

Bernhard Jacobs, Fachrichtung Erziehungswissenschaften der Universität des Saarlandes.
Email: b.jacobs@mx.uni-saarland.de
Version: 16.11.2010

Testfragen selbst beantworten oder Musterlösungen studieren?

Abstract

Ziel der Untersuchung war die empirische Überprüfung der Hypothese, ob das Testen mit Feedback im realen Universitätsbetrieb einen höheren langfristigen Lernerfolg bewirkt als die erneute gezielte Informationsaufnahme. Beide Übungsmethoden beinhalten dieselben Informationen. Während das Testen zunächst eine aktive Antwort des Lernenden erfordert, bevor eine Rückmeldung folgt, sieht der Lernende bei der alternativen Übungsmethode direkt die korrekte Antwort und soll sich diese Musterlösung verständlich machen. Die Übungen verfolgten den Zweck, wesentliche Kernpunkte einer vorangegangenen Seminarsitzung zu behandeln und legten besonderen Wert auf Verstehen und Anwendung. Die Auswirkungen der Übungsaufgaben wurden durch parallele Lernerfolgsaufgaben überprüft, um mindestens nahen Transfer zu erfassen. Den Ergebnissen zufolge erwiesen sich einmal das Testen mit Feedback und einmal das Studieren korrekt beantworteter Aufgaben als die günstigere Lernvariante. Ein weiterer Versuch ergab schließlich hoch vergleichbare Lernerfolgsmaße, so dass insgesamt gesehen beide Übungsmethoden als gleichwertig zu betrachten sind. Studierende schätzen das Testen mit Feedback jedoch eindeutig als die bessere Übungsmethode ein und würden es bei Wahlfreiheit vorziehen.

Schlagwörter: Testen, Feedback, Quiz, Üben, Musterlösung, Hochschuldidaktik

Einleitung, Problemstellung und Zielsetzung

Folgt im Anschluss an eine Instruktion ein Test, der die wichtigen Lehrziele der Instruktion erfasst, so wird das Behalten gestärkt (Hamaker, 1986). Die Fragen im Test fordern dazu auf, sich wichtige Informationen in Erinnerung zu rufen und dieses Abruftraining fördert das langfristige Behalten. Dieser Testeffekt tritt insbesondere dann sehr deutlich in Erscheinung, wenn viele Fragen richtig beantwortet wurden (z.B. Toppino & Cohen, 2009), hilft bei falschen Antworten aber nicht weiter. Durch informatives Feedback, z.B. in Form einer Rückmeldung der korrekten Antwort (= *Knowledge of correct response* [KCR]) erhält der Lernende die Chance, seinen Fehler zu korrigieren. Zwar bewirkt Feedback zu einem gewissen Teil auch eine Stabilisierung korrekter Antworten, vorwiegend bei unsicher gelösten Aufgaben (Butler, Karpicke & Roediger, 2008) oder nach einem längeren Zeitintervall zur Testung (Smith & Kimball, 2010), seine wesentliche Lernwirkung liegt jedoch in der Fehlerkorrektur. Wenn gleich hinreichende empirische Untersuchungen den Lerneffekt durch Testen und Feedback gegenüber einer No-Treatment-Kontrollgruppe deutlich belegen (Jacobs, 2008b), so stellt sich doch die Frage, ob nicht alternative Instruktionsmaßnahmen als echte Alternativen aufzufassen sind.

Liegen Lehrtexte als wesentliche Lernaneignungsinstruktionen zugrunde, so kommen im Prinzip alle pädagogischen Maßnahmen in Betracht, welche das Textverständnis verbessern könnten, z.B. selbst Fragen an den Text stellen, Überschriften zu Textabschnitten bilden, Zusammenfassungen schreiben, Concept-maps erstellen usw. Allerdings wurden bis jetzt nur wenige dieser Textverarbeitungsstrategien im Vergleich zu Testen mit Feedback einer systematischen empirischen Prüfung unterzogen.

Lehrtexte erneut lesen als Alternative zu Testen mit Feedback ?

Statt einen Lehrtext zu testen, könnte man die Studierenden auch auffordern, den Text ein zweites Mal durchzulesen. Einige empirische Studien konnten auch bei diesem Vergleich einen Vorteil des Testens nachweisen, der sich vor allem nach einem längeren Retentionsintervall heraus kristallisiert. (z.B. Callender & McDaniel, 2007 für schwache Lernende; Duchastel & Nungester, 1984; Nungester & Duchastel, 1982; Roediger & Karpicke, 2006). Es gibt jedoch auch Gegenbeispiele (z.B. Kester & Tabbers, 2008). Da ein lehrzielvalider Test aber lediglich eine repräsentative Stichprobe des gesamten Instruktionstextes erfasst bzw. erfassen soll, erhalten die Getesteten gegenüber den wiederholten Lesenden genauere und gezielte Hinweise darüber, was im Lernerfolgs-, Behaltens- oder Kriteriumstest letztlich von ihnen verlangt wird. Ein fairer Vergleich sieht daher vor, sowohl beim Testen als auch beim wiederholten Studieren, dieselben Informationen zugrunde zu legen. Der Unterschied zwischen beiden Instruktionsvarianten liegt lediglich darin, beim erneuten Studieren direkt die korrekte Antwort zur gestellten Frage zu präsentieren, während beim Testen erst die Frage beantwortet werden muss, bevor eine Rückmeldung folgt. Sofern beim erneuten gezielten Studieren oberflächliches Lernen vermieden und eine möglichst tiefe Verarbeitung angeregt wird, so sind auch deutliche Lern- und Behaltenseffekte zu erwarten.

Empirische Befunde zum Vergleich von Testen und gezieltem Studieren

Mittlerweile liegen relativ viele Laborstudien vor, welche die Überlegenheit des Testens mit und teilweise sogar ohne Feedback gegenüber der erneuten Präsentation der korrekten Antwort bei unterschiedlichen Lernarten nachgewiesen haben (z.B. Einüben von Vokabeln: Carrier & Pashler, 1992; Cull, 2000; Toppino & Cohen, 2009, Name-Face-Learning: Carpenter & DeLosch, 2005, Lernen aus Texten: Kang, McDermott & Roediger, 2007; LaPorte & Voss, 1975, Lernen mit Landkarten: Carpenter & Pashler, 2007; Jacobs, 2006; Rohrer, Taylor & Sholar, 2010). In letzter Zeit findet man sogar Untersuchungen, welche einen Testvorteil in realen schulischen Umwelten feststellten (Anderson, Weywadt & McDaniel, 2009; Larsen, Butler & Roediger, 2009; McDaniel, Anderson, Derbish & Morrisette, 2007). Der erwartete Behaltensunterschied fällt gelegentlich recht mager aus und kann meist nur durch sehr effiziente Wiederholungsdesigns statistisch gesichert werden. Besonders schwierig erscheint es, die Behaltensvorteile des Testens bei MC- Übungsaufgaben nachzuweisen. Schließlich lässt sich die Testüberlegenheit auch nicht in jeder Studie (z.B. Teilergebnisse aus den Studien von Kang et al., 2007; Pilotti, Chodorow & Petrov, 2009, Teilergebnisse von Anderson et al., 2008) zuverlässig belegen.

Eigene Vorarbeiten

Laborstudien

Die vier Laborstudien von Jacobs (2006, 2007a, 2008c, 2009b) beziehen sich auf das Einüben der Bundesstaaten der USA. Hierbei sollten Studierende die Namen der Bundesstaaten den entsprechenden Territorien auf der Landkarte zuordnen. Als reine Studiervarianten dienten entweder die direkte sukzessive Platzierung eines Staatsnamens in das zutreffende Gebiet oder die Präsentation der Landkarte der USA mit allen Staatsbezeichnungen in den zutreffenden Territorien. Als Testvarianten fungierten mehrere Aufgabentypen, die alle zunächst eine eigene Antwort des Lernenden vorsahen, bevor als Feedback die korrekte Antwort folgte. In den später erfassten Behaltentests, ca. eine Woche nach den Übungen, ließen sich zum Teil signifikante Vorteile für eine der Testvarianten gegenüber der Studiervariante nachweisen (2006, 2007a), zum Teil aber auch nicht (2008c, 2009b). Niemals trat der Fall ein, dass alle Testversionen besser als die Studiervariante abschnitten oder eine Studiervariante einer Tes-

tung überlegen war. Jacobs (2009b) kommt auf der Basis seiner bisherigen Befunde beim Einprägen der Bundesstaaten der USA zu der Schlussfolgerung, die möglichen Behaltensvorteile des Testens seien unter fairen Bedingungen letztlich zu gering und entbehrten hinreichender praktisch pädagogischer Bedeutsamkeit, um einer bestimmten Übungsmethode die eindeutige Priorität einzuräumen. Unabhängig von der objektiven Behaltensleistung der verschiedenen Übungsmethoden ließen sich aber teilweise Präferenzen auf subjektiv verbaler wie auch auf Verhaltensebene für einige Testformen nachweisen.

Feldexperiment

Das Feldexperiment von Jacobs (2008a) fand im Rahmen einer Vorlesung statt. Der Verfasser der vorliegenden Arbeit (Jacobs) bat den Professor der Vorlesung ‚Lehren und Lernen‘ erfolgreich um die Erlaubnis, den Studierenden der Vorlesung eine freiwillige Übungsmöglichkeit zum Thema klassische und operante Konditionierung via Internet anbieten zu dürfen. Jacobs entwickelte diese Übung vorwiegend auf der Basis seiner verfügbaren Fragen aus eigenen Veranstaltungen, orientierte sich zum Teil auch an den Folien der Vorlesung, war aber selbst gar nicht in der entsprechenden Vorlesung zugegen. Diese Unkenntnis lässt vermuten, dass die Übung nicht stringent Bezug auf die Schwerpunkte des in der Vorlesung vermittelten Wissens nahm. Mit Sicherheit waren die Vorlesung und die Übung nicht gezielt aufeinander abgestimmt worden.

Jacobs fand vergleichbare Behaltenswerte für das Studieren korrekt beantworteter Quizfragen und das Testen mit Feedback. Hierbei erforderte das Testen mit Rückmeldung aber eine längere Studierzeit, womit sich das Studieren korrekt beantworteter Quizfragen gegenüber dem Testen mit Feedback bei gleicher Lerneffektivität als lerneffizienter erwies. Zugleich widerspricht dieses Ergebnis den meisten empirischen Studien.

Ziel der nachfolgenden Untersuchung ist es daher, im realen Schulsetting weitere empirische Belege zum Vergleich der beiden Instruktionsmaßnahmen zu finden, um auf einer solideren Datenbasis besser abschätzen zu können, ob das Studieren korrekt beantworteter Testfragen in der Schulpraxis als ernsthafte Instruktionsalternative zu Testen mit Feedback betrachtet werden kann. Dabei sollte eine möglichst gute Abstimmung zwischen Wissensvermittlung während des Unterrichts, verfügbaren Lernmaterialien für die Studierenden und den in der Übung dargebotenen Aufgaben gewährleistet sein.

Begriffsfestlegung für gezieltes Studieren

Beim Studieren der korrekten Antwort werden lediglich Informationen präsentiert, die der Lernende zunächst einmal enkodieren muss. Beim Testen hingegen wird er unmissverständlich aufgefordert, selbst eine Antwort zu produzieren, wozu in der Regel ein Abruf aus dem Langzeitgedächtnis erforderlich erscheint. Je nachdem, welche Art von Lernen in den einzelnen Studien untersucht wurde, ergeben sich etwas unterschiedliche Fassungen der Studiervariante und eine Fülle verschiedener Bezeichnungen in der Literatur, wie etwa: *question and answer*, *study only*, *review-only practice*, *restudy*, *restudy-isolated-sentences* oder *focused re-exposure to the target information*.

Jacobs (2008a) übernahm für die Studiervariante von LaPorte und Voss (1975) die Bezeichnung *read statements*, die auch bei Kang et al. (2007) Verwendung findet. Die Autoren wandelten die Fragen einer Quizversion einfach in Aussagesätze um. Aus einer Testfrage „Wie heißt die erste Bundeskanzlerin der BRD?“ wird so die Aussage "Die erste Bundeskanzlerin der BRD heißt Angela Merkel." Read Statements repräsentieren folglich zutreffende Aussagesätze, die man aufmerksam lesen bzw. sich einprägen sollte. Jacobs (2008a) stellte diesen Aussagen allerdings die (rhetorischen) Fragen voraus.

Bei anspruchsvollen Problemlöseaufgaben würde man von *Lösungsbeispielen* sprechen. Lösungsbeispiele beziehen sich jedoch meist auf ein relativ einheitliches Problemfeld, z.B. auf Aufgaben zu Gleichungen eines bestimmten Typs (z.B. Cooper und Sweller, 1987) oder zur Kombinatorik (z.B. Jacobs 2004). Sie erfordern in der Regel mehrere Lösungsschritte.

