

B. Jacobs, Bildungswissenschaften der Universität des Saarlandes. [b.jacobs@mx.uni-saarland.de](mailto:b.jacobs@mx.uni-saarland.de)  
Version vom 1.3.2016

## Verpflichtende Probeklausur als Maßnahme zur Verbesserung der Klausurleistung und Reduktion aktueller Prüfungsangst

### Abstract

Im Rahmen universitärer Lehre wurde geprüft, ob die verpflichtende Teilnahme an einer Probeklausur die Klausurleistung verbessert und die Angst vor der Klausur senkt. Studierende aus 4 Seminaren wurden nach Zufall auf zwei Bedingungen aufgeteilt. Während ein Teil der Studierenden eine Woche vor der Klausur die Probeklausur bearbeitete, sollte die Kontrollgruppe zeitgleich die Seminarsitzung nutzen, um sich eigenverantwortlich selbst auf die Prüfung vorzubereiten. Studierende, die an der Probeklausur teilnehmen mussten, erzielten erwartungskonform bessere Klausurleistungen und traten teilweise mit geringerer aktueller Angst zur Klausur an. Die Effekte fielen nicht vollkommen konsistent, in ihrer Gesamtheit aber relativ eindeutig aus. Etliche erhobene subjektive Daten lassen sowohl die Leistungssteigerung wie auch die geringere Angst plausibel erscheinen. Durch die Probeklausur erhöhte sich die subjektive Transparenz der Prüfungsanforderungen, das erwartete Prüfungsergebnis, die Einschätzung, man könne sich besser auf die Klausur vorbereiten und das Vertrauen in die eigene Leistungsfähigkeit.

**Schlagworte:** Probeklausur, Testen, Feedback, Prüfungsangst, Hochschuldidaktik, mock exam, testing, feedback, test anxiety

### Einleitung, Zielsetzung und Hypothesen

Nachfolgende Studie versteht sich als Replikation einer früheren Untersuchung von Jacobs, Bernd und Fey (2004, siehe auch die Langfassung von Jacobs 2003) zur Wirkung einer Probeklausur auf Klausurleistung und aktuelle Angst vor der echten Klausur. Dort wurde in das Thema eingeführt, ein kurzer Literaturüberblick gegeben, sowie die wichtigsten Erwartungen an die Wirkungen einer Probeklausur gestellt, so dass sich die nachfolgenden Ausführungen dazu nur auf die wesentlichen Punkte beschränken. Leider blieb keine Zeit mehr, neuere Entwicklungen auf diesem Gebiet zu verfolgen oder etliche Behauptungen mit Hinweis auf Forschungsarbeiten sonstiger Autoren weiter zu belegen.

Nach Jacobs (2000) sollte die Probeklausur

- eine realistische Einschätzung des eigenen Wissens ermöglichen (diagnostischer Aspekt)
- nach Möglichkeit Hilfen anbieten, bestehende Defizite auszugleichen (lernfördernder Aspekt).
- gegebenenfalls zusätzlichen Arbeitseinsatz bis zur echten Klausur anregen (motivationaler Aspekt)

Neben seiner diagnostischen Funktion, die Leistung möglichst gut zu erfassen, fördert das Testen selbst eine Stabilisierung des Wissens. Das Bemühen, die Antwort aus dem Langzeitgedächtnis hervorzurufen und zu verarbeiten, festigt im Falle einer korrekten Antwort das Behalten (retrieval-effect). Folgt der Testung eine sachorientierte Rückmeldung, etwa über die korrekte Antwort und ihr Zustandekommen, gewährt sie dem Prüfling die Chance, seinen Fehler zu korrigieren und sich die korrekte Antwort verständlich zu machen. Innerhalb der letzten Jahre hat sich die Test- und Feedbackforschung vorwiegend im anglo-amerikanischen Raum rasant entwickelt. Hierbei konnte die Lernwirksamkeit des Testens mit Feedback meistens in Form von Laborforschung, mittlerweile aber auch in der Feldforschung sowie im realen Schulbetrieb nachgewiesen werden. Allerdings ergeben sich bei der Übertragung der Forschungsergebnisse in die universitäre Praxis etliche Komplikationen, ob und wie Tests als Maßnahmen

zur Förderung von Studierleistungen wirksam umgesetzt werden könnten (siehe dazu Jacobs 2010a).

Da die Bearbeitung einer Probeklausur auch einer Testung entspricht, sich Testen und sachorientiertes Feedback als wirksame Lernmethoden erwiesen haben, müssten Probeklausuren die gleichen Lerneffekte nach sich ziehen. Es ist aber nicht so einfach, diesen Effekt in einer Klausur nachzuweisen, weil alle Studierenden an guten Klausurleistungen interessiert sind und ohne Probeklausur dann eben andere, möglicherweise aufwändigere, aber ebenfalls wirksame Vorbereitungsmaßnahmen, ergreifen könnten. Wenn Studierende ohne Probeklausur den Effekt der Probeklausur aber durch selbst initiierte Vorbereitung kompensieren, dann findet man in der Klausur keinen Effekt der Probeklausur mehr.

Bei Jacobs et al. (2004) profitierten vor allem die schwächeren Studierenden von der Probeklausur. Auf Mittelwertebene ergab sich dort ein schwacher, üblichem Signifikanzniveau folgend, jedoch nicht signifikanter Klausurvorteil für die Probeklausur. Basierend auf dieser Erfahrung wird hier bestensfalls ein leichter Vorteil der Probeklausur im niedrigen Effektstärkebereich erwartet. Angesichts beschränkter Power wird das Signifikanzniveau bzgl. des erwarteten Unterschieds in der Klausurleistung einseitig auf 10% festgelegt. Ein der Konvention widerstrebendes höheres Alpha wird in Kauf genommen, um der Alternativhypothese überhaupt eine Realisierungschance zu geben. Es wird auch keine Alphaadjustierung durchgeführt. Aber, da alle Hypothesen an zwei unabhängigen Seminaren geprüft werden, kann die Zuverlässigkeit der Befunde einigermaßen gut eingeschätzt werden.

Nach Jacobs (1981) wird aktuelle Angst durch die Bedrohung hervorgerufen. Diese hängt ab von der Stärke der Aversivität negativer Konsequenzen und der Wahrscheinlichkeit ihres Eintretens. Eine Probeklausur kann natürlich nicht die tatsächlichen bzw. die erwarteten Konsequenzen des Durchfallens ändern, sondern lediglich die Wahrscheinlichkeit des Eintretens dieses unliebsamen Ereignisses reduzieren.

**Abbildung 1:** Theoretisch angenommener Weg von der Probeklausur zur Angstreduktion nach Jacobs, Bernd und Fey (2004)



Abbildung 1 veranschaulicht, auf welche Weise durch Bearbeitung der Probeklausur die Erwartung des Misserfolgs und infolge dessen die aktuelle Angst sinken sollte. Die Probeklausur schafft Transparenz, welche eine bessere Vorbereitung ermöglicht, dadurch ein günstigeres Prüfungsergebnis erwarten lässt, wodurch die Wahrscheinlichkeit eines möglichen Misserfolgs sinkt. Nur wenn dieser Weg an keiner Stelle unterbrochen wird, ist mit geringerer aktueller Angst unmittelbar vor der Klausur zu rechnen. Bei Jacobs et al. (2004) war dieser Weg nach der Wahrnehmung höherer subjektiver Transparenz zu Ende. Eine geringere Angst infolge der Probeklausur kann hier also nur dann angenommen werden, wenn das Modell in Abbildung 1 insgesamt bestätigt wird.

## Die empirische Untersuchung

### Probanden und Konzeption der Seminare

Als Probanden dienten Lehramtsstudierende (Alter 24, 70% Frauen) aus 4 Seminaren des Verfassers im WS\_15/16 und zwar 2 Seminare zur pädagogischen Diagnostik und Intervention - im Folgenden **Diagnostik** genannt - und 2 Seminare zum Themenkomplex "Grundlagen der Evaluation und pädagogischer Forschungsmethodik", im Folgenden **Methoden**, genannt. Die beiden Diagnostik - und die beiden Methodenseminare können jeweils als Parallelseminare aufgefasst werden.

Alle Seminare wurden überwiegend vom Seminarleiter gestaltet. Er trug den Lehrstoff unter Verwendung von Powerpointfolien dozierend sowie Fragen entwickelnd vor. In den Vortrag waren gelegentlich auch einige Aufgaben oder kurze Übungen integriert, die dann an alle Studierenden gerichtet, von Interessierten beantwortet und schließlich diskutiert wurden. Darüber hinaus bearbeiteten die Studierenden etliche Arbeitsaufträge und ergänzten das Seminar teilweise durch Kurzreferate oder Projekte.

Im Seminar Diagnostik empfahl der Seminarleiter, die umfangreiche Literaturliste auf der Homepage des Seminars nicht weiter zu beachten, sondern dem Seminar aufmerksam zu folgen und als wesentliche Grundlage zur intensiven Nacharbeitung nur die Folien zu verwenden, die nach jeder Seminarsitzung allen zur Verfügung standen. Während des Seminars bot der Seminarleiter den Studierenden 2 freiwillige anonyme Computerquiz mit Rückmeldungen zu den Themen zweier Seminarsitzungen an. Im Seminar Methoden diente ein Internettutorium als alleinige Lerngrundlage für die Nacharbeitung des Seminars. In diesem als Selbstlernprogramm konzipierten Onlinetutorium war der Lehrstoff textlich sehr prägnant formuliert. Zu jedem Kapitel konnte der Studierende elektronische Übungsaufgaben mit elaborierten Rückmeldungen bearbeiten. Auf die im Seminar vorgeführten Folien, die sich thematisch stark am Tutorium orientierten und pädagogische Themen akzentuierten, hatten die Studierenden keinen Zugriff.

### Bedeutung und Bewertung der Abschlussklausur

In allen Seminaren galt als Leistungsnachweis für das Prüfungsamt neben den Arbeitsaufträgen die erfolgreiche Teilnahme an einer Abschlussklausur, dort stets Abschlussquiz genannt. Da laut Prüfungsordnung für alle Seminare ein unbenoteter Leistungsnachweis vorgeschrieben ist, war lediglich das Bestehen der Abschlussklausur notwendig. Der Seminarleiter erklärte den Studierenden, die Bewertung der Klausur basiere auf einer kriteriumsorientierten Konzeption. Unabhängig von den Leistungen aller Studierenden sei allein wichtig, eine bestimmte Leistungsschwelle zu erreichen. Wo diese Grenze liege, würde er aber nicht mitteilen, weil er befürchte, manche Studierende würden so ermuntert, diese Grenze auszutesten. Er erwarte eine akzeptable Leistung und lasse unzureichende Leistungen nicht durchgehen. Alle im Verlauf des Seminars gestellte Anfragen der Studierenden darüber, wie viel Prozent sie richtig machen müssten, beantwortete der Seminarleiter mit der Floskel, er könne darüber nichts sagen. Die Bewertungskonzeption entspricht einer Pass/Fail-Variante mit nicht explizierter Grenze.

### Probeklausur und Abschlussklausur (Klausur)

Im Seminar und bei einzelnen Erhebungen war stets von Probequiz und Abschlussquiz die Rede. Diese Bezeichnungen ändern nichts an der Tatsache, dass man die Tests als Klausuren betrachten kann, weswegen diese hier auch durch den Begriff Klausur ersetzt werden. Beide

Klausuren umfassen jeweils 47 Aufgaben und verfügen über eine hinreichende Kontentvalidität, weil sie stets eine annähernd repräsentative Stichprobe aus dem Universum aller Lehrziele des entsprechenden Seminars erfassen. Hierbei galt als vorrangiges Lehrzielniveau Verstehen und Anwendung. Als Aufgabentypen der Computerklausuren kamen die gängigen objektiven Aufgabenformate Multiple Choice mit einer oder mehreren korrekten Antworten, Serien von True/False-, diverse Zuordnungs- sowie Short Answer-Aufgaben zum Einsatz. Der 100% objektiven Auswertung lag eine Zufallskorrektur zu Grunde, die aber nur bei Aufgaben mit hoher Ratewahrscheinlichkeit (z.B. True/False Aufgaben, MC mit mehreren korrekten Antworten) Verwendung fand.

Zu jeder Klausuraufgabe wurde eine parallele Probeklausuraufgabe konstruiert. Da die Aufgaben in Probeklausur und Klausur dieselben Lehrziele mit unterschiedlichen Aufgaben vergleichbar gut erfassen sollten, sind die parallelen Aufgaben nicht identisch, sondern ähnlich. Hierbei kamen unterschiedliche Maßnahmen zum Einsatz. (Paraphrasierung, Transformation, unterschiedliche Beispiele usw.) Je nach speziellem Lehrziel gelang die Parallelisierung mehr oder weniger gut. Sofern die Fragestellung in Probe- und Abschlussklausur gleich war oder eine sehr hohe Ähnlichkeit aufwies, unterschieden sich korrekte Antworten und die Distraktoren. Bei Anwendungsaufgaben sind Probe- und Klausuraufgaben in der Regel strukturgleich. Wenn der Studierende die Probeklausuraufgabe lösen konnte oder durch das Feedback im Nachhinein verstand, so war meistens noch eine niedrige Transferleistung notwendig, um die Klausuraufgabe erfolgreich zu beantworten. Im Anhang A findet der Leser einige Probeklausuraufgaben mit den analogen Abschlussklausuraufgaben.

Alle Aufgaben in Probeklausur und Klausur waren auf einer HTML-Seite untergebracht. In der Probeklausur folgte nach der Beantwortung bzw. Bestätigung einer Aufgabe unmittelbares Feedback (immediate item specific feedback) in nachfolgender Form

1. Knowledge of response : richtig oder falsch für jede Antwort.
2. Knowledge of correct response: die korrekte(n) Antwort(en).
3. elaboriertes Feedback: erklärendes Feedback unterschiedlicher Art.

Rückmeldungen 1 und 2 folgten jeder Aufgabe obligatorisch. Durch Anklicken auf den Button "korrekte Lösung?" konnte der Studierende nach eigener Aufgabenbestätigung die korrekte Beantwortung der Aufgabe erzwingen. Elaboriertes Feedback wurde sparsam und meist nur dann gewährt, wenn die Vermutung vorlag, der Studierende könne sich die korrekte Aufgabenlösung allein nicht verständlich machen. Ansonsten bestand die weitere Rückmeldung in der Aufforderung, die Aufgabenlösung anzufordern und sich diese selbst verständlich zu machen. Die Abschlussklausuren entsprachen einer reinen Testung ohne Feedback 1 bis 3.

