

Bernhard Jacobs, Fachrichtung Bildungswissenschaften der Universität des Saarlandes.
Email: b.jacobs@mx.uni-saarland.de
Version: 23.3.2012

Musterlösungen oder Testen mit Feedback beim Einüben von Aufgaben zur Wahrscheinlichkeitstheorie

Abstract

Auf der Grundlage des Fragebogens zur Wahrscheinlichkeitstheorie von Nachtigall & Wolf (2001) wurden die Auswirkungen von reinem Testen, Testen mit Feedback und der Bearbeitung von Musterlösungen auf das langfristige Behalten und den Transfer geprüft und in diesem Zusammenhang einige Fragen der Test- und Feedbackforschung analysiert. Studierende des Lehramts wurden ohne vorherigen Unterricht nach Zufall auf die 3 Übungsbedingungen zugeteilt und bearbeiteten die Aufgaben zu Hause via Internet ohne jede Zeitbegrenzung. Eine Woche später folgte überraschend der Behaltentest im Seminar. Die Ergebnisse deuten darauf hin, dass reines Testen keinerlei Behaltenswirkung zeigte, während Testen mit Feedback und die Bearbeitung von Musterlösungen bessere Behaltens- und Transferwerte als reines Testen verursachten. Wie erwartet fielen die Leistungsergebnisse von Musterlösungen und Testen mit Feedback vergleichbar aus und die Bearbeitung von Musterlösungen erforderte weniger Übungszeit. Die Befunde geben Anlass zu der Einschätzung, pädagogische Testverfahren könnten durchaus auch für Übungszwecke nützlich sein, wenn man adäquate Rückmeldungen zur Verfügung stellt oder die Aufgaben als Musterlösungen konzipiert.

Schlagworte: Testen, Feedback, Musterlösung, Übung

Zielsetzung und Hypothesen

Ziel der Studie war es, etliche praktisch pädagogische Fragen zur Behaltenswirksamkeit des Testens, Feedbacks und von Musterlösungen zu prüfen. Während die Laborforschung meist einfache Lernziele wie z.B. Vokabellernen oder triviales Faktenwissen bevorzugt und das Behaltenskriterium letztlich ein Memorieren der eingeübten Fakten prüft, wurde hier ein anspruchsvolleres Lerngebiet aus der Wahrscheinlichkeitsrechnung herangezogen und zusätzlich die Transferwirkung der Maßnahmen getestet. Allerdings fehlte eine klassische Lernaneignungsphase und die Übung begann direkt mit Fragen bzw. Musterlösungen. Das Vorwissen der Probanden basierte daher auf mehr oder weniger intensiven Lernepisoden des entsprechenden Lehrgebietes, die zu weiter zurück liegenden, unterschiedlichen Zeitpunkten oder auch niemals stattgefunden hatten. Als Übungstool diente ein Testverfahren, welches die Kompetenzen von Studierenden in einem Bereich der Wahrscheinlichkeitsrechnung misst, also vorwiegend diagnostischen Zwecken dient. Solche Testverfahren beinhalten eine relativ große Spannweite unterschiedlich schwieriger Aufgaben und zielen darauf ab, die Differenzierung zwischen Könnern und Nichtkönnern zu optimieren, wobei das eingesetzte Messinstrument aber auch Lerneffekte erfassen will. Klassische Testaufgaben eignen sie sich nicht besonders für Übungszwecke, weil ihre Funktion gar nicht darauf ausgerichtet ist, die Anforderungen wohldosiert an das Wissens- bzw. Leistungsniveau der Personen anzupassen.

Trotzdem werden in der Praxis öfter klassische Eignungs-, Schulleistungs- sowie Lernstandstests als Übungen bearbeitet. Der Verfasser dieser Arbeit hat selbst im Internet einige Übungsaufgaben aus zentralen Vergleichsarbeiten für Übungszwecke angeboten (Jacobs 2008). Hierbei stellen sich unter anderem die Fragen, ob eine reine Testung irgendeinen Effekt bewirkt oder ein unmittelbares Feedback im Anschluss an die Antwort ausreicht, zumindest einen kleinen Lernerfolg zu erzielen. Damit soll natürlich keinem Testcoaching oder "teaching to the test" Vorschub geleistet werden, da solche Maßnahmen lediglich eine Steigerung der Testperformance anstreben. Die geistige Auseinandersetzung mit den Aufga-

ben sollte vielmehr einen Lernprozess initiieren, das Verständnis im Sachgebiet fördern und die Kompetenz im entsprechenden Wissensbereich erhöhen. Wenngleich es völlig unrealistisch erscheint, durch ein bisschen Aufgabenpraxis anspruchsvolle Kompetenzen wirksam bewegen zu können, stellt die Bearbeitung von Aufgaben einen zentralen Kern jeden Trainings dar. Für ein erfolgreiches Lernen anspruchsvoller Lehrziele sind Musterlösungen, Aufgabenstellungen und Feedback höchstwahrscheinlich notwendig, aber natürlich nicht hinreichend.

Häufig führte die Bearbeitung eines Tests, also das reine Testen ohne Rückmeldung, zu einer Stabilisierung des Wissens (z.B. Hamaker 1986). Zuvor getestete Personen erzielten bessere Ergebnisse in einem Nachtest als zuvor nicht getestete Personen. Dieser auch aus der Versuchsplanung als Störfaktor der Versuchsplanvalidität bekannte Testeffekt tritt im Lernbereich allerdings vornehmlich dann in Erscheinung, wenn eine Lernphase zeitnah der Testung vorausgegangen war. Da hier die Lernaneignungsphase entfällt, muss man auch nicht zwingend mit einem einfachen Testeffekt rechnen.

Das Einüben bestimmter Aufgaben soll eine Lernwirkung hinterlassen, die sich nicht ausschließlich auf das Eingeeübte beschränkt, sondern einen gewissen Transfer nach sich zieht. Da der überwiegende Teil der Test- und Feedbackforschung einfache Lernformen bevorzugt, gab es bislang wenig Anlass, die Transferhypothese zu prüfen. Aber neuere Forschungen greifen diese Frage zunehmend auf (z.B. Jacobs, 2004, Chan et. al., 2006, Butler, 2010). Da das hier eingesetzte Testverfahren zwei Parallelförmigkeiten umfasst, erlaubte es eine Trennung von Übungs- und Lernerfolgsaufgaben. Nur wenn es gelingen sollte, bei den eingeübten Aufgaben im Nachtest Lernfortschritte nachzuweisen, ergäbe sich möglicherweise auch die Chance, Transfereffekte zu finden. Denn Transfer setzt solides Wissen voraus.

