

Autor: Bernhard Jacobs, Fachrichtung 5.1 Erziehungswissenschaft der Universität des Saarlandes
Version vom 13.4.2010

Url des Originals: http://www.phil.uni-sb.de/~jakobs/wwwartikel/tests_als_studierhilfen/testen_vs_nichts.html

Leistungssteigerung ohne Notendruck ? - Die Wirkung verpflichtender, unbenoteter Quiz auf die Studierleistung

Abstract

Ziel der Studie war es, in der Universitätspraxis die Auswirkung eines verpflichtenden, aber unbenoteten Onlinequiz mit unmittelbarem Feedback auf den längerfristigen Lernerfolg zu überprüfen. Studierende aus zwei Seminaren wurden nach Zufall auf 2 Gruppen aufgeteilt und nahmen auf der Basis eines Treatment-Reversal-Designs einmal am Onlinequiz teil, während sie das andere Mal als No-Treatment-Kontrollgruppe dienten. Quiz- und Lernerfolgsmaße enthielten parallele Items, um die vorrangig angestrebten Lehrziele Verstehen und Anwendung in Form eines nahen Transfers zu erfassen. Der Lernerfolg wurde im Rahmen eines unangekündigten, unbenoteten Abschlussquiz, ca. 4 Wochen nach dem letzten Quiz, erhoben. Tendenziell sprechen die Ergebnisse für die Lernwirksamkeit von verpflichtenden, unbenoteten Onlinequiz, wobei sich der Quizvorteil jedoch nur bei einem der beiden Quiz statistisch sichern ließ und im erwarteten geringen bis mittleren Effektstärkebereich liegt.

Schlagworte: Testen, Feedback, Quiz, Lernförderung, extrinsische Motivierung

Einleitung, Problemstellung und Zielsetzung

Testen im Anschluss an eine Instruktion beschränkt sich nicht auf die Diagnostik der Lehrziele, sondern bewirkt zugleich eine Veränderung im Lernenden. Die Fragen fordern dazu auf, sich wichtige Informationen in Erinnerung zu rufen und dieses Abruftraining fördert das langfristige Behalten. Der Testeffekt im Sinne eines Retrievaleffektes zeigt sich insbesondere dann sehr deutlich, wenn viele Fragen richtig beantwortet werden. Aber, je mehr Fehler auftreten, desto notwendiger erscheint informatives Feedback, z.B. in Form einer Rückmeldung der korrekten Antwort (=Knowledge of correct response [KCR]); siehe zusammenfassend Jacobs (2008).

Mittlerweile liegen viele empirische Studien vor, welche die Lernwirksamkeit des Testens mit und teilweise sogar auch ohne Feedback hinreichend empirisch belegten (z.B.: LaPorte & Voss 1975, Cull 2000, Butler et al. 2007, Roediger & Karpicke 2006, Kang et al. 2007). So konnte etwa Jacobs (2006) in einem Laborexperiment zum Erlernen der Bundesstaaten der USA nachweisen, bereits reines Testen ohne jedes Feedback verbessere langfristiges Behalten. Studierende, welche am Ende der Bundesstaatenübung einen Test zur korrekten Zuordnung der Staatsnamen zu den Regionen auf der Landkarte vornahmen, erzielten eine Woche später einen Behaltensvorteil von ca. 10 % [Effektstärke ca. $d = 0.5$] gegenüber Studierenden, welche diesen Test nicht absolvierten. Auch der zusätzliche Lernnutzen des Feedbacks gegenüber reiner Testung kann als gesichert gelten und wird durch die Metaanalysen von Bangert-Drowns et al. (1991) und sowie Kluger und DeNisi (1996) gestützt.

Allerdings sind die meisten Laborstudien mit allerlei Vereinfachungen und Restriktionen verbunden, die eine unproblematische Generalisierung der Befunde auf die pädagogische Praxis erschweren (siehe zusammenfassend dazu: Jacobs 2010). Im Labor findet die Testung meist unmittelbar nach einer relativ kurzen Instruktionsphase statt, während in der Praxis der Stoffumfang wesentlich höher ausfällt und der Zeitabstand zwischen Lernaneignung und Testung auch deutlich weiter auseinander liegt. Als beliebtes Forschungsparadigma gilt triviales Faktenwissen, das in der Regel jegliches Vorwissen der Lernenden ausschließt (z.B. Eskimo-Vokabeln, obskure Fakten) und den Lernerfolg meist als einfaches, relativ kurzfristiges Be-

halten des exakt zuvor Eingebübten erfasst. Unter solchen Bedingungen entfaltet sich dann "the power of testing memory" (Roediger & Karpicke 2006). In letzter Zeit findet man aber auch zunehmend Test- und Feedbackstudien innerhalb des realen Schulbetriebs (z.B. Dihoff et al. 2004, McDaniel et al. 2007, Brosvic & Epstein 2007). Diese Studien bestätigen die Lernwirksamkeit des Testens mit Feedback, lassen aber dennoch etliche pädagogische Fragen offen.

Im Universitätsbetrieb sollten vornehmlich anspruchsvolle Lehrziele vermittelt werden, die zumindest einen nahen Transfer erzeugen und das erworbene Wissen etwas längerfristiger konservieren. Jacobs (2009a) konnte solche Effekte durch Testen in einem Feldexperiment relativ deutlich bestätigen. Allerdings erwarteten die Studierenden ein benotetes Quiz. Der Autor wollte weniger den Test- und Feedbackeffekt untersuchen, sondern die durch Testung ausgelöste Wirkung des Notendrucks auf die Studierleistung erfassen und verglich deshalb die Auswirkungen von Quiz unter der Erwartung einer Benotung mit der unter der Erwartung einer unbenoteten, aber verpflichtenden Quizbearbeitung. Hierbei zeigten sich konsistent klare Leistungsvorteile für die benotete Variante sowohl in den Quiz wie in den quizanalogen Aufgaben einer unangekündigten unbenoteten Abschlussklausur gegen Ende des Semesters. Notendruck erhöhte somit die Studierleistung, führte aber nachweislich auch zu mehr Angst vor den Quiz.