Erst nachdem in Probeklausur sowie Abschlussklausur alle Aufgaben beantwortet worden waren, zeigte der Button "Klausur abgeben und auswerten lassen" eine entsprechende Wirkung. Jedwede Antwortverweigerung wurde demnach nicht geduldet. In beiden Klausuren folgte abschließend als Rückmeldung

4. Knowledge of performance: hier: Prozentsatz der korrekten Lösungen.

Nicht vorhersehbare Eingriffe des Rechenzentrums am Server verursachten Kompatibilitätsprobleme mit dem zuvor funktionierenden Probeklausurprogramm. Infolgedessen fiel bei zehn Studierenden des Seminars Methoden Rückmeldung 4 aus. Zudem wurden deren Probeklausurergebnisse nicht gespeichert. Diese 10 Studierenden bearbeiteten die Probeklausur immerhin

ordnungsgemäß zu Ende, erhielten auch die wichtigen Rückmeldungen 1 bis 3, sodass bei diesen bis auf "Knowledge of performance" alle Bedingungen einer adäquaten Probeklausur erfüllt waren. Tabelle 1 listet als Reliabilitätskennwerte den Alphakoeffizienten von Cronbach sowie die Retestreliabilität auf. Der Zeitabstand von Probe- zur Abschlussklausur beträgt genau eine Woche. Alle Werte bis auf .12 sind hochsignifikant.

**Tabelle 1: Reliabilitäten von Probeklausur und Klausur**

|            | Probe-<br>klausur<br>$\alpha$ (N) | Abschluss-<br>klausur<br>$\alpha$ (N) | Retest<br>(1 Woche)<br>$r_{tt}$ (N) |
|------------|-----------------------------------|---------------------------------------|-------------------------------------|
| Diagnostik | .89 (27)                          | .86 (56)                              | .58 (27)                            |
| Methoden   | .89 (19)                          | .72 (61)                              | .12 (18)                            |

## Die Erhebung subjektiver Daten

Unmittelbar vor der Probeklausur und unmittelbar vor der echten Klausur beantworteten die Studierenden noch einige Fragen, welche eine Überprüfung der vermuteten Wirkungsweise einer Probeklausur auf die Reduktion von aktueller Angst sowie einige Hinweise für die erhoffte Leistungssteigerung ermöglichen sollte. In diesem Zusammenhang wurde die subjektive Transparenz der Klausuranforderungen, das Vertrauen in die eigene Leistungsfähigkeit, Einschätzung der erwarteten Leistung und aktuelle Angst erhoben. Etliche Items stammen aus früheren Studien, z.B. Jacobs et al. (2004). Darüber hinaus kamen noch einige Fragen zum Einsatz, die konkrete und detaillierte Aussagen zur vermuteten Wirkung der Probeklausur erwarten ließen. Abbildung 2 verdeutlicht, wie diese Items erhoben wurden. Die Codierung basiert auf einer 7 stufigen Likertskala (stimmt ganz genau=7 .... stimmt überhaupt nicht =1). Als Testwert für die Kurzfragebogen gilt der Mittelwert der in richtige Richtung gepolten Items.

**Abbildung 2: Fragebogen und einzelne Items**

|    | <b>Transparenz der Prüfungsanforderungen</b>                                                         | stimmt<br>ganz<br>genau | stimmt<br>über-<br>wie-<br>gend | stimmt<br>eher        | weder/<br>noch        | stimmt<br>eher<br>nicht | stimmt<br>über-<br>wie-<br>gend<br>nicht | stimmt<br>über-<br>haupt<br>nicht |
|----|------------------------------------------------------------------------------------------------------|-------------------------|---------------------------------|-----------------------|-----------------------|-------------------------|------------------------------------------|-----------------------------------|
| 1. | Ich weiß genau, welche Anforderungen im Abschlussquiz gestellt werden.                               | <input type="radio"/>   | <input type="radio"/>           | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>   | <input type="radio"/>                    | <input type="radio"/>             |
| 2. | Mir ist hinreichend klar, welche Themen im Abschlussquiz geprüft werden.                             | <input type="radio"/>   | <input type="radio"/>           | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>   | <input type="radio"/>                    | <input type="radio"/>             |
| 3  | Ich kann sehr gut abschätzen, wie die Aufgaben im Abschlussquiz aussehen werden.                     | <input type="radio"/>   | <input type="radio"/>           | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>   | <input type="radio"/>                    | <input type="radio"/>             |
|    |                                                                                                      |                         |                                 |                       |                       |                         |                                          |                                   |
|    | <b>Vertrauen in die eigene Leistungsfähigkeit</b>                                                    |                         |                                 |                       |                       |                         |                                          |                                   |
| 4. | Ich werde im Abschlussquiz vermutlich ziemlich schlecht abschneiden. (-)                             | <input type="radio"/>   | <input type="radio"/>           | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>   | <input type="radio"/>                    | <input type="radio"/>             |
| 5. | Auch wenn im Abschlussquiz Schwierigkeiten auftreten sollten, werde ich ein gutes Ergebnis erzielen. | <input type="radio"/>   | <input type="radio"/>           | <input type="radio"/> | <input type="radio"/> | <input type="radio"/>   | <input type="radio"/>                    | <input type="radio"/>             |
|    |                                                                                                      |                         |                                 |                       |                       |                         |                                          |                                   |
|    | <b>Einzelne Items</b>                                                                                |                         |                                 |                       |                       |                         |                                          |                                   |

|    |                                                                                               |                       |                       |                       |                       |                       |                       |                       |
|----|-----------------------------------------------------------------------------------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|-----------------------|
| 6. | Ich kann meine Leistung im Abschlussquiz ziemlich genau einschätzen.                          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
|    | <i>nur in der Abschlussklausur</i>                                                            |                       |                       |                       |                       |                       |                       |                       |
| 7. | Ich zweifelte daran, ob das, was ich lernte, mir für das Abschlussquiz überhaupt etwas nützt. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 8. | Ich konnte mich sehr gut auf das Abschlussquiz vorbereiten.                                   | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

### Subjektive Kompetenz (erwartete Leistung)

|    |                                                                                                                                       |                                |
|----|---------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|
| A. | Schätzen Sie den Prozentsatz der korrekten Lösungen ein, den Sie im Abschlussquiz erzielen werden? (Zahl zwischen 0 und 100 eingeben) | <input type="text"/> % richtig |
|----|---------------------------------------------------------------------------------------------------------------------------------------|--------------------------------|

### Wahrscheinlichkeit, die angestrebte Leistung zu erzielen

|    |                                                                                                                                                                                                                                                     |                        |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|
| B. | Wie hoch schätzen Sie die Wahrscheinlichkeit [in Prozent] ein, im Abschlussquiz die Leistung zu erzielen, die sie wirklich zufrieden stellen würde? (Zahl zwischen 0 und 100 eingeben! : 0 = sicherlich nicht ..... 100 = mit absoluter Sicherheit) | <input type="text"/> % |
|----|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------|

Die aktuelle Angst vor der Probeklausur und der Klausur wurde mit dem aktuellen Angstfragenbogen (SPA) sowie dem Fearthermometer nach Jacobs (1981) in der [neuesten Fassung von Jacobs](#) gemessen. Angaben zu den Reliabilitäten der verwendeten Instrumente findet der Leser im Zusammenhang mit der Darstellung der Ergebnisse.

### Versuchsvorgehen und Versuchsplan

Ca. 3 Wochen vor der anstehenden Abschlussklausur konfrontierte der Seminarleiter die Studierenden mit der Idee, die Wirkung einer Probeklausur auf das Klausurergebnis überprüfen zu wollen. Eine vernünftige Überprüfung dieser Fragestellung sei aber nur möglich, wenn die eine Hälfte der Studierenden an der Probeklausur (=EG) teilnähme, die andere aber nicht (=KG) und, dass nicht der Studierende selbst, sondern der Zufall bestimme, wer welche Bedingung erhielte. Außerdem mache das Ganze aus statistischen Gründen überhaupt nur dann Sinn, wenn alle Studierenden aus allen 4 Seminaren damit einverstanden wären. Um mögliche Benachteiligungen der KG zu entschärfen, versprach der Seminarleiter der Kontrollgruppe, im Falle besserer Klausurergebnisse der EG, die Klausurmittelwertsdifferenz EG-KG der KG als Bonus zu gewähren, im Falle eines schlechteren Ergebnisses der EG aber keinen Malus für die KG zu vergeben. Nachdem keinerlei Einwand gegen diesen Vorschlag vorgetragen wurde, erhielten die Studierenden exakt eine Woche vor dem Termin der Probeklausur via Email die Mitteilung, welcher Bedingung sie zugeordnet worden waren.

Aus einer Erhebung während des Seminars war der Abiturnotendurchschnitt von ca. 90 Prozent der TeilnehmerInnen bekannt. Um eine vergleichbare schulische Leistungsfähigkeit in EG und KG zu garantieren, wurden die Studierenden innerhalb der Seminare nach Abiturnotendurchschnitt parallelisiert und dann nach Zufall auf EG und KG zugewiesen. Bei den wenigen Studierenden ohne Daten zum Abiturnotendurchschnitt kam das übliche Randomisierungsverfahren zur Anwendung. Die Parallelisierung erschien insbesondere im Fach Diagnostik angebracht, da sowohl in einer früheren Diagnostikvorlesungsklausur, wie auch in der Diagnostikseminarklausur des Sommersemesters 2015, die mit der Klausur in dieser Studie identisch ist, hochsignifikante Zusammenhänge zwischen Abiturnotendurchschnitt und Klausurergebnis ermittelt wurden ( $r = .38$  ;  $N=247$  bzw.  $r = .49$  ;  $N=48$ ).

Die vorletzte Seminarsitzung war für die Klausurvorbereitung vorgesehen. Während die TeilnehmerInnen der EG die Probeklausur im Computerraum unter kontrollierten Bedingungen bearbeiteten, wurden die Studierenden der KG zuvor via Email über einige Details der Klausur informiert. Ihnen wurde zudem der Vorschlag unterbreitet, sie sollten die verfügbare Seminarzeit der vorletzten Seminarsitzung für eine selbstkontrollierte Klausurvorbereitung nutzen (siehe genaue Anweisung im Anhang B). Man kann die Kontrollgruppe somit nicht als ganz triviale "No-Treatment-Kontrollgruppe" ansehen. Es bleibt allerdings gänzlich offen, wie ernsthaft die Kontrollgruppe die Anregungen umsetzte. Der Versuchsplan zur Erfassung der Klausurleistung (O) entspricht dem klassischen experimentellen Design ohne Vortest.

R   X<sub>EG: Probeklausur</sub>   O

R   X<sub>KG: keine Probeklausur</sub>   O

Hierbei sind bis auf die sozial bedingten Störfaktoren alle Störvariablen der internen Validität kontrolliert. Da vor der Probeklausur aber bei der EG auch subjektive Daten erhoben wurden, bei der KG aber nicht, ist beim Vergleich der subjektiven Daten zwischen EG und KG der Vortesteffekt streng genommen ebenfalls nicht zwingend kontrolliert.

### Ergebnisse zur Wirkung der Probeklausur auf das Klausurergebnis

Die Klausurergebnisse von Studierenden mit und ohne vorherige Probeklausur wurden mittels t-Test für unabhängige Stichproben bei einseitiger Testung auf dem 10% Niveau geprüft. Wie aus Tabelle 2 hervorgeht, erzielten die Studierenden unter Probeklausurbedingungen signifikant bessere Abschlussklausurergebnisse.

**Tabelle 2 :** Prozentsatz korrekter Lösungen in der Abschlussklausur für EG und KG

|                   | M    | s    | N  | t-Wert | p <sub>e</sub> | d   |
|-------------------|------|------|----|--------|----------------|-----|
| <b>Diagnostik</b> |      |      |    |        |                |     |
| mit Probeklausur  | 77.4 | 11.7 | 26 | 1.56   | .066           | .42 |
| ohne Probeklausur | 72.6 | 11.0 | 29 |        |                |     |
| <b>Methoden</b>   |      |      |    |        |                |     |
| mit Probeklausur  | 77.7 | 7.0  | 29 | 2.22   | .02            | .55 |
| ohne Probeklausur | 73.1 | 9.4  | 30 |        |                |     |

Das mit Cohens d erfasste Ausmaß des Effektes liegt im erwarteten niedrigen bis höchstens mittleren Effektstärkebereich. Eine verpflichtete Probeklausur bewirkte somit einen etwas höheren Klausurerfolg als die Empfehlung, die verfügbare Seminarzeit der letzten Seminarsitzung vor der Klausur für eine selbstkontrollierte Prüfungsvorbereitung zu nutzen.

### Leistungsveränderungen von Probeklausur zur Klausur

Viele Studierende konzentrieren ihre Prüfungsvorbereitung auf die letzten Tage vor der Klausur. Vorliegende Studie erlaubt die Einschätzung der Leistungsverbesserung im Verlauf der letzten Woche vor der Klausur sowohl für die Studierenden der Probeklausur, wie für die Studierenden der Kontrollgruppe, da beide Gruppen ja nach Zufall zustande kamen. Tabelle 3 erlaubt zunächst die Betrachtung des Leistungszuwachses von der Probe - zur Abschlussklausur für Studierende mit Probeklausur.

**Tabelle 3:** Prozentsatz korrekter Lösungen in Probeklausur und Abschlussklausur

|            |        | Probe-<br>klausur |      | Abschluss-<br>klausur |      | d    | $r_{tt}$ |
|------------|--------|-------------------|------|-----------------------|------|------|----------|
|            |        | M                 | s    | M                     | s    |      |          |
| Diagnostik | (N=26) | 61.3              | 14.7 | 77.4                  | 11.7 | 1.21 | .58      |
| Methoden   | (N=19) | 57.9              | 16.5 | 77.9                  | 5.9  | 1.61 | .12      |

Der Leistungszuwachs fällt mit Effektstärken über 1 erheblich aus. Diese vornehmlich im Fach Methoden erkennbare gewaltige Leistungssteigerung mag mit ein Grund für die fehlende Retest reliabilität sein. Eine derart schwache, nicht existierende, Retest reliabilität war bereits bei Jacobs (2003) im Seminar „Lehren und Lernen, im Folgenden Didaktik genannt, aufgefallen. Etliche Studierende waren dort ziemlich unvorbereitet zur Probeklausur angetreten und hatten dann in der letzten Woche den Prüfungsstoff massiv eingepaukt. Das Ergebnis im Probequiz Methoden korreliert mit dem Leistungszuwachs zur Methodenklausur  $r = -.93$ . Der analoge Wert für Diagnostik lautet  $r = -.64$ . Den größten Leistungsgewinn erzielten naturgemäß diejenigen Studierenden, deren Kenntnisse zu Beginn ziemlich gering ausgeprägt waren, weil deren bisheriger Arbeitseinsatz schwach war. Dann fällt aber auch das Zuwachspotenzial deutlich höher aus. Auch die Studierenden der Kontrollgruppen hatten in der letzten Woche vor der Klausur deutliche Leistungsfortschritte erzielt. Die Effektstärken der KG lassen sich für Diagnostik auf einen Wert um  $d = .87$  und für Methoden um ca.  $d = 1.13$  einschätzen. Bereits Jacobs (2003) fand im Seminar Statistik deutliche und im Seminar Didaktik gigantische Leistungszuwächse im Verlauf der letzten Woche bis zur Klausur.

Die Strategie der Studierenden, hauptsächlich kurz vor der Prüfung intensiv zu lernen (cramming), erscheint nicht unvernünftig. Zwar bewirkt verteiltes Lernen einen nachhaltigeren Lerngewinn, honoriert wird aber die Leistung zu einem bestimmten Zeitpunkt. Massives Lernen unmittelbar vor der Prüfung liefert eher bessere Erfolgchancen für die Prüfung als regelmäßiges Lernen, auch wenn dieses Wissen dann schneller dem Vergessen anheimfällt.