Die Metaanalyse von Kluger & DeNisi (1998) ergab deutliche Feedbackeffekte bei einfachem Lehrzielniveau bzw. reinen Memory-Aufgaben (Effektstärke=.69), aber nahezu keine Wirkung bei geistig anspruchsvollen, komplizierten Aufgaben (Effektstärke=0.03). Es bleibt schwer einzuschätzen, wo das Anforderungsniveau des hier eingesetzten Tests genau anzusiedeln ist. Aber, da der Test auch die Zielsetzung verfolgt, die Leistungsveränderungen durch die Teilnahme an einem Seminar zu erfassen, dürfte es recht schwierig werden, Personen ohne entsprechende Seminarbelegung mit ein paar Erklärungen nach der Aufgabebearbeitung die Prinzipien wirksam vermitteln zu können. Zu Beginn der Studie gab es eine gewisse Hoffnung, zumindest bei den eingeübten Aufgaben Behaltenseffekte durch Feedback oder Musterlösungen zu finden, es lag aber keine klare Erwartung für die Transferaufgaben vor.

Der strenge Testeffekt besagt, Testen mit Feedback sei dem Studieren einer präsentierten Aufgabenlösung (=Musterlösung) im Hinblick auf das langfristige Behalten überlegen. Er konnte in etlichen Laborstudien relativ konsistent für einfaches Faktenwissen (z.B.: Carrier & Pashler 1992; Cull, 2000), Lernen aus Texten (z.B.: LaPorte & Voss, 1975; Roediger & Karpicke 2006) oder Lernen mit Landkarten (z.B.: Jacobs 2006, Carpenter & Pashler, 2007; Rohrer et al. 2010) bestätigt werden. Auch einige Untersuchungen im realen Schulsetting belegten die Vorteile des Testens mit Feedback gegenüber dem Studieren vorgegebener Lösungen (z.B.: Clifton 2005, Larsen et al. 2008). Eigene Studien des Verfassers im natürlichen Lernumfeld mit universitären Lehrzielen ergaben hingegen ziemlich konsistent vergleichbare Behaltensleistungen von Testen mit Feedback und der Präsentation von Musterlösungen (Jacobs 2007, 2010, 2011). Deshalb liegt hier die Erwartung einer vergleichbaren Wirkung beider Instruktionmethoden zugrunde.

Übungsgrundlage und abhängige Variable

Als Übungsgrundlage und Kriteriumstest diente der Fragebogen zur Wahrscheinlichkeitstheorie (FWT) von Nachtigall & Wolf (2001), der von den Autoren auch als [FWT-Computerversion](#) im Internet angeboten wird. Angaben der Autoren zufolge ist der FWT "ein Instrument zur Erfassung des Wissens und der Wissensveränderung im Bereich elementarer Stochastik". Er wurde allerdings nicht exakt entsprechend der Handanweisung dargeboten, sondern für die Untersuchungszwecke modifiziert eingesetzt. Wegen der zwischenzeitlichen Einführung des Euro mussten die Aufgabenstellungen für DM-Münzen auf Euromünzen angepasst werden. Im Übungsteil unterscheiden sich Testanweisung und Durchführung von der Standardvariante vorwiegend durch den Einsatz einer eigenen Computerversion, der Veränderung der Aufgabenreihenfolge und einer zeitunbeschränkten Bearbeitung. Die experimentellen Bedingungen in der Übungsphase beziehen sich neben einer reinen Testung auf eine Testung mit unmittelbarem Feedback sowie einer Modifikation des FWT als Musterlösung. Um die Auswirkungen der Übungsvarianten zu prüfen, kam der FWT eine Woche später während des Seminars in der Papierfassung zum Einsatz. Die Bearbeitungszeit wurde dort entsprechend der Handanweisung auf 15 Minuten beschränkt, die Aufgabenreihenfolge jedoch größtenteils zufällig angeordnet, wobei strukturähnliche Aufgaben nicht direkt aufeinander folgen durften. Der FWT besteht aus 14 Aufgaben und setzt sich aus den Parallelförmigen A und B zusammen, die jeweils 7 Aufgaben umfassen. Parallelförmige A diente als Übungsbasis für die experimentelle Phase und als identisches Behaltenskriterium. Durch die Parallelförmige B sollte die Transferwirkung der Übungsvarianten überprüft werden. Nachfolgende Übersicht in Tabelle 1 bezeichnet die eingesetzten Messvariablen und ordnet sie dem FWT zu.

Tabelle 1: Übersicht der eingesetzten Übungs- und Kriteriumstests

FWT_A1 = Parallelförmige A, Übungsphase ohne Zeitbegrenzung, Zeitpunkt 1, online zu Hause
 FWT_A2 = Parallelförmige A, Behaltentest, Zeitpunkt 2, Papierfassung im Seminar
 FWT_B2 = Parallelförmige B, Transfertest, Zeitpunkt 2, Papierfassung im Seminar
 FWT2 = FWT_A2+FWT_B2, Papierfassung im Seminar

Die Aufgaben des FWT basieren auf dem Short Answer Aufgabenformat. Hierbei muss der Proband meist eine Dezimalzahl, einen Bruch oder einen mathematischen Term (z.B. $1/6 * 1/6$) eingeben, der als richtig (= 1 Punkt) oder falsch (= 0 Punkt) bewertet wird. Der Testwert ergibt sich durch Aufsummieren der gelösten Aufgaben. Er wird hier jedoch stets als Prozentsatz der korrekten Lösungen dargestellt. Angaben zu den Testgütekriterien sind dem Artikel von Nachtigall und Wolf (2001) zu entnehmen. In Tabelle 2 werden die Reliabilitäten der hier eingesetzten Varianten dargestellt. Sie beziehen sich auf alle Probanden, die an den jeweiligen Tests teilgenommen hatten. Die Korrelation der beiden Testhälften von $r=.80$ entspricht recht gut den ermittelten Werten von Nachtigall und Wolf (2001).

Tabelle 2: Ermittelte Reliabilitäten der hier erhobenen Messinstrumente des FWT

| | α | split half | Anzahl Items | N |
|--------|----------|---------------|-----------------|-----|
| FWT_A1 | .68 | .68 | 7 | 63 |
| FWT_A2 | .71 | .71 | 7 | 100 |
| FWT_B2 | .67 | .73 | 7 | 100 |
| FWT2 | .83 | .89 | 14 | 100 |

Retestreliabilität: Korrelation FWT_A1 mit FWT_A2 = .57 (N=58, Zeitabstand eine Woche)

Paralleltestreliabilität: Korrelation FWT_A2 mit FWT_B2 = .80 (N=100)