Hier kommt die Methode jedoch für alle möglichen Fragen und Lehrzielniveaus in Betracht, die in einem Quiz Verwendung finden können. So muss ein Studierender etwa im Quiz „Statistische Grundlagen der pädagogischen Diagnostik“ wissen, was ein Mittelwert bedeutet, Auswirkungen bestimmter Daten auf Mittelwert und Median erkennen, aus einigen Testwerten Modus, Median und Mittelwert berechnen, aus mehreren Datensätzen die Größe der Streuung abschätzen, prozentuale kumulative Häufigkeiten ermitteln oder verschiedene Eigenschaften von Korrelationen einschätzen.

Für die Instruktionmethode der Präsentation der korrekten Antwort eignet sich als deutscher Begriff vermutlich am besten das *Studieren korrekt beantworteter Quiz- oder Testfragen*. Synonym zu dieser umfangreichen Bezeichnung wird im Folgenden der Begriff *Musterlösung* verwendet. **Eine Musterlösung** umfasst

1. die Fragestellung
2. die Präsentation der korrekten Antwort
3. alle über die korrekte Antwort hinausgehenden Rückmeldungen

Den Studierenden gegenüber wurde die Musterlösung als Statement bezeichnet, was der früheren Terminologie anzulasten ist. Der Begriff Statement taucht gelegentlich in Fragebögen und Beispielen auf, bedeutet aber dasselbe wie Musterlösung. Folglich gilt:

Musterlösung = Studieren korrekt beantworteter Quizfragen = Statement

Anforderungen an die Untersuchung

Die Studie strebt, geleitet von theoretischen Erwartungen und den Ergebnissen der empirischen Laborforschung, eine ernsthafte Überprüfung der Lerneffektivität zweier Instruktionmethoden in der Schulwirklichkeit an. Diese unterscheidet sich in etlichen Punkten von der Laborsituation, etwa im Umfang des Lehrstoffs, dem Lehrzielanspruch, dem Zeitintervall zwischen Instruktionsphase und Testung, der Länge des Retentionsintervalls oder der Eigensteuerung und Motivierung des Lernens. Im Vergleich zur Laborforschung werden bestimmte Anforderungen und Einschränkungen wirksam, welche als Voraussetzung für eine faire Testung beider Methoden angesehen werden, einen realisierbaren Einsatz in der Schulpraxis ermöglichen und ein pädagogisch anspruchsvolles Ziel festlegen.

1. **Die Übungsbearbeitung sollte innerhalb des schulischen Umfelds in praktikabler Weise hinreichend angeregt werden.** Mögliche Methodenunterschiede sollten im Wesentlichen durch die Übungsbearbeitung und nicht durch unterschiedliche Nutzung sonstiger Lernmöglichkeiten (z.B. Vorbereitung) zustande kommen, die von solchen Methoden in unterschiedlicher Weise indirekt angeregt werden könnten. Damit wird modellhaft ein direkt realisierbarer Weg aufgezeigt, wie die Übungsmethoden im Schulalltag zur Anwendung kommen könnten.
2. **Die Lehrziele der Übung sollten sich weniger auf Faktenwissen, sondern vornehmlich auf Verstehen und Anwendung beziehen.** Im Mittelpunkt der Übungsanforderungen stehen demnach keine Memorierungsleistungen wie etwa beim Vokabellernen, welche den überwiegenden Teil der Laborforschung ausmachen. Ziel ist es vielmehr, das

Verständnis zu fördern. Deshalb müssen die Aufgaben in der Übung und im Kriteriumstest, welcher die Auswirkung der Übung auf den Lernerfolg misst, die gleichen Lehrziele mit nicht identischen Aufgaben erfassen. Dies wurde durch die Konstruktion paralleler Aufgaben in Quiz und Kriterium erreicht. D.h. Jede Kriteriumsaufgabe bezieht sich auf das Lehrziel einer bestimmten Übungsaufgabe, erfasst dieses Lehrziel aber in modifizierter Form. Hierbei kamen verschiedene Methoden zum Einsatz: Bei sehr einfachen Aufgaben meist Paraphrasierung und Transformation, bei anspruchsvolleren Verstehensaufgaben z.B. neue Beispiele und/oder bei Anwendungsaufgaben verschiedene Daten oder leicht veränderte Anforderungen. (siehe einige Aufgabenbeispiele in diesem Text). Im Vergleich zur Verwendung identischer Items erschwert diese Art der Lernerfolgskontrolle den Nachweis eines Test- und Feedbackvorteils, wie etwa die Ergebnisse von Anderson et al. (2009) belegen.

3. **Der angestrebte Lernerfolg sollte eine gewisse Lernstabilität bzw. Nachhaltigkeit aufweisen.** Da eher langfristige Lerneffekte interessierten, wurde ein Retentionsintervall von mindestens vier Wochen angestrebt.

Untersuchungsansatz

An der Untersuchung nahmen Studierende des Lehramts teil, die ein vom Verfasser dieser Arbeit geleitetes Proseminar zur Pädagogischen Diagnostik im Wintersemester 2009/2010 belegten. Das erste Seminar fand montags statt, das zweite Seminar donnerstags, beide zur gleichen Zeit. Beide Seminare können als Parallelklassen aufgefasst werden, da sie vom selben Dozenten unterrichtet wurden und nach identischem Seminarplan den gleichen Lehrstoff mit denselben Lehrmaterialien zu bearbeiten hatten. Die Planung der Seminare sah vor, die experimentellen Übungen als normale Übungsbestandteile zu nutzen. Deshalb wurden die Studierenden nicht darüber informiert, dass sie an einem Experiment teilnahmen. Erst nach Abschluss aller Erhebungen erhielten die Studierenden aufklärende Informationen über Sinn und Zweck der Untersuchung.

Der Seminarleiter gestaltete alle für die experimentellen Übungen relevanten Seminarsitzungen weitgehend dozierend, meist im Sinne des darstellenden Unterrichts mit einigen interaktiven Elementen. Hierbei bemühte er sich, einen möglichst gleichartigen Unterricht in beiden untersuchten Seminaren zu halten. Die Thematik wurde gründlich und ausführlich dargelegt und durch eine umfangreiche Powerpoint-Folie unterstützt, die den Studierenden unmittelbar nach der Seminarsitzung verfügbar gemacht wurde und als einzig notwendige Literaturgrundlage bezeichnet wurde.

Die Überprüfung der Fragestellung basiert auf zwei unterschiedlichen Versuchsplänen in zwei verschiedenen Lernsettings. Zum einen präsentierte der Seminarleiter auf der Grundlage eines vorexperimentellen Versuchsplans eine Übung mit Unterstützung von Computer und Beamer innerhalb des Seminars (Studie 1), zum anderen bearbeiteten die Studierenden im Rahmen eines Feldexperimentes zwei Online-Übungen individuell zu Hause (Studie 2). Der Kriteriumstest fand in Form einer unangekündigten und unbenoteten Abschlussklausur statt und war so terminiert, dass eine zusätzliche Vorbereitung des Lehrstoffs unwahrscheinlich erschien.

Tabelle 1 verdeutlicht den Zeitpunkt der Übungen sowie des Kriteriumstests, der die langfristigen Auswirkungen aller Übungen auf das Behalten erfassen sollte. Alle hier berichteten Zeitangaben beziehen sich auf das Seminar 1. Für Seminar 2 sind jeweils drei Tage hinzu zu addieren.

Tabelle 1: Zeitpunkte für die Übungen und den Lernerfolgstest

2.11.2009	27.11.2009	16.12.2009	11.1.2010
Übung im Seminar	Online Übung 1	Online Übung 2	Kriterium, Behaltenstests

Für die statistische Prüfung des Lernerfolgs wurde ein Signifikanzniveau von 10 Prozent zugrunde gelegt, da mit kleinen bis höchstens mittleren Effekten gerechnet wurde und die Testpower bei der gegebenen Stichprobengröße der unabhängigen Gruppen ansonsten zu gering ausfallen würde.

Studie 1: Testen bzw. Musterlösungen in einer Seminarsitzung

Das Thema der dritten Seminarsitzung lautete „Aufgabentypen und Aufgabenformen“. In der unmittelbar nachfolgenden Seminarsitzung, also sieben Tage später, setzte der Seminarleiter völlig überraschend eine spezielle Übung zum Inhalt der vorausgehenden Sitzung an. Seminar 1 bearbeitete das Quiz „Aufgabenformen und Aufgabentypen“ auf eine besondere Art des Testens mit Feedback, während dem Seminar 2 dieselben Fragen als korrekt beantwortete Quizfragen (= Musterlösungen) vorgeführt wurden. Die Auswirkungen der beiden Methoden wurden dann in einem unangekündigten unbenoteten Behaltenstest (siehe Tabelle 1) geprüft. Versuchsplantechnisch liegt diesem Vorgehen der vorexperimentelle statistische Gruppenvergleich zugrunde.

N Testen mit Feedback	0
N Musterlösung	0

Die Aussagekraft des Designs hängt ganz entscheidend von der Vergleichbarkeit der beiden vorgegebenen, formal nicht zwingend äquivalenten, Gruppen ab. Zu Beginn des Semesters drängten sich keine offenkundig plausiblen Gründe für entsprechende Unterschiede auf, da alle Studierenden ein entsprechendes Seminar belegen mussten und die Zuteilung der Studierenden zu den Seminaren weitgehend von einem Gremium bestimmt wurde. Tabelle 2 belegt zudem hoch vergleichbare Ergebnisse beider Seminare hinsichtlich Alter, Geschlecht und Abiturnotendurchschnitt.

Tabelle 2: Vergleichbarkeit der Seminare 1 und 2 (N pro Seminar 28)

	Seminar 1		Seminar 2		
	M	s	M	s	
Abiturnotendurchschnitt:	2.23	0.59	2.15	0.56	ns
Alter:	22.9	4.4	22.4	3.0	ns
Geschlecht:	75% w		64% w		ns

Seminarplan, Seminaranforderungen und Lernmaterialien waren für beide Seminare identisch. Derselbe Dozent achtete darauf, den experimentell relevanten Unterricht möglichst vergleichbar zu gestalten, womit entscheidende [Störfaktoren der Interaktion zwischen Selektion und den übrigen Störfaktoren](#) sehr unwahrscheinlich sind. Die vorgebrachten Argumente und Daten sprechen trotz des formal schwachen Versuchsplans für eine recht hohe Aussagekraft.

Testen mit Feedback im Seminar

Das Quiz „Aufgabenformen“ wurde im Seminar via Computer des Dozenten über Beamer Aufgabe für Aufgabe auf der Leinwand dargeboten. Zuvor wurden einige Regeln vermittelt, wie die Studierenden die verschiedenen Aufgaben (True-False-, MC-, Zuordnungs- und Short-Answer-Aufgaben) beantworten sollten. Der Seminarleiter teilte rote und grüne Kärtchen aus, welche als Signale für die True (grün)- False (rot) Aufgaben dienten. MC-Aufgaben mit einer korrekten Antwort sollten durch die Anzahl der hoch gehaltenen Finger beim Heben der Hand ausgedrückt werden. Bei zwei möglichen korrekten Antworten durch analoges Vorgehen beim Hochhalten beider Hände. Bei mehr als 3 korrekten Antworten sollten wenigstens zwei Antworten gegeben werden. Zuordnungsaufgaben waren als Serien von MC-Aufgaben definiert und konnten wie einfache MC-Aufgaben sukzessiv beantwortet werden. Short-Answer-Aufgaben erforderten lediglich Zahlen als Ergebnis. Die Zahl sollte durch große Ziffern verdeutlicht auf einem Blatt Papier hochgehalten werden.

Der Seminarleiter kontrollierte das Vorgehen, in dem er die auf der Leinwand sichtbare Frage laut vorlas und nach einer ihm angemessenen Bedenkzeit ein Zeichen zur Beantwortung für alle Studierenden gab. Hierbei achtete er darauf, dass nach dem Zeichen alle Studierenden durch Handzeichen aktiv Stellung zur Aufgabe bezogen. Bei Nichteinhaltung mahnte er die entsprechenden Studierenden an. Nach der Beantwortung aktivierte der Seminarleiter einen Button, was die Markierung der korrekte(n) Lösung(en) aktivierte (=KCR) und bei schwierigeren Aufgaben im Rückmeldefenster elaborierte Rückmeldungen zeigte. Was der Computer visuell verdeutlichte, wurde dann vom Seminarleiter zusätzlich akustisch verbal bestätigt, sowie gegebenenfalls Fragen von Studierenden beantwortet oder diskutiert.

Professionelle *Response Cards* oder gar elektronisch unterstützte *Personal Response Systeme*, z.B. *Clickers*, standen nicht zur Verfügung (siehe z.B. Mayer, Stull, DeLeeuw, Almeroth, Bimber, Chun, Bulger, Campbell, Knight & Zhang, 2009). Obgleich derartige Systeme die Prozedur deutlich vereinfachten und teilweise mehr Gestaltungsmöglichkeiten beinhalteten, erscheint die hier angewandte Methode den angestrebten Zweck einer offenen, aktiven Aufgabenbearbeitung hinreichend zu erfüllen. Einer Metaanalyse von Randolph (2007) zufolge, führt Response Card im Übrigen zu besseren Leistungsergebnissen als freiwilliges Melden.