Die hier betrachteten Klausuren erfassen keine stabilen Fähigkeiten wie Intelligenz oder Pisakompetenz. Der überwiegende Teil des Lehrstoffs kann erlernt, verstanden und zumindest kurzfristig auch behalten werden. Über die dazu notwendige Lernfähigkeit verfügen vornehmlich Studierende mit guten Abiturnoten. Wie oben bereits erwähnt, konnten bisher etliche signifikante Korrelationen zwischen dem Abiturnotendurchschnitt und den Ergebnissen von Diagnostikklausuren im Bereich zwischen  $r = -.30$  bis  $r = -.49$  festgestellt werden, während analoge Ergebnisse im Fach Methoden bisher fehlten. In dieser Studie korrelierte der Abiturnotendurchschnitt in Diagnostik mit der Probeklausur  $r = -.58$  (N=25) und mit der Klausur  $r = -.35$  (N=50). Für das Seminar Methoden lauten die analogen Ergebnisse  $r = .02$  (N=18) und  $r = -.39$  (N=55). Unter der EG-Bedingung des Seminars Methoden liefert die Abiturnote hierbei keine Prognose für das Abschlussklausurergebnis ( $r = -.05$ ), in der KG hingegen eine recht gute Vorhersage ( $r = -.61$ ). Der z-Wert des Korrelationsdifferenzentest erwies sich als signifikant ( $p_z = 0.02$ ). Im Seminar Methoden profitierten im Übrigen nur Studierende mit schwächerer Abiturnote von der Probeklausur. In Diagnostik zeigten sich derartige Befunde allerdings nicht.

### Die Wirkung der Probeklausur auf die aktuelle Angst vor der Klausur

Obwohl eine Probeklausur laut Theorie unter günstigen Bedingungen aktuelle Angst vor der Klausur reduzieren sollte, ergaben die bisherigen empirischen Befunde von Jacobs et. al. (2004)

in Statistik keine Angstunterschiede und im Fach Didaktik (Jacobs 2003) entgegen der Erwartung signifikant höhere Angst unter der Bedingung Probeklausur.

**Tabelle 4:** Ergebnisse zur aktuellen Angst unmittelbar vor der Klausur für EG (Probeklausur) und KG (keine Probeklausur)

| <b>Diagnostik Klausur</b>                                       |    | M    | s    | t (54) | p <sub>e</sub> | d    | sign |
|-----------------------------------------------------------------|----|------|------|--------|----------------|------|------|
| Aktuelle Prüfungsangst<br>(SPA) $\alpha = .96$ ; $r_{tt} = .87$ | EG | 20.2 | 10.0 | -5.03  | .000           | 1.34 | *    |
|                                                                 | KG | 32.1 | 7.7  |        |                |      |      |
| Fearthermometer (FT)<br>$r_{SPA,FT} = .88$ ; $r_{tt} = .78$     | EG | 2.4  | 2.1  | -4.38  | .000           | 1.18 | *    |
|                                                                 | KG | 4.9  | 2.1  |        |                |      |      |

| <b>Methoden Klausur</b>                                         |    | M    | s    | t (57) | p <sub>e</sub> | d   | sign |
|-----------------------------------------------------------------|----|------|------|--------|----------------|-----|------|
| Aktuelle Prüfungsangst<br>(SPA) $\alpha = .97$ ; $r_{tt} = .87$ | EG | 25.5 | 13.7 | -1.26  | .107           | .33 |      |
|                                                                 | KG | 29.8 | 12.9 |        |                |     |      |
| Fearthermometer (FT)<br>$r_{SPA,FT} = .80$ ; $r_{tt} = .9$      | EG | 3.1  | 2.5  | -1.17  | .125           | .30 |      |
|                                                                 | KG | 3.8  | 2.5  |        |                |     |      |

**Anmerkung:**  $\alpha$  = Reliabilität nach Cronbach. Als untere Grenze der Reliabilitätsschätzung  $\alpha$  für das FT wurde die Korrelation mit SPA herangezogen.  $r_{tt}$  = Retest (1 Woche Zeitabstand) nur EG; \*= signifikant

Wie aus Tabelle 4 hervorgeht, liegen nun alle Ergebnisse zumindest in der erwarteten Richtung geringerer Klausurangst nach einer Probeklausur. Während im Seminar Diagnostik die Probeklausur eine sehr deutliche Angstreduktion unmittelbar vor der Klausur bewirkte, fällt der entsprechende Vorteil der Probeklausur im Fach Methoden numerisch zu gering aus, um die Signifikanzhürde zu nehmen. Die Hypothese, eine Probeklausur führe zu einer Angstreduktion kann somit partiell gestützt werden. Würde man beide Seminare zusammenlegen, dann ergäbe sich im Gesamtmittelwert eine signifikante Angstreduktion für die Bedingung Probeklausur.

In beiden Seminaren fällt die aktuelle Angst unmittelbar vor der Probeklausur (SPA: 21,4 bzw. 22,0; FT: 2,0 bzw. 1,8) deutlich geringer aus als bei Jacobs et al. 2004 bzw. Jacobs (1999, siehe [Online-Angsttest](#).) Allerdings wurden die dort analysierten Klausuren in üblicher Weise mit den Noten 1 bis 5 bewertet. Die divergierenden Ergebnisse im Vergleich zur sehr hohen Angst in der Probeklausur "Statistik" bei Jacobs et al. (2004) könnten darauf zurückgehen, dass die Probeklausur Statistik völlig unerwartet zu Beginn einer Seminarsitzung angesetzt wurde, während die Studierenden in dieser Studie eine Woche vor der Probeklausur über den Termin der Probeklausur informiert worden waren. Manche Studierende haben hier die Zeit möglicherweise genutzt, sich auf die Probeklausur vorzubereiten.

Die aktuelle Angst in der Diagnostikklausur der Kontrollgruppe von SPA=32,1 bzw. FT= 4.9 entspricht recht gut den bisher bekannten Mittelwertsergebnissen aus mehreren Klausuren sowie der Angst [unmittelbar vor der mit Noten 1 bis 5 bewerteten Vorlesungsklausur "Pädagogische Diagnostik"](#). Somit bestätigt sich eine Vermutung von Jacobs (2010b), auch Klausuren mit einer Bewertung „bestanden/nicht bestanden“ könnten ähnlich hohe aktuelle Angst auslösen wie klassisch benotete Klausuren.

Betrachtet man die aktuelle Angst der EG in Probeklausur und Klausur, so stellt man für das Seminar Diagnostik keinerlei Veränderung fest. D.h. Die Studierenden der EG traten zur Probeklausur und Klausur mit hoch vergleichbarem, ziemlich geringem Angstlevel in SPA und FT an. Im Methodenseminar findet man jedoch für beide Angstmaße einen signifikanten Angstanstieg von der Probeklausur zur Klausur ( $p$  jeweils  $<0.05$ ). Alle Befunde stehen im Widerspruch zu vergleichbar hohen Angstwerten in einer Statistikprobe- und Abschlussklausur bei Jacobs et al (2004), was die Autoren damals zu der These veranlasste, Probeklausuren seien Stresssituationen und würden in gleichem Maße Angst erzeugen wie echte Klausuren. Wie die Ergebnisse hier aber übereinstimmend aufzeigen, erzeugten die Probeklausuren recht wenig Angst, echte Klausuren ohne vorherige Probeklausur hingegen deutlich mehr Angst.

### Theoretisch angenommener Weg zu einer Angstreduktion

Gemäß dem Modell der Wirkungsweise einer Probeklausur auf die Angst in Abbildung 1 sollte die Teilnahme an der Probeklausur zunächst die Transparenz der Prüfungsanforderungen steigern und so den Glauben an ein besseres Ergebnis in der Klausur stärken. Höhere Erfolgserwartungen bedeuten zugleich geringere Misserfolgsantizipationen. Da die Erwartung des Misserfolgs als eine tragende Säule der Angstentstehung aufgefasst wird, müsste die Angst mit sinkender Misserfolgswahrscheinlichkeit abnehmen.

Wie aus Tabelle 5 erkennbar, fällt die Transparenz der Prüfungsanforderungen bei den Studierenden mit Probeklausur deutlich höher aus als bei der KG. Zudem liegen auch die Ergebnisse einzelner Items in der erwarteten Richtung. Das eigene Leistungsergebnis kann besser eingeschätzt werden. Damit wird gegenüber der KG etwas Ungewissheit genommen, die häufig auch als Angst auslösende Bedingung angesehen wird. Die Transparenz schaffende Wirkung der Probeklausur trägt offenbar auch zur Erkenntnis bei, dass man sich besser auf die Prüfung vorbereiten konnte. Eine wertvolle Hilfe dafür wäre die zielgerechte Vorbereitung. Daran zweifeln eher die KG-Studierenden, wenngleich der erwartete Unterschied nur im Seminar Methoden auf dem üblichen 5% Niveau signifikant wurde.

**Tabelle 5:** Ergebnisse von EG (=Probeklausur) und KG (keine Probeklausur) für etliche subjektive Einschätzungen unmittelbar vor der Klausur

| <b>Diagnostik Klausur</b>                                                 |    | M    | s    | t (54) | $p_e$ | d    | sign |
|---------------------------------------------------------------------------|----|------|------|--------|-------|------|------|
| Transparenz der Prüfungsanforderungen                                     | EG | 5.7  | 0.9  | 3.87   | .001  | 1.05 | *    |
|                                                                           | KG | 4.7  | 1.1  |        |       |      |      |
| Vertrauen in die eigene Leistungsfähigkeit                                | EG | 4.8  | .78  | 2.53   | .007  | .68  | *    |
|                                                                           | KG | 4.2  | .93  |        |       |      |      |
| "Ich kann meine Leistung im Abschlussquiz ziemlich genau einschätzen."    | EG | 4.7  | 1.0  | 2.17   | .018  | .58  | *    |
|                                                                           | KG | 4.1  | 1.2  |        |       |      |      |
| "Ich zweifelte daran, ob mir das, was ich lernte, überhaupt etwas nützt." | EG | 2.9  | 1.2  | -1.57  | .061  | .42  |      |
|                                                                           | KG | 3.4  | 1.3  |        |       |      |      |
| "Ich konnte mich sehr gut auf das Abschlussquiz vorbereiten."             | EG | 5.0  | 1.2  | 2.55   | .007  | .68  | *    |
|                                                                           | KG | 4.1  | 1.3  |        |       |      |      |
| Subjektive Kompetenz (erwartete % richtig)                                | EG | 61.4 | 11.8 | 1.99   | .025  | .54  | *    |
|                                                                           | KG | 54.8 | 12.9 |        |       |      |      |

|                                                                   |    |      |      |      |      |     |   |
|-------------------------------------------------------------------|----|------|------|------|------|-----|---|
| Wahrscheinlichkeit (%),<br>die angestrebte Punktzahl zu erzielen. | EG | 73.6 | 17.5 | 2.49 | .008 | .67 | * |
|                                                                   | KG | 61.6 | 18.3 |      |      |     |   |

| <b>Methoden Klausur</b>                                                   |    | M    | s    | t (57) | p <sub>e</sub> | d   | sign |
|---------------------------------------------------------------------------|----|------|------|--------|----------------|-----|------|
| Transparenz der Prüfungsanforderungen                                     | EG | 5.5  | 1.0  | 2.57   | .007           | .68 | *    |
|                                                                           | KG | 4.8  | 1.1  |        |                |     |      |
| Vertrauen in die eigene Leistungsfähigkeit                                | EG | 4.47 | 0.97 | 1.98   | .027           | .52 | *    |
|                                                                           | KG | 3.93 | 1.09 |        |                |     |      |
| "Ich kann meine Leistung im Abschlussquiz ziemlich genau einschätzen."    | EG | 4.7  | 1.0  | 2.74   | .004           | .71 | *    |
|                                                                           | KG | 3.9  | 1.3  |        |                |     |      |
| "Ich zweifelte daran, ob mir das, was ich lernte, überhaupt etwas nützt." | EG | 3.3  | 1.4  | -1.89  | .037           | .49 | *    |
|                                                                           | KG | 4.0  | 1.4  |        |                |     |      |
| "Ich konnte mich sehr gut auf das Abschlussquiz vorbereiten."             | EG | 4.9  | 1.4  | 2.43   | .009           | .63 | *    |
|                                                                           | KG | 4.0  | 1.5  |        |                |     |      |
| Subjektive Kompetenz (erwartete % richtig)                                | EG | 55,9 | 10.7 | 1.90   | .031           | .50 | *    |
|                                                                           | KG | 49,8 | 13.4 |        |                |     |      |
| Wahrscheinlichkeit (%), die angestrebte Punktzahl zu erzielen.            | EG | 64.8 | 20.2 | 2.12   | .019           | .55 | *    |
|                                                                           | KG | 54.4 | 17.7 |        |                |     |      |

Der Vorteil der Probeklausur manifestiert sich letztlich auch in der konkreten Erwartung eines besseren Leistungsergebnisses (höhere subjektive Kompetenz) sowie der Einschätzung, das angestrebte Leistungsergebnis eher erreichen zu können. Ähnliches gilt für das Vertrauen in die eigene Leistungsfähigkeit. Wie die beiden Items des Vertrauens in die eigene Leistungsfähigkeit in Abbildung 2 verdeutlichen, rechneten die Studierenden der EG mit einem weniger schlechten Ergebnis und der Chance, bei auftretenden Problemen in der Klausur noch handlungsfähig zu bleiben. In ihrer Gesamtheit deuten die Ergebnisse der Tabelle 5 klar in die Richtung, den Angst bedingenden Faktor eines möglichen Misserfolgs eingedämmt zu haben.

### Veränderungen subjektiver Einschätzungen von Probeklausur zur Klausur

Der experimentelle Versuchsplan begründet bereits hinreichend, die nachgewiesenen Vorteile der Bedingung Probeklausur wären durch die Probeklausur hervorgerufen worden. Daher bleibt anzunehmen, vor Bearbeitung der Probeklausur hätte die Einschätzung der EG hohe Ähnlichkeit mit der Einschätzung der KG unmittelbar vor der Klausur. Um diese plausible Annahme zu prüfen, werden im Folgenden noch die Veränderungen der EG von Probeklausur zur Klausur analysiert, in dem sich freilich auch ein möglicher Vortesteffekt verbirgt. Zudem werden Reliabilitätsangaben der erhobenen subjektiven Daten vorgestellt, sofern das möglich ist.

Wie eine Analyse nach Augenschein ergab, entsprechen die Mittelwerte der EG vor der Probeklausur in etwa den Mittelwerten der KG unmittelbar vor der Klausur. Tabelle 6 zeigt jedoch direkt die Veränderung der EG von der Probeklausur zur Klausur als Differenz "Abschlussklausur - Probeklausur" in Effektstärke d. Positive Werte bedeuten Zuwachs. So steigern die Studierenden im Seminar Diagnostik ihre subjektive Transparenz unmittelbar vor der Probeklausur bis unmittelbar vor der Abschlussklausur in einem Effektstärkeausmaß von  $d = 1.33$ .