Untersuchungsvorgehen

An der Studie nahmen Studierende des Lehramts aus 4 Seminaren teil, die der Verfasser im WS 11/12 leitete [zwei Seminare zu Grundlagen der pädagogischen Diagnostik und 2 Seminare zu Grundlagen der pädagogischen Forschungsmethodik und Evaluation]. Das Durchschnittsalter betrug 23 Jahre. Ca. 2/3 der Probanden waren Frauen. Die Probanden studieren eine große Breite unterschiedlicher Studienhauptfächer und ihr Wissen im Bezug auf den Übungstest dürfte stark variieren. Den Studierenden wurde im Vorfeld mitgeteilt, sie sollten an einer Onlineerhebung teilnehmen, die den Zweck verfolgte, Anschauungsmaterial und Datengrundlage für etliche diagnostische Berechnungen sowie die Evaluation von Maßnahmen zu liefern. Etwa 90% aller SeminarteilnehmerInnen beteiligten sich an der Untersuchung. In einer kurzen Vorerhebung, die den Studierenden als Versuch deklariert wurde, die Funktionsfähigkeit des Onlineablaufs zunächst zu überprüfen, wurde der Abiturnotendurchschnitt erfragt, um ihn als Parallelisierungsvariable für den Versuch zu nutzen. 85% machten hierzu entsprechende Angaben. 85% aller Probanden wurden zunächst nach Abiturnotendurchschnitt in Tripel eingeteilt und innerhalb jedes Tripels nach Zufall den experimentellen Bedingungen zugewiesen. Die verbliebenen 15 % wurden nach dem klassischen Randomisierungsverfahren zugeteilt.

Die Organisation der Studie wurde über Email und Internetbrowser abgewickelt. Hierbei gewährte der Versuchsleiter den Probanden einen gewissen zeitlichen Rahmen, die Übung zu Hause durchzuführen. Entsprechend der experimentellen Zuordnung bekamen die Probanden zunächst eine Email zugesandt, welche URL, Username und Passwort für die betreffende Übung enthielt sowie den Zeitrahmen für die Übungsrealisation absteckte. Die Onlineübung zur Wahrscheinlichkeitstheorie war integriert in eine Übungsserie zum Erlernen von Vokabeln (siehe genaue Darstellung bei Jacobs, 2012, S.7). In der Seminarsitzung, eine Woche später teilte der Seminarleiter für alle Studierenden völlig überraschend den vollständigen Fragenbogen zur Wahrscheinlichkeitstheorie [FWT2] aus und kündigte ein Zeitlimit von 15 Minuten an. Da die zuvor anberaumte Übung einen gewissen zeitlichen Freiheitsspielraum zur Bearbeitung gewährte, variiert der Zeitabstand zwischen Übung und Behaltenstestung für einzelne Probanden. Die statistischen Kennwerte lauten: Median 7 Tage, Mittelwert 6,7 Tage, Streuung 1,2 Tage. Der Einfachheit halber gilt hier die Festlegung: Retentionsintervall = eine Woche.

Tabelle 3: Versuchsplanformalisierung

| | Übung | Behalten |
|---|----------------------------|----------|
| R | Testen | ○ |
| R | Testen mit Feedback | ○ |
| R | Musterlösung | ○ |

Wie aus der Formalisierung des experimentellen Versuchsplans in Tabelle 3 hervorgeht, wird das Testen der beiden ersten Bedingungen vornehmlich als Übungsbestandteil interpretiert, wenngleich man es natürlich auch als Vortestung betrachten und nutzen kann.

Unabhängige Variablen - die experimentellen Bedingungen

Die Probanden wurden zunächst darüber informiert, im Folgenden würden Sie ohne Zeitbegrenzung ein kleines Quiz zur Wahrscheinlichkeitsrechnung bearbeiten und dadurch die Möglichkeit erhalten, zu erfahren, was Sie zu diesem Thema wüssten. Es gab keinerlei Hinweis für eine erneute Testung zu einem späteren Zeitpunkt. Die 7 Aufgaben des FWT_A1

waren auf einer HTML-Seite untergebracht und durch einen Leerraum voneinander getrennt, so dass jede Aufgabe ohne Sicht auf die restlichen Aufgaben auf dem Bildschirm platziert war. Die experimentellen Bedingungen in Tabelle 4 beziehen sich auf die Variation der Instruktion bei der Bearbeitung von 7 Aufgaben des Fragebogens zur Wahrscheinlichkeitstheorie [FWT_A1].

Tabelle 4: Beschreibung der Übungsvarianten

| | |
|--|--|
| Übungsvarianten | [siehe auch die Bildschirmkopie in Anhang] |
| Testen ohne Feedback | |
| Die Methode beschränkt sich auf eine reine Testung. Der Proband schreibt sein Ergebnis in das Antwortfeld und klickt anschließend auf den Button "Aufgabe bestätigen", der sich dann zum Button "erledigt" wandelt. | |
| Testen mit Feedback, hier meist abgekürzt bezeichnet als Feedback | |
| entspricht der reinen Testmethode mit dem Unterschied, dass nach Bestätigung der Aufgabe Feedback folgt: Knowledge of response: richtig oder falsch Knowledge of correct response: die korrekte Antwort Elaboriertes Feedback: Knappe Erklärung des Lösungsweges. Dieses Feedback konnte beliebig oft eingefordert werden. | |
| Musterlösung | |
| Die Musterlösung umfasst analog dem Testen die Fragestellung, beinhaltet jedoch zugleich die korrekte Antwort und alle elaborierten Rückmeldungen, die beim Testen mit Feedback gewährt wurden. | |

Zu Beginn jeder Übungsvariante erhielten die Probanden einige Tipps, wie sie an die Aufgaben herangehen sollten. Ziel dieser Maßnahme war es, das Bemühen um Verständnis anzuregen und Selbsterklärungen zu provozieren. Um das Auslassen oder Ignorieren mancher Aufgaben zu verhindern, mussten die Probanden stets die Bearbeitung jeder Aufgabe bzw. Musterlösung durch Mausklick auf einen Button bestätigen. Erst danach konnte die Übung beendet werden. Für die Testvarianten bedeutete diese Regelung, im Falle des Unwissens, die Aufgabe ohne Beantwortung oder mit einer vagen Schätzung zu bestätigen. Bei der Übungsmethode Testen mit Feedback löste diese Bestätigung zwingend das Feedback aus. Rückmeldungen zum Prozentsatz der korrekten Lösungen waren weder vorgesehen noch angekündigt worden.

Ergebnisse

Vergleichbarkeit der experimentellen Gruppen

Durch die vor dem Versuch realisierte Parallelisierung nach Abiturnotendurchschnitt und der anschließenden Randomisierung sollte die Vergleichbarkeit der schulischen Leistungsfähigkeit aller experimentellen Gruppen sicher gestellt werden. Tabelle 5 bezieht sich auf diejenigen Probanden, welche letztlich den gesamten Versuch auch vollständig absolviert hatten und zuvor die entsprechenden Angaben gemacht hatten, also nicht auf alle am Versuch beteiligten Probanden. Sie belegt die hohe Ähnlichkeit hinsichtlich Abiturnotendurchschnitt und Mathematikleistung (Punktesystem).

