweise nachweisen, dass gezielte Studierfragen in einem recht umfangreichen Text mit der Erwartung, zu diesem Text werde ein benotetes Quiz anstehen, im Vergleich zu einer tatsächlichen benoteten Testung derselben Fragen im Seminar oder den Hausaufgaben sogar bessere Behaltenswerte nach sich zog. Auch Ryan (2006) konnte aufzeigen, dass präzise Arbeitsaufträge zu einem Lehrtext (focus worksheets), die zu Hause zu bearbeiten waren und benotet wurden, ähnlich hohe Lerngewinne erbrachten als die Androhung bzw. Durchführung einer benoteten Testung des Lehrtextes ohne solche Fragen.

## Schlussfolgerungen und Konsequenzen

Das optionale Angebot an die Studierenden, zur Unterstützung der Lehre bzw. als Vorbereitung auf ein Examen Onlinequiz bearbeiten zu können, ist auf die Freiwilligkeit der Studierenden angewiesen und mag denjenigen helfen, die es ausgiebig nutzen. Allerdings sind mir keine soliden Studien bekannt, welche den Nachweis erbracht hätten, ein entsprechendes Angebot bewirke im Mittel mehr Examenserfolg als der Verzicht auf ein solches. Etliche Studien berichten zwar von signifikant positiven Korrelationen zwischen der Häufigkeit der freiwilligen Onlinequiznutzung und dem Examenserfolg, dieser Zusammenhang ist aber kausal nicht interpretierbar, weil der Versuchsplan auf der Selbstselektion der Studierenden basiert. Nicht nur die Hypothese, mehr freiwillige Tests bewirkten einem höheren Examenserfolg, klingt plausibel. Auch die umgekehrte Deutung, die guten Studierenden nähmen mehr Tests in Anspruch, erscheint nachvollziehbar. Maurer (2006) fand beispielsweise relativ konsistent positive Korrelationen zwischen der Häufigkeit der Quiz-Nutzung und dem Examenserfolg in Seminaren mit der Möglichkeit, Kreditpunkte durch Onlinequiz zu erzielen, aber keine Mittelwertsunterschiede im Examenserfolg zwischen den Seminaren mit und ohne diese Quizmöglichkeiten. Leider klären auch viele Studien, die keinen Effekt optionaler Onlinequiz gefunden haben, die Frage nicht endgültig, da sie meist auf sehr schwachen Versuchsplänen basieren. (z.B. Harter & Harter 2004).

Wie etliche Untersuchungen aufgezeigt haben, ist mit höheren Testwerten in zeitnahen Prüfungen zu rechnen, wenn auf der Bearbeitung der Quiz bestanden wird (z.B. Daniel & Broida 2004, Narloch, Garbin, & Turnage 2006, Hautau et al. 2006a, b, Jacobs 2009, Hodge 2009, Johnson & Kiviniemi 2009). Das kann selbst dann gelingen, wenn die verpflichtenden Quiz nicht benotet werden (z.B. Jacobs 2009, McDaniel 2008) oder Mastery Quiz sooft wiederholt werden dürfen bzw. müssen, bis ein gesetztes Kriterium erzielt ist. Manchmal beschränkt sich der Vorteil des Quiz auf die kurzfristige Leistung und/oder zeigt sich nicht mehr in einem späteren Abschlussexamen (z.B. Azorlosa & Renner 2006, Urtel et al. 2006). Durch das Androhen und Durchführen von benoteten Tests ließ sich häufiger die Anwesenheit bzw. Mitarbeit im Seminar oder die Seminarvorbereitung bzw. die Studierzeit erhöhen (Wilder, Flood & Stromsnes 2001, Marcell 2005, Azorlosa & Renner 2006, Ruscio 2001, Hautau et al. 2006, Jacobs 2009). Der Einsatz benoteter Quiz bewirkt somit nachweislich einen höheren Arbeitseinsatz für diese Tests. Benotete Quiz bzw. leistungsabhängige Bonusquiz führen in der Regel auch zu höheren Examensleistungen oder Nachtests. (z.B.: Hautau et al. 2006b, Sansgiry et al. 2006, Narloch, Garbin, & Turnage 2006, Jacobs 2009, Hodge 2009). Bei benoteten 'offiziell nicht kontrollierten' Onlinetests muss möglicher Missbrauch aber weitgehend ausgeschlossen werden (Daniel & Broida 2004). Theoretisch erhöhen Quiz die Wissensstabilität, was man am besten in unangekündigten und unbenoteten Leistungsprüfungen erfassen kann. Letztlich steigern benotete Tests indirekt die Anforderungen an die Studierenden, indem sie zeitweise ein aufwändigeres Studium einfordern. Soll man diesen Aufwand einfordern und wenn ja, könnte man diesen auch ohne Androhung von Noten erreichen?