**Tabelle 6:** Reliabilitäten der subjektiven Daten und Veränderungen der Studierenden, die an der Probeklausur teilgenommen haben.

| <b>Diagnostik</b>                                                              | Probe-<br>klausur<br>(P) | Absch.<br>klausur<br>(A) | Abstand<br>1 Woche | Veränderung<br>[A-P] in Ef-<br>fektstärke d |
|--------------------------------------------------------------------------------|--------------------------|--------------------------|--------------------|---------------------------------------------|
| N                                                                              | 27                       | 56                       | 26                 |                                             |
| Reliabilität                                                                   | $\alpha$                 | $\alpha$                 | Retest             |                                             |
| Transparenz der<br>Prüfungsanforderungen                                       | .71                      | .87                      | .43                | 1.33*                                       |
| Vertrauen die eigene<br>Leistungsfähigkeit                                     | .89                      | .85                      | .54                | 0.77*                                       |
| "Ich kann meine Leistung im<br>Abschlussquiz ziemlich ge-<br>nau einschätzen." |                          |                          | .45                | 1.06*                                       |
| Subjektive Kompetenz<br>(erwartete % richtig)                                  |                          |                          | .45                | 0.83*                                       |
| Wahrscheinlichkeit (%),<br>die angestrebte Punkt-<br>zahl zu erzielen.         |                          |                          | .63                | 0.57*                                       |

| <b>Methoden</b>                                                                | Probe-<br>klausur<br>(P) | Ab-<br>sch.<br>klau-<br>sur<br>(A) | Ab-<br>stand<br>1 Wo-<br>che | Verände-<br>rung<br>[A-P] in<br>Ef-<br>fektstärke<br>d |
|--------------------------------------------------------------------------------|--------------------------|------------------------------------|------------------------------|--------------------------------------------------------|
| N                                                                              | 29                       | 61                                 | 29                           |                                                        |
| Reliabilität                                                                   | $\alpha$                 | $\alpha$                           | Retest                       |                                                        |
| Transparenz der<br>Prüfungsanforderungen                                       | .74                      | .84                                | .30                          | 1.10*                                                  |
| Vertrauen die eigene<br>Leistungsfähigkeit                                     | .68                      | .76                                | .50                          | .73*                                                   |
| "Ich kann meine Leistung im<br>Abschlussquiz ziemlich ge-<br>nau einschätzen." |                          |                                    | -.13                         | .86*                                                   |
| Subjektive Kompetenz<br>(erwartete % richtig)                                  |                          |                                    | .66                          | 1.02*                                                  |
| Wahrscheinlichkeit (%),<br>die angestrebte Punkt-<br>zahl zu erzielen.         |                          |                                    | .38                          | .52*                                                   |

\* = signifikante Veränderung auf dem 10% Niveau.

Alle Veränderungen sind signifikant auf dem 10% Niveau. Die p-Werte selbst liegen häufig unter .001. Wie nicht anders zu erwarten, bestätigen die Ergebnisse zu den EG-Veränderungen von Probeklausur zur Klausur die Befunde des unabhängigen Vergleichs zwischen EG und KG zum Zeitpunkt unmittelbar vor der Klausur. Derart klare Effekte können nicht allein durch einen möglichen Vortesteffekt verursacht worden sein und deshalb besteht kein ernsthafter Zweifel daran, die Probeklausur habe auf subjektiver Ebene bedeutsame Veränderungen bewirkt, die eine geringere Angst vor der Klausur begünstigen müssten.

## Zusammenfassung und Diskussion

Forschung im natürlichen Feld lässt sich auf die Komplexität der Praxis ein. Diese wird von einem Geflecht vielfältiger Wirkfaktoren beeinflusst, so dass pauschale Aussagen darüber, welche Konsequenzen Tests und Rückmeldungen im Schulalltag hinterlassen, gar nicht möglich sind (siehe Jacobs 2010a). Um in der natürlichen Umwelt überhaupt zu interpretierbaren Ergebnissen kommen zu können, sind stringente Versuchspläne noch wichtiger als in der Laborforschung.

Der in dieser Studie für Feldforschung ungewöhnlich rigorose experimentelle Versuchsplan ist im Hinblick auf die interne Validität aber gegenüber den sozial bedingten Störfaktoren unkontrolliert. Da die Probeklausur Passwort geschützt in den Computerräumen der Universität bearbeitet wurde und Notizen vor Verlassen des Computerraumes eingezogen wurden, dürfte die Probeklausur oder nennenswerte Teile davon nicht in die Hände der KG gelangt sein. Es bleibt allerdings unklar, ob und wie viele EG- und KG-Studierende vor der Klausur real oder über Medien miteinander in Kontakt traten, um sich über die Probeklausur auszutauschen. Allerdings ist zu vermuten, solch vereinzelte potenzielle Detailauskünfte hätten im Klausurergebnis angesichts von 47 teils umfangreichen Aufgaben kaum eine nennenswerte Wirkung hinterlassen. Wenn dennoch bedeutsame Informationen von der EG in die KG durchgedrungen sein sollten, dann hätten sie den Nachweis der erwarteten Effekte eher erschwert.

### **Probeklausur kann die Klausurleistung steigern**

Trotz anfänglicher Skepsis, durch pädagogische Maßnahmen Klausurergebnisse im natürlichen universitären Umfeld kaum bedeutsam verbessern zu können, ist es gelungen, durch das verpflichtende Angebot einer Probeklausur die Klausurleistung signifikant in einem Ausmaß niedriger bis mittlerer Effektstärke zu steigern. Von der Höhe der Effektstärke her entsprechen die Ergebnisse dem Leistungseffekt der Probeklausur im Fach Statistik ( $d=.43$ ; Jacobs et. al 2004), auch wenn dort wegen des strengeren Signifikanzniveaus die Signifikanzhürde nicht generell, wohl aber für die schwächeren Studierenden, übersprungen wurde.

Mit Recht behauptete Jacobs (2003): "Die positive Wirkung einer Probeklausur auf Klausurleistung und Angst hängt mit davon ab, wie stark ihre Vorteile gegenüber sonstigen Realisierungen einer guten Prüfungsvorbereitung ausfallen." Hierbei spielen mehrere Faktoren eine Rolle, unter anderem die Strukturiertheit der Seminare hinsichtlich des relevanten Lehrstoffs, die Durchsichtigkeit der erwarteten Anforderungen, die Verständlichkeit der Lernmaterialien, das Angebot sonstiger pädagogischer Hilfen. Diesbezüglich boten die Seminare auch der Kontrollgruppe schon eine gute Orientierung. Allerdings mag man kritisieren, die Kontrollgruppe hätte nicht als echte Konkurrenz zur Experimentalgruppe fungiert, weil diese lediglich eine pädagogische Empfehlung an die selbstkontrollierte Klausurvorbereitung erhielt, während die Kontrollgruppe bei Jacobs et al. (2004) bzw. Jacobs (2003) eine verpflichtende Seminarsitzung absolvierte, die dem alleinigen Ziel diene, möglichst gut auf die Klausur vorzubereiten (genaue Angaben siehe: Jacobs 2003, S. 9). Solch wesentlich anspruchsvollere Kontrolle mag mit ein Grund dafür gewesen sein, warum im Fach Didaktik keine Leistungsunterschiede in der Klausur zwischen EG und KG festzustellen waren.

Der hier erzielte Klausurvorteil durch der Probeklausur kann auf mehreren Ursachen beruhen, die in dieser Studie allerdings nicht voneinander getrennt werden können. Als potenzielle Gründe kommen hauptsächlich folgende Faktoren in die nähere Auswahl

- Die Probeklausur operationalisiert die prüfungsrelevanten Lehrziele, präzisiert die Anforderungen und reduziert den Umfang potenzieller Prüfungsthemen.
- Testen und sachorientiertes Feedback in der Probeklausur sind wirksame direkte Instruktionsmaßnahmen. Korrekte Antworten stärken das Behalten, falsche Antworten werden als falsch erkannt und können durch die Rückmeldung der korrekten Antwort revidiert werden.
- Sachorientiertes Feedback einschließlich der Rückmeldung des Prozentsatzes der Lösungen offenbaren Stärken, Schwächen und den aktuellen Leistungsstand des Studierenden.

Insgesamt liefert die Probeklausur wertvolle Informationen, die noch verfügbare Vorbereitungszeit entsprechend der Zielsetzung des Studierenden bestmöglich zu nutzen. Diese Argumentation wird durch die Einschätzung der Studierenden gestützt. Unter der Bedingung Probeklausur fällt die subjektive Transparenz der Prüfungsanforderungen und die Einschätzung, man könne sich gut auf die Klausur vorbereiten, deutlich höher aus als unter Kontrollbedingung. Teilnehmer an der Probeklausur zweifeln auch weniger am Nutzen Ihrer Vorbereitung. Obwohl die EG im Vergleich zur KG unmittelbar vor der Klausur subjektiv zudem den Eindruck hatte, ihre Leistung im Abschlussquiz besser einschätzen zu können, trifft dies im Hinblick auf die Genauigkeit allerdings objektiv nicht zu. Erwartete und tatsächliche Klausurleistung korrelieren zwar in beiden Seminaren signifikant um ca.  $r=.30$ , die Studierenden in allen Seminaren und Gruppen unterschätzten unmittelbar vor der Klausur ihre tatsächliche Leistung in der Klausur jedoch massiv um durchschnittlich 10 bis 18 Prozent korrekter Antworten. Hierbei fiel die Unterschätzung für die EG-gruppen im Vergleich zu den KG-gruppen zwar numerisch um ca. 4 Prozent geringer aus, dieser Unterschied reichte aber für ein signifikantes Ergebnis nicht aus.

### **Probeklausur kann Angst senken**

Faktoren, welche die objektive Leistung steigern, scheinen auch geeignet, Angst abzubauen, sofern sie subjektiv das Gefühl einer Kompetenzerhöhung nach sich ziehen. In der Klausur Diagnostik empfanden die Studierenden mit vorausgehender Probeklausur deutlich weniger Angst als Studierende ohne Probeklausur. In der Methodenklausur deutete sich der Angst reduzierende Effekt einer Probeklausur an, lies sich aber statistisch nicht belegen. Fasst man beide Seminare zusammen, dann ergibt sich im Mittel ein statistisch deutlicher Angstabbau durch Probeklausur in einem Ausmaß mittlerer bis hoher Effektstärke. In den bei Jacobs et al. (2004) bzw. Jacobs (2003) analysierten Klausuren fand man für Statistik allerdings gar keinen Angst-reduktionseffekt und in Didaktik entgegen der Erwartung sogar eine signifikante Angsterhöhung durch die Probeklausur.

Hier ließ sich der theoretisch angenommene Weg von der Probeklausur zu einer Angstreduktion relativ überzeugend bestätigen. Durch die Probeklausur wurde subjektiv Transparenz erhöht, das Vertrauen in die eigene Leistungsfähigkeit gestärkt, ein besseres Prüfungsergebnis erwartet und eher damit gerechnet, das angestrebte Prüfungsergebnis auch zu erreichen. Diesen Angst senkenden Faktoren zufolge hätte man in beiden Seminaren eigentlich auch eine ähnlich hohe Angstreduktion erwarten müssen.

In der Studie von Jacobs et al. (2004) gelang es ebenfalls, durch die Probeklausur die subjektive Transparenz sowie die subjektive Vorbereitbarkeit einer Statistik Klausur zu erhöhen. Diese

Vorteile führten aber nicht zu einer höheren subjektiven Kompetenz. Die Studierenden der EG glaubten nicht daran, eine höhere Punktzahl zu erreichen oder eher ihre angestrebte Note zu erzielen. Die Probeklausur stärkte auch nicht das Vertrauen in die eigene Leistungsfähigkeit. Wenngleich unklar bleibt, warum die verbesserte Vorbereitbarkeit der Klausur keine optimistischeren Klausurergebniserwartungen nach sich zog, war daraus zumindest die fehlende Angstreduktion in der Statistikklausur verständlich. Denn trotz Probeklausur konnte die Misserfolgserwartung nicht gesenkt werden. Transparenz der Prüfungsanforderungen kann den Lernerfolg keineswegs garantieren. Das gilt insbesondere bei schwierigen Lehrstoffen. So führten etwa im Saarbrücker Schulangstprojekt Transparenz schaffende Maßnahmen (z.B. die Explizierung der prüfungsrelevanten Lehrziele) in fähigkeitsabhängigen Schulfächern (Fremdsprachen und Mathematik) zwar zu einer Erhöhung der subjektiven Transparenz. Aber ähnlich wie in Statistik zog diese keine verbesserte Leistungserwartung in der Klassenarbeit, und damit auch nicht weniger Angst nach sich. (Jacobs (1987 bzw. [Studie 3](#)).

Im seminar Didaktik schließlich ließ sich durch die Probeklausur nicht einmal die subjektive Transparenz erhöhen (Jacobs 2003), geschweige denn die Vorbereitbarkeit der Klausur oder eine Verbesserung der erwarteten Note. Eine mögliche Erklärung für die **durch die Probeklausur bewirkte signifikant höhere aktuelle Angst** vor der Didaktikklausur lieferten die ausgesprochen schwachen Ergebnisse in der Probeklausur. In der Probeklausur Didaktik erzielte ca. die Hälfte der Studierenden Leistungsergebnisse, die in der Klausur als "durchgefallen" gegolten hätten, vermutlich auch deshalb, weil die Probeklausur für die Studierenden völlig unerwartet und überraschend angesetzt wurde. In vorliegender Studie hingegen erreichten bzw. übertrafen fast alle Studierende ein Leistungsergebnis in der Probeklausur von 50% korrekter Lösungen, das in der späteren Klausur ganz locker einem "Bestanden" gleich gekommen wäre. Die unterschiedlichen Ergebnisse einzelner Seminare verdeutlichen, eine Probeklausur würde nicht zwingend positive Effekte zeitigen, sondern nur unter bestimmten günstigen Bedingungen.