Studieren korrekt beantworteter Quizfragen im Seminar

Aus dem Quiz „Aufgabenformen“ wurden Musterlösungen konstruiert, welche die korrekten Antworten in den Aufgaben bereits enthielten. Bei einer einfachen MC-Variante entsprach die Musterlösung damit exakt der Testfassung, wobei die korrekte Alternative bereits markiert war. Der Seminarleiter bot eine Musterlösung via Laptop über den Beamer an, las die Frage in der Regel laut vor und forderte die Studierenden auf, sich die präsentierte Lösung verständlich zu machen. Nach einer ihm ausreichenden Bedenkzeit blendete er die elaborierte Rückmeldung ein und unterstützte die Erklärungen akustisch verbal, analog der Vorgehensweise beim Testen mit Feedback. (siehe Abbildung 1)

Ergebnisse

Bearbeitungszeiten in den Seminaren

Die Übung „Aufgabenformen“ umfasste insgesamt zwölf Aufgaben unterschiedlichsten Aufgabenformates (MC mit einer oder mehreren Antworten, Zuordnungsaufgaben, Richtig-Falsch-Aufgaben und zwei Short-Answer-Aufgaben), die alle insgesamt 57 Antworten ver-

langten. Die Bearbeitungszeiten für beide Übungen fielen hochvergleichbar aus: 36 Minuten für das Testen mit Feedback und 35 Minuten für die Musterlösungen. Die zuvor gehegte Erwartung einer kürzeren Bearbeitungszeit für die Musterlösungen bestätigte sich nicht. Die Aufgaben zum Thema „Aufgabenformen“ wurden nur bei den Übungen im Seminar eingesetzt und waren den Studierenden danach nicht mehr zugänglich.

Abbildung 1. Aufgabe 1 für Quiz, Statement =Musterlösung und Lernerfolgstest in der Seminarübung

<p>1. Seminarleiter liest die Frage vor.</p> <p>2. Seminarleiter fordert nach einer ihm angemessenen Bearbeitungszeit zur Beantwortung aller Studierenden auf. -hier mussten die Studierenden die Zahl mit großen Ziffern auf ein Blatt schreiben und das Blatt dann hochhalten.</p> <p>3. Seminarleiter aktiviert die korrekte Lösung und die elaborierte Rückmeldung, nennt das korrekte Ergebnis, kommentiert die Rückmeldung und geht gegebenenfalls auf Fragen ein.</p>	<p>Aufgabe 1 Quiz</p> <p>Wie hoch ist die Wahrscheinlichkeit, eine einfache MC-Aufgabe mit 3 Antwortoptionen durch Zufall korrekt zu beantworten ? Hinweis: Die Wahrscheinlichkeit p muss zwischen 0 und 1 liegen.</p> <p>$p =$ <input type="text"/></p> <p><input type="button" value="Aufgabe bestätigen"/> <input <="" p="" type="button" value="Korrekte Lösung?"/> </p>
<p>1. Seminarleiter liest die Frage und Antwort vor - lässt Zeit zum Nachdenken</p> <p>2. Seminarleiter aktiviert "Aufgabe bestätigen", was die elaborierte Rückmeldung in einem Fenster erscheinen lässt, kommentiert die Rückmeldung und geht gegebenenfalls auf Fragen ein.</p>	<p>Statement 1: Studieren korrekt beantworteter Quizfragen</p> <p>Wie hoch ist die Wahrscheinlichkeit, eine einfache MC-Aufgabe mit 3 Antwortoptionen durch Zufall korrekt zu beantworten ? Hinweis: Die Wahrscheinlichkeit p muss zwischen 0 und 1 liegen.</p> <p>$p =$ <input type="text" value="0.33"/></p> <p><input type="button" value="Aufgabe bestätigen"/></p>
<p>elaborierte Rückmeldung zu Aufgabe 1 Quiz und Statement 1: 1 richtige Antwort 3 mögliche Antworten $p = 1/3 = .333$ Allerdings beantworten Schüler MC-Aufgaben nicht nach Zufall. Insofern ist die theoretische Wahrscheinlichkeit einer korrekten Schülerantwort durch den Schüler keineswegs .33.</p>	<p>Aufgabe 1 Lernerfolgstest/Kriterium</p> <p>Wie hoch ist die Wahrscheinlichkeit, eine einfache MC-Aufgabe mit 4 Distraktoren und einer korrekten Antwort durch Zufall korrekt zu beantworten ? Hinweis: Die Wahrscheinlichkeit p muss zwischen 0 und 1 liegen.</p> <p>$p =$ <input type="text"/></p>

Behaltenswerte im Kriteriumstest

Der Lernerfolgstest fand zehn Wochen später als Teil der unangekündigten, unbenoteten Abschlussklausur statt. Er umfasste zehn Aufgaben mit insgesamt 46 Antwortanforderungen und erzielte eine Konsistenz von $\alpha = .61$.

Tabelle 3: Prozentsatz korrekter Lösungen im Kriteriumstest Aufgabenformen

	M	s	N	t	
Testen	59,6	14,6	28		
Musterlösung	59,0	17,6	28	.14	ns

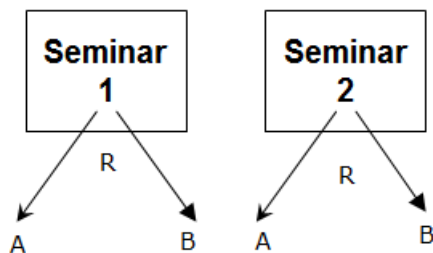
Wie aus Tabelle 3 hervorgeht, ließ sich kein Unterschied zwischen beiden Treatments feststellen. Für den langfristigen Lernerfolg spielte es demnach keine Rolle, ob die Studierenden im

Seminar offen aktiv die Quizfragen beantworteten und erst nach ihrer Antwort eine informative Rückmeldung erhielten oder ob sie die bereits korrekt beantworteten Quizfragen lediglich durchzulesen hatten, wobei sie sich um ein Verständnis bemühen sollten. Beide Instruktionmethoden führten zu hoch vergleichbaren Behaltenswerten im Sinne eines nahen Anforderungstransfers.

Studie 2: Testen bzw. Musterlösungen als verpflichtende Online-Übungen

Zu Beginn des Semesters wurden die Studierenden innerhalb jedes der beiden Seminare nach Zufall auf die Gruppen A und B zugeteilt.

Abbildung 2: Zufallszuteilung auf die Gruppen A und B



Bei der statistischen Analyse und Auswertung wurden dann die A- bzw. B- Gruppen der einzelnen Seminare zu einer gemeinsamen A- bzw. B-Gruppe zusammengefasst. Wie die vergleichbaren Ergebnisse beider Gruppen hinsichtlich Abiturnotendurchschnitts, Alter und Geschlecht in der Tabelle 4 aufzeigen, war die Randomisierung sehr erfolgreich.

Tabelle 4: 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

Die Zuteilung der Studierenden zu den Gruppen A und B entspricht einer stratifizierten Randomisierung (R) und liefert die Grundlage für den in Tabelle 5 dargestellten experimentellen Versuchsplan.

Tabelle 5: Versuchsplan der Onlineübungen

	Online Übung 1 27.11.09	Online Übung 2 16.12.09	Abschlussquiz Kriteriumstest 11.1.10
R A: Musterlösung		Testen	O
R B: Testen		Musterlösung	O

Als erste Übung absolvierte Gruppe A die Variante Musterlösung zum Thema „Statistische Grundlagen der pädagogischen Diagnostik“, während Gruppe B zur selben Zeit das entsprechende Quiz beantwortete und unmittelbares Feedback erhielt. Bei der zweiten Übung, ca.

drei Wochen später, stand das Thema „Aufgabenanalyse“ an. Nun wurden die Übungsbedingungen für beide Gruppen getauscht. Gruppe A bearbeitete die Übung als Quiz mit unmittelbarem Feedback und Gruppe B als Variante Studieren korrekt beantworteter Quizfragen. Unmittelbar vor und nach jeder Übung beantworteten die Studierenden anonym ein paar Fragen, die einige wichtige Informationen zur Interpretation der Ergebnisse beisteuern sowie eine Bewertung der Übungsmethoden aus Sicht der Studierenden ermöglichen sollten. Ca. vier Wochen nach der zweiten Übung nahmen alle Studierenden an der unangekündigten unbenoteten Abschlussklausur teil, welche für jede Übung zehn parallele Aufgaben enthielt, die zusammengefasst das jeweilige Behaltenskriterium einer bestimmten Übung messen sollten.

Im Seminarplan, der via Internet jederzeit eingesehen werden konnte, waren die Termine der Online-Übungen exakt festgelegt worden (z.B.: 27.11. Gruppe A und B: Online-Übung „Statistische Grundlagen“). Die Teilnahme an den Online-Übungen war verpflichtend und wurde elektronisch überprüft. Zu Beginn einer Übung wussten die Studierenden zwar nicht, welche Übungsform sie zu bearbeiten hatten, das Thema war jedoch bekannt. Es bezog sich auf die Inhalte der vorausgehenden Seminarsitzung(en). Unmittelbar nach einer übungsrelevanten Seminarsitzung informierte der Seminarleiter die Studierenden via Email, die Powerpoint-Folie aus der vorausgegangenen Seminarsitzung sei jetzt von der Homepage des Seminars erreichbar und es sei ratsam, diese aufmerksam zu studieren, da man so von der Übung besonders profitieren könne. Der Seminarleiter erklärte zu Beginn sowie mehrmals im Verlauf des Seminars, dass die Online-Übung absolut verpflichtend sei, das Ergebnis aber nicht benotet werde. Das Studieren korrekt gelöster Aufgaben entzieht sich zudem einer Benotung, da die Studierenden ja keinerlei Ergebnisse vorweisen müssen. Auf eine Mindestleistung bei der Bearbeitung eines Quiz wurde bewusst verzichtet, um jegliche Art einer Benotung kategorisch auszuschließen. Einige wenige Studierenden lieferten sehr schwache Leistungen ab, was aber zunächst keinerlei extrinsische, zertifikationsrelevante Konsequenz nach sich zog.

Es wurden etliche Empfehlungen gegeben, die Onlineübungen möglichst sorgfältig zu bearbeiten („Handy aus! ICQ und ähnliche Programme sowie Emailsystme auf jeden Fall schließen! Übung konzentriert in einem Zug durcharbeiten, vorbereitet antreten usw.) Die Übungen selbst wurden im Vollbildmodus gestartet, was sonstige weitere unerwünschte Computeraktionen während der Übung bei den meisten Studierenden ausschloss. Es wurde mitgeteilt, das Computerprogramm könne missbräuchliche Nutzung des Programms erkennen.

Testen mit Feedback

Das Online-Quiz 1 sowie die meisten Aufgaben des Online-Quiz 2 wurden von Jacobs (2009a) übernommen. Alle Aufgaben waren auf einer HTML-Seite untergebracht. Vor den Aufgaben erhielten die Studierende nähere Tipps, wie sie Aufgaben bearbeiten sollten. Der Studierende musste jede Aufgabe beantworten bzw. bestätigen, wurde aber nicht, wie in etlichen Studien zu Lösungsbeispielen, dazu gezwungen, sich unter der Kontrollbedingung „eigene Problemlösung“ eine festgelegte Mindestzeit lang mit einer Aufgabe zu beschäftigen und bei Fehlern erst mal weitere, häufig unproduktive Lösungsversuche auszuprobieren (siehe Jacobs 2003). Die Aufgaben gewährten dem Studierenden hier eher das Angebot, die Aufgabe selbst zu beantworten, überließen ihm aber die Entscheidung, wie viel Zeit er dafür in Anspruch nehmen wollte. Sieht der Studierende keine echte Lösungschance, so kann er direkt die Aufgabe bestätigen und das Feedback einsehen.

Unmittelbar nach der Aufgabenbestätigung hatte der Übende Zugriff auf folgende Rückmeldungen

- Knowledge of Response (richtig oder falsch)
- Knowledge of Correct Response (die korrekte Antwort)
- elaboriertes Feedback, [nur dann wenn es als förderlich angesehen wurde.]

Erst nach Beantwortung aller Aufgaben hatte der Studierende die Möglichkeit, die Übung zu beenden und damit die notwendige Zertifizierung für die Übungsteilnahme zu erhalten. Im Anschluss an die Übungsbeendigung folgte dann noch das Feedback „Prozentsatz der korrekt gelösten Aufgaben“.

Musterlösungen in den Online-Übungen

Um die entscheidende Information einer Testfrage in eine Musterlösung zu transformieren, bieten sich mehrere Möglichkeiten an. Wünschenswert wäre eine Transformation, welche die Stärken der Musterlösung möglichst klar und prägnant hervortreten lässt und unnötige Features von Testfragen ausschließt. Wie bereits oben erwähnt, formulierten LaPorte und Voss (1975) Short-Answer-Fragen in Aussagesätze um. Das Weglassen der Fragestellung reduziert die Informationsmenge gegenüber einer Testung und erwies sich dann als effizient, wenn der Fragestellung selbst keinerlei Lernwert beizumessen wäre. Die Fragestellung wurde hier jedoch stets beibehalten, weil ihr eher eine positive Lernwirkung bzw. zumindest keine Lernverschlechterung unterstellt wurde und darüber hinaus ein Weglassen die Informationsmenge nur geringfügig reduziert hätte.