Theoretisch verbessern Seminar begleitende Quiz die abschließenden Prüfungsleistungen eher bei aufeinander aufbauenden Lerninhalten sowie vornehmlich bei relativ schwachen Studierenden oder solchen Studenten, die ohne Quiz keine ernsthafte Vor- bzw. Nachbearbeitung des Lehrstoffs durchführen, ihre Examensvorbereitung dann aber zu spät in Angriff nehmen, so dass die verfügbare Lernzeit nicht mehr ausreicht (z.B. hohes Procrastination; Tuckman 1998).

Insgesamt hat sich Testen mit Feedback als praktikable lernwirksame Studiermethode erwiesen. Im Hinblick auf eine objektive Leistungssteigerung lohnen sich Tests als Studierhilfen für die Studierenden aber letztlich nur dann, wenn sie im Vergleich zu ernsthaften al-

alternativen Maßnahmen entweder bessere Examensleistungen bewirkten, oder bei gleicher Lerneffektivität weniger Lernaufwand erforderten. Jedoch selbst bei fehlendem objektiven Leistungsvorteil in einer Abschlussklausur, sind Studierende von der Nützlichkeit der Tests überzeugt und fühlen sich besser auf das Examen vorbereitet, was dann insgesamt einen positiven Eindruck bei der meist subjektiv gestalteten Evaluation von Lehre hinterlassen mag.

## Literatur

- Abraham, J. & Sherman, T. (2009). A Comparison of Writing Strategies and Multiple Choice Tests on Text Comprehension and Retention. International Society for Exploring Teaching and Learning. <http://www.isetl.org/conference/presentation.cfm?pid=1013> [14.12.2009]
- Agarwal, P. K., Karpicke, J. D., Kang, S.H.K., Roediger, H.L., & McDermott, K.B. (2008). Examining the testing effect with open- and closed-book tests. *Applied Cognitive Psychology*, 22, 861-876.
- Azorlosa, J. L & Renner, C. H. (2006). The Effect of Announced Quizzes on Exam Performance. *Journal of Instructional Psychology*, 33 (4), 278-283.
- Bangert-Drowns, R., Kulik, J., & Kulik, C. (1991). Effects of frequent classroom testing. *Journal of Educational Research*, 85 (2), 89-99.
- Brosvic, M. & Epstein, M. L. (2007). Enhancing learning in the introductory course. *Psychological Record*, 57, 391-408.
- Brown, M. J. & Sternberg, C. (2009). The Effects of Prelecture Quizzes on Test Anxiety and Performance in a Statistics Course. Brooklyn College - City University of New York [http://homepage.mac.com/psychresearch/Sites/site2/publications/manuscripts/prelecturequizzes\\_mjbrown.pdf](http://homepage.mac.com/psychresearch/Sites/site2/publications/manuscripts/prelecturequizzes_mjbrown.pdf) [16.12.2009]
- Burns, D. J. & Vinchur, A. J. (1992). Effects of evaluative quizzes on test performance. *Journal of Instructional Psychology*, 19 (3), 148-154.
- Daniel, D. B. & Broida, J. (2004). Using Web-based quizzing to improve exam performance: Lessons learned. *Teaching of Psychology*, 31, 207-208
- Deck, D. W. (1998). THE EFFECTS OF FREQUENCY OF TESTING ON COLLEGE STUDENTS IN A PRINCIPLES OF MARKETING COURSE. Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University. <URL:http://scholar.lib.vt.edu/theses/available/etd-110298-195932/unrestricted/dis.pdf> [8.11.2001]
- Gier, V. S. & Kreiner, D. S. (2009). Incorporating Active Learning With PowerPoint-Based Lectures Using Content-Based Questions. *Teaching of Psychology* 36 (2), 134 - 139.
- Graham, R. B. (1999). Unannounced quizzes raise test scores selectively for mid-range students. *Teaching of Psychology*, 26 (4), 271-273.
- Hamaker, Ch. (1986). The Effects of Adjunct Questions on Prose Learning. *Review of Educational Research*, 56 (2), 212-242.
- Hautau, B., Turner, H.C., Carroll, E., Jaspers, K., Krohn, K., Megan, P., Williams, R.L. (2006a). Differential Daily Writing Conditions and Performance on Major Multiple-Choice Exams. *Journal of Behavior Education*, 15, 171-181.
- Hautau, B., Turner, H.C., Carroll, E., Jaspers, K., Parker, M. · Krohn, K., Williams, R.L. (2006b). Differential Daily Writing Contingencies and Performance on Major Multiple-Choice Exams. *Journal of Behavior Education*, 15, 259-276.
- Harter, C. L. & Harter, J. F. R. (2004) TEACHING WITH TECHNOLOGY: DOES ACCESS TO COMPUTER TECHNOLOGY INCREASE STUDENT ACHIEVEMENT? *Eastern Economic Journal*, 30 (4), 507-514.