### Alternative Vorbereitungshilfen für die Klausur

Als sehr einfache Vorbereitungsmethode bietet sich an, den Studierenden die Lehrziele aller Klausuraufgaben mitzuteilen. Unter ganz besonderen, sehr günstigen Bedingungen, gelang es so, in Klassenarbeiten die Leistung zu steigern und die Angst deutlich zu senken (Jacobs, Bedersdorfer & Bohse-Wagner (1983) siehe auch [Studie 2](#)). Normalerweise funktioniert diese Methode im normalen Schulbetrieb bei fähigkeitsabhängigen Schulfächern allerdings nicht (Jacobs, 1987). Der Verfasser hatte diese Maßnahme in den letzten Semestern regelmäßig als Klausurvorbereitung im Methodenseminar eingesetzt. Unter Zuhilfenahme der Kontrollgruppe "Methoden" in vorliegender Studie lies sich die Wirksamkeit der Explizierung von Klausurlehrzielen überprüfen, leider nur auf der Basis eines ziemlich schwachen Versuchsplans. Die Ergebnisse deuten im gegebenen Fall darauf hin, die Klausurleistung auf diese Weise nicht verbessern zu können. (siehe Anhang C)

Eigene empirische Studien zur Wirkung von Tests und Rückmeldungen auf das Behalten umfassten als Kontrolle die Präsentation korrekt beantworteter Übungsaufgaben. Statt die Fragen wie in der Bedingung "Testen mit Feedback" selbst zu beantworten und erst danach die korrekten Lösungen mit Erklärungen einzusehen, betrachteten die Studierenden der Kontrollgruppe "Musterlösungen" direkt die bereits korrekt beantworteten Aufgaben einschließlich des elaborierten Feedbacks. Sie wurden aufgefordert, sich die Aufgabenlösung verständlich zu machen. In der Gesamtsicht aller durchgeführten eigenen Studien erwiesen sich beide Übungsmethoden als vergleichbar lernwirksam. Musterlösungen erforderten aber weniger Lernzeit. (Jacobs, 2010c, 2010d, 2011, Jacobs & Sparfeldt, 2014). Vermutlich könnten Musterlösungen ähnliche

Auswirkungen auf das Klausurergebnis und die Klausurangst wie Probeklausuren haben. Ihre Nutzung ist zudem ökonomischer, da unter anderem keine Auswertung anfällt.

### **Einige praktische Probleme beim Einsatz von Probeklausuren**

Wer zu einer Klausur noch eine Probeklausur erstellen will, hat die doppelte Arbeit, beim konkreten Einsatz im Falle von Papier und Bleistift-Varianten auch noch den doppelten Korrekturaufwand. Rein ökonomisch betrachtet, kommen Probeklausuren daher vornehmlich bei wiederkehrenden Seminaren mit relativ konstanten Lehrzielen (etwa Grundlagenfächer) in Betracht, die mit Hilfe von Computern durchgeführt und ausgewertet werden. Die Hauptaufgabe des Dozenten bestünde dann in der Konstruktion geeigneter Aufgaben, informativer Rückmeldungen und der Aufgabenauswahl für die Klausur. Den Rest übernimmt ein Computerprogramm. Hilfskräfte kontrollieren die Durchführung. Bei intelligenter Handhabung eines großen Aufgabenpools ließen sich langfristig, gestützt auf bisher durchgeführte Probe- und Abschlussklausuren, Aufgabenanalysen durchführen, mit deren Hilfe man unbrauchbare Aufgaben eliminieren, Klausuren verbessern und Paralleltests für Probeklausuren erzeugen könnte.

Manch einer mag Probeklausuren auf Grund pädagogischer Grundüberzeugung ablehnen, etwa deshalb, weil damit das Universum aller möglichen Prüfungsaufgaben in Frage gestellt wird. Das Argument noch etwas verschärfend, könnte man analog der Ablehnung des Prinzips "teaching for the test" die Ansicht vertreten, es mache pädagogisch wenig Sinn, "den Prüfling erst mal vorzukauen zu lassen, was ihm dann in der Klausur später vorgesetzt wird". Dem kann aber entgegen gehalten werden, in der Klausur würden zwar die gleichen Lehrziele, diese aber mit unterschiedlichen Aufgaben geprüft werden. Ziel einer Klausur sollte es meiner Meinung nach nicht sein, alle möglichen, sondern die wichtigsten Lehrziele eines Seminars abzuprüfen. Mit einer Probeklausur als Vortest und der Abschlussklausur als Nachtest werden diese wichtigen Lehrziele zweimal getestet. Durch eine Probeklausur könnte so nicht nur das Ergebnis in der Abschlussklausur besser ausfallen, sondern das erworbene Wissen auch etwas nachhaltiger behalten werden.

### **Literatur**

Jacobs, B. (1981). Angst in der Prüfung - Beiträge zu einer kognitiven Theorie der Angstentstehung in Prüfungssituationen. Fischer Verlag. Frankfurt.

Jacobs, B., Bedersdorfer H.-W., Bohse-Wagner, N. (1983). Angstabbau durch Transparenz - Eine quasiexperimentelle Felduntersuchung zur ökologischen Validierung eines theorieorientierten Interventionsprogramms zur Reduktion von Angst in der Prüfung. Arbeitsbericht Nr. 18 aus der Fachrichtung Allgemeine Erziehungswissenschaft. Universität des Saarlandes.

Jacobs, B. (1987). Die Auswirkungen transparenzschaffender Maßnahmen auf die aktuelle Angst vor einer Klassenarbeit - Eine Metaanalyse zum Saarbrücker Schulangstprojekt. Empirische Pädagogik 1(2). 139-160

Jacobs, B. (1999). Löst eine Probeklausur aktuelle Angst aus?  
<http://www.phil.uni-sb.de/~jakobs/seminar/vpl/klausur/angst.htm>

Jacobs, B. (2000). [Feedback nach Bearbeitung einer Probeklausur - Wird es genutzt](http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/probefeedback.htm) <http://www.phil.uni-sb.de/~jakobs/wwwartikel/feedback/probefeedback.htm>  
 [24.2.2000]

Jacobs, B., Bernd, H. & Fey, A. (2004). Die Wirkung einer Probeklausur auf Klausurleistung und Angst in einer Statistikklausur.

URN: [urn:nbn:de:bsz:291-psydok-2720](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-2720)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2004/272/>

Jacobs, B. (2003). Die Wirkung einer Probeklausur auf Lernleistung und aktuelle Angst in einer echten Klausur.

URL des Originals [wegen Serverschließung nicht mehr erreichbar]:

<http://www.phil.uni-sb.de/~jakobs/wwwartikel/probeklausur/index.htm>

created: 21.11.2003 = Langfassung von Jacobs, B., Bernd, H. & Fey, A. (2004).

Kopie zurzeit verfügbar unter:

<http://bildungswissenschaften.uni-saarland.de/personal/jacobs/artikel/probeklausur.pdf>

Jacobs, B. (2010a). Tests als Maßnahmen zur Förderung von Studierleistungen ?

URN: [urn:nbn:de:bsz:291-psydok-26022](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-26022)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2602/>

Jacobs, B. (2010b). Bestanden/nicht bestanden im Vergleich zum traditionellen, mehrstufigen Benotungssystem.

URN: [urn:nbn:de:bsz:291-psydok-35203](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-35203)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2012/3520/>

Jacobs, B. (2010c). Gezieltes Studieren gelöster Aufgaben als alternative Übungsmethode zu Testen mit Feedback.

URN: [urn:nbn:de:bsz:291-psydok-26022](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-26022)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2010/2602/>

Jacobs, B. (2010d). Testfragen selbst beantworten oder Musterlösungen studieren?

URN: [urn:nbn:de:bsz:291-psydok-26934](https://nbn-resolving.org/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](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-27127)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2011/2712/>

Jacobs, B. & Sparfeld, J. (2014). Musterlösungen und Testen mit Feedback als vergleichbar lernwirksame Übungsmethoden in der universitären Lehre

URN: [urn:nbn:de:bsz:291-psydok-51683](https://nbn-resolving.org/urn:nbn:de:bsz:291-psydok-51683)

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2014/5168/>

# Anhang

## Anhang A

### Aufgabenbeispiele für Probeklausur und Klausur

#### Diagnostik Aufgabenbeispiele

##### Aufgabe 1 [Probeklausur]

Dagobert wirkt recht ängstlich, wenn er aufgerufen wird, an die Tafel kommen soll, eine Klassenarbeit geschrieben oder zurück gegeben wird. Ein strenger Persönlichkeitstheoretiker erklärt diese Ängste von Dagobert am ehesten mit

- 1.) ☐ der hohen allgemeinen Ängstlichkeit von Dagobert
- 2.) ☐ den Stress auslösenden Bedingungen in Schule und Elternhaus.
- 3.) ☐ der übertriebenen Leistungsorientierung der Gesellschaft.
- 4.) ☐ der Angst auslösenden Persönlichkeit des Lehrers.

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

*Rückmeldung nach Aufgabenbestätigung : richtig = 1*  
Im Gegensatz zur Verhaltenstheorie, bestimmen nicht Reize und Konsequenzen der Umwelt, sondern die Persönlichkeit des Betroffenen sein Verhalten.

##### Aufgabe 1 [Klausur] richtig = 2

Peter beteiligt sich nicht mehr am Unterricht, vernachlässigt zunehmend die Hausaufgaben und bereitet sich nicht mehr auf Prüfungen vor. Ein strenger Verhaltenstheoretiker erklärt diese geringe schulische Leistungsbereitschaft von Peter am ehesten mit

- 1.) ☐ einem sehr geringen Leistungsmotiv von Peter.
- 2.) ☐ ungünstigen Reiz-, Kontroll- und Verstärkungsbedingungen in Schule und Elternhaus.
- 3.) ☐ der Null-Bock-Mentalität eines großen Teils der Jugendlichen.
- 4.) ☐ mangelnden Wertschätzung schulisch angemessenen Verhaltens.

|                    |
|--------------------|
| Aufgabe bestätigen |
|--------------------|

## Aufgabe 20 Probeklausur

**Eine Schulklasse mit 28 Schülern erzielt in der Mathe- und Deutscharbeit folgende Noten.**

Die Zahlen hinter den Fächern repräsentieren die Häufigkeiten [z.B. 3 Schüler erhalten in Deutsch eine 5]

| Note    | 1 | 2 | 3  | 4  | 5 | 6 |
|---------|---|---|----|----|---|---|
| Deutsch | 1 | 2 | 10 | 10 | 3 | 2 |
| Mathe   |   | 2 | 14 | 12 |   |   |

|     | Entscheide jeweils, ob die Aussage zutrifft oder nicht!                       | a<br>richtig          | b<br>falsch           |
|-----|-------------------------------------------------------------------------------|-----------------------|-----------------------|
| 1.) | Die Deutschnoten fallen im Durchschnitt besser aus als die Mathenoten         | <input type="radio"/> | <input type="radio"/> |
| 2.) | Die Noten in Deutsch streuen stärker als die Noten in Mathe                   | <input type="radio"/> | <input type="radio"/> |
| 3.) | Der Modus in Mathe lautet 14                                                  | <input type="radio"/> | <input type="radio"/> |
| 4.) | Der Median fällt in Mathe höher aus als in Deutsch                            | <input type="radio"/> | <input type="radio"/> |
| 5.) | Der Mittelwert in Mathe ist höher als der Median in Mathe                     | <input type="radio"/> | <input type="radio"/> |
| 6.) | Über die Hälfte der Klasse erzielt in Mathe mindestens die Note 3 oder besser | <input type="radio"/> | <input type="radio"/> |

|                     |                    |
|---------------------|--------------------|
| Aufgaben bestätigen | Korrekte Lösungen? |
|---------------------|--------------------|

*Rückmeldung nach Aufgabenbestätigung : richtig : 1b 2a 3b 4b 5a 6a*

Klicke gegebenenfalls auf korrekte Lösung  
und versuche die korrekte Antwort zu verstehen.

## Aufgabe 20 Klausur [richtig: 1a 2b 3b 4a 5a 6b]

**Eine Schulklasse mit 28 Schülern erzielt in der Mathe und Deutscharbeit folgende Noten**

Die Zahlen hinter den Fächern repräsentieren die Häufigkeiten [z.B. 10 Schüler erhalten in Deutsch eine 3]

| Note    | 1 | 2 | 3  | 4  | 5 | 6 |
|---------|---|---|----|----|---|---|
| Deutsch | 1 | 0 | 10 | 14 | 3 | 0 |
| Mathe   | 4 | 2 | 12 | 5  | 2 | 3 |

|     | Entscheide jeweils, ob die Aussage zutrifft oder nicht!                                           | a<br>richtig          | b<br>falsch           |
|-----|---------------------------------------------------------------------------------------------------|-----------------------|-----------------------|
| 1.) | Der Median in Mathe ist kleiner als der Median in Deutsch                                         | <input type="radio"/> | <input type="radio"/> |
| 2.) | Die Noten in Deutsch streuen stärker als die Noten in Mathe                                       | <input type="radio"/> | <input type="radio"/> |
| 3.) | Der Modus in Deutsch lautet 14                                                                    | <input type="radio"/> | <input type="radio"/> |
| 4.) | In Mathe fällt der Median kleiner aus als der Mittelwert                                          | <input type="radio"/> | <input type="radio"/> |
| 5.) | Der Modus in Deutsch ist größer als der Modus in Mathe                                            | <input type="radio"/> | <input type="radio"/> |
| 6.) | Mindestens 90% aller Schüler erzielen in Deutsch mindestens eine 4 (bzw. eine Note besser als 5). | <input type="radio"/> | <input type="radio"/> |

|                     |
|---------------------|
| Aufgaben bestätigen |
|---------------------|

**Aufgabe 23 [Probeklausur]**

Welches Testgütekriterium wird jeweils geprüft ?

**Hinweis:** Es genügt das allgemeine Gütekriterium. z.B. Objektivität statt Auswertungsobjektivität

| Nr. | Beschreibung                                                                                                                                                                                | Fachbegriff |
|-----|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| 1.  | Aus allen möglichen Additionsaufgaben wird eine repräsentative Stichprobe von 25 Aufgaben für den Tests gezogen.                                                                            |             |
| 2.  | Erst- und Zweitkorrektor haben die Arbeiten bewertet. Es ergab sich eine Übereinstimmung von 91% vergleichbarer Notenbewertungen                                                            |             |
| 3.  | Wer in der Klassenarbeit gut abgeschnitten hatte, erzielte auch im Schulleistungstest gute Leistungen. Schlechte Klassenarbeitsergebnisse gingen mit schwachen Schulleistungswerten einher. |             |
| 4.  | Die Schüler werden zweimal mit einem Schulleistungstest gemessen. Es ergaben sich hoch vergleichbare Ergebnisse ( $r=.92$ )                                                                 |             |

Aufgabe bestätigen    Korrekte Lösung?

*Rückmeldung nach Aufgabenbestätigung*

1. Inhaltsvalidität
2. Auswertungsobjektivität
3. Kriteriumsvalidität
4. Retestreliabilität

**Aufgabe 23 [Klausur]** [richtig: 1.Relibilität 2.Objektivität 3.Validität, 4.Validität]

Welches Testgütekriterium wird jeweils geprüft ?

| Nr. | Beschreibung                                                                                                                                              | Fachbegriff |
|-----|-----------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| 1.  | Studenten wurden im Abstand von 4 Wochen erneut mit demselben Konzentrationstest gemessen. Zwischen beiden Messungen ergab sich eine Korrelation von .77. |             |
| 2.  | Zwei Lehrer markieren unabhängig voneinander die Fehler in einem Diktat und erzielen eine Übereinstimmung von 95%                                         |             |
| 3.  | Es wird die Korrelation zwischen dem Studieneingangstest und dem Abschlusszeugnis im Studienfach ermittelt.                                               |             |
| 4.  | Für einen Englischtest wird von den 2000 wichtigsten Vokabeln eine repräsentative Stichprobe von 50 nach Zufall ausgewählt.                               |             |

Aufgabe bestätigen

**Aufgabe 33 [Probeklausur]**

Nachfolgende Alternativen geben den Prozentsatz der Schüler an, welche eine Aufgabe richtig gelöst haben.

Bei welcher Alternative ist die Aufgabenstreuung am höchsten ?