Bei etlichen, vornehmlich komplexeren Aufgabentypen, treten weitere Probleme auf, was man aus der Testfrage streichen sollte oder auch nicht. Die hier verwandte Musterlösungen bei der Seminarübung „Aufgabenformen“ und der ersten Online-Übung „Statistische Grundlagen“ sieht genauso aus wie die entsprechenden Testen-Varianten mit dem einzigen Unterschied, dass in der Musterlösung die korrekten Antworten und die elaborierten Rückmeldungen sichtbar sind. Dadurch enthält die Musterlösung exakt die gleichen Informationen wie die Testen-Variante. Ihr größter Vorteil liegt in der einfachen Konstruktion, weil keinerlei Änderungen gegenüber der Testen-Variante vorgenommen werden müssen.

Bei der zweiten Online-Übung wurden gegenüber der Testen-Variante gewisse Kürzungen vorgenommen. Wenn möglich, sollte auf die Mitteilung der Distraktoren bei MC- oder True-False-Aufgaben verzichtet werden. Eine klassische MC-Aufgabe beim Testen sieht dann in der Musterlösung wie eine beantwortete Short Answer Aufgabe aus. Statt den Lernenden wie beim Testen z.B. entscheiden zu lassen, welche Testkonstrukte oder -beispiele eher auf einen homogenen oder heterogenen Test hinweisen, wurde als Musterlösung eine Tabelle erstellt, welche alle Beispiele korrekt den Begriffen zuordnete und so einen direkten Vergleich im Überblick erlaubte. Die veränderte Fassung der Musterlösung beinhaltet so insgesamt weniger Text als die Testen-Version. Allerdings hält sich die Textreduktion in Grenzen, weil die Fragestellungen und elaborierten Rückmeldungen auch bei den Musterlösungen beibehalten wurden.

Vor der Präsentation der Musterlösungen (=Statements) wurden Ratschläge erteilt, wie man diese möglichst gut bearbeiten sollte.

1. Lesen Sie die Statements aufmerksam durch und versuchen Sie, sich die Lösungen verständlich zu machen.
2. Zur eigenen Kontrolle könnten Sie die vorgegebenen Fragen mit geschlossenen Augen und eigenen Worten beantworten bzw. begründen.
3. Suchen Sie nach Möglichkeiten, die dargebotenen Informationen mit eigenen Erfahrungen zu verbinden.

Bei den Online-Musterlösungen fehlen unmittelbare, harte Kontrollen (etwa Videoanalysen), die einschätzen ließen, ob der Lernende die Informationen überhaupt zur Kenntnis genommen bzw. aktiv verarbeitet hat. Einige Forscher versuchten dieses Problem dadurch zu lösen, den Lernenden jeweils aufzufordern, die korrekte Antwort oder irgendeine Form der Kenntnisnahme oder Bewertung durch Mausklick zu bestätigen (z.B. Hodge, 2009). Auch wenn derartige Kontrollen keine Garantie für aufmerksames Durcharbeiten gewähren, so dürften sie eine gewisse Disziplinierung erreichen. Als objektives Bearbeitungskriterium konnte hier nur die Zeitspanne ermittelt werden, welche die Übungsseite auf dem Computerbildschirm des Studierenden vorhanden war. Man muss man praktisch darauf vertrauen, die Studierenden hätten die Übung im Mittel in konstruktiver Weise bearbeitet. Das gilt jedoch in ähnlicher, wenngleich nicht so offensichtlicher Weise auch für die Testen-Version, da der Studierende dort bestimmte Aufgaben auch einfach ungelesen bestätigen kann.

Zusammenfassung der wichtigsten Rahmenbedingungen für die Online-Quiz

Die Lernwirksamkeit von Tests bzw. Übungen hängt von etlichen Konstellationen ab, welche teilweise über die Eigenschaften der speziellen Übungen hinausgehen. Ob bzw. wie eine Instruktionsmaßnahme wirkt, wird auch von sonstigen Faktoren bestimmt. (siehe z.B. Jacobs, 2010a). Um die Einordnung dieser Bedingungen zu erleichtern, gewährt Tabelle 6 einen Überblick der wichtigsten Faktoren.

Tabelle 6 : Überblick der wichtigsten sonstigen Bedingungen für die Onlinequiz.

Übungstermin	klar festgelegt und lange bekannt, jeweils vier bis fünf Tage nach übungsrelevanter Seminarsitzung
Übungsteilnahme	verpflichtend, notwendig für die Zertifizierung des Seminars
Übungsgrundlage	vorausgehende Seminarsitzung (en) und Powerp.-Folie des Dozenten.
Übungsrelevante Lehrziele	transparent. (Übungsbeispiele in der Powerpointfolie)
Übungsart	für Studierende nicht vorhersehbar: Quiz oder Musterlösung.
Übungsbewertung	unbenotet, bereits zu Seminarbeginn als unbenotet angekündigt.
Aufgabenbearbeitung online zu Hause	bei Quiz: jede Aufgabe musste aktiv bestätigt werden, bei Musterlösung: gesamte Übung musste bestätigt werden.
Ergebnisrückmeldung	bei Quiz: Prozentsatz korrekter Lösungen bei Musterlösung: keine, da nicht möglich
Lernerfolg/Kriterium in Abschlussklausur	unangekündigt, unbenotet, parallele Aufgaben, kontrollierte Testung im Seminar

Aufgabenbeispiele für die Onlineübungen

Hinweis:

Computerfassungen der nachfolgenden, sowie weiterer Beispiele können beim Autor angefordert werden.

Statement = Musterlösung

Beispiel 1 aus der Übung statistische Grundlagen der pädagogischen Diagnostik

<p>Aufgabe 2 Quiz Gegeben sind folgende 7 Daten: 1, 1, 1, 2, 3, 5, 8</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">Nr.</th> <th style="width: 85%;">Bestimme die Statistiken</th> <th style="width: 10%;"></th> </tr> </thead> <tbody> <tr> <td>1.</td> <td style="text-align: right;">Modus</td> <td><input style="width: 20px;" type="text"/></td> </tr> <tr> <td>2.</td> <td style="text-align: right;">Median</td> <td><input style="width: 20px;" type="text"/></td> </tr> <tr> <td>3.</td> <td style="text-align: right;">arithmetischer Mittelwert</td> <td><input style="width: 20px;" type="text"/></td> </tr> </tbody> </table> <div style="display: flex; justify-content: space-around; margin-top: 5px;"> Aufgabe bestätigen Korrekte Lösung? </div> <p style="font-size: small; margin-top: 5px;"><i>Aufgabe bestätigen</i> aktiviert KOR und die <u>elaborierten Rückmeldungen</u> [siehe bei Statement 2] <i>Korrekte Lösung</i> schreibt die korrekten Antworten in die Antwortfelder</p>	Nr.	Bestimme die Statistiken		1.	Modus	<input style="width: 20px;" type="text"/>	2.	Median	<input style="width: 20px;" type="text"/>	3.	arithmetischer Mittelwert	<input style="width: 20px;" type="text"/>	<p>Aufgabe 2: Lernerfolgstest Gegeben sind folgende 6 Daten: 1, 1, 1, 2, 3, 4</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">Nr.</th> <th style="width: 85%;">Bestimme die Statistiken</th> <th style="width: 10%;"></th> </tr> </thead> <tbody> <tr> <td>1.</td> <td style="text-align: right;">Modus</td> <td><input style="width: 20px;" type="text"/></td> </tr> <tr> <td>2.</td> <td style="text-align: right;">Median</td> <td><input style="width: 20px;" type="text"/></td> </tr> <tr> <td>3.</td> <td style="text-align: right;">arithmetischer Mittelwert</td> <td><input style="width: 20px;" type="text"/></td> </tr> </tbody> </table> <div style="display: flex; justify-content: center; margin-top: 5px;"> Aufgabe bestätigen </div>	Nr.	Bestimme die Statistiken		1.	Modus	<input style="width: 20px;" type="text"/>	2.	Median	<input style="width: 20px;" type="text"/>	3.	arithmetischer Mittelwert	<input style="width: 20px;" type="text"/>
Nr.	Bestimme die Statistiken																								
1.	Modus	<input style="width: 20px;" type="text"/>																							
2.	Median	<input style="width: 20px;" type="text"/>																							
3.	arithmetischer Mittelwert	<input style="width: 20px;" type="text"/>																							
Nr.	Bestimme die Statistiken																								
1.	Modus	<input style="width: 20px;" type="text"/>																							
2.	Median	<input style="width: 20px;" type="text"/>																							
3.	arithmetischer Mittelwert	<input style="width: 20px;" type="text"/>																							
<p>Statement 2: Studieren korrekt beantworteter Quizfragen</p> <p>Gegeben sind folgende 7 Daten: 1, 1, 1, 2, 3, 5, 8</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th style="width: 5%;">Nr.</th> <th style="width: 85%;">Bestimme die Statistiken</th> <th style="width: 10%;"></th> </tr> </thead> <tbody> <tr> <td>1.</td> <td style="text-align: right;">Modus</td> <td style="text-align: center;">1</td> </tr> <tr> <td>2.</td> <td style="text-align: right;">Median</td> <td style="text-align: center;">2</td> </tr> <tr> <td>3.</td> <td style="text-align: right;">arithmetischer Mittelwert</td> <td style="text-align: center;">3</td> </tr> </tbody> </table> <ul style="list-style-type: none"> • Modus = Testwert, der am häufigsten vorkommt. Hier 1, da die 1 dreimal vorkommt, die anderen Werte jedoch nur einmal. • Median = mittlerer Wert; eine Hälfte unterhalb, die andere oberhalb. Bei ungerader Anzahl der Probanden, wie hier, gibt es einen Testwert, der genau in der Mitte liegt, hier 2. Bei gerader Probandenanzahl liegt der Median zwischen den beiden mittleren Werten. • Mittelwert = $(1+1+1+2+3+5+8)/7 = 3$ 		Nr.	Bestimme die Statistiken		1.	Modus	1	2.	Median	2	3.	arithmetischer Mittelwert	3												
Nr.	Bestimme die Statistiken																								
1.	Modus	1																							
2.	Median	2																							
3.	arithmetischer Mittelwert	3																							

Beispiel 2 aus der Übung statistische Grundlagen der pädagogischen Diagnostik

Hinweis: Der Text ist hier stark zusammengedrängt, um das Beispiel überhaupt auf einer Seite zeigen zu können. In den Übungen ist Vieles übersichtlicher gestaltet.

Bei Aufgabe 6 waren analog dem ersten Beispiel (oben) die Buttons Aufgabe bestätigen und Korrekte Lösung angebracht, deren Funktionsweise oben bereits geschildert wurde.

Aufgabe 6 Quiz; Genauso wie Statement 6, nur als Testaufgabe

Statement 6: Studieren korrekt beantworteter Quizfragen

Sie sehen unten die Häufigkeitsverteilung von Noten aus den Klassen 1,2 und 3.

Lesebeispiel: 5 Schüler aus Klasse 1 erzielen die Note 3.

Welche Klasse erzielt die niedrigste, die mittlere und die höchste Notenstreuung ?

Versuche, die korrekten Antworten nachzuvollziehen !

Klasse	Häufigkeitsverteilung der Noten in den Klassen 1,2 und 3	a höchste Streuung	b mittlere Streuung	c geringste Streuung
1	Noten: 1 2 3 4 5 6 Häufigkeiten: 1 3 5 4 2 1	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>
2	Noten: 1 2 3 4 5 6 Häufigkeiten: 1 10 11 1 0 0	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>
3	Noten: 1 2 3 4 5 6 Häufigkeiten: 4 3 5 3 4 3	<input checked="" type="radio"/>	<input type="radio"/>	<input type="radio"/>

Man muss für jede Klasse den Mittelwert abschätzen und dann schauen, wie stark wie viele Noten von diesem Mittelwert abweichen. Klasse 2 hat die geringste Streuung, weil die Noten nicht sehr stark vom Mittelwert von ca 2,5 abweichen. Bei Klasse 3 ist die Streuung am höchsten, weil es bei einem Mittelwert von ca 3.5 viele gute und viele schwache Noten gibt.

Aufgabe 6: Lernerfolgstest

Der Lehrer einer Klasse mit 17 Schülern hat 16 Klassenarbeiten korrigiert und die unten dargestellte Häufigkeitsverteilung für die Noten dieser 16 Schüler erhalten. Die Klassenarbeit von Emil wurde aber noch nicht korrigiert und benotet.. Wie muss die Note von Emil ausfallen, um die Notenstreuung der Klasse (also aller 17 Klassenarbeiten) 1.) zu maximieren und 2.) zu minimieren.