Tabelle 5: Abiturnotendurchschnitt und Mathematikabitur

| | Abitur | | | | |
|--------------|------------------------|-----|------|------------------------------|-----|
| | Durchschnitt (Note) | | | Mathematik (Punktesystem) | |
| | N | M | s | M | s |
| Testen | 27 | 2.2 | 0.57 | 9.5 | 4.0 |
| Feedback | 25 | 2.4 | 0.58 | 9.7 | 2.4 |
| Musterlösung | 25 | 2.1 | 0.73 | 9.6 | 3.9 |

keinerlei signifikante Unterschiede zwischen allen Gruppen

Nachfolgende Analyse in Tabelle 6 greift auf alle verfügbaren Probanden zu, die am Kriteriumstest teilgenommen hatten. Sie lässt erkennen, dass eine Parallelisierung nach dem Mathematikabitur noch etwas effizienter gewesen wäre.

Tabelle 6: Korrelationen zwischen FWT und Abiturleistungen (N jeweils 89)

| | FWT_A2 | FWT_B2 | FWT2 |
|---------------------|--------|--------|--------------|
| Abi-Durchschnitt | -.28** | -.15 | -.23* |
| Abi-Mathe (1..15) | .35** | .30** | .34** |
| Abi-Deutsch (1..15) | -.06 | -.02 | -.05 |

Erwartungsgemäß korreliert der FWT mit dem Mathematikabitur signifikant in erwarteter Richtung, während kein Zusammenhang mit dem Deutschabitur festgestellt werden konnte. Die Korrelationsbeziehungen sehen ähnlich aus, wenn man nur die 77 Studierenden einbezieht, welche sowohl an Übung und Kriterium teilgenommen hatten als auch ihre Abiturleistungsdaten angegeben hatten. Hier sind alle Beziehungen mit Mathematik signifikant, während die Zusammenhänge mit dem Abiturnotendurchschnitt zwar stets in erwarteter Richtung liegen, aber die Signifikanzhürde nicht überwinden konnten. Deutsch korreliert wie in Tabelle 6 stets negativ, aber insignifikant.

Vorwissen der Probanden

Die beste Schätzung für das Vorwissen aller am Versuch beteiligten Gruppen liefern die Ergebnisse der Bedingung Testen, da die entsprechenden Probanden ausschließlich getestet wurden und alle Studierenden nach Zufall den Bedingungen zugewiesen wurden. Die Feedbackgruppe unterzog sich zwar ebenfalls einer Testung, erhielt aber unmittelbare Rückmeldungen zu den Aufgaben, die möglicherweise Hinweise zur Lösung anderer Aufgaben hätten liefern können. Allerdings erscheinen solche Übungseffekte durch das sachorientierte Feedback für weitere Aufgaben des Tests sehr unwahrscheinlich, weil - etwa im Gegensatz zu einer Menge von Analogieaufgaben in einem Intelligenzsubtest - jede Aufgabe des FWT_A1 ein anderes Lehrziel erfasste. Die unmittelbare Rückmeldung nach der Aufgabenbeantwortung könnte aber die Motivation und das Interesse gesteigert und über diesen Umweg eine sorgfältigere und höher konzentrierte Aufgabenbearbeitung angeregt haben. Zudem erhält der Proband durch das KOR-Feedback eine höhere Sicherheit im Hinblick auf korrekte Eingaben, z.B. , dass die Wahrscheinlichkeit nicht als Prozentwert angegeben werden darf.

Tabelle 7 verdeutlicht die Ergebnisse der Übungsphase für die Bedingungen Testen sowie Testen mit Feedback.

Man kann davon ausgehen, dass die zum Versuch angetretenen Probanden im Durchschnitt ein Wissensniveau von ca. 45% korrekter Aufgaben im FWT aufwiesen. Diese Leistung muss im Übrigen höher (niedriger) eingeschätzt werden als die, welche die Autoren bei Erhebungen

an mehreren Universitäten unmittelbar vor (nach) einem Statistikseminar für Psychologen ermittelt hatten.

Tabelle 7: Ergebnisse in der Übung für den FWT_A1

| | M | s | N | t | p _z |
|----------|------|------|----|-------|----------------|
| Testen | 42.9 | 25.7 | 31 | | |
| Feedback | 50.3 | 24.4 | 27 | -1.13 | .26 |

Die Feedbackgruppe erzielte zwar numerisch etwas höhere Testwerte, beide Gruppen unterscheiden sich aber nicht signifikant voneinander. Beim FWT_A1 ergab sich die Möglichkeit, den Zusammenhang zwischen dem Prozentsatz der korrekten Lösungen und der frei gewählten Bearbeitungszeit zu berechnen, der nach der Elimination von 7 Ausreißern $r = .28$ ($p_e = 0.048$) beträgt und darauf hinweist, dass die Leistungen etwas höher ausfielen, wenn mehr Zeit investiert wurde.

Übungszeiten

Unterschiede im Lernerfolg sind nur bei annähernd vergleichbaren Übungszeiten der experimentellen Übungsvarianten klar zu interpretieren. Zu Beginn des Versuchs lag die Erwartung zugrunde, bei freier Aufgabenbearbeitungszeit beanspruchten Musterlösungen weniger Zeit als Testen mit Feedback, da bei einer Musterlösung die Testung entfällt und ansonsten die gleichen Informationen zur Verfügung stehen. Außerdem wurde die Hypothese durch einige Studien des Verfassers bereits mehrfach empirisch belegt (z.B. Jacobs 2008, 2011). Da etliche Studierende überlange Zeiten aufwiesen, wurde die Hypothese auf zweifache Weise getestet. Die Berechnung des Medians und des U-Tests basiert auf den Daten aller Probanden, die den Versuch vollständig durchgeführt hatten, für die Berechnung von Mittelwert, Streuung und t-Tests wurden 7 Ausreißer aus dem Datensatz entfernt.

Tabelle 7: Übungszeiten der Übungsmethoden in Sekunden

| | N | Median | N | M | s |
|--------------|----|------------|----|------------|-----|
| Testen | 31 | 448 | 25 | 402 | 157 |
| Feedback | 27 | 416 | 26 | 433 | 191 |
| Musterlösung | 31 | 285 | 31 | 345 | 182 |

Die in Tabelle 7 dargestellten Ergebnisse lassen sowohl im Median wie auch dem arithmetischen Mittel kürzere Übungszeiten der Musterlösung gegenüber dem Testen mit Feedback erkennen. Die Unterschiede zwischen Musterlösung und Feedback konnten sowohl mittels t-Test ($t(55) = -1.77$; $p_e = 0.04$) wie durch den U-Test ($z = -2.27$ $p_e = 0.012$) und zusätzlich mit den logarithmisch [LN] transformierten Zeiten aller Probanden einschließlich Ausreißer ($p_e = 0.018$) statistisch bestätigt werden und erreichten im Ausmaß ca. $d = 0.5$ Effektstärke.