- 1.) ☐ 10%
- 2.) ☐ 31%
- 3.) ☐ 56%
- 4.) ☐ 85%

Aufgabe bestätigen    Korrekte Lösung? *Rückmeldung nach Aufgabenbestätigung: richtig:3*

Die Aufgabenstreuung hängt zwingend von der Aufgabenschwierigkeit  $S$  bzw. von der Lösungswahrscheinlichkeit  $p$  ab.  $s^2 = p * (1-p)$  Die Beziehung ist umgekehrt u-förmig. Bei mittlerer Aufgabenschwierigkeit ist die Streuung am höchsten.

**Aufgabe 33 [Klausur] richtig=2**

Nachfolgende Alternativen geben den Prozentsatz der Schüler an, welche eine Aufgabe **falsch** gelöst haben.

Bei welcher Alternative ist die Aufgabenstreuung am höchsten ?

1.) ☐ 10%

2.) ☐ 42%

3.) ☐ 85%

4.) ☐ 95%

**Aufgabe 47 [Probeklausur]**

**Berechnen Sie die Kennwerte auf der Basis der nachfolgenden Tabelle**

|             |       | Kriterium |    | Summe |
|-------------|-------|-----------|----|-------|
|             |       | +         | -  |       |
| <b>Test</b> | +     | 36        | 24 | 60    |
|             | -     | 4         | 36 | 40    |
|             | Summe | 40        | 60 | 100   |

Als Antwort werden Brüche, z.B. 50/100 oder Dezimalzahlen im Bereich von 0 bis 1 erwartet.

| Nr. |                                           |                      |
|-----|-------------------------------------------|----------------------|
| 1.  | <b>Sensitivität des Tests</b>             | <input type="text"/> |
| 2.  | <b>Spezifität des Tests</b>               | <input type="text"/> |
| 3.  | <b>positiver Vorhersagewert des Tests</b> | <input type="text"/> |
| 4.  | <b>negativer Vorhersagewert des Tests</b> | <input type="text"/> |
| 6.  | <b>Grundquote/Prävalenz</b>               | <input type="text"/> |

*Rückmeldung nach Aufgabenbearbeitung*

Sensitivität 36/40

Spezifität 36/60

positive Vorhersage 36/60

negative Vorhersage 4/40

Grundquote 40/100

**Aufgabe 47 [Klausur] Analoge Aufgabe mit anderen Zahlen in der Tabelle**

## Methoden Aufgabenbeispiele

### Aufgabe 1 [Probeklausur]

Welche der nachfolgenden Sätze sind verifizierbar, welche falsifizierbar, welche sind im erfahrungswissenschaftlichen Sinne weder verifizierbar noch falsifizierbar?

|     |                                                                                                                   | a<br>verifizierbar    | b<br>falsifizierbar   | c<br>weder/noch       |
|-----|-------------------------------------------------------------------------------------------------------------------|-----------------------|-----------------------|-----------------------|
| 1.) | Je mehr ein Lehrer seinen Unterricht vorbereitet, desto besser fallen die Klassenarbeiten aus.                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.) | Aus den katastrophalen Folgen atomarer Zerstörung lässt sich ableiten, dass kein Staat Atombomben einsetzen kann. | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.) | Es gibt außerirdische intelligente Wesen, die Kontakt zu den Menschen aufnehmen könnten.                          | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.) | Wenn Geld für gute Noten gegeben wird, dann erhöht sich die Lernfreude der Schüler.                               | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5.) | Wissenschaftliche Ergebnisse sprechen dagegen, Lernbehinderte besonders zu fördern.                               | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

|                     |                    |
|---------------------|--------------------|
| Aufgaben bestätigen | Korrekte Lösungen? |
|---------------------|--------------------|

Rückmeldung nach Aufgabenbestätigung : richtig = 1b 2c 3a 4b 5c

- Allgemeine Hypothesen sind falsifizierbar.
- Unbeschränkte nicht lokalisierende Existenzaussagen sind nur verifizierbar
- Ethisch, moralische oder ästhetische Behauptungen sind weder verifizierbar noch falsifizierbar.

### Aufgabe 1 [Klausur] richtig: 1b 2b 3a 4c 5c

Welche der nachfolgenden Sätze sind verifizierbar, welche falsifizierbar, welche sind im erfahrungswissenschaftlichen Sinne weder verifizierbar noch falsifizierbar?

|     |                                                                                          | a<br>verifizierbar    | b<br>falsifizierbar   | c<br>weder/noch       |
|-----|------------------------------------------------------------------------------------------|-----------------------|-----------------------|-----------------------|
| 1.) | Je intransparenter der Prüfungsstoff, desto schlechter fallen die Prüfungsergebnisse aus | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.) | Wenn die Noten abgeschafft werden, dann erhöht sich die intrinsische Lernmotivation.     | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.) | Es gibt Menschen mit einem IQ von 200.                                                   | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.) | Es wäre ein Segen für die Menschheit, wenn es keine Kriege mehr gäbe.                    | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 5.) | Je geringer das Wachstum, desto mehr muss der Staat in die Wirtschaft investieren.       | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |

|                     |
|---------------------|
| Aufgaben bestätigen |
|---------------------|

### Aufgabe 10 [Probeklausur]

|                     | Die Variable Angst kann bei den einzelnen Alternativen am ehesten aufgefasst werden als                                                       | a<br>UV<br>zuge-<br>wiesen | b<br>UV<br>aktiv      | c<br>AV               | d<br>SV               |
|---------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|----------------------------|-----------------------|-----------------------|-----------------------|
| 1.)                 | Der Forscher setzt in zufälligen Intervallen Elektroschocks ein, um die Auswirkungen der Angst auf das Wohlbefinden zu untersuchen.           | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.)                 | Bei der Überprüfung, ob strenge Benotung die Examensleistung verbessert, spielt möglicherweise die Angst der Kandidaten im Examen eine Rolle. | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.)                 | Ein Forscher überprüft, welche Auswirkungen die Schwierigkeit der Aufgaben auf die Angst der Schüler ausübt.                                  | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.)                 | Ein Sozialarbeiter will nachweisen, dass Jugendliche, die in stark Angst auslösenden Gegenden wohnen, aggressiv oder depressiv werden.        | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Aufgaben bestätigen |                                                                                                                                               | Korrekte Lösungen?         |                       |                       |                       |

Rückmeldung nach Aufgabenbestätigung : richtig 1b 2d 3c 4b

- Angst ist UV, wenn sie etwas verursachen soll.  
aktiv, wenn irgendwelche Angst manipulativ induziert wird,  
zugewiesen, wenn sie in der Person oder Umwelt der Person angelegt ist
- Angst die AV, wenn sie Effekt einer Maßnahme sein soll,
- Störvariable, wenn sie neben der UV unerwünscht auch die AV beeinflusst.

### Aufgabe 10 [Klausur] richtig 1b 2c 3d 1a

|                     | Die Variable Lärm kann bei den einzelnen Alternativen am ehesten aufgefasst werden als                                             | a<br>UV<br>zuge-<br>wiesen | b<br>UV<br>aktiv      | c<br>AV               | d<br>SV               |
|---------------------|------------------------------------------------------------------------------------------------------------------------------------|----------------------------|-----------------------|-----------------------|-----------------------|
| 1.)                 | Der Forscher setzt Staubsauger, Bohrmaschine und Discomusik ein, um die Auswirkungen von Lärm auf das Wohlbefinden zu untersuchen. | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 2.)                 | Ein Forscher überprüft, welche Auswirkungen bestimmte Spiele auf den Lärm von Kindern ausüben.                                     | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 3.)                 | Bei einer Versuchsreihe zur Wirkung von Entspannungsmusik auf die Konzentration hört man gelegentlich tief fliegende Düsenjets.    | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| 4.)                 | Ein Mediziner will nachweisen, dass Personen, die in stark lärmbelasteten Gebieten wohnen, mehr Herzbeschwerden aufweisen.         | <input type="radio"/>      | <input type="radio"/> | <input type="radio"/> | <input type="radio"/> |
| Aufgaben bestätigen |                                                                                                                                    |                            |                       |                       |                       |

### Aufgabe 34 [Probeklausur]

Es werden die Englisch -und Physikleistungen von Jungen und Mädchen mit Hilfe von Schulleistungstests erfasst. Bei welcher Alternative liegt eine Wechselwirkung zwischen Geschlecht und Schulfach vor ?

|                       | J   | M   |                       | J   | M   |                       | J   | M   |                       | J   | M   |
|-----------------------|-----|-----|-----------------------|-----|-----|-----------------------|-----|-----|-----------------------|-----|-----|
| Engl                  | 100 | 110 | Engl                  | 100 | 110 | Engl                  | 110 | 100 | Engl                  | 100 | 105 |
| Phys                  | 90  | 100 | Phys                  | 110 | 120 | Phys                  | 90  | 100 | Phys                  | 95  | 100 |
| <input type="radio"/> |     |     | <input type="radio"/> |     |     | <input type="radio"/> |     |     | <input type="radio"/> |     |     |
| Aufgabe bestätigen    |     |     | Korrekte Lösung?      |     |     |                       |     |     |                       |     |     |

Rückmeldung nach Aufgabenbestätigung : richtig = 3

Tipp: Wechselwirkung = keine parallele Linien bei einer Präsentation der Daten im Liniendiagramm.

**Aufgabe 34 [Klausur]** richtig=1

Es werden die Deutsch -und Matheleistungen von Jungen und Mädchen mit Hilfe von Schulleistungstests erfasst. Bei welcher Alternative liegt mit Sicherheit **keine** Wechselwirkung zwischen Geschlecht und Schulfach vor ?

|                       |     |     |                       |     |     |                       |     |     |                       |     |     |
|-----------------------|-----|-----|-----------------------|-----|-----|-----------------------|-----|-----|-----------------------|-----|-----|
| Deutsch               | J   | M   | Deutsch               | J   | M   | Deutsch               | J   | M   | Deutsch               | J   | M   |
| Math                  | 100 | 110 | Math                  | 100 | 110 | Math                  | 100 | 120 | Math                  | 100 | 110 |
|                       | 90  | 100 |                       | 110 | 100 |                       | 100 | 110 |                       | 110 | 110 |
| <input type="radio"/> |     |     | <input type="radio"/> |     |     | <input type="radio"/> |     |     | <input type="radio"/> |     |     |

Aufgabe bestätigen

**Aufgabe 47 [Probeklausur]**

In 4 Studien wurde untersucht, ob Methode A oder B zu höheren Erfolgen führt. Jede Studie erfasste die AV allerdings mit einer anderen Variablen. Nachfolgende Tabelle zeigt für A und B die Mittelwerte M, sowie die gemeinsame Streuung s. Wie hoch fällt jeweils die Effektstärke aus? **Hinweis:** Man kann auch einen Bruch eingeben, z.B. 80/120

| Nr. | Testverfahren      | A<br>M | B<br>M | Streu-<br>ung<br>s |                      |
|-----|--------------------|--------|--------|--------------------|----------------------|
| 1.  | korrekt in Prozent | 80     | 75     | 10                 | <input type="text"/> |
| 2.  | normierter Test    | 115    | 105    | 15                 | <input type="text"/> |
| 3.  | Noten              | 2.9    | 2.5    | 1                  | <input type="text"/> |
| 4.  | erzielte Punktzahl | 21     | 19     | 6                  | <input type="text"/> |

Aufgabe bestätigen

Korrekte Lösung?

Rückmeldung nach Aufgabenbestätigung : richtig 1= .5, 2= .66, 3= .4, 4= .33  
Effektstärke = (M1-M2)/s

**Aufgabe 47 [Klausur]** richtig: 1= .2, 2= .1, 3= .3, 4= .33

Bei Schülern der 10. Klassen in Deutschland und Österreich wurden verschiedene Daten erhoben. Nachfolgende Tabelle zeigt die Mittelwerte M, sowie die gemeinsame Streuung s. Wie hoch fällt jeweils die Effektstärke aus?

| Nr. | Testverfahren      | Deutsch-<br>land<br>M | Öster-<br>reich<br>M | Streu-<br>ung<br>s |                      |
|-----|--------------------|-----------------------|----------------------|--------------------|----------------------|
| 1.  | Pisa               | 515                   | 495                  | 100                | <input type="text"/> |
| 2.  | Schulleistungstest | 102                   | 101                  | 10                 | <input type="text"/> |
| 3.  | Noten              | 3.2                   | 2.9                  | 1                  | <input type="text"/> |
| 4.  | Intelligenz        | 107                   | 102                  | 15                 | <input type="text"/> |

Aufgabe bestätigen

## Anhang B

### Email an alle Studierenden der Kontrollgruppe

Hallo SeminarteilnehmerInnen der Gruppe

"eigenständige Vorbereitung auf das Abschlussquiz Diagnostik"

Das Abschlussquiz umfasst insgesamt 47 Aufgaben, welche die Lehrziele des Seminars ziemlich repräsentativ erfassen. Als beste Vorbereitungsvorlage dienen die Folien des Seminars. Dort finden Sie auch etliche Aufgaben, wie sie in etwa im Abschlussquiz vorkommen könnten. Als Aufgabentypen kommen vorwiegend Zuordnungsaufgaben, klassische MC-, sowie Serien von True/false Aufgaben zum Einsatz. Gelegentlich müssen Sie auch eine Short Answer-Aufgabe beantworten, die in der Regel eine Zahl oder ein Wort als Antwort verlangt.

Sie müssen keine Formeln auswendig lernen. Es wird aber erwartet, dass Sie einfache statistische Kennwerte (z.B. Mittelwert) berechnen können, wobei dies wegen der einfachen Zahlen in der Regel via Kopfrechnen möglich ist. Es wird allerdings verlangt, dass Sie einige Kennwerte auch richtig verstehen und so ohne Ausrechnen zumindest abschätzen können. Lernen Sie nicht stur auswendig, sondern bemühen Sie sich um Verständnis.

Die Seminarsitzung am Montag, den 25.1 ist dafür vorgesehen, Ihnen die Gelegenheit zu geben, die verfügbare Seminarzeit optimal für die eigene Vorbereitung des Abschlussquiz zu nutzen. Sie sollten daher im Vorfeld überlegen, welche Lernstrategie aus Ihrer Sicht für die Vorbereitung am effektivsten erscheint. Der Seminarraum steht zur Verfügung, um gemeinsam mit den Studierenden des Seminars bestimmte Themen des Seminars zu besprechen. Sie können die Zeit aber auch individuell außerhalb des Seminarraums nutzen, um sich alleine oder mit Kommilitonen vorzubereiten. Wer dem Seminar aufmerksam folgte und die Folien ernsthaft bearbeitete, hat hohe Chancen, das Abschlussquiz erfolgreich abzuschließen.