	Noten	1	2	3	4	5	6
	Anzahl der Schüler		3	10	3		
1	Note Emils für maximal mögliche Streuung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
2	Note Emils für minimal mögliche Streuung	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Aufgaben bestätigen

Beispiel 1 aus der Übung Aufgabenanalyse

Hinweis: Die elaborierte Rückmeldung wurde hier nur bei der Musterlösung dargestellt. Beim Quiz wird exakt dieselbe elaborierte Rückmeldung nach Anklicken auf *Aufgabe bestätigen* automatisch aktiviert

Aufgabe 1 Quiz

Unter dem Schwierigkeitsindex einer dichotomen Aufgabe (0=falsch, 1=richtig) versteht man

- 1.) den Prozentsatz aller Probanden, welche die Aufgabe richtig gelöst haben.
- 2.) die Anzahl der Probanden, welche diese Aufgabe gelöst haben.
- 3.) den Prozentsatz derjenigen Probanden, welche die Aufgabe falsch beantwortet haben.
- 4.) die objektiv kognitive Belastung bei der Bearbeitung der Aufgabe.

Aufgabe bestätigen

Korrekte Lösung?

Statement 1: Studieren korrekt beantworteter Quizfragen

Was versteht man unter dem Schwierigkeitsindex einer dichotomen Aufgabe (0=falsch, 1=richtig) ?

den Prozentsatz aller Probanden, welche die Aufgabe richtig gelöst haben.

Der Schwierigkeitsindex S ist kontraintuitiv, da ein höherer Prozentsatz korrekter Lösungen inhaltlich auf eine leichtere Aufgabe schließen lässt.

Die relative Häufigkeit korrekter Lösungen entspricht dem Schwierigkeitsindex geteilt durch 100 und liefert eine Schätzung für die Lösungswahrscheinlichkeit der Aufgabe, die wesentlich besser verständlich ist.

Beispiel: 30 % der Probanden lösen eine Aufgabe:

Schwierigkeitsindex = 30%

Lösungswahrscheinlichkeit = 0.3

Aufgabe 1 Lernerfolgstest

Sie erhalten die Information, dass 75 % aller Probanden Aufgabe 1 gelöst haben

Diese Aufgabeninformation entspricht am ehesten

- 1.) dem Schwierigkeitsindex
- 2.) der Trennschärfe
- 3.) der Homogenität
- 4.) der Konsistenz
- 4.) dem Prozentrang

Aufgabe bestätigen

Beispiel 2 aus der Übung Aufgabenanalyse

Hinweis: Auf die Mitteilung der elaborierten Rückmeldung wurde hier aus Platzgründen vollständig verzichtet.

Aufgabe 5 Quiz

Eine trennscharfe Aufgabe in einem Leistungstest..

- 1) ermöglicht eine relativ gute Prognose für die gesamte Testleistung.
- 2) hat meistens einen hohen Schwierigkeitsindex.
- 3) hat meistens einen niedrigen Schwierigkeitsindex.
- 4) wird eher von Leistungsfähigen als von Leistungsschwachen korrekt beantwortet.
- 5) korreliert substantiell positiv mit dem Ergebnis des Tests.
- 6) kann in der Regel von den meisten Probanden nicht gelöst werden.

3 Antworten treffen zu !

Aufgabe bestätigen	Korrekte Lösung?
--------------------	------------------

Statement 5: Studieren korrekt beantworteter Quizfragen

Was ist eine trennscharfe Aufgabe ?

Eine trennscharfe Aufgabe in einem Leistungstest

- ermöglicht eine relativ gute Prognose für die gesamte Testleistung
- wird eher von Leistungsfähigen als von Leistungsschwachen korrekt beantwortet
- korreliert substantiell positiv mit dem Ergebnis des Test

Aufgabe 5: Lernerfolgstest

	Eine trennscharfe Aufgabe in einem Leistungstest..	a richtig	b falsch
1.)	hat in der Regel einen ziemlich hohen Schwierigkeitsindex	<input type="radio"/>	<input type="radio"/>
2.)	wird eher bei mittlerer Itemsstreuung erreicht.	<input type="radio"/>	<input type="radio"/>
3.)	korreliert bedeutsam positiv mit dem Leistungstestwert	<input type="radio"/>	<input type="radio"/>
4.)	wird eher von leistungsstarken als von leistungsschwachen Schülern korrekt gelöst	<input type="radio"/>	<input type="radio"/>
5.)	wird in der Regel nur von sehr wenigen Schülern korrekt gelöst	<input type="radio"/>	<input type="radio"/>
6.)	fördert das Lösungsverhalten bei den meisten Schülern	<input type="radio"/>	<input type="radio"/>

Aufgaben bestätigen

Um die Testmotivation bei der unbenoteten Abschlussklausur möglichst hoch zu halten, wurde auf eine besondere pädagogische Motivierungsmaßnahme zurückgegriffen, die sich bei Jacobs (2009c) gegenüber einem finanziellen Anreiz von mindestens zehn Euro als vergleichbar wirksam erwies.

Ergebnisse

Leistungsergebnisse und Testgütekriterien der eingesetzten Tests

Als Leistungsdaten bzw. Testwerte der Studierenden in Quiz und Kriterium dient der Prozentsatz der korrekten Lösungen, der wie folgt berechnet wurde:

$$\% \text{ korrekt} = (\text{Anzahl aller erzielten Punkte} / \text{Anzahl möglicher Punkte}) \cdot 100$$

Tabelle 7 zeigt die deskriptiven Ergebnisse der beiden Quiz aus vorliegender Untersuchung. Die Reliabilitäten der Quiz und der Lernerfolgsmaße sind zufrieden stellend. Bei der Interpretation der durchaus ansprechenden Parallelretestreliabilitäten ist zu beachten, dass die Quiz mehr Aufgaben enthalten als die Kriterien.

Tabelle 7: Deskriptive Ergebnisse zu den experimentellen Quiz und Kriterien

	Anzahl Aufgaben	Reli α	M	s	N
Quiz 1	17	.66	72.3	12.9	30
Quiz 2	15	.76	67.1	21.1	26
Kriterium 1	10	.70	64.1	19.9	56
Kriterium 2	10	.71	53.1	21.4	55

Parallelretest	Quiz 1 - Kriterium 1	r = .57 (N=29)
Parallelretest	Quiz 2 - Kriterium 2	r = .66 (N=25)
Korrelation zwischen	Kriterium 1 und 2	r = .63 (N=25)

Die Prozentsätze der korrekten Lösungen für ein unbenotetes Quiz können als recht akzeptabel eingeschätzt werden. Hohe Erfolgsquoten stärken den Retrieval-Effekt des Testens. Zudem bleibt hinreichender Spielraum offen, durch das angebotene Feedback (unmittelbares KCR und gegebenenfalls elaboriertes Feedback nach jeder Aufgabenbearbeitung) einige Fehler korrigieren zu können. Wegen der Zufallszuteilung zu den Übungsmethoden ist davon auszugehen, dass die Studierenden unter der Variante Musterlösungen mit einem vergleichbaren Wissen zu den Übungen angetreten waren.

Bearbeitungszeiten der Online-Übungen

Da das individuelle Durcharbeiten bereits korrekt beantworteter Quiz weniger Zeit erfordert als die eigenständige Beantwortung von Quiz, war hier mit kürzeren Bearbeitungszeiten für die Musterlösungen gerechnet worden, zumal die Untersuchungen von Jacobs (2007b, 2008a) diese Hypothese übereinstimmend empirisch eindeutig belegten. Zunächst wurden aus Übung 1 mittels Tukeys Boxplot vier Ausreißer mit überlangen Zeiten aus der Testen mit Feedback-Gruppe, sowie zwei Studierende mit ultrakurzen Übungszeiten aus der Bedingung Musterlösungen ausgeschlossen. Bei der zweiten Übung mussten zwei Ausreißer mit überlangen Zeiten (Einer aus der Quiz-, einer aus der Gruppe Musterlösungen) aus dem Datensatz genom-

men werden. Mit den so modifizierten Daten wurde ein t-Test für unabhängige Stichproben berechnet (siehe Tabelle 8).

Tabelle 8: Übungszeiten in Minuten für beide Übungen

	M	s	N	t	p	d
Testen 1	29.9	14.1	26			
Musterlösung 1	16.0	9.6	24	4.1	.00	1.15
Testen 2	16.3	7.1	25			
Musterlösung 2	14.4	8.6	28	0.83	.39	-

Obwohl die Ergebnisse hinreichend evident erscheinen, wurde zusätzlich der Mann Withney U-Test mit allen Bearbeitungszeiten (also einschließlich Ausreißer) durchgeführt, der die hoch signifikant höheren Quizbearbeitungszeiten bei Übung 1 (z-Wert von 4.6, $p < .001$) sowie den nicht signifikanten Unterschied in Übung 2 (z-Wert von 0.96) bestätigte. Im Median bearbeiteten die Studierenden bei Übung 1 das Quiz 28,3 Minuten und die Musterlösungen 13,3 Minuten lang. In Übung 2 unterscheiden sich auch die Mediane (Quiz 14,7; Musterlösungen 13,4 Minuten) kaum voneinander.

Während das Ergebnis in Übung 1 der Erwartung voll entspricht, fallen die hoch vergleichbaren Lernzeiten beider Übungs-Varianten in Übung 2 aus dem Rahmen. Möglicherweise lässt sich das unerwartete Ergebnis mit einer Intervention des Seminarleiters erklären, der drei Studierende nach sehr kurzen Übungszeiten (kleiner zwei Minuten) anwies, die Übung erneut in Angriff zu nehmen und eine gesonderte Email an alle Studierende unter der Bedingung Musterlösungen sandte, die mit Nachdruck dazu aufforderte, die Übung ernsthaft zu bearbeiten, auch wenn diesmal keine Fragen zu beantworten wären.

Ergebnisse im Lernerfolgstest

Beide Kriteriumsmaße sind Untertests der unangekündigten, unbenoteten Abschlussklausur. In die nachfolgende Auswertung wurden nur solche Studierende aufgenommen, die auch an den jeweiligen Übungen teilgenommen hatten. Viele Studien lassen bessere Ergebnisse für das Testen mit Feedback erwarten. Die eigene Studie (Jacobs 2008a) spricht eher für eine vergleichbare Übungseffektivität. Deshalb wird konservativ zweiseitig getestet (siehe p_z). Tabelle 9 verdeutlicht die deskriptiven Daten sowie das Ergebnis des t-Tests.

Tabelle 9: Prozentsatz der korrekten Lösungen im Lernerfolgstest, 8 bis 4 Wochen nach der Übung

Gruppe	Exp. Bedingung	M	s	N	t	p_z	d
B	Testen 1	62.1	18.6	30			
A	Musterlösung 1	66.3	21.5	26	0.79	0.43	-
A	Testen 2	58.4	21.5	30			
B	Musterlösung 2	47.8	20.0	25	2.05	0.045	0.51

Die Ergebnisse fallen unterschiedlich aus. In Übung 1 erzielten beide Übungsmethoden vergleichbare Behaltenswerte. Ganz anders sieht die Situation in Übung 2 aus. Hier bewirkte das eigene Beantworten der Quizfragen mit anschließendem Feedback einen signifikant höheren Lernerfolg. Diese Unterschiedlichkeit der Lernerfolgsergebnisse für beide Übungen wird auch

durch Interaktion zwischen Übungszeitpunkt (Übung 1, 2) als Within-Faktor und Gruppe (A,B) als Between-Faktor von $F(1,52) = 3,54$ $p_z = 0.065$ statistisch gestützt. Fasst man den Lernerfolg beider Übungen jedoch zusammen, so sprechen die Ergebnisse insgesamt eher für einen vergleichbaren Lernerfolg, was wie folgt begründet wird. Zunächst wurden die Testwerte jedes Kriteriumstests in z-werte transformiert, um die unterschiedlichen Schwierigkeiten beider Kriteriumstests zu egalisieren. Somit lagen für jeden Studierenden ein z-Wert für die Übung Testen und ein z-Wert für die Übung Musterlösung vor, die dann mittels t-test für abhängige Stichproben auf Signifikanz geprüft werden konnten.

Tabelle 10 z-Mittelwerte und Streuungen im Lernerfolgstest für Testen und Musterlösung

	M	s	t(53)	p_z
Testen	.09	.99		
Musterlösung	-.07	.99	1.38	0.174

Tabelle 10 lässt im Mittel der Übungen einen leichten numerischen Vorteil für das Testen erkennen, der aber für das gesetzte Signifikanzniveau von 10% nicht ausreicht.

Zusammenfassend lassen sich die Ergebnisse der Onlineübungen wie folgt bewerten. In Übung 1 erzielten Testen und Musterlösungen vergleichbare Behaltenswerte. Zieht man die deutlich kürzere Übungszeit für die Musterlösungen in Betracht, so erweist sich das Durcharbeiten korrekt beantworteter Aufgaben im Vergleich zu einer aktiven Quizbearbeitung bei vergleichbarer Lerneffektivität als deutlich lerneffizienter. Bei Übung 2 bewirkte das eigene Beantworten der Quizfragen mit anschließendem Feedback im Vergleich zu Musterlösungen einen signifikant höheren Lernerfolg bei vergleichbarer Lernzeit. Insgesamt erwiesen sich so einmal das Studieren korrekt beantworteter Quizfragen und das andere Mal die eigene Quizbearbeitung als die bessere Übungsmethode. Im Durchschnitt beider Übungen sind jedoch keine Lernerfolgsunterschiede nachzuweisen und eher etwas günstigere Lernzeiten für Musterlösungen anzunehmen.