Die Androhung einer Benotung verschärft in der Regel die Studieranforderungen, erhöht den Stress und begünstigt die extrinsische Motivation. Eine pädagogische Herausforderung lautet, ob die Bearbeitung eines Quiz auch ohne jegliche Form von Notenandrohung einen Lernvorteil bewirken kann. Im Mittelpunkt der Arbeit standen daher die Fragen, ob ein unbenotetes Quiz in der Schulpraxis als effektive Lernchance tatsächlich genutzt wird, eine nachhaltige Behaltenswirkung hinterlässt und letztlich besser als gar nichts einzuschätzen ist. Das theoretische Interesse gilt dann auch weniger der Lernmotivierung durch Tests, sondern eher dem Test -und Feedbackeffekt, der allerdings auch eine gewisse Lernmotivierung voraussetzt. Vorliegende Untersuchung knüpft an eine Teilstudie von Jacobs (2009a) an, die den Behaltenserfolg des unbenoteten Testens im realen Universitätsbetrieb partiell bestätigen konnte, aber gewisse methodische Unzulänglichkeiten aufwies, die diesmal vermieden werden sollten.

Anforderungen an die Studie

Durch die Bearbeitung eines unbenoteten Übungstests mit Feedback - im folgenden Quiz genannt - sollte die Lernleistung nachhaltig gesteigert werden, was durch einen später erhobenen Lernerfolgs- bzw. Kriteriumstest nachgewiesen werden sollte. Um hierbei eine ernsthafte pädagogische Relevanz für die Praxis zu erlangen, sind meiner Meinung nach einige pädagogische Anforderungen an die Studie zu erfüllen.

1. Die Quizbearbeitung sollte innerhalb des schulischen Umfelds in praktikabler Weise hinreichend angeregt werden. Insbesondere sollte sicher gestellt sein, dass alle Studierenden die Übung auch tatsächlich bearbeiten.
2. Die Lehrziele der Quiz sollten sich weniger auf Faktenwissen, sondern vornehmlich auf Verstehen und Anwendung beziehen. Im Mittelpunkt der Übungsanforderungen stehen demnach keine reinen Memorierungsleistungen, sondern Verständnis und Anwendung.
3. Durch die Quiz sollte zumindest ein minimaler Lerntransfer gefördert werden. D.h. Die Aufgaben in der Übung und im Kriteriumstest müssen die gleichen Lehrziele mit nicht identischen Aufgaben erfassen.
4. Der Lernerfolg sollte vornehmlich die experimentellen Auswirkungen erfassen und eine gewisse Lernstabilität aufweisen. Da eher langfristige Lerneffekte interessierten, wurde ein Retentionsintervall von mindestens 4 Wochen angestrebt.

Praktikable Motivierung zur Übungsbearbeitung

Hinreichende Untersuchungen im pädagogischen Feld belegen die geringe Nutzung und mäßige Bearbeitungsqualität freiwilliger Übungen (siehe einige Angaben dazu in Jacobs 2009a). Dies wurde durch eigene Lehrerfahrung immer wieder bestätigt und erreichte im WS09/10 seinen nicht mehr zu überbietenden Höhepunkt, als kein einziger Student einen fakultativen Arbeitsauftrag [Ankündigung freiwillig: Ergebnisse werden nicht kontrolliert, individuelle Musterlösung wird angeboten] überhaupt zu Kenntnis nahm, geschweige denn bearbeitete.

Mir sind keine soliden Studien bekannt, welche etwa den Nachweis erbracht hätten, ein optionales Angebot von Onlinequiz führe im Mittelwert zu besseren Studierleistungen. Demgegenüber lassen sich etliche Studien finden, welche von besseren Leistungen berichten, wenn auf der Bearbeitung von Arbeitsaufträgen oder Tests bestanden wird (Jacobs 2010). Die meisten Studierenden haben im Studium vermutlich wenig Zeit, sich aus Interesse, Neugier oder besonderem Kompetenzstreben unmittelbar Noten irrelevanten Aufgaben oder Problemen zu widmen, sondern müssen sich auf die vielen zertifikationsrelevanten Anforderungen innerhalb eines Semesters konzentrieren. Deshalb wurden die Studierenden schon zu Beginn des Seminars darauf hingewiesen, bestimmte, - die experimentellen- Übungen, seien absolut verpflichtend und Voraussetzung für die Bestätigung eines erfolgreichen Seminarabschlusses. Zwingend notwendig sei die Quizbearbeitung. Auf eine Bewertung bzw. Benotung der Leistung werde verzichtet. Der Seminarleiter erwarte aber ein angemessenes Bemühen. Diese Quizverpflichtung aktiviert natürlich auch eine Form extrinsischer Motivierung, verzichtet aber auf das Androhen massiver Konsequenzen.

Aufgaben in Quiz und Kriterium

Viele Laborstudien verwandten als Aufgaben triviales Faktenwissen. Im späteren Kriteriumstests (=Lernerfolgstest) wurde dann geprüft, wie viele dieser Fakten die Probanden noch behalten hatten. Auf unterstem Lehrzielniveau "Wissen Kenntnisse" müssen Übungs- und Kriteriumsaufgaben weitgehend identisch sein. Es macht schlichtweg wenig Sinn, die Frage nach einem Transfer zu stellen, weil es im Wesentlichen nur um Erinnerung geht. Werden im Quiz- und Kriteriumstest anspruchsvollere, aber dieselben Fragen gestellt, - wie in vielen Studien innerhalb der Schulpraxis (z.B. Brosvic & Epstein 2007), so ist ein Abgrenzen der Lehrziele von Erinnerung sehr schwierig. Denn das Lehrziel Anwenden erfordert beispielsweise den Nachweis, bisher noch nicht realisierte Phänomene oder Beispiele entsprechend einer Regel oder Prozedur anzuwenden. Spätestens ab dem Lehrzielniveau Verstehen sollten Quiz- und Kriteriumsaufgaben nicht mehr identisch sein. Im realen Schulsystem hat jede etwas anspruchsvollere Übung zumindest einen minimalen Transfer zum Ziel.

Deshalb unterscheiden sich in dieser Untersuchung die Übungsaufgaben von den Kriteriumsaufgaben. Diese Forderung gilt im Übrigen für alle Aufgaben, nicht nur für den überwiegenden Anteil der Aufgaben mit Lehrzielniveau Verstehen und Anwendung, sondern auch für die einfachsten Fragen, die somit fast ausnahmslos mindestens minimales Verstehen erforderten. Ziel der Aufgabenkonstruktion war es, zu jeder Quizaufgabe eine annähernd vergleichbare, parallele Kriteriumsaufgabe zu konzipieren. Hierbei kamen verschiedene Methoden zum Einsatz: Bei sehr einfachen Aufgaben meist Paraphrasierung und Transformation, bei MC-Aufgaben unter anderem die Wahl neuer Distraktoren, bei anspruchsvolleren Verstehensaufgaben z.B. neue Beispiele, bei Anwendungsaufgaben verschiedene Daten oder leicht veränderte Anforderungen. Diese Anforderungen gehen damit über das Vorgehen von Jacobs (2009a) hinaus, der nur zur Hälfte identische und parallele Aufgaben im Kriteriumstest verwandte. Die Interpretierbarkeit der Untersuchungsergebnisse hängt ganz entscheidend davon ab, wie gut diese Form der Aufgabenkonstruktion hier gelungen ist. Denn zum einen sollte ein gewisser Transferanspruch berechtigt sein, damit potentielle Treatmentunterschiede auch eine pädagogische Relevanz besitzen, zum andern würden zu hohe Transferanforderungen bei beiden Methoden keine nennenswerten Lernvorteile erwarten lassen. Auf der Html-Seite

http://www.phil.uni-sb.de/~jakobs/wwwartikel/tests_als_studierhilfen/aufgabenbeispiele.html werden einige Quiz- sowie die dazugehörigen parallelen Lernerfolgs bzw. Kriteriumsaufgaben dargestellt (siehe auch hier im Anhang).