Hallo SeminarteilnehmerInnen der Gruppe

"eigenständige Vorbereitung auf das Abschlussquiz Methoden"

Das Abschlussquiz umfasst insgesamt 47 Aufgaben, welche die Lehrziele des Seminars ziemlich repräsentativ erfassen. Als beste Vorbereitungsvorlage dienen die im Seminar besprochenen Kapitel des Tutoriums. Eine Aufgabe bezieht sich auf die Folie "Effektstärke", die von der Homepage des Seminars aus erreichbar ist. Etliche Kapitel des Tutoriums sind nicht Gegenstand des Abschlussquiz, weil Sie nicht hinreichend im Seminar besprochen wurden, etwa: Statistische Validität, Konstruktvalidität, Kenntnis spezieller Varianten von Wechselwirkungen. In den Kapiteln des Tutoriums sind meistens Übungsaufgaben enthalten. Die Aufgaben im Abschlussquiz sehen ähnlich aus, sind teilweise aber kompakter in Form von Zuordnungsaufgaben konzipiert. In seltenen Fällen müssen Sie auch Short Answer-Aufgaben mit einem Wort, Symbol oder einer Zahl beantworten. Offene Fragen (Essay-Test-Aufgaben) kommen gar nicht vor.

Lernen Sie nicht stur auswendig, sondern bemühen Sie sich um Verständnis. Der überwiegende Teil der Fragen erfordert mindestens Verstehen. Die Seminarsitzung am Donnerstag, den 28.1 ist dafür vorgesehen, Ihnen die Gelegenheit zu geben, die verfügbare Seminarzeit optimal für die eigene Vorbereitung des Abschlussquiz zu nutzen. Sie sollten daher im Vorfeld überlegen, welche Lernstrategie aus Ihrer Sicht für die Vorbereitung am effektivsten erscheint. Der Seminarraum steht zur Verfügung, um gemeinsam mit den TeilnehmerInnen des Seminars bestimmte Themen des Seminars zu besprechen. Sie können die Zeit aber auch individuell außerhalb des Seminarraums nutzen, um sich alleine oder mit KommilitonInnen vorzubereiten.

Wer dem Seminar aufmerksam folgte, und die Kapitel des Tutoriums ernsthaft bearbeitete, hat sehr hohe Chancen, die Anforderungen des Abschlussquiz zu erfüllen.

## Anhang C

### Exkurs

#### Verbessert die Explizierung der Prüfungslehrziele das Ergebnis in einer Abschlussklausur?

Im Rahmen einer Studie zur Wirkung transparenzschaffender Maßnahmen zum Abbau von Angst vor Klassenarbeiten (Jacobs, Bedersdorfer & Bohse 1983) zeigte sich neben der beabsichtigten Angstreduktion als willkommener Zusatzeffekt von Transparenz auch eine signifikante Leistungsverbesserung in der Klassenarbeit. Als entscheidender Faktor Transparenzschaffender Maßnahmen wird hierbei die Explizierung der prüfungsrelevanten Lehrziele angesehen. Allerdings treten diese Effekte nur unter sehr günstigen Bedingungen zu Tage. In der Regel dürfte es nicht gelingen, auf solche Weise im Schulsystem Angst zu senken oder Leistung zu steigern (Jacobs 1987). Diese Erkenntnis hat den Verfasser dieses Exkurses allerdings nicht davon abgehalten, in etlichen Seminaren den Studierenden rechtzeitig vor der Klausur die Lehrziele der Abschlussklausur mitzuteilen. Sie sollten sich so auf die wichtigsten Themen des Seminars konzentrieren, diese verstärkt bearbeiten und so auch bessere Leistungen erbringen.

#### Methodisches Vorgehen

Der Verfasser leitete bereits seit geraumer Zeit ein Seminar mit dem Schwerpunkt "Grundlagen der Evaluation und pädagogischer Forschungsmethodik", im Folgenden **Methoden**, genannt. Ab dem WS 13/14 hat sich die Konzeption des Methodenseminars nicht mehr verändert. Die Unterrichtsinhalte wurden überwiegend vom Seminarleiter in Form eines lehrerzentrierten Unterrichts meist dozierend sowie Fragen entwickelnd unter Verwendung von Powerpointfolien vorgetragen und sollten von den Studierenden auf der Basis eines Internettutoriums mit ausführlichen Texten und etlichen interaktiven Elementen nachgearbeitet werden. Daneben waren die Studierenden angehalten, Arbeitsaufträge zu erledigen sowie ein Gruppenreferat- oder Projekt vorzutragen. Am Ende des Seminars stand eine kontrollierte Computerabschlussklausur an, welche die Lehrziele des Seminars ziemlich repräsentativ in Form von 47 Aufgaben unterschiedlichsten Aufgabenformates erfasste (siehe hier Anhang A, Seminar Methoden). Ab dem WS 2013/2014 wurden die Klausuraufgaben nicht mehr verändert. Um den Studierenden die Vorbereitung auf die Klausur zu erleichtern, stellte der Seminarleiter ca. eine Woche vor der Klausur auf der Homepage des Seminars eine Seite mit den prüfungsrelevanten Lehrzielen des Seminars dar und machte die Studierenden auf diese Seite aufmerksam.

#### Die prüfungsrelevanten Lehrziele

Zunächst erhielten die Studierenden die Information, dass in der Klausur ausschließlich die nachfolgenden Lehrziele geprüft werden und die Klausur somit aus 47 Aufgaben besteht. Es folgte dann eine nach Prüfungsthemen geordnete Aufstellung der Lehrziele für jede Klausuraufgabe.

### Abbildung 1: Die Lehrziele für die ersten 5 Aufgaben der Klausur.

#### Theorien und Hypothesen

1. Erkennen können, welche vorgegebenen Sätze verifizierbar, falsifizierbar oder weder verifizierbar noch falsifizierbar sind.
2. Erkennen können, welche der vorgegebenen Fragen empirische Wissenschaften nicht entscheiden können.
3. Aus der konkreten Beschreibung eines Gesetzes bzw. einer Hypothese entscheiden können, unter welchen Bedingungen diese widerlegt ist.
4. Den empirischen Gehalt von unterschiedlicher Hypothesen einschätzen können.
5. Auf Grund vorgegebener Information eine Erklärung begründen können.

Wie aus Abbildung 1 hervorgeht, beziehen sich die Lehrziele der Klausur vorwiegend auf Verstehen und Anwendung. Aus der Kenntnis des Lehrziels kann der Studierende zwar die Thematik sowie die verlangte Anforderung herauslesen, aber keineswegs die konkrete Aufgabe ableiten. [siehe dazu die Aufgabe 1 der Aufgabenbeispiele Methoden aus Anhang A]

#### Versuchsplan

Während die prüfungsrelevanten Lehrziele vom WS 13/14 bis SS 15 stets bekannt gegeben wurden, gab es im WS 15/16 aus experimentellen Gründen keine Explizierung der Lehrziele. In diesem Seminar sollte nämlich die Wirkung einer Probeklausur auf das Ergebnis der Abschlussklausur getestet werden. Deshalb wurden ca. 60 Studierende nach Abiturnotendurchschnitt parallelisiert und nach Zufall auf zwei Gruppen aufgeteilt. Im Folgenden wird nur noch diejenige Kontrollgruppe betrachtet, welche das Seminar bis auf das Fehlen der Lehrziele in gleicher Weise bearbeitete wie die früheren Seminare. Durch die Zufallszuweisung ist garantiert, dass diese verbliebene Kontrollgruppe von 29 Studierenden repräsentativ für das Methodenseminar im WS 15/16 ist.

Das Design folgt einer Variante des klassischen vorexperimentellen Gruppenversuchs mit nicht zwingend äquivalenten Gruppen (N= nonequivalent) und kann wie folgt gedeutet werden:

### Abbildung 2: Versuchsplanformalisierung.

|   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| N | X | ○ |   |   |   |   |
| N |   |   | X | ○ |   |   |
| N |   |   |   | X | ○ |   |
| N |   |   |   |   | X |   |
| N |   |   |   |   |   |   |
| N |   |   |   |   |   |   |
|   |   |   |   |   |   | ○ |

X= Explizierung der prüfungsrelevanten Lehrziele der Klausur.

○= Klausur

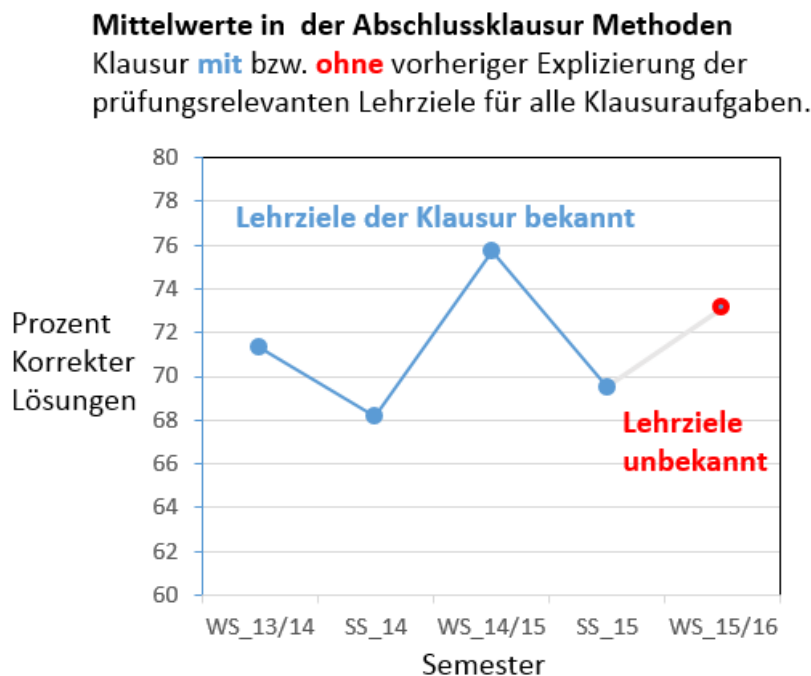
In den Seminaren der Semester WS 13/14 bis SS 15 hatten die Studierenden vor der Abschlussklausur Zugriff auf die prüfungsrelevanten Lehrziele der Klausur (X). Im WS 15/16 fehlte X, womit die Kontrollgruppe dieses Semester auch als Kontrolle für den Versuchsplan gilt. Der Versuchsplan entspricht einer retrospektiven Analyse, weil für die aktuelle Kontrollgruppe erst im Nachhinein die EG-Seminare aus den früheren Zeiten gesucht wurden. Damit rückt er schon fast in die Nähe eines nicht experimentellen Versuchsplans.

Es gibt zwar wenige plausible Anhaltspunkte dafür, dass sich die Studierenden in den einzelnen Semestern bedeutsam voneinander unterscheiden oder die Seminare im Verlauf der Zeit deutliche Unterschiede aufweisen, aber die Äquivalenz der Probanden und Seminare ist natürlich nicht so stringent einzuschätzen wie bei einer Zufallszuweisung innerhalb eines Seminars. Dieser in der pädagogischen Interventionsforschung öfter angewandte Versuchsplan hat durchaus sein Tücken. Jacobs (2009, siehe Kopie im Anhang D) konnte schon mal eine offensichtliche Fehldeutung aus einem solchen Plan eindrucksvoll nachweisen.

## Ergebnisse

Eine Inspektion der Klausurergebnisse im Zeitverlauf der Semester deutet relativ klar in Richtung Unwirksamkeit der Lehrzielexplikation im Hinblick auf verbesserte Klausurleistungen.

### Abbildung 3



Wie aus der Abbildung 3 hervorgeht, schwanken die Klausurergebnisse bis SS 15 um den Wert 71% korrekter Antworten herum. Das rot markierte Klausurergebnis im WS 15/16 fügt sich zwanglos in den Trend bzw. die Zufallsschwankung der früheren Klausurergebnisse ein. Fasst man die Ergebnisse aller 4 Klausuren mit vorheriger Lehrzielexplikation zusammen, so erzielten die entsprechenden Studierenden im gewichteten Mittelwert 71,25% korrekter Lösungen. Dieser Gesamtmittelwert unter Lehrzielexplikation fällt geringer aus als der Mittelwert der Kontrollgruppe im WS 15/16 (siehe dazu auch Tabelle 2 auf Seite 7). Im WS 15/16 gab es keine Liste der prüfungsrelevanten Lehrziele, aber einen Klausurmittelwert der Kontrollgruppe von 73,1% korrekter Lösungen. Auf der Basis des vorliegenden Designs muss man eindeutig zu dem Ergebnis kommen, die Explizierung der prüfungsrelevanten Lehrziele habe das Ergebnis in der Methodenklausur in keiner nennenswerten Weise beeinflusst. Die Studierenden, welche im WS 15/16 jedoch an der Probeklausur teilgenommen hatten, erzielten mit durchschnittlich 77.7% korrekter Lösungen zumindest ein signifikant besseres Ergebnis im Vergleich zum Gesamtmittelwert unter Lehrzielexplikation, was darauf hindeutet, die Probeklausur habe im gegebenen Fall auch mehr Leistung bewirkt als die Mitteilung der prüfungsrelevanten Lehrziele.

## Diskussion

Die hier erkennbare Unwirksamkeit der Mitteilung von prüfungsrelevanten Lehrzielen dürfte zu einem großen Teil der speziellen Seminarkonzeption angelastet werden. Sie kann keineswegs auf alle Seminare oder Vorlesungen generalisiert werden. Insbesondere in Vorlesungen oder Seminaren mit hoher Stofffülle, umfangreichen, notwendig zu erarbeitenden, Literaturquellen, gering strukturiertem Lehrmaterial, diffusen Lehrzielen und/oder Klausuren geringer Kontentvalidität (= keine repräsentative Aufgabenstichprobe aus dem Universum aller möglichen Lehrziele), rechne ich bei Transparenz mit signifikant besseren Leistungsergebnissen.

Hier waren in den Folien des Seminarleiters öfter Aufgabenstellungen eingebunden, welche die Studierenden im Seminar beantworten sollten. Nicht selten führte der Seminarleiter die Aufgabe mit dem Statement ein. "So könnte eine Aufgabe in der Abschlussklausur aussehen". Nach der Aufgabenbearbeitung wurden die korrekten Lösungen dargeboten und begründet. Als einzige Literaturquelle diente das Internettutorium, das als Selbstlernprogramm konzipiert war und zu allen Texten viele Aufgaben mit elaborierten Rückmeldungen zur freiwilligen Nutzung anbot. Die Klausur selbst umfasste ziemlich repräsentativ alle im Seminar behandelten Themen. Auch Studierende ohne die Liste der prüfungsrelevanten Lehrziele konnten mit relativ hoher Transparenz antizipieren, welche Aufgabeninhalte geprüft werden und wie Aufgaben in der Klausur in etwa konzipiert sein würden.

Der objektive Vorteil einer klausurrelevanten Lehrzielliste liegt vornehmlich darin, den Umfang der möglichen Aufgaben einzugrenzen. Was nicht als Lehrziel definiert war, konnte in der Klausur auch nicht geprüft werden. Insofern bleibt zu vermuten, die Studierenden hätten sich mit der Lehrzielliste effizienter, d.h. weniger Zeit raubend, vorbereiten können. Man kann diesen Vorteil auch als pädagogisch unerwünscht kritisieren, weil so der verbindliche Lehrstoff reduziert wird. Der interessierte Leser sei in diesem Zusammenhang auf meinen Beitrag: [Einige Thesen pro und Contra Transparenz](#) verwiesen.