Subjektive Daten

Die Erhebung subjektiver Daten unmittelbar vor (Angstmaße) sowie nach (sonstige Maße) der Übung, diente dem Zweck, die Vergleichbarkeit der Gruppen im Hinblick auf wichtige Lern- und Motivationsvariablen zu prüfen, die möglicherweise durch die Methode selbst bedingt ebenfalls Einfluss auf den Kriteriumstest ausüben könnten, sowie am Rande einige Antworten auf mehr oder weniger plausible Erwartungen an mögliche Unterschiede zwischen beiden Übungsmethoden zu prüfen. Hierbei kamen einige Skalen zur Anwendung, deren Items und Testgütekriterien bei Jacobs (2009a S. 10 ff) näher dargestellt sind. Erst beim Starten der Online-Übung wurden die Studierenden darüber informiert, an welcher Übungsform sie teilnehmen und welchen Sinn und Zweck diese Übungsart erfüllen sollte. Danach beantworteten die Studierenden den [aktuellen Angstfragebogen](#) (Jacobs, 1981). Tabelle 11 zeigt die Ergebnisse, die keinerlei signifikante Unterschiede zu Tage brachten.

Man hätte vermuten können, unmittelbar vor der Übung löse die Erwartung einer unbenoteten Testung mehr aktuelle Angst aus als die Erwartung, korrekt beantwortete Aufgaben durchzuarbeiten. Denn auch bei unbenoteten Testungen kann man versagen, zumal den Studierenden versprochen wurde, sie erhielten später auch Rückmeldung zum Prozentsatz der korrekten Lösungen. Die Daten sprechen aber gegen eine solche Vermutung, da in beiden Übungen keine praktisch bedeutsamen Treatment-Unterschiede in beiden aktuellen Angstmaßen ins Auge

springen. Die aktuellen Angstwerte reflektieren offenbar eher eine milde Anspannung, welche auf die Neuartigkeit der Online-Übung zurückgeht, die sich tendenziell im Verlauf weiterer Übungen etwas legt und die insgesamt zumutbar erscheint.

Tabelle 11: Mittelwerte und Streuungen subjektiver Daten zu zwei Quiz und den analogen Musterlösungen (N pro Bedingung 22-29) [Reliabilität α : Übung1; Übung 2]

Gruppe	Übung 1		Übung 2	
	B Quiz	A Musterlösung	A Quiz	B Musterlösung
aktuelle Prüfungsangst SPA [.97 ; .98]	20.0 13.2	17.9 9.0	16.2 11.2	15.5 10.2
Fearthermometer FT [.86 ; .90]	1.8 2.4	1.6 1.5	1.0 1.6	1.3 1.9
Zeit für Übungsvorbereitung in Minuten	34.6 28.8	29.0 22.2	26.3 21.5	29.8 24.3
Übungsvorbereitung [.56 ; .78]	3.2 0.9	3.5 0.8	3.1 1.0	3.5 0.8
Bedeutsamkeit guter Leistungen [.54 ; .76]	3.6 0.8	3.5 0.7	3.4 0.8	3.4 0.8
Anstrengungsbereitschaft während Übung [.84 ; .76]	4.0 0.6	3.9 0.6	3.6 0.9	3.7 0.6
Akzeptanz der Übung [.79 ; .87]	3.7 0.7	3.4 0.7	3.5 0.7	3.4 0.8
Note für die Übung	2,5 0.7	2.7 0.9	2,6 0.9	2.9 1.0
Item:Übung anstrengend	2.8 0.9	2.5 1.2	2.4 1.0	2.3 0.7

Mögliche Leistungsunterschiede zwischen den Treatments sollten hauptsächlich durch die Bearbeitung der Übungen bewirkt werden und nicht durch unterschiedliche Vorbereitung des Lehrstoffs oder durch Motivationsunterschiede während der Übungsbearbeitung. Eine Analyse belegt über beide Übungen hinweg eine sehr hohe Vergleichbarkeit im Hinblick auf die angegebene Vorbereitungszeit einer Übung sowie hohe Ähnlichkeit im Hinblick auf subjektiven Angaben zur Güte der Vorbereitung. Beide Übungsvarianten wurden mit der gleichen Anstrengungsbereitschaft bearbeitet. Auch ein gutes Ergebnis in einem potentiellen Test zum Thema der Übung hat für beide Gruppen eine ähnlich hohe Bedeutung. Aus subjektiver Sicht liegt demnach eine vergleichbare Übungsvorbereitung und Motivation für beide Übungstypen zugrunde. Die durchaus plausible Annahme, Testen werde anstrengender erlebt als korrekt gelöste Quizfragen durcharbeiten, konnte nicht bestätigt werden. Alle Übungen erscheinen im Mittel weder zu leicht noch zu schwer.

Die Studie von Jacobs (2008a, Langfassung) ergab etwas höhere Akzeptanzwerte für das Testen mit Feedback, sowie teilweise eine signifikant höhere Benotung für das Testen mit Feedback. Die numerischen Unterschiede fallen hier jedoch viel zu gering aus, um die Signifikanzhürde zu nehmen. Die so angestrebte Analyse lässt somit keine klare Schlussfolgerung zu, welche Übungsform Studierende präferieren. Um diesbezüglich sensitivere Aussagen zu

erhalten, kam ein weiterer Untersuchungsansatz zur Anwendung, der den direkten Vergleich beider Methoden aus der Sicht der Studierenden thematisiert.

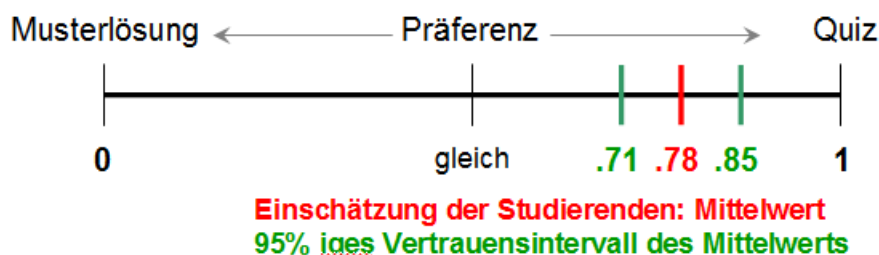
Direkter Vergleich beider Übungsmethoden aus Sicht der Studierenden

Es erscheint relativ schwierig, eine gerade durchgeführte Übung zu bewerten, wenn keine klaren Vergleichskriterien vorgegeben werden, anhand derer die Bewertung vorzunehmen ist. Wie oben aufgezeigt, ließen sich so aus der Sicht der Studierenden keine subjektiven Unterschiede zwischen den Übungsmethoden aufweisen. Beide Übungsvarianten wurden im Schnitt mit einer Note von ca. 2.7 bewertet. Nachdem alle Studierenden im Verlauf des Seminars Erfahrungen mit Testen und Feedback sowie den Musterlösungen gemacht hatten, ergab sich die Möglichkeit, die zwei Übungsmethoden im direkten Kontrast gegeneinander einschätzen zu lassen. Dazu wurde am 04.01.2010 ein kurzer Fragebogen im Seminar zur Beantwortung vorgelegt.

Aufgabe der Studierenden war es, bestimmte Fragen bzw. Bewertungen komparativ für Quiz oder Musterlösung (=Statement) zu entscheiden, wobei als dritte Alternative auch eine Unentschiedenheit möglich war. Auf die Frage „Welche Übungsmethode bewirkt ein besseres Verständnis“ konnte der Studierende sich für *Quiz*, *Statement* oder *weder Quiz noch Statement* entscheiden. Die Antwortkodierung *Quiz* = 1, *Statement* = 0, *weder Quiz noch Statement* = 0.5 erlaubt eine direkte Interpretation der Bevorzugung einer bestimmten Methode. Je näher der Wert bei Eins liegt, desto eher sprechen die Ergebnisse für eine Bevorzugung von Quiz. Die Skala *Präferenz der Übung* [Cronbachs α = .86] setzt sich aus dem Mittelwert von sechs Items zusammen, die inhaltlich auf eine bessere Bewertung der Methode im Hinblick auf Lernwirksamkeit, Verständnis und Interesse sowie eine Bevorzugung der entsprechenden Methode bei Wahlfreiheit hindeuten. Die restlichen Items thematisieren spezifische Stärken bzw. Schwächen der Methoden, die auf offensichtlichen Plausibilitäten basieren.

Die in Abbildung 3 zusammengefassten Ergebnisse zur Präferenz der Übungen sprechen eindeutig für eine Favorisierung der Quiz.

Abbildung 3. Ergebnisse zur Präferenz beider Übungsmethoden im Direktvergleich



Die Präferenz ($M = .78$, $SD = .26$) für das Quiz ist nicht nur statistisch hochsignifikant nach konservativem Binomialtest ($p < .001$), sondern auch von hoher praktischer Bedeutsamkeit. Der hypothetische Populationsmittelwert schwankt mit 95-prozentiger Wahrscheinlichkeit im Intervall von .71 bis .85.

Tabelle 12 stellt für einige Items die prozentualen Häufigkeiten der Entscheidungen für Quiz und Musterlösung sowie die Mittelwerte gemäß der Kodierung dar. Die ersten beiden Items gehören zur Präferenz der Übungsmethoden und konkretisieren in aller Klarheit die höhere Bewertung und Bevorzugung der Quiz.

Tabelle 12: Prozentuale Entscheidungen für Quiz oder Musterlösung [Statement], Mittelwert nach Kodierung (N= 54-55)

Item Nr.	Welche Übungsmethode.....	Quiz %	State ment %	Mittelwert Kodierung
1	ist insgesamt als die bessere Übungsmethode einzuschätzen	80	4	.88
2	würde ich bei Wahlfreiheit selbst auswählen	66	15	.75
3	erfordert eine größere geistige Anstrengung	87	7	.91
4	ermöglicht eine bessere Einschätzung des eigenen Wissens	91	2	.95
5	verleitet eher zu oberflächlichem Arbeiten	9	86	.12
6	erfordert weniger Lernzeit und Lernaufwand	7	64	.22

Alle Entscheidungen der Studierenden fielen statistisch – nach Binomialtest - sehr überzeugend aus (p jeweils $<.001$) und bestätigen die Einschätzung, Testen sei aufwendiger, anstrengender, ließe mehr Verarbeitungstiefe erwarten und liefere eine bessere Einschätzung des eigenen Wissens.

Zusätzlich sollten die Studierenden den pädagogischen Nutzen beider Methoden auf einer Ratingskala zwischen $0 = unbrauchbar$ und $10 = hervorragend$ sowie die Qualität jeder Übungsmethode anhand der üblichen Notenskala von Eins bis Sechs einschätzen. Dieselbe Noteneinschätzung wurde bereits früher in jeder Übung vorgenommen. Hier jedoch konnte der Studierende Quiz und Musterlösung direkt hintereinander bewerten.

Tabelle 13 Mittelwert und Standardabweichung der Ratings zum pädagogischem Nutzen und der Qualität der Übungsmethoden, t-test für abhängige Stichproben (N=55) und Effektstärke.

	Quiz		Musterlösung		t	d
	M	s	M	s		
Pädagogischer Nutzen (0-10):	7.2	1.8	4.7	2.3	6.9	1.26
Qualität via Note (1-6):	2.2	0.7	3.6	1.1	-8.6	1.55

Die Ergebnisse der Tabelle 13 weisen in die gleiche Richtung wie die Befunde zur Präferenz und bestätigen sehr überzeugend die eindeutig bessere Bewertung der Quiz gegenüber den Musterlösungen.

Die spontane Bevorzugung der Quiz gilt offenbar auch für die Übungsform im Seminar via Beamer und Handzeichen. Der Seminarleiter erklärte den Studierenden des nachfolgenden Wintersemesters (WS 2010/11) zunächst die beiden Übungsformen und ließ danach abstimmen, welche der beiden Methoden nun im Seminar zum Thema Aufgabentypen und Aufgabenformen tatsächlich zur Anwendung kommen sollte. Ca. 90% aller Studierenden (49 von 54) entschieden sich für das Testen mit Rückmeldung.