Art und Zeitpunkt des Kriteriumstests

Der Kriteriumstest selbst umfasst für jedes Quiz 10 parallele Aufgaben, die insgesamt jedoch deutlich mehr als 30 Antworten verlangen und als Subtests des Kriteriumstest den Lernerfolg einer Übung möglichst zuverlässig messen sollten. Der Abstand zwischen den beiden Übungen und dem Kriteriumstest variiert in der Länge, beträgt aber wenigstens 4 Wochen. Der Kriteriumstest wurde weder angekündigt, noch benotet und so terminiert, dass eine vom Studierenden selbst initiierte Vorbereitung oder Wiederholung der in den Übungen vermittelten Lehrziele sehr unwahrscheinlich war, damit der Lernerfolg möglichst eindeutig auf die Wirkung der Quiz zurückgeführt werden kann. Die Parallelität der Quiz- und Kriteriumsaufgaben und die Länge des Behaltensintervalls lassen keine großen Effekte erwarten. Aber das Anstreben kleiner Effekte mit hoher pädagogischer Relevanz erscheint mir sinnvoller als die Jagd nach hoher Effektstärke ohne pädagogischen Belang.

Untersuchungsansatz

An der Untersuchung nahmen Studierende des Lehramts teil, die ein vom Verfasser dieser Arbeit geleitetes Proseminar zur Pädagogischen Diagnostik im WS 09/10 belegten. Das erste Seminar fand montags statt, das zweite Seminar donnerstags zu gleichen Zeit. Beide Seminare können als Parallelklassen aufgefasst werden, da sie nach identischem Seminarplan den gleichen Lehrstoff mit denselben Lehrmaterialien zu bearbeiten hatten und vom selben Dozenten unterrichtet wurden. Alle hier berichteten Zeitangaben beziehen sich auf das Seminar 1. Für Seminar 2 sind jeweils 3 Tage hinzu zu addieren

Zu Beginn des Semesters wurden die Studierenden innerhalb jedes Seminars nach Zufall auf die Gruppen A und B zugeteilt, bei der statistischen Analyse und Auswertung dann die A- bzw. B Gruppen der einzelnen Seminare zu einer gemeinsamen A- bzw. B-Gruppe zusammengefasst. Die Zuteilung der Studierenden zu den Gruppen A und B entspricht einer stratifizierten Randomisierung und liefert die Grundlage für den in Tabelle 1 dargestellten, experimentellen Versuchsplan.

Tabelle 1: Versuchsplan der Onlineübungen

| | Übung 1 | Übung 2 | Abschlussquiz |
|------|----------|----------|---------------|
| | 13.11.09 | 11.12.09 | 11.1.10 |
| R A: | Quiz 1 | | O |
| R B: | | Quiz 2 | O |

Gruppe A bearbeitete das Quiz 1 "Skalen und Stichproben", während Gruppe B als No-Treatment-Kontrollgruppe fungierte. Dann wurden die Bedingungen getauscht. Nun bearbeitete nur die Gruppe B das Quiz 2 "Testgütekriterien und Normierung". Gruppe A nahm nicht an diesem Quiz teil. Ca. 4 Wochen nach dem zweiten Quiz nahmen alle Studierenden am unangekündigten und unbenoteten Abschlussquiz im Seminar teil, welches für jedes Quiz 10 parallele Aufgaben enthielt, die zusammengefasst das jeweilige Behaltenskriterium eines bestimmten Quiz messen sollten. Wie die vergleichbaren Ergebnisse beider Gruppen hinsichtlich Abiturnotendurchschnitts, Alter und Geschlecht in der Tabelle 2 aufzeigen, war die Randomisierung sehr erfolgreich.

Tabelle 2: Vergleichbarkeit der Gruppen A und B (N für jede Gruppe jeweils 29)

| | A | | B | | |
|--------------------------|------|------|------|------|----|
| | M | s | M | s | |
| Abiturnotendurchschnitt: | 2.19 | 0.62 | 2.22 | 0.57 | ns |
| Alter: | 22.6 | 3.1 | 22.7 | 4.2 | ns |
| Geschlecht: | 76%w | | 62%w | | ns |
| Quiz Aufgabentypen: | 60,1 | 18.0 | 58,6 | 14.2 | ns |

Im unangekündigten, unbenoteten Abschlussquiz wurden weitere Lehrziele überprüft, welche die hier interessierende Fragestellung nicht tangieren, u. a. zum Seminarthema Aufgabentypen und Aufgabenformen, welches vor den Themen der experimentellen Quiz anstand. Wie man erkennt, erzielten beide Gruppen hoch vergleichbare Prozentsätze korrekter Lösungen, was auch für eine vergleichbare Lernaneignung beider Gruppen während des Seminars spricht.

Versuchsablauf

Die Seminare waren so geplant, dass die Untersuchung zugleich wesentliche Übungsbestandteile der Seminare abdeckte. Aus diesem Grunde wurden die Studierenden nicht darüber informiert, dass sie zugleich an einem Experiment teilnahmen. Erst 2 Wochen nach Abschluss aller Erhebungen erhielten die Studierenden aufklärende Informationen über Sinn und Zweck dieser Untersuchung. Im Seminarplan, der via Internet jederzeit eingesehen werden konnte, waren die Termine der Onlinequiz festgelegt worden. Das Proseminar Pädagogische Diagnostik gehört zu den Pflichtveranstaltungen der Lehramtsstudierenden. Wenngleich der benotete Schein für dieses Seminar abgeschafft wurde, müssen die Studierenden gewisse Leistungen erbringen, um eine erfolgreiche Teilnahme am Seminar beim Prüfungsamt belegen zu können. Zu diesen verpflichtenden Arbeitsanforderungen gehörten neben einigen Gruppenaufträgen oder Kurzreferaten auch die experimentellen Quiz. Das Thema der Quiz war durch explizite Angabe im Seminarplan, z.B. "Gruppe A: Onlineübung Skalen- und Stichproben" bekannt gegeben worden. Es bezog sich auf die Inhalte der vorausgehenden Seminarsitzung(en). Der Seminarleiter gestaltete diese Sitzungen weitgehend dozierend im Sinne des darstellenden sowie Fragen entwickelnden Unterrichts. Hierbei bemühte er sich, einen möglichst gleichartigen Unterricht in beiden untersuchten Seminaren zu halten. Die Thematik wurde sehr gründlich und ausführlich dargelegt und durch eine umfangreiche Powerpoint-Folie unterstützt.