- Haynie, W. J. (1997). [Effects of Anticipation of Tests on Delayed Retention Learning](#). Journal of Technology Education. Volume 9. Number 1.  
<http://vega.lib.vt.edu/ejournals/JTE/jte-v9n1/haynie.html>  
<http://scholar.lib.vt.edu/ejournals/JTE/v9n1/pdf/haynie.pdf>
- Haynie, W. J. (2003a). Effects of Take-Home Tests and Study Questions on Retention Learning in Technology Education. Journal of Technology Education, 14 (2), 6-18  
<http://scholar.lib.vt.edu/ejournals/JTE/v14n2/pdf/haynie.pdf>
- Haynie, W. J. (2007). Effects of Test Taking on Retention Learning in Technology Education: A Meta-Analysis. Journal of Technology Education, 18 (2)  
<http://scholar.lib.vt.edu/ejournals/JTE/v18n2/haynie.html>
- Hodge, G. K. (2009, March). Pedagogical best practices: The power of quizzing. Invited panelist at the 3rd Annual Redesign Alliance Conference, National Center for Academic Transformation, Orlando, FL. [NCAT Orlando S09 fo#1DA43F1.ppt](#)  
[http://www.center.rpi.edu/RedesignAlliance/2009\\_Conference/Presentation\\_Slides\\_09/Hodges\\_UNMPowerofQuizzingPanel.pdf](http://www.center.rpi.edu/RedesignAlliance/2009_Conference/Presentation_Slides_09/Hodges_UNMPowerofQuizzingPanel.pdf)
- Jacobs, B. (1999). Feedback nach Bearbeitung einer Klausur - Wird es genutzt ?  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/klausurfeedback.htm>
- Jacobs, B. (1999b). Was wird durch Aufgaben mit Feedback gelernt ?  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/transfer.htm>
- Jacobs, B. (2000). Feedback nach Bearbeitung einer Probeklausur - Wird es genutzt?  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/probefeedback.htm>
- Jacobs, B. (2002). Lerneffekte häufigen Testens in pädagogischen Umwelten  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/frequenttest.htm>
- Jacobs, B. (2003). Die Wirkung einer Probeklausur auf Lernleistung und aktuelle Angst in einer echten Klausur (Kurzfassung)  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/probeklausur/kurzfassung.htm>
- Jacobs, B. (2005). Selbst Fragen generieren oder vorgegebene Fragen beantworten ?  
[http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/selfquestion\\_vs\\_lehrerquestion.html](http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/selfquestion_vs_lehrerquestion.html)
- Jacobs, B. (2007a). Geld und Noten als extrinsische Motivatoren zur Verbesserung kognitiver Leistungen.  
 URN: urn:nbn:de:bsz:291-psydok-9644  
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2007/964/>
- Jacobs, B. (2007b). Gezieltes Studieren oder Aufgaben mit Feedback als Übungsmethoden - Erfahrungsbericht eines missglückten Internetexperimentes.  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/testread/testread.html>
- Jacobs, B. (2008a). Einige Überlegungen zu unmittelbarem und verzögertem Feedback  
[http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/unmittelbar\\_vs\\_delayed.html](http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/unmittelbar_vs_delayed.html)
- Jacobs, B. (2008b). Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback  
 URN: <urn:nbn:de:bsz:291-psydok-15597>  
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2008/1559/>
- Jacobs, B. (2008c). Was wissen wir über die Lernwirksamkeit von Aufgabenstellungen und Feedback. In: Thonhauser, J. (Hrsg.) 2008: Aufgaben als Katalysatoren von Lernprozessen. Waxmann, Münster, 99-113.
- Jacobs, B. (2009) Leistungssteigerung durch Notendruck ? - Die Wirkung der Benotung auf die Studierleistungen in einem Seminar.  
 URN: <urn:nbn:de:bsz:291-psydok-25299>  
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2009/2529/>
- Jacobs, B. (1998-2007). Forschungsprojekt: Aufgaben stellen und Feedback geben.  
<http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/projekt.htm>