Zusammenfassung und Diskussion

Bei einer Online-Übung erwies sich das Studieren korrekt beantworteter Testfragen im Vergleich zum Testen dieser Fragen einschließlich Feedback bei gleicher Lerneffektivität als deutlich lerneffizienter. Dieses Ergebnis entspricht sehr gut den früheren Befunden von Jacobs (2008a). Bei der zweiten Onlineübung erzielte das Testen gegenüber den Musterlösungen bei gleicher Lernzeit einen signifikant höheren Lernerfolg, was die öfter berichteten Vorteile für das Testen zu belegen scheint. Während bei der zweiten Onlineübung keine befriedigende Erklärung für die Lerneffektivität des Testens gefunden werden konnte, geht die erwartungswidrige vergleichbare Lernzeit beider Methoden höchstwahrscheinlich auf die Intervention des Seminarleiters zurück, der einige Studierende anwies, die Übung wegen sehr kurzer Bearbeitungszeit zu wiederholen und alle Studierenden dieser Bedingung in einer Sonderemail mitteilte, aus gegebenem Anlass darauf hinweisen zu müssen, die Übung sorgfältig durchzuführen. Die Überprüfung beider Methoden im Rahmen einer Seminarsitzung führte schließlich zu sehr hoch vergleichbaren Lernergebnissen für beide Methoden. **Insgesamt fiel der Wettstreit im Bezug auf die Lernförderung letztlich unentschieden aus und damit bleibt das Studieren von Musterlösungen eine ernst zu nehmende Instruktionalternativen zu Testen mit Feedback.** Auch wenn durch das Testen der Abruf relevanter Informationen aus dem Langzeitgedächtnis geradezu herausgefordert wird, war diese Art der Übung hier für den langfristigen Lernerfolg offenbar nicht zwingend erforderlich bzw. deutlich vorteilhafter. Im Gegensatz zu relativ einfachem Faktenwissen reicht erfolgreicher Retrieval allein auch keineswegs aus, die Aufgaben hinreichend lösen zu können. Für den Lernerfolg wichtiger erscheint vielmehr, wie ernsthaft und elaboriert der Studierende die ihm verfügbare Übungsgrundlage selbst durcharbeitet.

Haben beide Übungsmethoden überhaupt einen Lerneffekt?

Man könnte wohlwollend unterstellen, beide Übungsmethoden bewirkten einen höheren Lernerfolg als gar keine Maßnahme. Wenngleich theoretische Überlegungen zum Test- und Feedbackeffekt sowie die bisherigen empirischen Befunde der Test- und Feedbackforschung klare Belege für ihre Lernwirksamkeit gegenüber einer No-Treatment-Bedingung vorweisen können, fehlte hier eine solche Kontrollgruppe, welche einen entsprechenden Zweifel daran restlos ausräumen würde. Ein beruhigender Verweis auf die bisherigen Ergebnisse der Test- und Feedbackforschung allein ist aber insofern problematisch, als die hier geforderten Bedingungen (anspruchsvolle Lehrziele, parallele Aufgaben, langes Retentionsintervall, Einbettung in der Schulpraxis) wesentlich höhere Anforderungen darstellen und von daher schon geringere Effekte erwarten lassen. Zudem hängt die Wirkung einer Übungsmethode immer auch vom vorangegangenen Unterricht und sonstigen Bedingungen ab, die hier im Einzelnen gar nicht alle abgehandelt werden können.

Weitere Studien im Verlauf der hier untersuchten Seminare, bei denen zwei verpflichtende, unbenotete Online-Testungen gegen gar keine Maßnahme getestet wurden, bestätigten lediglich partiell den Vorteil eines Online-Quiz mit vielfältigen Rückmeldungen gegenüber keiner Maßnahme (Jacobs, 2010b). Auch Jacobs (2009a) konnte die Wirkung von unbenoteten Online-Quiz gegenüber keiner Maßnahme nur teilweise belegen, fand allerdings auch für das hier eingesetzte Online-Quiz 2 (Aufgabenanalyse) signifikant bessere Lernerfolgswerte. Diese Befunde geben begründeten Anlass zu der Hoffnung, beide Übungsmethoden hätten tatsächlich einen echten Übungseffekt im niedrigen Effektstärkebereich erzielt.

In der Studie von Jacobs (2008a) erreichten beide Übungsgruppen einen sehr hohen Behaltensvorteil im Vergleich zu einer No-Treatment-Kontrollgruppe, der einer Effektstärke in der

Nähe von $d=1.5$ entspricht und wegen der hohen Probandenanzahl (insgesamt 221 Studierende) auch sehr zuverlässig einzuschätzen ist. Leider ist dieser riesige Lernvorteil beider Übungsgruppen gegenüber der No-Treatment-Kontrollgruppe bei Jacobs (2008a) schwer interpretierbar, weil die Übungen dort auf Freiwilligkeit basierten und keine Zufallszuteilung zur Kontrollbedingung vorlag. Jedoch erscheint es gänzlich unwahrscheinlich, einen derart massiven Unterschied allein den unterschiedlichen Personeneigenschaften der Studierenden anzulasten. Ein Großteil des Lernerfolgsunterschieds geht vermutlich auf die fehlende Abstimmung zwischen Instruktion und Übung sowie auf die Nachlässigkeit etlicher Kontrollschüler zurück, die zum Teil die relevante Vorlesungssitzung gar nicht besuchten und weder die Powerpoint –Folien noch die empfohlene Literatur überhaupt anschauten, dann aber beim unangekündigten Lernerfolgstest völlig überrascht wurden.

Hier wurde ganz besonderer Wert darauf gelegt, den Studierenden im Seminar die Lehrziele verständlich zu vermitteln, den Unterricht möglichst gut auf die Anforderungen der Übungen abzustimmen, die erforderlichen Lehrmaterialien auf die allen direkt verfügbaren Powerpointfolien zu beschränken und die Anwesenheit im Seminar durch Kontrolle von Anwesenheitslisten weitgehend zu garantieren. Dann bleibt vermutlich wenig Spielraum übrig, durch die angebotenen Übungen noch bedeutend große Lerngewinne zu erreichen.

Um den Übungseffekt zu maximieren, müsste man den Unterricht verschlechtern, z.B. etwas mehr Vagheit reinbringen, die Schwerpunkte des Seminars auf Aspekte lenken, die gar nicht geprüft werden, Intransparenz der Prüfungsanforderungen fördern, mit Nachdruck auf eine riesige Literaturliste notwendig zu erarbeitender Quellen verweisen usw. Den größten Lerngewinn durch Übungen, z.B. Testen mit Feedback erzielt man vermutlich damit, Probanden vor dem Test gar nicht zu unterrichten und dann durchweg mit Fragen zu konfrontieren, welche diese mit an Sicherheit grenzender Wahrscheinlichkeit gar nicht lösen können. Etliche Laborexperimente der Feedbackforschung waren so konzipiert und die Forscher berichten von riesigen Effektstärken. Wie immer derartige Experimente theoretisch auch begründet sein mögen, wobei es durchaus auch gute Gründe dafür geben kann, derartige Vorgehensweisen zur Sicherung eines Treatmenteffektes will und kann man sich im realen pädagogischen Feld natürlich nicht erlauben.

Übungen wie z.B. Quiz oder Musterlösungen sind vermutlich dann recht nützlich bzw. deutlicher lernwirksam, wenn es, aus welchen Gründen auch immer, gar nicht möglich ist, alle notwendigen Lehrziele in einem hinreichenden Ausmaß verständlich zu vermitteln.

Grenzen verpflichtender Übungen

Im Idealfall mag man theoretisch den konstruktiven Lernenden im Auge haben, der eine angebotene Übung pädagogisch intentionsgemäß bearbeitet. Allerdings trifft diese Vision im pädagogischen Alltag selten zu. So wurden die Studierenden hier dazu verpflichtet, die Übung durchzuführen, unabhängig davon, ob sie dies wollten oder nicht. Etliche Studierende könnten dieser Forderung mit der Einstellung begegnet sein: „Hier kommt jetzt schon wieder so eine Anforderung auf mich zu, die ich irgendwie hinter mich bringen muss.“ Natürlich bleiben Zweifel, ob es gelungen war, den Sinn der Übungen allen Studierenden hinreichend verdeutlicht zu haben. Jedenfalls deuten einige sehr kurze Bearbeitungszeiten bei Musterlösungen darauf hin, die entsprechenden Studierenden wären in dieser Zeit nicht einmal in der Lage gewesen, die Hälfte der angebotenen Informationen überhaupt zu lesen. Einige wenige Studierende lieferten aber auch unzumutbar schwache Online-Quiz Ergebnisse ab, die in der Nähe von Zufallsantworten liegen. Der Pädagoge im realen Schulalltag kann sich eben nicht nur diejenigen „Probanden“ aussuchen, die an der „Untersuchung“ teilnehmen wollen. Die offensichtliche Unwilligkeit dieser wenigen Studierenden reicht meines Erachtens nicht aus, das Gesamtergebnis dieser Studie ernsthaft anzuzweifeln.

Generalisierungsprobleme der Befunde

Möglicherweise kommen ganz andere Ergebnisse heraus, wenn die Studierenden im Vorfeld bereits genau wüssten, welche Übung sie erwartet. Dann bestünde die Gefahr, Studierende sähen bei Musterlösungen keinerlei Veranlassung, sich ordentlich vorzubereiten. Sie würden womöglich solche Übungsformen ähnlich auffassen wie eine verbale Aufforderung, eine Powerpoint-Folie durchzuarbeiten oder einen Text zu lesen. In solchen Fällen stellt sich natürlich die Frage, ob bzw. auf welche Weise eine Kontrolle ausgeübt bzw. Maßnahmen ergriffen werden könnten, um eine ordentliche Vorbereitung oder zumindest eine Übungsbearbeitung zu garantieren. Derartiges gelingt natürlich eher, wenn man auf eine anstehende echte Klausur verweisen könnte, die ähnliche Aufgaben beinhaltet.

Bei der Übung innerhalb des Seminars wurden die Studierenden durch die kontrollierte Vorgabe der Aufgaben via Beamer dazu veranlasst, beide Übungsvarianten in annähernd gleicher Weise mental aktiv zu bearbeiten. Unter weitgehender Gestaltungsfreiheit einer individuellen Online-Übung ohne ersichtliche Konsequenzen sieht die Lage anders aus. Während selbst das unbewertete Testen durch eine unmittelbare Ergebnisrückmeldung nach der Aufgabenbearbeitung quasi auf natürliche Art die Eigenaktivität anregt und eine gewisse Art von Kontrolle bei der Testbearbeitung ausübt, fehlten derartige Instanzen bei den Musterlösungen, was aber nicht die Schlussfolgerung zulässt, die Studierenden hätten das Übungsangebot überhaupt nicht ernst genommen. Das Studieren korrekt beantworteter Quizfragen erfordert aber ohne Frage eine deutlich höhere Selbstkontrolle.

Bei einer verpflichtenden Übung mit hohem Freiheitsspielraum und geringer Bearbeitungskontrolle ist man auf das konstruktive Wohlwollen der Studierenden angewiesen. Die Zufallszuweisung der Probanden müsste jedoch sicher gestellt haben, dass beide Übungsgruppen mit gleichem Interesse zu ihrer Übung antraten. Auch die Angaben zur Anstrengungsbereitschaft während der Übung waren hoch vergleichbar. Die Ergebnisse der hier untersuchten Online-Übungen führten dennoch zu unterschiedlichen Ergebnissen, aber die Gründe dafür sind nicht hinreichend klar. Es gibt wenige Argumente, die Unterschiedlichkeit mit der Übungsform „Online“ zu begründen, aber es waren bei etlichen Studierenden Tendenzen erkennbar, bei der zweiten Übung die Musterlösungen schnell hinter sich zu bringen. Deshalb sollten Maßnahmen ergriffen werden, auch die Bearbeitung von Musterlösungen irgendwie besser zu kontrollieren.

Bei der zweiten Übung sollten die Musterlösungen optimiert werden, in dem unnötig erscheinende Textteile, welche jedoch beim Testen notwendig waren, heraus genommen wurden. Vielleicht lernen Studierende besser, wenn sie z.B. auch die Distraktoren sehen, die bei den Musterlösungen der Onlineübung 2 fehlten. Dennoch erscheint es gewagt, die unterschiedlichen Ergebnisse beider Onlineübungen damit zu begründen, da derartige Vereinfachungen stellenweise auch bei Jacobs 2008a vorkamen und bestimmte Distraktoren Fehler durchaus auch begünstigen können. Aus Gründen der Einfachheit und besseren Vergleichbarkeit der Experimente wäre es vermutlich eher angebracht, erst dann mit Optimierungen zu beginnen, wenn sich beim Grundtyp der Musterlösung eine gewisse Konsistenz der Befunde abzeichnet.

Subjektive Einschätzungen

Die Analysen zur subjektiven Einschätzung beider Übungsmethoden erbrachten zwei wesentliche Erkenntnisse:

1. Aus methodischer Sicht erscheint es umstritten, subjektive Bewertungen einer Methode ohne Erfahrung mit der Bewertungsalternative miteinander zu vergleichen. Nach dieser Vorgehensweise wurden Studierende unmittelbar nach der von ihnen bearbeiteten Übung gefragt, wie sie diese einschätzten und anschließend die Einschätzungen unter beiden Übungsvarianten mit einem t-Test für unabhängige Stichproben gegeneinander getestet. Dabei konnten überhaupt keine signifikanten Unterschiede in der subjektiven Bewertung beider Übungsmethoden gefunden werden. Durch eine derartige eher konservative Analyse sind höchstwahrscheinlich nur unübersehbare, sehr prägnante Methodenunterschiede statistisch zu sichern. Hatten die Studierenden jedoch bereits Erfahrung mit beiden Methoden gemacht und sollten diese Methoden vergleichend bewerten, so ergaben sich eindeutige Unterschiede.
2. Im direkten Vergleich beider Methoden favorisieren Studierende eindeutig das – **unbenotete** - Testen mit Feedback gegenüber den Musterlösungen. Subjektiv betrachtet wird das Testen zwar als anstrengender und zeitaufwendiger beurteilt, verleitet aber zu weniger oberflächlichem Lernen und ermöglicht eine bessere Einschätzung des eigenen Wissens. Die Präferenzen für das Testen mit Feedback liegen im Vergleich zu den Musterlösungen im sehr hohen Effektstärkebereich, sollten aber nicht überbewertet werden, weil hier primär die Zuverlässigkeit des Unterschieds und nur teilweise das Ausmaß der Unterschieds erfasst wurde.