Unmittelbar nach einer quizrelevanten Seminarsitzung informierte der Seminarleiter alle Studierenden via Email, die Powerpointfolie zur Thematik der vorausgegangenen Seminarsitzung sei jetzt von der Homepage des Seminars aus erreichbar. Im Seminar wurde öfter darauf hingewiesen, es sei ratsam, die Folien aufmerksam zu studieren, da man so von den Quiz besonders profitieren könne. Darüber hinaus wurde keine weitere Literatur oder Lerngrundlage empfohlen. Der Seminarleiter erklärte zu Beginn sowie mehrmals im Verlauf des Seminars, das Onlinequiz sei absolut verpflichtend, das Ergebnis werde aber nicht benotet. Einige wenige Studierenden lieferten Leistungen ab, die man als mangelhaft oder ungenügend bewerten könnte, was aber keinerlei direkte extrinsische, zertifikationsrelevante Konsequenz nach sich zog. Zwei Studierende entschuldigten sich spontan für Ihre schwachen Leistungen und gaben mehr oder weniger plausible Gründe dafür ab. Ihnen wurde versichert, das Ergebnis werde nicht benotet. Ohne die Studierenden vorab darauf hinzuweisen, wurde auf eine Mindestleistung beim Quiz bewusst verzichtet, um jegliche Art einer Benotung kategorisch auszuschließen. Erst nach Abschluss der gesamten Untersuchung wurden Studierende mit unzumutbar schwachen Leistungen dazu aufgefordert, in einem Sonderarbeitsauftrag das entsprechende Thema nachzuarbeiten und einzureichen.

Die Onlinequiz wurden 4 Tage nach der quizrelevanten Seminarsitzung für 2 Tage frei geschaltet. Der Studierende erhielt via Email die Information über die URL sowie den Benut-

zernamen und das Passwort und sollte das Quiz von zu Hause aus via Internet bearbeiten. Zu Beginn wurden die Lernvorteile eines Quiz vermittelt, sowie einige pädagogisch hilfreiche Empfehlungen gegeben, wie man das Quiz bearbeiten sollte. Die Quizbearbeitung unterlag keinerlei Zeitbegrenzung. Im Median bearbeiteten die Studierenden beide Quiz jeweils 26 Minuten lang. (Quiz 1: 25,7 ; Quiz 2: 26,1)

Alle Aufgaben waren auf einer Html-Seite untergebracht und umfassten ein weites Spektrum unterschiedlicher Aufgabenformate (MC mit einer oder mehreren korrekten Lösungen, Serien von True-false Aufgaben, Zuordnungsaufgaben, Short Answer Aufgaben). Meistens erforderte eine Aufgabe mehrere Antworten, so dass die Gesamtanzahl der Antworten deutlich die Anzahl der Aufgaben übersteigt. Nach Bestätigung einer Aufgabenbearbeitung, erhielt der Studierende die Rückmeldungen "richtig/falsch" (KOR), die korrekte(n) Lösung(en) (KCR) sowie in vielen Fällen elaboriertes Feedback. Die korrekten Lösungen einer bearbeiteten Aufgabe konnten beliebig oft angefordert werden (siehe Aufgabenbeispiele). Erst nach Bearbeitung aller Aufgaben hatte der Studierende die Möglichkeit, das Quiz zu beenden und seine Teilnahme am Quiz zu bestätigen. Danach erhielt er eine Rückmeldung zum Prozentsatz der korrekten Lösungen. Tabelle 3 stellt die wesentlichen Quizbedingungen im Überblick dar.

Tabelle 3: Überblick zu den wichtigsten Quiz- und Kriteriumsbedingungen.

Quizankündigung: klar festgelegt und lange bekannt.

Basis der Quizlehrziele: Seminarsitzung und Powerpoint-Folie des Dozenten.

Quizfunktion: Überprüfung des vermittelten -und selbst nachzuarbeitenden - Lehrstoffs

Quizrelevante Lehrziele: transparent, (einige Aufgabenbeispiele in der Powerpointfolie)

Quiztermin: vier bis fünf Tage nach der Seminarsitzung

Quizteilnahme: verpflichtend, relevant für Seminarschein

Quizablauf: Zeit unbeschränktes, extern unkontrolliertes Online-Quiz von zu Hause via Internet

Quizbewertung: unbenotet, bereits zu Seminarbeginn als unbenotet angekündigt, keine Mindestleistung erforderlich

Aufgabenbearbeitung: jede Aufgabe musste aktiv bestätigt werden.

Feedback: unmittelbar nach Beantwortung einer Aufgabe: KOR, KCR, gegebenenfalls elaboriertes Feedback

Ergebnisrückmeldung: unmittelbar nach Quizbeendigung: Prozentsatz korrekter Lösungen

Abschlussquiz: Kriterium: unangekündigt und unbenotet, parallele Aufgaben

Abschlussquizablauf: streng kontrolliert im Seminar

Der Lernerfolgstest fand ca. 8 Wochen nach Quiz 1 und ca. 4 Wochen nach Quiz 2 in Form eines Abschlussquiz statt. Die meisten Studierenden belegten neben dem Proseminar Pädagogische Diagnostik auch die Vorlesung "Pädagogische Diagnostik", in der ähnliche Themen, wenn auch in unterschiedlicher Akzentuierung und Intensität zur Sprache kamen. Am Ende des Semesters stand für die TeilnehmerInnen der Vorlesung eine benotete Abschlussklausur an. Das Abschlussquiz des Proseminars wurde deshalb 3 Wochen vor dieser Abschlussklausur angesetzt, um potenzielle Übungseffekte, die sich aus der möglichen Vorbereitung auf die Vorlesungsklausur hätten ergeben können, weitgehend auszuschließen. Da man sich auch nicht auf das Abschlussquiz des hier untersuchten Proseminars vorbereiten sollte, wurde dieses überraschend ohne Vorankündigung im Seminar ausgeteilt. Die Korrelation zwischen dem Gesamtergebnis des unbenoteten Abschlussquiz mit der späteren benoteten Abschlussklausur der Vorlesung beträgt im Übrigen $r=.54$ (persönliche Mitteilung von Prof. S. Buch). Um eine hinreichende Motivierung zur Bearbeitung des Abschlussquiz zu gewährleisten, wurde in Anlehnung an Jacobs (2009b) eine besondere pädagogische Motivierungsmaßnahme eingeleitet, welche die Lernwirksamkeit einer Testung propagierte, den Nutzen des Testens für die Klausur der Vorlesung herausstellte sowie neben der Rückmeldung des Leistungsergebnisses, persönliche Rückmeldungen des Seminarleiters für Leistungsverbesserungen gegenüber den vorherigen Quiz in Aussicht stellte.