Normalerweise wäre auch zu erwarten gewesen, dass Testen mit Feedback mehr Zeit beansprucht als reines Testen. Leider konnten beim Testen mit Feedback die Zeiten nicht getrennt für das Testen und die Feedbacknutzung erhoben werden. Jedenfalls sollte man aus dem geringen Zeitunterschied der beiden Methoden nicht voreilig folgern, dem Feedback sei kaum Zeit gewidmet worden. Möglicherweise hatten die Teilnehmer der reinen Testgruppe länger gewartet, bis sie sich zur Bestätigung einer Aufgabe durchgerungen hatten.

Ergebnisse im Nachtest

Eine Woche nach der Übung teilte der Seminarleiter für die Studierenden überraschend den FWT im Papierformat aus und kündigte eine Bearbeitungszeit von 15 Minuten an. Die Antworten der Probanden wurden später in die Computerfassung übertragen und automatisch ausgewertet. Tabelle 8 stellt die wichtigsten Ergebnisse zusammen. Die Ergebnisse bestätigen weitgehend die Erwartungen hinsichtlich des Behaltens. Aber auch die Befunde zum Transfer liegen in der gleichen Richtung. Der Gesamttest fasst beide Aspekte zusammen und liefert natürlich keine weitere Information.

Tabelle 8: Ergebnisse im Fragebogen zur Wahrscheinlichkeitstheorie

| | FWT_A2 Behalten | | FWT_B2 Transfer | | FWT2 Gesamttest | | N |
|------------------------------|--------------------------------------|------|-------------------------------------|------|--------------------------------------|------|----|
| Testen | 40,1 | 22.6 | 42.9 | 23.0 | 41.5 | 21.4 | 31 |
| Feedback | 55.6 | 27.0 | 55.6 | 26.4 | 55.5 | 24.9 | 27 |
| Musterlösung | 51.6 | 28.7 | 53.0 | 29.1 | 52.3 | 27.8 | 31 |
| Kontrast (F+M): T | t(86)=2.31 p _e = 0.012 | | t(86)=1.95 p _e =0.027 | | t(86)=2.24 p _e = 0.014 | | |
| Einzelvergleiche nach LSD | | | | | | | |
| | p _e | d | p _e | d | p _e | d | |
| F > T | 0.014 | .62 | 0.035 | .51 | 0.018 | .60 | |
| M > T | 0.044 | .45 | 0.066 | .39 | 0.046 | .44 | |
| F > M | 0.285 | | 0.355 | | 0.312 | | |

Zunächst einmal verdeutlicht ein Kontrast zwischen solchen Methoden, welche instruktionale Unterstützung bieten, und dem reinen Testen, dass Erklärungen zum Lösungsweg mehr Lerngewinn bewirkten als reines Testen. Dieser Unterschied konnte auch bei den Einzelvergleichen statistisch belegt werden. Denn sowohl die Musterlösung, wie auch Testen mit Feedback erzielten bessere Testwerte als die einfache Testung und die Lernvorteile entsprechen in etwa einer mittleren Effektstärke.

Zugleich verdeutlichen die Mittelwerte sowie die letzte Zeile aus Tabelle 8 die vergleichbare Lernwirkung von Testen mit Feedback und dem Durcharbeiten einer Musterlösung. Selbst bei einseitiger Testung der Hypothese, Testen mit Feedback sei einer Musterlösung überlegen, liefen die LSD-Tests Wahrscheinlichkeitswerte, welche die Signifikanzhürde nicht nehmen können.

Leistungsentwicklung der Test – und Feedbackgruppe

Durch die Randomisierung sollten die Ausgangsbedingungen aller experimentellen Gruppen vergleichbar ausgefallen sein, so dass die in Tabelle 8 aufgezeigten Unterschiede bereits hinreichend die positive Wirkung des Feedbacks untermauern. Andererseits garantiert die Randomisierung keine absolute Vergleichbarkeit, auch wenn durch die Parallisierung des Abiturnotendurchschnitts die Vergleichbarkeit der schulischen Leistungsfähigkeit aller Gruppen gegeben ist. Aus Tabelle 7 ging nun hervor, dass die Ergebnisse der Feedbackgruppe im FWT_A1 zumindest numerisch besser ausfielen als die Ergebnisse der Testgruppe. Die Hypothese zum Vergleich der Leistungsentwicklungen beider Methoden von der Übung zum Behaltenstest geht davon aus, die Leistungen der Feedbackgruppe müssten sich mehr verbessern als die der reinen Testgruppe.

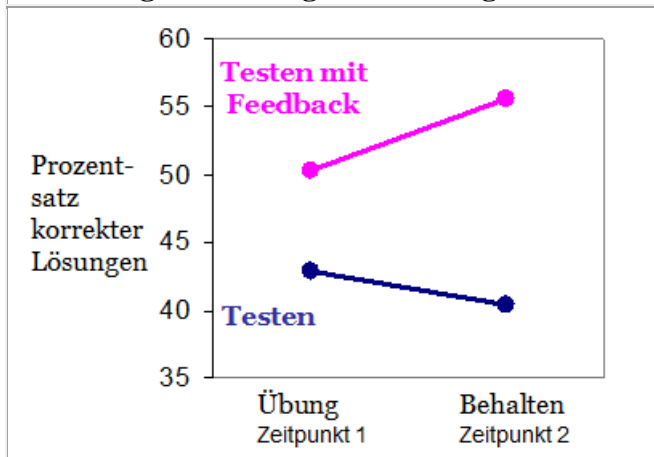
Abbildung 1: Leistungsentwicklung im FWT_A

Abbildung 1 deutet die erwartete Wechselwirkung an. Eine zweifaktorielle Varianzanalyse mit dem Zeitpunkt (1,2) als within subject factor und den Übungsmethoden (Testen, Testen mit Feedback) als between subject factor erbrachte aber keine deutlich bessere Leistungsentwicklung zugunsten von Testen mit Feedback, da die Interaktion zwischen Zeitpunkten und Übungsmethoden $F(1,56)=1,7$; $p_e=0,10$ die übliche Signifikanzhürde von 5% nicht überspringen konnte. Allerdings widerlegt die insignifikante Interaktion auch nicht zwingend die Wirksamkeit der Feedbackmethode, weil die Messung des FWT_A1 bereits von der Übungsmethode beeinflusst gewesen sein könnte. Denn unter Feedbackbedingungen folgten unmittelbare Konsequenzen (z.B.: „richtig/falsch“), die möglicherweise eine höherer Sorgfalt gegenüber reinem Testen bewirkt haben könnten. Andererseits scheint die didaktische Funktion des sachorientierten Feedbacks im Anschluss an die Aufgabenbearbeitung weniger überzeugend ausgefallen zu sein.