Schlussfolgerungen für die pädagogische Praxis

Abschließend komme ich zu nachfolgender, vorläufiger pädagogischer Gesamteinschätzung:

- Wenn die Lernenden eine hinreichende Lernmotivation aufweisen, die Übung angemessen zu bearbeiten, dann dürften sie im Normalfall von beiden Übungsmethoden in ähnlicher Weise profitieren, für das Durcharbeiten korrekt beantworteter Quizfragen in der Regel aber weniger Zeit benötigen. Diese Deutung stützt sich auf die Ergebnisse von Jacobs (2008a), bei der nur interessierte Studierende überhaupt an den Übungen teilgenommen hatten.
- Wenn es gelingt, im Unterricht die Studierenden zur aktiven Teilnahme an einer Übung zu motivieren, sind ebenfalls vergleichbare Ergebnisse zu erwarten. Hierfür sprechen die Befunde dieser Untersuchung zum Übungseinsatz während des Seminars und dieses Ergebnis konnte im SS 2010 eindeutig repliziert werden.
- Bei verpflichtenden Übungen mit Verzicht auf direkte leistungsabhängige Sanktionen muss man immer damit rechnen, dass einige Studierende die Übungen nur halbherzig bearbeiten. Das gilt freilich nicht nur für das Studium von Musterlösungen, sondern auch für Testen mit Feedback. Die auf diese Weise angeregte, vermutlich schwächere Lernmotivierung beeinflusst alle möglichen Übungen, nicht nur Onlineübungen ohne direkte Überwachung, weswegen auch unter diesen Bedingungen vergleichbare Ergebnisse für Testen und Musterlösung herauskommen sollten, die hier im Durchschnitt auch bestätigt werden konnten. Für Musterlösungen in Onlineübungen ohne ernsthafte extrinsische Konsequenzen lassen sich aber noch bessere Bearbeitungsaufforderungen einführen, als dies hier geschehen ist.
- Es erscheint wesentlich einfacher, Tests mit wirksamen Kontrollen zu verbinden als das Studieren von Musterlösungen. Somit lässt sich in der Regel eine angeordnete Übungs-

verpflichtung durch Tests deutlich bequemer und effizienter gestalten. Zudem erhalten die Studierenden eine objektive Rückmeldung über ihre Stärken, Schwächen sowie ihr gesamtes Leistungsniveau. Möglicherweise bewirkt die Einführung einer erforderlichen Mindestleistung im Test eine höhere Lern- und Testmotivierung und über diesen Weg auch bessere Leistungsergebnisse.

- Zur gezielten Prüfungsvorbereitung dürfte ein hinreichendes Interesse bestehen, auch freiwillig Musterlösungen zu bearbeiten. Viele Studierende bereiten sich auf solche Weise auf Prüfungen vor, sofern sie auf entsprechende Unterlagen, meist erstellt von Kommilitonen früherer Semester, zurückgreifen können. Wie die eigene Erfahrung zeigt, kann man sorgfältig konstruierte Testaufgaben sehr einfach in Musterlösungen umwandeln. Die bisherigen Ergebnisse geben Anlass zu der Vermutung, auf diese Weise etwas mehr Lehrstoff als durch Testen bewältigen zu können.
- Forschungsergebnisse zu Lösungsbeispielen lassen eine Interaktion zwischen den Übungsmethoden und der Aufgabenschwierigkeit in dem Sinne erwarten, dass eine Testung bei leichteren Anforderungen und Musterlösungen bei schwierigen Anforderungen günstiger erscheinen (expertise reversal effect, Kalyuga et al. ,2003). Letztlich handelt es sich bei den hier untersuchten Instruktionsmaßnahmen nicht um echte Alternativen, sondern um sinnvolle Ergänzungen, bei denen das eigentliche Problem darin besteht, den richtigen Zeitpunkt für die jeweilige Übungsmethode zu finden (Jacobs 2003). Die Quiz waren hier so konzipiert, dass der Lernende im Prinzip selbst entscheiden konnte, ob er die Aufgabe selbst ausdauernd beantworten wollte oder die korrekte Lösung einfach anforderte und damit eine Falschlösung in Kauf nahm. Obgleich unklar bleibt, ob der Lernende eine solche Entscheidung in vernünftiger Weise vornimmt, dürfte diese Wahlmöglichkeit dem natürlichen Lernen eher entgegenkommen als eine zwanghafte Reglementierung. Die relativ hohen Erfolgsquoten in den Quiz deuten darauf hin, der Studierende habe relativ selten von der Möglichkeit, seine Antwort ohne echten Lösungsversuch zu bestätigen, Gebrauch gemacht. Die Gesamtbearbeitungszeit legt die Vermutung nahe, er habe sich andererseits aber auch nicht unnötig lang mit unergiebigem Lösungsversuchen gequält.
- Studierende bewerten **unbenotete Tests mit Rückmeldung** gegenüber dem Studieren korrekt gelöster Aufgaben deutlich positiver und geben an, bei Wahlfreiheit würden sie eher die Testmethode wählen.

Anmerkung: Ich danke dem Reviewer A der Zeitschrift für Pädagogische Psychologie für einige wertvolle Anmerkungen.

Literatur

- Anderson, J. L., Weywadt, C., & McDaniel, M. A. (2008, October). *Repeated quizzing facilitates learning of core content in an undergraduate neuroscience course*. Poster presented at the Harvard Medical School Education Day, Boston, MA [PDF]. Verfügbar unter: http://psych.wustl.edu/memory/TELC/Anderson_et_al_Poster_2008.pdf [8.3.2010].
- Butler, A. C., Karpicke, J. D. & Roediger III, H. L. (2008). Correcting a metacognitive error: Feedback increases retention of low-confidence correct responses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34 (4), 918–928.
- Callender, A. A. & McDaniel, M. A. (2007). The benefits of embedded question adjuncts for low and high structure builders. *Journal of Educational Psychology*, 99 (2), 339–348.
- Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name-learning. *Applied Cognitive Psychology*, 19, 619-636.
- Carpenter, S. K., & Pashler, H. (2007). Testing beyond words: Using tests to enhance visuospatial map learning. *Psychonomic Bulletin & Review*, 14, 474-478.
- Carrier, M. & Pashler, H. (1992). The influence of retrieval on retention. *Memory & Cognition*, 20, 632-642.
- Cooper, G. & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem solving transfer. *Journal of Educational Psychology*, 79, 347-362.
- Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology*, 14, 215-235.
- Duchastel, P. C. & Nungester, R. J. (1984). Adjunct question effects with review. *Contemporary Educational Psychology*, 9 (2), 97-103.
- Hamaker, Ch. (1986). The effects of adjunct questions on prose learning. *Review of Educational Research*, 56 (2), 212-242.
- Hodge, G. (2009, März 24). *Pedagogical best practices: the power of quizzing*. The Redesign Alliance Third Annual Conference, Orlando [PDF]. Verfügbar unter: http://www.thencat.org/RedesignAlliance/2009%20Conference/Presentation%20Slides%2009/Hodges_UNMPOWERofQuizzingPanel.pdf [29.12.2009].
- Jacobs, B. (1981). *Angst in der Prüfung - Beiträge zu einer kognitiven Theorie der Angstentstehung in Prüfungssituationen*. R. G. Fischer-Verlag, Frankfurt/Main
Online Angsttest: URN: urn:nbn:de:bsz:291-psydok-26924;
URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2692/>
- Jacobs, B. (2003). *Feedback mit oder ohne eigene Aufgabenbearbeitung?* URL <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/direktesfeedback.htm> [09.07.2010].
- Jacobs, B. (2004). *Die Wirkung von Lösungsbeispielen, Aufgaben und Feedback auf das Lösen von Kombinatorikproblemen*. URN: urn:nbn:de:bsz:291-psydok-3105 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2004/310/>
- Jacobs, B. (2006). *Erneutes Studieren oder Testen mit Feedback beim Einüben von Faktenwissen am Beispiel des Erlernens der Bundesstaaten der USA*. URL <http://psydok.sulb.uni-saarland.de/volltexte/2006/599/> [09.07.2010].
- Jacobs, B. (2007a). *Die Behaltenswirksamkeit wiederholten Einprägens im Vergleich zu Computer- und selbst gesteuertem Testen mit Feedback*. URN: urn:nbn:de:bsz:291-psydok-26913; URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2691/>
- Jacobs, B. (2007b). *Gezieltes Studieren oder Aufgaben mit Feedback als Übungsmethoden - Erfahrungsbericht eines missglückten Internetexperimentes*. URL <http://www.phil.uni-sb.de/~jakobs/wwwartikel/testread/testread.html> [09.07.2010].

- Jacobs, B. (2008a). *Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback*. URL <http://psydok.sulb.uni-saarland.de/volltexte/2008/1559/> [09.07.2010], Langfassung: URL <http://www.phil.uni-sb.de/~jakobs/wwwartikel/testread/testread2.html> [09.07.2010].
- Jacobs, B. (2008b). Was wissen wir über die Lernwirksamkeit von Aufgabenstellungen und Feedback. In J. Thonhauser (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen* (99-113). Münster: Waxmann.
- Jacobs, B. (2008c). Führt selbst gesteuertes Testen mit Feedback zu höheren Behaltensleistungen als das Einprägen mit Hilfe einer Landkarte? <http://www.phil.uni-sb.de/~jakobs/wwwartikel/teststudy/teststudy3.html>.
- Jacobs, B. (2009a). *Leistungssteigerung durch Notendruck ? - Die Wirkung der Benotung auf die Studierleistungen in einem Seminar*. URL <http://psydok.sulb.uni-saarland.de/volltexte/2009/2529/> [09.07.2010].
- Jacobs, B. (2009b). *Die Wirkung der Übungsverteilung beim Studieren und Testen auf das Behalten der Bundesstaaten der USA*. URL <http://psydok.sulb.uni-saarland.de/volltexte/2009/2354/> [09.07.2010].
- Jacobs, B. (2010a). *Tests als Maßnahmen zur Förderung von Studierleistungen?* URL <http://psydok.sulb.uni-saarland.de/volltexte/2010/2602/> [07.07.2010].
- Jacobs, B. (2010b). Leistungssteigerung ohne Notendruck ? -Die Wirkung verpflichtender, unbenoteter Quiz auf die Studierleistung. URN: urn:nbn:de:bsz:291-psydok-26899 ; URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2689/>
- Jacobs, B. (2009c): *Geld zur Steigerung der Testmotivation- und Leistung in einer unbenoteten Abschlussklausur?* URL <http://psydok.sulb.uni-saarland.de/volltexte/2009/2360/> [09.07.2010].
- Kalyuga, S., Ayres, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist* 38, 23–31.
- Kang, S. H. K., McDermott, K. B., & Roediger, H. L., III. (2007). Test format and corrective feedback modulate the effect of testing on memory retention. *European Journal of Cognitive Psychology*, 19, 528-558.
- Kester, L., & Tabbers, H. K. (2008). The effect of intervening tests on text retention. In J. Zumbach, N. Schwartz, T. Seufert, & L. Kester (Hrsg.), *Beyond knowledge: The legacy of competence* (183-187). Dordrecht, The Netherlands: Springer.
- LaPorte, R. E., & Voss, J. F. (1975). Retention of prose materials as a function of postacquisition testing. *Journal of Educational Psychology*, 67, 259–266.
- Larsen, D. P., Butler, A. C., Roediger III, H. L. (2009). Repeated testing improves long-term retention relative to repeated study: a randomised controlled trial. *Medical Education*, 43, 1174–1181.
- 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.
- 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.
- Nungester, R. J., & Duchastel, P. C. (1982). Testing versus review: Effects on retention. *Journal of Educational Psychology*, 74 (1), 18-22.
- Pilotti, M., Chodorow, M. & Petrov, R. (2009). The usefulness of retrieval practice and review-only practice for answering conceptually related test questions. *Journal of General Psychology*, 136 (2), 179–203.
- Randolph, J. J. (2007). Meta-analysis of the research on response cards: effects on test achievement, quiz achievement, participation, and off-task behavior. *Journal of Positive Behavior Interventions*, 9 (2), 113–128.

- Roediger, H. L. & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255.
- Rohrer, D., Taylor, K., & Sholar, B. (2010). Tests enhance the transfer of learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 233–239.
- Smith, T. A. & Kimball, D. R. (2010). Learning from feedback: spacing and the delay–retention effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition 36* (1), 80-95.
- Toppino T. C, Cohen M. S. (2009). The testing effect and the retention interval: questions and answers. *Experimental Psychology, 56* (4), 252-257.