Ergebnisse

Die Aufgaben aus den beiden Onlinequiz wurden von Jacobs 2009a übernommen, die entsprechenden Aufgaben des Kriteriumstests überwiegend neu konstruiert. Trotz der großen Leistungsstreuungen in den Quiz und den Lernerfolgstests ergab eine Analyse mit Tukeys-Boxplot keine Ausreißer, so dass alle verfügbaren Testwerte in der Analyse verblieben.

Tabelle 4 zeigt die deskriptiven Ergebnisse aller Studierenden für beide Quiz in der Onlineversion und im späteren Kriteriumstest in der Papier- und Bleistift-Version. Die via Cronbachs α ermittelten Konsistenzreliabilitäten der Onlinequiz fallen mit .78 und .76 sehr zufrieden stellend aus und entsprechen von der Höhe den Ergebnissen von Jacobs (2009a). Auch die Reliabilitäten der Kriteriumstests im Abschlussquiz erreichen eine zufrieden stellende Höhe und erfüllen damit eine deutliche Verbesserung gegenüber den schwachen Konsistenzen aus der Studie von Jacobs 2009a. Der angestrebte Reliabilitätsgewinn in dieser Studie geht im Wesentlichen auf die Verdoppelung der Aufgabenanzahl zurück. Nur bei den Studierenden unter der jeweiligen Quizbedingung lassen sich Korrelationen zwischen Quiz und Kriterium ermitteln, die als untere Schätzung der Parallelretestreliabilität aufzufassen sind. Sie lauten für Quiz1: $r_{tt} = .79$ ($N=27$; Zeitabstand ca. 8 Wochen) und für Quiz 2 $r_{tt} = .53$ ($N=29$; Zeitabstand ca. 4 Wochen). Hierbei gilt zu beachten, dass die Kriteriumstests nur eine, wenn auch sehr große Teilmenge der Quizaufgaben in Form von parallelen Aufgaben enthalten. Ob so die schwächere Retestreliabilität von Quiz 2 erklärt werden kann, sei dahin gestellt. Die Korrelationen zwischen allen Quizaufgaben und den 10 jeweils für die Parallelisierung ausgewählten Aufgaben der Quiz fallen mit .96 für das Quiz 1 und .90 für das Quiz 2 hinreichend hoch aus. Der Vollständigkeit halber sei noch die Korrelation zwischen beiden Lernerfolgsmaßen im Abschlussquiz von $r = .64$ ($N=58$) genannt.

Tabelle 4: Deskriptive Ergebnisse zu den experimentellen Quiz und Kriteriumsmaßen

| Test | Anzahl Aufgaben | Reli α | M | s | N |
|-------------|--------------------|------------------|------|------|----|
| Quiz 1 | 15 | .78 | 67.1 | 16.1 | 30 |
| Quiz 2 | 24 | .76 | 66.8 | 14.5 | 29 |
| Kriterium 1 | 10 | .80 | 53.4 | 21.6 | 58 |
| Kriterium 2 | 10 | .67 | 53,7 | 18.8 | 58 |

Ich schätze die Prozentsätze der korrekten Lösungen für beide unbenoteten Onlinequiz von ca. 66% als recht akzeptabel ein. Zum einen belegen sie, dass sich die Studierenden im Mittel ernsthaft mit dem Thema beschäftigt hatten und der Testeffekt, der ja theoretisch vornehmlich bei korrekter Antwort das Behalten stärkt, hier auch wirken kann. Zum anderen lassen die Prozentsätze noch hinreichenden Spielraum offen, durch das angebotene Feedback einige Fehler korrigieren zu können. Zu beachten gilt, dass die Onlinequiz unkontrolliert bearbeitet wurden. Die Vermutung, durch diese Art von Testung unrealistisch hohe Werte zu erhalten, erscheint aber eher unbegründet. Quiz 2 wurde bereits im WS08/09 unter kontrollierten Bedingungen im Seminar durchgeführt. Die Testwerte der Probanden unter gleicher experimenteller Bedingung wie in dieser Studie, erzielten hierbei die Werte $M = 62,1$, $s = 16,2$, $N = 30$ und unterscheiden sich nicht signifikant gegenüber den Testwerten unter unkontrollierter Online-Bedingung. Hierbei bleibe aber nicht unerwähnt, dass Vergleiche über Semesterjahrgänge hinweg durchaus ihre Tücken haben können (Jacobs 2009c).

Ergebnisse im Abschlussquiz (Lernerfolgstest, Kriteriumstest)

Die oben dargelegten Überlegungen einschließlich der bisherigen empirischen Befunde lassen eindeutig bessere Lernerfolgsmaße für die jeweils getesteten Gruppen erwarten, weswegen die Signifikanzprüfung einseitig vorgenommen wird. Da der erwartete Unterschied im niedrigen bis höchstens mittleren Effektstärkebereich vermutet wird und als Testverfahren der t-Test für unabhängige Stichproben zur Anwendung kommt, lege ich wegen des relativ hohen Betafehlers das Signifikanzniveau auf 10% fest. Tabelle 5 stellt die deskriptiven Daten sowie das Ergebnis des t-Tests dar, Abbildung 1 veranschaulicht die jeweiligen Mittelwerte unter EG - und KG-Bedingung.