Zu beachten gilt weiterhin, dass die Übung zeitunbeschränkt und ohne Kontrolle zu Hause durchgeführt wurde und der Behaltenstest mit Zeitbeschränkung in kontrollierter Weise ablief. Unter kontrolliert strengeren Durchführungsbedingungen sind vermutlich etwas geringere und zudem etwas validere Testwerte zu erwarten.

Meistens bewirkt reines Testen keinen Lernfortschritt, sondern lediglich eine Behaltensstabilisierung. Die Leistungsentwicklung der Testgruppe in Abbildung 1 deutet aber eher darauf hin, reines Testen habe im Mittel weder einen Lern- noch einen Stabilisierungseffekt nach sich gezogen. Für den strengen Nachweis dieser These fehlt allerdings eine unabhängige No-Treatment-Kontrollgruppe, die ausschließlich zum Zeitpunkt 2 getestet worden wäre und dann gegenüber der hier realisierten Testgruppe hoch vergleichbare Ergebnisse erzielt hätte. Aus rein diagnostischer Perspektive sprechen die Befunde der Testgruppe für das Testinstrument, weil eine Testbearbeitung zum Zwecke der Diagnose auch keinen Übungseffekt bewirken soll.

Zusammenfassung und Diskussion

Vorliegende Untersuchung ist vornehmlich als Laborstudie aufzufassen, auch wenn die experimentellen Übungen in der natürlichen Umwelt stattfanden. Im Gegensatz zum normalen schulischen Unterricht fehlte die Lernaneignungsphase. Die Studierenden wurden unter dem Vorwand einer Erhebung zur Bearbeitung eines Lehrstoffs animiert, der nichts mit den curricularen Lehrzielen im Studium zu tun hatte. Das Experiment hat mehrere für die Test-

und Feedbackforschung wesentliche Fragen aufgegriffen und ist zu folgenden Ergebnissen gekommen:

1. Das reine Testen in der experimentellen Übungsphase hatte im vorliegenden Fall im Mittel überhaupt keine Lernwirkung im Behaltenstest. Die hier gefundene Leistungsveränderung [Nachttest-Vortest] von $d=-0.12$ widerspricht dem durchschnittlichen Vorteil einer Wiederholungsmessung bei kognitiven Fähigkeitstests aus der Metaanalyse von Hausknecht et al. (2007). Allerdings widerlegt der Befund keineswegs generell die lernfördernde Wirkung eines Testens, sondern ist auch der besonderen Versuchskonstellation geschuldet. Denn üblicherweise geht der Testung im pädagogischen Bereich eine zeitnahe Instruktionsphase voraus. Die erfolgreiche Erinnerungsbemühung im Test an das zuvor Gelernte bewirkt dann eine Stabilisierung des Wissens und gleicht bei hohen Erfolgsquoten einer gezielten Stoffwiederholung. Hier fungierte der Test aber hauptsächlich als Diagnoseinstrument zur Erfassung des verfügbaren Wissens zur Wahrscheinlichkeitstheorie, dessen möglicher Erwerb lange zurück lag. Lernen während der Testung erweist sich als sehr schwierig, da die Aufgaben eine kurze freie Beantwortung verlangten, im Gegensatz zu MC-Aufgaben eines Fähigkeitstests auch keinerlei Hinweise auf die mögliche Lösung hergaben (siehe dazu etwa: Marsh et al., 2009) und jede Aufgabe ein ganz spezifisches Lehrziel erfasste. Sensibilisierungseffekte, den im Test angesprochenen Problemen weiter nachzugehen, sind nicht ausgeschlossen, aber relativ unplausibel. Obwohl die reine Testung höchstwahrscheinlich keinen Behaltenseffekt nach sich zog, kann man ihr nicht automatisch jedes pädagogische Potenzial absprechen. Dieses könnte sich z.B. dann entfalten, wenn der entsprechende Lehrstoff nun ausführlicher vermittelt werden würde. Denn die Vortestung eines bisher unbekanntes Lehrstoffs unmittelbar vor einer Lernphase kann in bestimmten Fällen das in der Lernphase vermittelte Wissen stärken (z.B. Hamaker, 2006).

2. Testen mit Feedback bewirkte eine höhere Behaltensleistung als reines Testen. Dieser Feedbackeffekt konnte zwar schon oft bestätigt werden, allerdings überwiegend bei einfachem Faktenwissen und seltener bei anspruchsvolleren Aufgaben. Es ist jedoch schwer abzuschätzen, wie stark hierfür das sachorientierte Feedback im Anschluss an die Testung verantwortlich ist, weil ein Teil des Lerngewinns auch auf die gegenüber reiner Testung möglicherweise verbesserte Aufgabenbearbeitung zurück gehen könnte. Das Feedback beschränkte sich hier nicht auf KOR und KCR, sondern beinhaltete elaboriertes Feedback, welches eine sehr knappe Erklärung zur Aufgabenlösung gewährte. Zu Beginn der Übung wurde der Proband durch entsprechende Lerntipps aufgefordert, sich um ein aktives Verständnis zu bemühen. Die alleinige Rückmeldung der korrekten Antwort lieferte zwar im Falle der eigenen Lösung eine gewisse Bestätigung für den Lösungsweg, hätte bei einem Fehler im Normalfall aber nicht ausgereicht, um den der Lösung zugrunde liegenden Ansatz nachvollziehen zu können. Natürlich kann nicht ausgeschlossen werden, die Feedback bedingten Vorteile beim FWT_A2 beruhten überwiegend auf einfachen Erinnerungseffekten der Aufgabenergebnisse, da die meisten Aufgaben auch nur kurze Antworten, wie „ $2/3$ “ oder „ $1/2$ “, erforderten. Um solche Effekte zumindest einzuschränken, wurde die Reihenfolge im Behaltenstest auch durch Hinzuziehung der neuen Aufgaben völlig umgestellt. Zieht man das Behaltensintervall von einer Woche ins Kalkül, so scheint mir ganz oberflächliche Erinnerung unwahrscheinlich und eine gewisse Art des Verstehens notwendig, um selbst bei bereits beantworteten Aufgaben bessere Leistungen zu erbringen.

3. Testen mit Feedback bewirkte höheres Transferwissen als reines Testen. Jacobs (2004) fand vergleichbare Ergebnisse für recht sparsames Feedback beim Erlernen von Kombinatorikproblemen. Die Transfervorteile der Feedbackgruppe gegenüber der reinen Testung legen die Vermutung nahe, die Rückmeldungen hätten gewisse Lernprozesse angestoßen, die das

Verständnis für die Aufgabenlösung teilweise gefördert hätten. Möglicherweise hat sich der Proband auch nur an das Lösungsschema erinnert, das bei der bereits bekannten Aufgabe dann ohne aufwändige Denkarbeit zur Anwendung kam, was aber auch eine gewisse Denkleistung voraussetzt. Natürlich kann hier nur von einem minimalen Transfer die Rede sein. Es lässt sich sogar trefflich darüber streiten, ob überhaupt ein nennenswerter Transfer vorliegt, weil Paralleltests ja letztlich dasselbe messen sollen. Wohl aber darf man mit Recht behaupten, das Lehrzielniveau Anwendung ernsthaft überprüft und mehr als triviales Wissen gefördert zu haben. Dies ist angesichts der relativ geringen Übungszeit umso erstaunlicher.