- Jacobs, B., Bedersdorfer H.-W., Bohse-Wagner, N. (1983). Angstabbau durch Transparenz - Eine quasiexperimentelle Felduntersuchung zur ökologischen Validierung eines theorieorientierten Interventionsprogramms zur Reduktion von Angst in der Prüfung. Arbeitsbericht Nr. 18 aus der Fachrichtung Allgemeine Erziehungswissenschaft. Universität des Saarlandes. Kurzbericht dazu: <http://www.phil.uni-sb.de/~jakobs/trans/studie2.html>
- Johnson, B. C. & Kiviniemi, M. T. (2009). The Effect of Online Chapter Quizzes on Exam Performance in an Undergraduate Social Psychology Course. *Teaching of Psychology*, 36 (1) 33-37.
- Khalaf, A. S. S. & Hanna, G. S. (1992). The impact of classroom testing frequency on high-school-students achievement. *Contemporary Educational Psychology*, 17 (1), 71-77.
- Kibble, J. (2007). Use of unsupervised online quizzes as formative assessment in a medical physiology course: effects of incentives on student participation and performance. *Advances in Physiological Education*, 31, 253 - 260.
- Kling, N., McCorkle, D, Miller C. & Reardon, J. (2005). The impact of testing frequency on student performance in a marketing course. *Journal of Education for Business* (81) 67-72, zitiert nach Kuo, T. & Simon, A. (2009).
- Kluger, A. N. & DeNisi, A. (1996). The Effects of Feedback Interventions on Performance: A Historical Review, a Meta-Analysis, and a Preliminary Feedback Intervention Theory. *Psychological Bulletin*, 119 (2), 254-284.
- Krohn, K. R., Parker, M. R., Foster L. N., Aspiranti, K. B. , McCleary, D. F. & Williams, R. L. (2009). Effects of Writing-Related Contingencies on Both Quality of Writing and Multiple-Choice Exam Performance in Large College Courses. *The Behavior Analyst Today* , 9 (2,3), 184-197.  
<http://baojournal.com/VOL-9/BAT-9.3-9.4.pdf>
- Kuo, T. & Simon, A. (2009). How many Tests do we Really Need? *College Teaching* 57 (3), 56-160.
- Marcell, M. (August, 2005). Effectiveness of online quizzing in increasing class preparation and participation. Poster presented at the annual meeting of the American Psychological Association, Washington, D.C.  
<http://www.cofc.edu/~marcellm/quiz%20research/APA%20summary%20-%20quiz%20research.htm>  
auch: Marcell (2008) Effectiveness of Regular Online Quizzing in Increasing Class Participation and Preparation *International Journal for the Scholarship of Teaching and Learning*, 2 (1)
- McDaniel, M. (2008) How Quizzes Can Enhance Student Learning. *Iteach 2008*.  
[http://iteach.wustl.edu/ITeach2008/documents/iteach2008\\_mcdaniels\\_quizzes.pdf](http://iteach.wustl.edu/ITeach2008/documents/iteach2008_mcdaniels_quizzes.pdf) [27.8.2009]
- Maurer, T. W. (2006). DAILY ONLINE EXTRA CREDIT QUIZZES AND EXAM PERFORMANCE. *Journal of Teaching in Marriage and Family*, 6, 227-238.
- Mayer, R. E., Stull, A., DeLeeuw, K., Almeroth, K., Bimber, B., Chun, D., Bulger, M., Campbell, J., Knight, A., & Zhang, H. (2009). Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary Educational Psychology*, 34, 51-57.
- Morrison, G. R., Ross, S. M., Gopalakrishnan, M. & Casey, J. (1995). The Effects of Feedback and Incentives on Achievement in Computer-Based Instruction. *Contemporary Educational Psychology*, 20, 32-50.
- Myers, C.B. & Myers, S.M. (2007). Assessing Assessment: The Effects of Two Exam Formats on Course Achievement and Evaluation *Innov High Educ* 31: 227-236
- Narloch, R., Garbin, C. P. & Turnage, K.D. (2006). Benefits of Prelecture Quizzes. *Teaching of Psychology*. 33 (2), 109-112.
- Oliver, R. & Williams, R. L. (2005). Direct and Indirect Effects of Completion Versus Accuracy Contingencies on Practice-Exam and Actual-Exam Performance. *Journal of Behavioral Education*, 14 (2), 141-152.