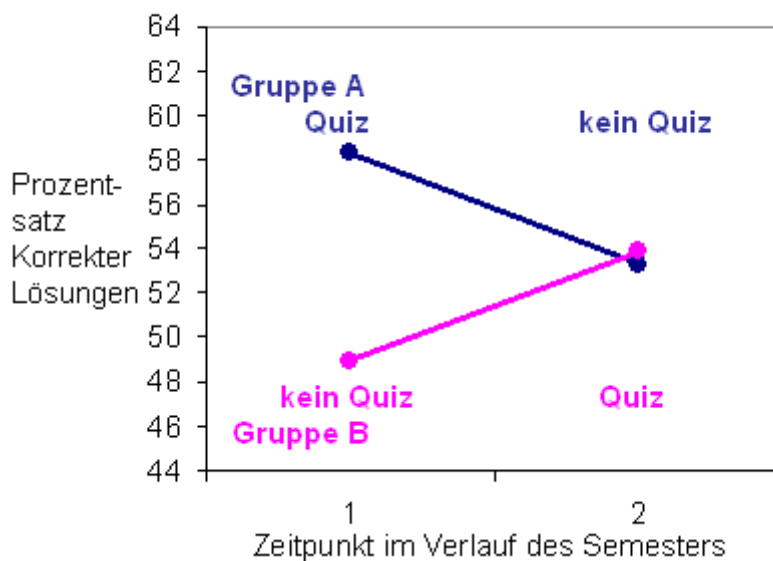
Tabelle 5: Ergebnisse im Lernerfolgstest, ca. 8 bis 4 Wochen nach der Übung

| Gruppe | Exp. Bedingung | M | s | N | t | p _e | d |
|--------|----------------|------|------|----|------|----------------|-----|
| A | Quiz 1 | 58,3 | 23,2 | 28 | 1,67 | 0,051 | .44 |
| B | kein Quiz 1 | 48,9 | 19,4 | 30 | | | |
| B | Quiz 2 | 53,9 | 18,6 | 30 | 0,08 | 0,47 | - |
| A | kein Quiz 2 | 52,5 | 19,4 | 28 | | | |

Die Ergebnisse fallen unterschiedlich aus. Im Quiz 1 erzielten Studierende mit Onlinequiz erwartungsgemäß bessere Lernerfolgsmaße als Studierende ohne Quiz. Der Lernvorteil entspricht einer Effektstärke von $d = .44$ und liegt damit im Rahmen der Erwartungen. Durch die Teilnahme am Quiz 2 hingegen lässt sich kein nachhaltiger Test -und Feedbackeffekt nachweisen.

Abbildung 1:

Ergebnisse im Lernerfolgstest, 8 bis 4 Wochen nach den Quiz



Zwar tendieren beide Gruppen im Zeitverlauf jeweils in die richtige Richtung, was auch die in der Graphik der Abbildung 1 erkennbare Interaktion unterstreicht. [signifikante Interaktion zwischen experimenteller Bedingung im Zeitverlauf als Within-Faktor und Gruppe als Between-Faktor ($F(1,56)=4,8$ $p_z=0,033$; partielles Eta-Quadrat =.079). Man könnte diese Interak-

tion als Beleg dafür werten, unter Quizbedingungen erzielten beide Gruppen jeweils das bessere Ergebnis. Aber der Effekt scheint mir insgesamt schwach. Man hätte hier eine deutliche disordinale Interaktion erwarten müssen. Der Test -und Feedbackeffekt konnte somit nur partiell bestätigt werden.

Hinweis: Wenn man für alle Probanden die Differenz (Zeitpunkt 1- Zeitpunkt 2) bildet, und mit diesen Differenzen einen t-Test für unabhängige Stichproben (Gruppe A, B) auf Signifikanz prüft, so erhält man dasselbe Ergebnis wie die Interaktion. ($t(56)=2.19$; $p_z = 0.033$; $d=.57$). Durch die ausschließliche Betrachtung der Veränderungen wird die hohe Personenvarianz eliminiert und eine sensitivere Testung ermöglicht. Sie kann aber nicht darüber hinwegtäuschen, dass beim Quiz 2 der erwartete Effekt ausblieb.

Abschließende Zusammenfassung und Diskussion

Ähnlich wie bei Jacobs (2009a) konnte die Lernwirksamkeit eines verpflichtenden, aber unbenoteten Onlinequiz hier nur teilweise bestätigt werden. Ein unbenotetes, verpflichtendes Quiz kann durchaus einen lernfördernden Effekt nach sich ziehen, tut dies aber offenbar nur unter bestimmten Bedingungen. Es ist müßig, über mögliche Gründe für die unterschiedlichen Ergebnisse zu spekulieren, weil kaum hinreichende Belege für derartige Vermutungen vorgebracht werden können. An der relativ langen Zeitspanne zwischen Quiz und Kriterium kann es kaum gelegen haben, weil das Retentionsintervall beim erfolgreichen Quiz 1 noch länger ausfiel. Vielleicht war die Transferdistanz zwischen Quiz und Kriterium beim zweiten Quiz zu hoch, so dass die Bearbeitung des Quiz nicht ausreichte, das erforderliche Verständnis beim Kriterium in hinreichendem Ausmaß zu erzielen. Um solche Fragen beantworten zu können, bräuchte man valide Kennwerte von Transferdistanzen, was aber angesichts der immer noch vorhandenen Lehrzielheterogenität der Aufgaben pro Quiz in der Praxis ziemlich unrealistisch ist. Möglicherweise fiel der Druck, sich anständig vorzubereiten bzw. das Quiz aufmerksam zu bearbeiten, beim zweiten Quiz etwas schwächer aus, weil die Studierenden in der Zwischenzeit irgendwie mitbekamen, dass selbst bei absolut schwachen Ergebnissen keinerlei Konsequenzen folgten.

Die Studie war nicht dafür konzipiert, diffizile theoretische Hypothesen abzuklären. So bleibt offen, worauf der Effekt beim erfolgreichen Quiz 1 letztlich zurückzuführen ist und wie stark daran Test- oder Feedbackeffekt beteiligt waren. Ein Teil des Effektes könnte z.B. auch darauf basieren, dass die Studierenden in Erwartung eines Quiz die Quizgrundlage intensiver vorbereiteten. Eine derartige Hypothese entzieht sich hier einer empirischen Überprüfung, weil keine weitere Gruppe zur Verfügung stand, die in Erwartung eines Quiz dieses dann aber nicht hätte bearbeiten dürfen. Ebenso fehlen im Kriteriumstest Aufgaben, die im Quiz nicht einmal im entfernten Sinne eingeübt wurden, aber zum lehrzielrelevanten Teil des Seminars bzw. der Powerpoint-Folie gehören. Denn dann hätte man den Effekt der Vorbereitungsanstrengungen vom Test -und Feedbackeffekt trennen können.