4. Testen mit Feedback und das Durcharbeiten von Musterlösungen haben sich als vergleichbar gute Instruktionmethoden erwiesen. Musterlösungen beanspruchten aber etwas weniger Lernzeit und sind daher etwas lerneffizienter. Dieser Befund gleicht allen bisherigen Studien von Jacobs im realen Schulsetting, widerspricht aber dem Ergebnis der Laborstudie zum Kombinatoriklernen. Die Studie von Jacobs (2004) ergab eine Überlegenheit des Testens mit Feedback gegenüber der Musterlösung, ist aber nur bedingt vergleichbar, da sie eine Lernanpassungsphase beinhaltete und ausschließlich unmittelbares Behalten erfasste. Da hier aber keinerlei einführender Unterricht stattgefunden hatte, hätte man analog den Forschungsbefunden zu Lösungsbeispielen für die meisten Studierenden sogar bessere Erfolgsquoten für die Musterlösungen erwarten können.

Weitere Erkenntnisse der Studie

Im Idealfall sollte sich Umfang und Ausgestaltung des Feedbacks am Wissens- und Fähigkeitsniveau des Übenden sowie an der Schwierigkeit der Aufgabe orientieren. Für relativ leichte Aufgaben haben die knappen Erklärungen sowie die Musterlösungen ausgereicht, das Lösungsprinzip zu durchschauen. Die Leistungsvorteile der beiden experimentellen Übungsmethoden gegenüber dem reinen Testen gingen im Wesentlichen auf 3 Lehrinhalte zurück: Laplace-Wahrscheinlichkeit, Kombinatorik, und Wahrscheinlichkeit stochastisch unabhängiger Ereignisse. Fasst man nur die insgesamt 6 Items dieser Lehrziele zu einem Test zusammen (Cronbachs Alpha= .79), so liegen die Erfolgsquoten beider Übungsmethoden gegenüber der Testgruppe nicht nur hochsignifikant, sondern mehr als 20% höher und kommen an das Ausmaß heran, welches normalerweise durch ein Statistikseminar erzielt wurde (Nachtigall & Wolf (2001, S.11).

Von den Übungsmethoden Feedback und Musterlösung profitierten höchstwahrscheinlich vorwiegend die Studierenden mit geringem Vorwissen. Da das Ausgangswissen aber nur bei der Feedbackgruppe erhoben wurde, kann diese These auch nur für diese Gruppe überprüft werden. Sie lässt sich aus den signifikant negativen Korrelationen zwischen Ausgangstest und Lernzuwachs für die einzelnen Tests vermuten. So korreliert der FWT_A1 mit der Differenz $\text{FWT_A2} - \text{FWT_A1}$ $r = -.35$ und mit der Differenz $(\text{FWT_B2} - \text{FWT_A1})$ $r = -.55$ (N jeweils 27).

Mit wachsender Komplexität der Aufgaben oder unzureichendem Vorwissen genügen einfache Erklärungen nicht mehr. Im FWT gehörten dazu offenbar die beiden Aufgaben zur Gegenwahrscheinlichkeit, z.B.: „Wie wahrscheinlich ist es, dass Sie bei viermaligem Würfeln mindestens eine eins werfen?“ Beim Lehrziel Gegenwahrscheinlichkeit erzielten die Probanden, auch solche mit hohen Vortestwerten, trotz erklärendem Feedback oder einer Musterlösung weder im Behaltens- noch im Transfertest einen erkennbaren Lerngewinn. Aber auch in Statistikseminaren von Psychologen erwies sich die Vermittlung dieses Themengebietes als wenig erfolgreich.

Schlussfolgerungen

Da hier Musterlösungen und das Feedback dieselben Informationen enthielten und zu vergleichbaren Ergebnissen führten, könnte man folgern, die Testung selbst habe bei der Methode Testen mit Feedback keinen entscheidenden zusätzlichen Lernvorteil nach sich gezogen. Bei einfachen Lernzielen, etwa Vokabellernen, entfaltet der reine Testeffekt seine stabilisierende Wirkung nur bei korrekten Lösungen. Die Erinnerungsbemühung soll im Falle einer erfolgreichen Lösung eine höhere Behaltenseffektivität bewirken als das einfache Einprägen einer präsentierten Lösung. Unklar bleibt, ob diese Erkenntnis so einfach auf anspruchsvollere Aufgaben übertragbar ist. Testen fördert das Wissen eher dadurch, durch erfolgreiches Erinnern das Vergessen hinaus zu zögern. Durch reines Testen lernt man im Normalfall aber nichts Neues hinzu. Wegen der relativ mäßigen Erfolgsquoten im Übungstest konnte sich der erhoffte positive Effekt des Testens kaum bemerkbar machen. Zudem geht es beim stochastischen Wissen eher um grundlegendes Verständnis. Wer dieses Verständnis erworben hatte, wird es auch ohne Testung vermutlich nicht so schnell vergessen wie einfaches Faktenwissen.

Auch unter didaktischer Perspektive würde man einen bisher nicht vermittelten Lehrstoff wohl eher in Form von Musterlösungen anbieten als sofort eine Testung anzusetzen und den Lernprozess durch das Feedback zu initiieren. Im Vergleich zu etlichen Experimenten, die beim Testen im Falle einer Falschlösung erst mehrere Versuche einforderten, war der Proband hier nicht gezwungen, besonders lange unergiebigere Lösungsansätze zu versuchen, bevor er das Feedback anfordern durfte. Er konnte vielmehr selbst entscheiden, ob er die Aufgabenstellung als Testung oder Musterlösung auffassen wollte und im Falle hoher Unwissenheit direkt die Aufgabe als konsequenzenlose Falschlösung bestätigen, um möglichst schnell als Feedback die Musterlösung einzusehen. Je geringer das Wissen, desto mehr direkte instruktionale Vorgabe und Unterstützung erscheint ratsam. Ab einer gewissen Komplexität bzw. Schwierigkeit reichen einmalige Erklärungen jedoch nicht aus, das Verständnis hinreichend aufzubauen. Mit wachsender Expertise erhöht sich die Chance, erfolgreich eigene Lösungen zu entwickeln und dann könnte das Testen mehr Vorteile erbringen als das Durcharbeiten vorgefertigter Lösungen (The expertise reversal effect; Kalyuga et al. 2003). In einem relativ großen Mittelbereich des Wissensniveaus scheinen beide Methoden eine vergleichbare Lernwirkung zu entfalten, wenn man den Lernenden einen hinreichenden Freiheitsspielraum bei der Aufgabenbearbeitung gewährt.