- Poljičanin, A., Čarić, A., Vilović, K., Košta, V., Guić, M.M., Aljinović, J. & Grković, I. (2009). Daily Mini Quizzes as Means for Improving Student Performance in Anatomy Course. *Croatian Medical Journal*, 50 (1), 55-60.  
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2657562/>
- Rakes, G. C. (2008). Open Book Testing in Online Learning Environments. *Journal of Interactive Online Learning*, 7 (1), 1-9.
- Roediger, H. L. & Karpicke, J. D. (2006). Test-Enhanced Learning: Taking Memory Tests Improves Long-Term Retention. *Psychological Science*, 17 (3), 249-255.
- Ruscio, J. (2001). Administering quizzes at random to increase students reading. *Teaching of Psychology*, 28 (3), 204-206.
- Ryan, N. (2006). Motivating Novice Students to Read their Textbooks. *Journal of Instructional Psychology*, 33 (2), 136-140.
- Ryan, C. S. & Hemmes, N. S. (2005). Effects of the contingency for homework submission on homework submission and quiz performance in a college course. *Journal of Applied Behavior Analysis*. 38, 79-88.
- 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>  
<http://www.ajpe.org/aj7005/aj7005103/aj7005103.pdf>
- Shapiro, A. An empirical study of personal response technology for improving attendance and learning in a large class. *Journal of the Scholarship of Teaching and Learning*, 9 (1), 13 - 26.
- Sun, J. & McDaniel, M. A. (2008, November). The testing effect: Experimental evidence from a college course. Poster presented at the Midstates Consortium for Math and Sciences, Chicago, IL. [\[PDF\]](#)  
[http://psych.wustl.edu/memory/TELC/Sun\\_McDaniel\\_Poster\\_2008.pdf](http://psych.wustl.edu/memory/TELC/Sun_McDaniel_Poster_2008.pdf)
- Tuckman, B. L. (1996). The Relative Effectiveness of Incentive Motivation and Prescribed Learning Strategy in Improving College Students' Course Performance. *Journal of Experimental Education*, 64 (3), 197-210.
- Tuckman, B. L. (1998). Using Tests as an Incentive to Motivate Procrastinators to Study. *Journal of Experimental Education*, 1998, 66 (2), 141-147.
- Tuckman, B. W. (2001) USING FREQUENT TESTING TO INCREASE STUDENTS' MOTIVATION TO ACHIEVE.  
 URL: <http://all.successcenter.ohio-state.edu/belgium-paper/BWT-belgium-paper.htm> [21.9.2001]
- Urtel, M., Bahamonde, R.E, Mikesky A.E., Udry E.M., Vessely J.F. (2006). On-line Quizzing and its Effect on Student Engagement and Academic Performance. *Journal of Scholarship of Teaching and Learning*, 6 (2), 84 - 92.
- Wellmann, G. S. (2005) Comparing Learning Style to Performance in On-Line Teaching: Impact of Proctored v. Un-Proctored Testing. *Journal of Interactive Online Learning*. 4 (1), 20-38.
- Wilder, D.A., Flood, W.A., & Stromsnes, W. (2001). The use of random extra credit quizzes to increase student attendance. *Journal of Instructional Psychology*, 28 (2), 117-120.