Eine analog Jacobs (2009a) durchgeführte Erhebung subjektiver Einschätzungen nach den Quiz ergab beim zweiten gegenüber dem ersten Quiz eine geringere aktuelle Angst ($p_e=0,037$) und eine geringere Anstrengung während der Quizbearbeitung ($p_e=0,014$) sowie eine um ca. 0,45 Effektstärken schwächere Vorbereitungsgüte ($p_e=0,04$) auf das Quiz. Die durchschnittlichen Studiervorbereitungszeiten für beide Quiz fielen mit jeweils 26 Minuten hoch vergleichbar aus. Quiz 2 basierte allerdings auf einem Lehrstoff, der von der Seminarsitzungszeit und vom Ausmaß der Powerpointfolie ca. doppelt so umfangreich war. Diese Ergebnisse geben zumindest einen Hinweis für die These eines geringeren Engagements bei der Vorbereitung und der Bearbeitung des zweiten Quiz.

Mögliche Verbesserungen könnten nachfolgende Punkte beachten:

- Festlegung einer Mindestleistung im Quiz: Durch die Einführung eines Minimalkriteriums (z.B. 50% korrekter Lösungen) in Verbindung mit deutlichem Nachbearbeitungsaufwand bei Nichterfüllung, könnten unzumutbar schwache Leistungen im Quiz höchstwahrscheinlich vermieden werden. Dadurch führt man jedoch eine der Benotung ähnliche Bewertung der Art "bestanden/nicht bestanden" ein (Jacobs 2010b), was vermutlich mit einer erhöhten Leistungsangst selbst für solche Studierenden verbunden sein könnte, welche diese Hürde locker nehmen würden. Um diese unnötige Angst möglichst auszuschalten, böte es sich an, sehr deutlich zu machen, dass "nicht bestanden" keineswegs "durchgefallen" bedeutet, aber mit einem erheblichen zusätzlichen Arbeitsaufwand verbunden sei.
- Mehrfache Bearbeitung eines Quiz: Sinn der hier realisierten einmaligen Quizbearbeitung war es auch, Studierende zu einer angemessenen Nacharbeitung des Seminars bzw. Vorbereitung auf das Quiz zu animieren, bevor sie das Quiz in Angriff nehmen. Bei stetem Zugriff auf vielfältige Quizaufgaben aus einem großen Itempool besteht die Gefahr, sorgfältiges Studieren durch zu viel Testen zu ersetzen. Manche Forscher bieten so genannte Mastery Quiz an. Hierbei muss der Studierende im Idealfall sooft ein paralleles Quiz bearbeiten bis er ein hohes Kriterium, z.B. 90% korrekter Antworten, erreicht. Zumindest bleibt überlegenswert, bei relativ schwachen Leistungen im ersten Test weitere Testungen bis zu einem akzeptablen Niveau verpflichtend einzufordern. Grundsätzlich würde eine verpflichtende mehrfache Testung, jeweils unterbrochen durch ein angemessenes Zeitintervall, nach der Theorie des verteilten Lernens eine stabilere Lernleistung erwarten lassen.
- Nutzung des Feedbacks anregen und kontrollieren: Es wurden keinerlei Erhebungen darüber vorgenommen, ob und wie überlegt Studierende die Fragen bearbeiteten, das Feedback nutzten und sich um eine Korrektur ihrer Fehler im Quiz bemühten. Höchstwahrscheinlich genügt es nicht, vor den Quiz einige verbale Empfehlungen zum sinnvollen Umgang mit den Aufgaben zu geben. In diesem Zusammenhang stellt sich die Frage, welche Maßnahmen geeignet wären, die Be- bzw. Nacharbeitung eines Quiz effektiver zu gestalten.

Trotz fehlgeschlagener Replikation des Quizeffektes stimmen die Ergebnisse insgesamt hoffnungsvoll, weil sowohl die pädagogischen wie auch die statistischen Anforderungen an einen empirischen Nachweis in dieser Studie ziemlich hoch gesteckt waren. Schließlich sollte durch die Quiz im realen Schulsetting naher Transfer erreicht und über eine ordentliche Zeitspanne hinweg erhalten bleiben. Dabei sollte dieser Effekt, wenn auch extrinsisch motiviert, so doch ohne großen äußeren Druck zustande kommen. Zu bedenken gilt weiterhin, dass die Größe eines Test- und Feedbackeffektes auch von den Bedingungen der Lernaneignungsphase sowie den Lernmaterialien abhängt, die sehr gut aufeinander abgestimmt waren und selbst ohne Quiz gute Lernmöglichkeiten boten. Unter solchen Voraussetzungen sind selbst kleine Effektstärken recht bedeutsame Ergebnisse. Das versuchsplanerische Niveau bzw. die interne Validität der Studie erreichte durch die erfolgreiche Randomisierung im natürlichen Setting ein sehr hohes methodisches Niveau. Das statistische Design mit unabhängigen Gruppen setzt hohe Hürden, solche kleinen Effekte zuverlässig nachzuweisen. Fast alle mir bekannten Test- und Feedbackstudien verwandten reine Repeated-Measurement-Designs, welche die riesige Personenvarianz im Zaume halten und nur die Veränderungen analysieren. Eine nur auf die Veränderungen bezogene Analyse ergab auch hier in Form der signifikanten Interaktion ein deutliches signifikantes Ergebnis im mittleren Effektstärkebereich.

Unter reinen Effektivitätsaspekten ist die Lernwirkung verschiedener Quizanregungen gegenüber keiner Maßnahme in der Schulpraxis wie folgt einzuschätzen

1. freiwillige Quiz: Effekt bisher nicht nachgewiesen
2. verpflichtende, unbenotete Quiz: möglicher, geringer Effekt
3. benotete Quiz: deutlicher Effekt

In diesen Effektivitätsunterschieden spiegelt sich die Bedeutung der Lernmotivation für den Lernerfolg wider. Je höher die Kontrolle und der Druck, desto höher die Leistung. Äußerer Druck bewegt vornehmlich solche Studierende zum Arbeiten, die ohne solchen Druck kaum Lernaktivität zeigen würden. Aus der deutlichen Lernwirkung benoteter Quiz darf allerdings nicht einfach gefolgert werden, Studierende müssten dann auch in einer benoteten Abschlussklausur besser abschneiden. Denn dann sorgt der Notendruck der Abschlussklausur auch bei Studierenden ohne vorherige Quiz dafür, sich intensiv vorzubereiten und den meisten Studierenden gelingt es dann, die Versäumnisse während des Seminars durch verstärkten Lerneinsatz nachholen.