Die Ergebnisse geben einen gewissen Anlass zu der Hoffnung, es sei nicht ganz unnützlich, verfügbare pädagogische Testverfahren für Übungszwecke mit adäquaten Rückmeldungen zu versehen, um sie von interessierten Lernenden bearbeiten zu lassen. Zwar leisten Musterlösungen den gleichen Lerneffekt in kürzerer Zeit und sind deshalb auch als echte Alternative zu einer aufwändigen Testung zu betrachten, aber unbenotete Tests motivieren eher und könnten darüber hinaus ein objektives Bild der eigenen Leistungsfähigkeit im betreffenden Sachgebiet gewähren. Für didaktische Zwecke wäre es sehr hilfreich, auf einen großen Pool von Aufgaben mit entsprechenden statistischen Aufgabenkennwerten zugreifen zu können, um eine gezielte Aufgabenauswahl vornehmen zu können. Da das ausschließliche Testen keinen Lernzuwachs bewirkte, stellt sich die Frage nach den best möglichen Hilfen vor, während und nach der Aufgabenbearbeitung. Zwar hat sich die Mitteilung der korrekten Antwort relativ zuverlässig als lernwirksam erwiesen. Eine besondere Herausforderung für die Feedbackforschung lautet aber, ob bzw. welche darüber hinaus gehende Anregungen sach- oder lernstrategischer Art Lernprozesse noch nachhaltiger fördern könnten.

Literatur

- Butler, A. C. (2010). Repeated testing produces superior transfer of learning relative to repeated studying. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36, 1118-1133.
- 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.
- Chan, J. C. K., McDermott, K. B., & Roediger, H. L., III (2006). Retrieval induced facilitation: Initially nontested material can benefit from prior testing of related material. *Journal of Experimental Psychology: General*, 135, 553-571.
- Clifton, K. S (2005) The Testing effect: using retrieval practice in the classroom. Thesis submitted to Marshall University In partial fulfillment of the Requirements for the degree of Master of Arts Psychology
<http://www.marshall.edu/etd/masters/clifton-karen-2005-ma.pdf> [19.10.2005]
- Cull, W. L. (2000). Untangling the Benefits of Multiple Study Opportunities and Repeated Testing for Cued Recall. *Appl. Cognit. Psychol.* 14: 215-235
- Hamaker, Ch. (1986). The Effects of Adjunct Questions on Prose Learning. *Review of Educational Research*, 56 (2) 212-242.
- Hausknecht, J. P., Halpert, J. A., Di Paolo, N. T., & Moriarty Gerrard, M. O. (2007). Retesting in selection: A meta-analysis of coaching and practice effects for tests of cognitive ability. *Journal of Applied Psychology*, 92, 373-385.
- 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.
 URN: urn:nbn:de:bsz:291-psydok-5992
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2006/599/>
- Jacobs, B. (2007). Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback.
 URN: urn:nbn:de:bsz:291-psydok-15597
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2008/1559/>
- Jacobs, B. (2008). Übungsaufgaben aus zentralen Vergleichsarbeiten
<http://bildungswissenschaften.uni-saarland.de/personal/jacobs/paedpsych/aufgaben/vergleichsarbeiten/index.html>

- Jacobs, B. (2010). Testfragen selbst beantworten oder Musterlösungen studieren?
 URN: urn:nbn:de:bsz:291-psydok-26934
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2693/>
- Jacobs, B. (2011). Musterlösungen durcharbeiten als Alternative zu Testen mit Feedback - Eine Replikationsstudie.
 URN: urn:nbn:de:bsz:291-psydok-27127
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2011/2712/>
- Jacobs, B. (2012). Auf der vergeblichen Suche nach dem Testeffekt - Studieren oder Testen mit Feedback beim Vokabellernen.
 URN: urn:nbn:de:bsz:291-psydok-33106
 URL: <http://psydok.sulb.uni-saarland.de/volltexte/2012/3310/>
- Kalyuga, S., Ayres, P., Chandler, P. & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist* 38, 23–31.
- 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.
- Larsen D.P., Butler A.C., Roediger H.L. III. (2008) Test-enhanced learning in medical education. *Medical Education*. 42, 959–66.
- Marsh, E. J., Agarwal, P. K. & Roediger, H. L. (2009). Memorial Consequences of Answering SAT II Questions. *Journal of Experimental Psychology: Applied*, 15 (1) 1–11.
- Nachtigall, C. & Wolf, A. (2001). Fragebogen zur Wahrscheinlichkeitstheorie (FWT)
 URL: www.metheval.uni-jena.de/materialien/reports/report_2001_03.pdf [20.12.2011]
- 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.

Anhang: Bildschirmausschnitte der Übungsmethoden

Aufgabe 1

Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf eine gerade Augenzahl geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl erbracht hat, die größer als drei ist?

p =

Aufgabe bestätigen

Proband gibt die falsche Antwort ein und bestätigt die Aufgabe.

Testen

Aufgabe 1

Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf eine gerade Augenzahl geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl erbracht hat, die größer als drei ist?

p =

*** erledigt ***

Aufgabe 1

Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf eine gerade Augenzahl geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl erbracht hat, die größer als drei ist?

p =

Aufgabe bestätigen

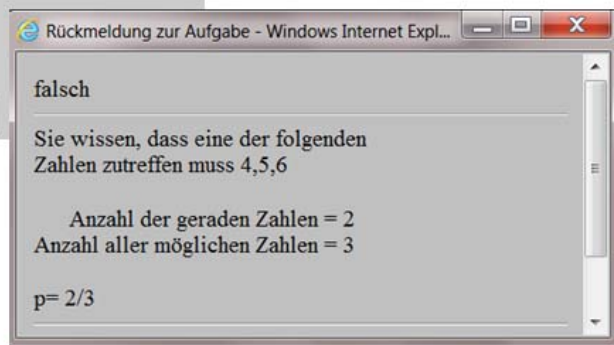
Korrekte Lösung?

Proband gibt die falsche Antwort ein und bestätigt die Aufgabe.

Dann erscheint das Rückmeldefenster.

Anklicken auf Korrekte Lösung? schreibt 0.66 in das Antwortfeld.

Testen mit Feedback

**Musterlösung 1**

Wie wahrscheinlich ist es, dass Sie bei einem Würfelwurf eine gerade Augenzahl geworfen haben, wenn Sie bereits wissen, dass Ihr Wurf eine Augenzahl erbracht hat, die größer als drei ist?

p = $2/3$

Sie wissen, dass eine der folgenden Zahlen zutreffen muss 4,5,6

Anzahl der geraden Zahlen = 2
Anzahl aller möglichen Zahlen = 3

p= $2/3$

Musterlösung bestätigen

Musterlösung

Proband inspiziert die Musterlösung und klickt danach auf Musterlösung bestätigen.

Dann wandelt sich der Button in den Button ok