Die allesamt hohen Streuungen der Testwerte in Quiz und Kriterium unterstreichen bei den relativ hohen Reliabilitäten die große Bedeutung der Person für das Leistungsergebnis. Im Vergleich zum Effekt der Personen (partielles $\eta^2 = .90$) fällt die Wirkung der Instruktionsmaßnahme Quiz (partielles $\eta^2 = 0,079$) verschwindend gering aus. Leistungsfähige Studierende erzielen - transparente Lehrziele und vernünftige Lehrmaterialien vorausgesetzt - auch ohne besondere pädagogische Maßnahmen gute Ergebnisse, da sie vermutlich über effiziente Studiertechniken verfügen und diese auch einsetzen. Bei leistungsschwächeren oder unwilligen Studierenden scheinen auch aufwändige pädagogische Maßnahmen keine großen Effekte zu erbringen. Insofern sollte man auch keine zu hohen Erwartungen an die Wirkung von Onlinequiz stellen. Diese Erwartungszurückhaltung gilt freilich nicht nur für Quiz, sondern empfiehlt sich für die meisten pädagogischen Maßnahmen.

Literatur

- Bangert-Drowns, R.L., Kulik, C., Kulik, J.A., & Morgan, M.T. (1991). The instructional effect of feedback in test-like events. *Review of Educational Research*, 61, 213-238.
- Brosvic, M. & Epstein, M. L. (2007). Enhancing learning in the introductory course. *Psychological Record*, 57, 391-408
- Butler, A. C., Karpicke, J. D. & Roediger III, H. L. (2007). The Effect of Type and Timing of Feedback on Learning from Multiple-Choice Tests. *Journal of Experimental Psychology: Applied*, 13 (4), 273-281.
- Cull, W. L. (2000). Untangling the Benefits of Multiple Study Opportunities and Repeated Testing for Cued Recall. *Applied Cognitive Psychology*, 14, 215-235.
- Dihoff, R. E., Brosvic, G. M., Epstein, M. L., Cook, M. J. (2004). Provision of Feedback During Preparation for Academic Testing: Learning is Enhanced by Immediate but Not Delayed Feedback. *Psychological Record*, 54 (2), 207-231.
- Jacobs, B. (2006). Erneutes Studieren oder Testen mit Feedback beim Einüben von Faktenwissen am Beispiel des Erlernens der Bundesstaaten der USA.
URN: urn:nbn:de:bsz:291-psydok-5992
URL: <http://psydok.sulb.uni-saarland.de/volltexte/2006/599/>
- Jacobs, B. (2008). 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. (2009a) 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. (2009b): Geld zur Steigerung der Testmotivation- und Leistung in einer unbenoteten Abschlussklausur ?
 URN: urn:nbn:de:bsz:291-psydok-23600
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2009/2360/>
http://www.phil.unisb.de/~jakobs/wwwartikel/extrinsische_motivierung/cash4quiz.html
- Jacobs, B. (2009c). Vergleich von Bedingung A im aktuellen mit Bedingung B in einem früheren Jahrgang
http://www.phil.uni-sb.de/~jakobs/seminar/vpl/experiment/exkurs/vorexperiment_variante.htm
- Jacobs, B. (2010). Tests als Maßnahmen zur Förderung von Studierleistungen?
 URN: urn:nbn:de:bsz:291-psydok-26022
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2602/>
http://www.phil.uni-sb.de/~jakobs/wwwartikel/tests_als_studierhilfen/tests_als_studierhilfen.pdf
- Jacobs, B. (2010b). Bestanden/nicht bestanden im Vergleich zum traditionellen, mehrstufigen Benotungssystem
http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/notendruck/pass_fail.html
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test for mat and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology* 19, 528-558.
- 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.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, 67, 259-266.
- McDaniel, M. A., Anderson, J. L., Derbish, M.H., & Morrisette, N. (2007). Testing the testing effect in the classroom. *European Journal of Cognitive Psychology*, 19, 494-513.
- Roediger, H. L. & Karpicke, J. D. (2006). The Power of Testing Memory. *Basic Research and Implications for Educational Practice. Perspective on Psychological Science*, 1 (3), 181-210.

Aufgabenbeispiele zu den experimentellen Quiz

Quiz1: Skalen und Stichproben

Aufgabe 2 Quiz

Gebe das höchstmögliche Skalenniveau an, welches die Skala annehmen kann.

| | Skala | a nominal | b ordinal | c intervall | d rational |
|-----|---------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1.) | Haarfarbe | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.) | Nationalität | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.) | Bildungsniveau | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.) | Persönlichkeitstest | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5.) | Soziale Schicht | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 6.) | Vornamen | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 7.) | Regenmenge pro qm | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 8.) | Gehalt in Euro | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Aufgaben bestätigen

Korrekte Lösungen?

Aufgabe 2 Lernerfolgstest/Kriterium

Gebe das höchstmögliche Skalenniveau an, welches die Skala annehmen kann.

| | Skala | a nominal | b ordinal | c intervall | d rational |
|-----|-------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 1.) | Hunderasse | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.) | Bundesland | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.) | Platzierung beim Pferderennen | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.) | Psychologischer Eignungstest | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5.) | Deutschnote | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 6.) | Gemüse | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 7.) | Alkoholgehalt im Getränk | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 8.) | Verkaufspreis in Dollar | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

Aufgaben bestätigen

Quiz 2: Testgütekriterien

Aufgabe 2 Quiz

Nenne jeweils den allgemeinen Fachbegriff für nachfolgende Beschreibungen

| Nr. | Beschreibung | Fachbegriff |
|--------------------|---|----------------------|
| 1. | Das Testergebnis soll unabhängig vom Auswerter sein. Alle Auswerter kommen zum gleichen Ergebnis | <input type="text"/> |
| 2. | Der Test soll zuverlässig sein, d.h. Das Testergebnis soll die Leistung bzw. das Merkmal relativ schwankungsfrei wiedergeben. | <input type="text"/> |
| 3. | Der Test soll das gewünschte Merkmal bzw. Lehrziel und nicht etwas anderes erfassen. | <input type="text"/> |
| Aufgabe bestätigen | | Korrekte Lösung? |

Aufgabe 2 Lernerfolgstest

Nenne jeweils den allgemeinen Fachbegriff für nachfolgende Beschreibungen

| Nr. | Beschreibung | Fachbegriff |
|--------------------|---|----------------------|
| 1. | Der Test soll ein Merkmal genau und zuverlässig erfassen. | <input type="text"/> |
| 2. | Es soll keine Rolle spielen, wer die Testanweisung gibt, die Testung überwacht, den Test auswertet oder interpretiert. | <input type="text"/> |
| 3. | Der Test soll genau das messen, was er zu messen vorgibt. z.B. ein Kreativitätstest soll Kreativität und nicht etwa Intelligenz messen. | <input type="text"/> |
| Aufgabe bestätigen | | |