## Literatur:

Jacobs, B., Bedersdorfer H.-W., Bohse-Wagner, N. (1983). Angstabbau durch Transparenz - Eine quasiexperimentelle Felduntersuchung zur ökologischen Validierung eines theorieorientierten Interventionsprogramms zur Reduktion von Angst in der Prüfung. Arbeitsbericht Nr. 18 aus der Fachrichtung Allgemeine Erziehungswissenschaft. Universität des Saarlandes. Kurzbericht dazu: Studie 2:  
Angstabbau durch Transparenz - Ein theoriegeleitetes Quasiexperiment  
<http://psydok.sulb.uni-saarland.de/volltexte/2004/131/html/studie2.html>

Jacobs, B. (1987). Die Auswirkungen transparenzschaffender Maßnahmen auf die aktuelle Angst vor einer Klassenarbeit - Eine Metaanalyse zum Saarbrücker Schulangstprojekt. Empirische Pädagogik 1(2). 139-160. Kurzbericht dazu [Studie 3](#):

Jacobs, B. (2009). Vergleich von Bedingung A im aktuellen mit Bedingung B in einem früheren Jahrgang.  
[http://www.phil.uni-sb.de/~jakobs/seminar/vpl/experiment/exkurs/vorexperiment\\_variante.htm](http://www.phil.uni-sb.de/~jakobs/seminar/vpl/experiment/exkurs/vorexperiment_variante.htm)  
hier als Kopie im Anhang D

## Anhang D

Hinweis: Dies ist eine Kopie nachfolgender HTML-Seite aus einem verschollenen Server  
[http://www.phil.uni-sb.de/~jakobs/seminar/vpl/experiment/exkurs/vorexperiment\\_variante.htm](http://www.phil.uni-sb.de/~jakobs/seminar/vpl/experiment/exkurs/vorexperiment_variante.htm)

|                                                                           |                         |
|---------------------------------------------------------------------------|-------------------------|
| <a href="#">Überblick   Vorexperimentelle Versuchspläne</a>               | Jacobs: Versuchsplanung |
| <b>Exkurs:</b> kein notwendiger Bestandteil des Tutoriums Versuchsplanung |                         |

Vergleich von Bedingung A im aktuellen mit Bedingung B in einem früheren Jahrgang

In der angewandten pädagogischen Forschung wird gelegentlich auf die Ergebnisse aus früheren Jahrgängen zurückgegriffen, um die Wirkung einer neuen Maßnahme beim aktuellen Jahrgang zu bewerten. Eine typische Fragestellung lautet etwa. "Bewirken Onlinetests im Rahmen einer Lehrveranstaltung bessere Ergebnisse im Abschlussexamen?" Das zugrunde liegende Design kann als eine spezielle erweiterte Variante des statistischen Gruppenvergleichs ohne Vortest gedeutet und beispielsweise wie in Tabelle 1 formalisiert werden:

**Tabelle 1:** Mögliche Formalisierung eines erweiterten vorexperimentellen Gruppenvergleichs ohne Vortest

|   |       |         |
|---|-------|---------|
| N | $X_k$ | ○       |
| N |       | $X_e$ ○ |

Im aktuellen Semester werden Onlinetests (=experimentelles Treatment  $X_e$ ) zur Unterstützung der Lehre eingesetzt. Im gleichen Seminar des Vorjahres gab es keine Onlinetests (= Kontrollbedingung  $X_k$ ), aber der Unterricht, die Lehrmaterialien und das Abschlussexamen waren hoch vergleichbar. Gegenüber dem klassischen statistischen Gruppenvergleich ohne Vortest sind die Bedingungen der beiden Gruppen hier zeitlich versetzt. Das Ganze ähnelt einer Zeitwandelanalyse mit veränderten Bedingungen zu den verschiedenen Zeiten. Kommt der Versuch von Anfang an gut geplant und sorgfältig kontrolliert zum Einsatz, so lassen sich die theorierelevanten Bedingungen weitgehend standardisieren und bestimmte Kontrollvariablen erheben. Außerdem könnte der Versuchsleiter nach Zufall bestimmen, welcher Jahrgang die Experimental - und welcher die Kontrollbedingung erhält. Von einer eher retrospektiven Analyse würde man hingegen sprechen, wenn erst im Nachhinein Seminare aus früheren Zeiten gesucht werden, für die entsprechende Daten, etwa Abschlussklausurergebnisse, verfügbar sind, die man dann als Vergleich heranziehen könnte. Potenzielle Störfaktoren sind so natürlich schwer oder gar nicht zu kontrollieren. Signifikant bessere Abschlussexamen für die Studierenden unter der experimentellen Bedingung werden - voreilig - den Onlinetests zugeschrieben, nicht signifikante Unterschiede -voreilig- als Bestätigung der Ineffizienz von Onlinetests interpretiert. Die Güte des Versuchs und damit seine Interpretierbarkeit hängen aber ganz entscheidend von der Vergleichbarkeit der Studierenden und der Ähnlichkeit derjenigen Bedingungen ab, die außer der UV die AV ebenfalls beeinflussen könnten. Die wichtigsten Störvariablen bei lernrelevanten Studien, bisherige Studierleistungen, Abiturnotendurchschnitt oder kognitive Vergleichstest sollten auf jeden Fall zusätzlich erhoben werden, um sich ein realistisches Bild der Vergleichbarkeit beider Probandengruppen machen zu können.

## Praktisches Beispiel für einen Vergleich der Methode mit einem experimentellen Versuchsplan

Als Probanden dienten Studierende des Lehramts, die im WS 08 oder im WS 09 an einem Seminar zur Pädagogischen Diagnostik teilnahmen. Fast alle Studierenden absolvierten den gleichen Studiengang (Bachelor). Alle Studierenden wurden vom selben Dozenten, dem Verfasser dieses Textes, in der gleichen Form unterrichtet, erhielten die gleichen Lehr/Lernmaterialien und bearbeiteten dasselbe Quiz.

### Hypothese: Die Erwartung einer Quizbenotung führt zu besseren Quizergebnissen.

$X_1$ : Studierende erwarten ein benotetes Quiz

$X_2$ : Studierende erwarten ein unbenotetes Quiz

O: Prozentsatz korrekter Lösungen im Quiz

Erwartung: Das Quizergebnis O unter  $X_1$  fällt besser aus als das unter  $X_2$

### Tabelle 2: spezieller erweiterter statistischer Gruppenvergleich ohne Vortest (vorexperimentell)

|       |              |   |       |         |
|-------|--------------|---|-------|---------|
| WS 08 | Quiz Skalen: | N | $X_1$ | O       |
| WS 09 | Quiz Skalen: | N |       | $X_2$ O |

Dieselbe Hypothese wurde allerdings zuvor experimentell untersucht (Jacobs 2009a):

### Tabelle 3: Experimenteller Versuchsplan ohne Vortest

|       |              |   |       |   |
|-------|--------------|---|-------|---|
| WS 08 | Quiz Skalen: | R | $X_1$ | O |
| WS 08 | Quiz Skalen: | R | $X_2$ | O |

## Ergebnisse

### Tabelle 4a: Prozentsatz der korrekten Lösungen im Quiz Skalen

|       |                       | M    | s    | N  |
|-------|-----------------------|------|------|----|
| WS 08 | Quiz Skalen benotet   | 70,0 | 15,1 | 29 |
| WS 09 | Quiz Skalen unbenotet | 67,1 | 16,1 | 30 |
| <hr/> |                       |      |      |    |
| WS 08 | Quiz Skalen unbenotet | 54,9 | 23,4 | 29 |

Wie man aus Tabelle 4a erkennt, wäre man unter Anwendung des vorexperimentellen Versuchsplans (Tabelle 2) zu der Schlussfolgerung gelangt, die Erwartung einer Quizbenotung übe keinen Einfluss auf die Quizleistung aus, da die Studierenden unter Notenerwartung im WS 08 mit 70% korrekter Lösungen nicht besser abschnitten als die Studierenden im WS 09 ohne Notenerwartung mit 67,1 % korrekter Lösungen. ( $t(57) = 0.76$  p einseitig = 0.23, ns). Der vorexperimentelle Versuchsplan geht ungeprüft von einer Vergleichbarkeit der Jahrgänge unter der Kontrollbedingung aus. Ihm bleibt verborgen, dass die Studierenden des WS 08 gegenüber den Studierenden des WS 09 im unbenoteten Quiz signifikant schwächere Ergebnisse erzielten. ( $t(57) = -2,24$ ; p zweiseitig = 0.023).

Der experimentelle Versuchsplan in Tabelle 3 [mit strenger Randomisierung innerhalb der Seminare im WS 08] ergab hingegen eindeutige Bestätigungen für die Hypothese höherer Quiz-

leistungen unter Notenerwartung. Beide Gruppen des WS 08 unterscheiden sich hochsignifikant und der Leistungsvorteil der Gruppe unter Notenerwartung entspricht einer Effektstärke von  $d = .77$ . Weitere Überprüfungen auf der Basis strenger Versuchspläne ergaben konsistent eindeutige Bestätigungen für höhere Quizleistungen unter Notenerwartung. (Jacobs 2009a), die insgesamt einer hohen praktischen Bedeutsamkeit entsprechen. Die Vergleichbarkeit der Studierenden und der Seminardurchführung für beide Bedingungen ist im experimentellen Plan des WS 08 durch die nachgeprüft effiziente Randomisierung hergestellt worden, weswegen der experimentelle Versuchsplan in Tabelle 3 intern valide Ergebnisse erbringt.

Die Ergebnisse beim Quiz Skalen waren offenbar kein Ausrutscher. Denn der Vergleich im Quiz Statistik zeigt Ergebnisse, welche in die gleiche Richtung weisen, wie die Tabelle 4a aufzeigt.

**Tabelle 4b:** Prozentsatz der korrekten Lösungen im Quiz Statistik

|       |                          | M    | s    | N  |
|-------|--------------------------|------|------|----|
| WS 08 | Quiz Statistik benotet   | 74,9 | 12,6 | 29 |
| WS 09 | Quiz Statistik unbenotet | 72,3 | 12,9 | 30 |
| <hr/> |                          |      |      |    |
| WS 08 | Quiz Statistik unbenotet | 65,7 | 18,2 | 25 |

Ähnliche Ergebnisdivergenzen in der Anwendung der beiden Versuchsplanarten ließen sich hinsichtlich der Erwartung aktueller Angst unmittelbar vor dem Quiz ermitteln. Die theoretisch gut begründete Hypothese, benotete Quiz lösten mehr aktuelle Angst aus als unbenotete Quiz, ließ sich Tabelle 5 zufolge im vorexperimentellen Versuchsplan augenscheinlich nicht, im experimentellen Randomisierungsplan hingegen eindeutig und darüber hinaus mehrfach bestätigen.

**Tabelle 5:** [Situative aktuelle Prüfungsangst](#) unmittelbar vor dem Quiz (N pro Gruppe 25-29)

|       |                       | M    | s    |
|-------|-----------------------|------|------|
| WS 08 | Quiz Skalen benotet   | 25,6 | 10,7 |
| WS 09 | Quiz Skalen unbenotet | 21,9 | 10,5 |
| <hr/> |                       |      |      |
| WS 08 | Quiz Skalen unbenotet | 17,2 | 8,2  |

Lediglich bei einigen Variablen, die ganz deutliche Unterschiede im experimentellen Versuchsplan erbrachten [ $d > 1$ ], liefert auch der vorexperimentelle Plan die erwarteten Ergebnisse. So fällt etwa die Qualität und Quantität der Quizvorbereitung im benoteten Quiz des WS 08 immer noch signifikant und deutlich höher aus als die vor dem unbenoteten Quiz im WS 09.

Den Ursachen dieser Versuchsplan bedingten Ergebnisdivergenzen kann an dieser Stelle nicht vertiefend nachgegangen werden. Selbstselektion der Studierenden kann sich kaum auswirken, da das Verteilungsverfahren für die Seminarplätze gleich geblieben ist. Beide Gruppen unter der Bedingung unbenotetes Quiz unterscheiden sich nicht im Abiturnotendurchschnitt ( $t(53) = 1,4$ ;  $p$  zweiseitig = 0.16). Die Seminargestaltung der Jahrgänge weist sehr hohe Ähnlichkeit auf und lief im Hinblick auf die Vermittlung der quizrelevanten Lehrziele nahezu identisch ab. Lediglich die Seminaranforderungen waren 2009 insgesamt etwas geringer, weil der benotete Leistungsnachweis für dieses Seminar abgeschafft wurde.

## Resümee

Meiner Meinung nach hat die Analyse im gegebenen Beispiel die Schwäche des speziellen vor-experimentellen Versuchsplans eindrucksvoll belegt. Das bedeutet freilich nicht, schwache Versuchspläne würden vornehmlich theoriwidrige Ergebnisse produzieren. Ihre eigentliche Gefahr liegt eher im Bereich der Scheinbewährung, d.h. im Vortäuschen von Hypothesenbestätigungen und der Erzeugung überhöhter Effektstärken. Schließlich liefern sie manchmal zum Teil aber auch Daten der zutreffenden Art. So führte etwa bei einem vermutlich noch schwächeren Versuchsplan eine Überprüfung der Hypothese, die Erwartung einer Klausurbenotung müsse bessere Klausurergebnisse nach sich ziehen als die Erwartung einer unbenoteten Klausur, zu einem deutlich theoriekonformen Ergebnis in erwarteter Effektstärkengröße (Jacobs 2009b). In diesem Versuch legten die Studierenden lange vor der für alle verpflichtenden Abschlussklausur selbst fest, ob sie einen benoteten oder unbenoteten Schein erwerben wollten (= Selbstselektion zu den experimentellen Bedingungen). Die Erwartung höherer Angst in der benoteten Klausur konnte dort hingegen statistisch auch nicht bestätigt werden.

Das Problem schwacher Versuchspläne besteht weniger darin, möglicherweise theoriwidrige Ergebnisse zu produzieren, sondern Daten zu liefern, die unabhängig von ihrem Ergebnis Zweifel an ihrer Interpretierbarkeit aufkommen lassen. Deshalb müssen bei solchen Versuchsplänen die Experimentalphasen sorgfältig beobachtet und kontrolliert werden, begründete Alternativhypothesen zur UV argumentativ ausgeschlossen werden, sowie theorierelevante Kontrolldaten erhoben werden, was je nach Untersuchung mehr oder weniger gut gelingen kann.

---

## Literatur:

- Jacobs, B. (2009a). Leistungssteigerung durch Notendruck ? - Die Wirkung der Benotung auf die Studierleistungen in einem Seminar.  
[http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische\\_motivierung/notendruck/notenanreiz.html](http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/notendruck/notenanreiz.html)
- Jacobs, B. (2009b). Die Wirkung der Klausurbenotung auf Klausurleistung, Angst und Lernmotivation. - Ein Replikationsversuch auf der Basis eines schwachen Versuchsplans  
[http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische\\_motivierung/notendruck/notenanreiz\\_selbstselektion.html](http://www.phil.uni-sb.de/~jakobs/wwwartikel/extrinsische_motivierung/notendruck/notenanreiz_selbstselektion.html)

---

created 25.11.2009; last update 16.12. 2009; Bernhard Jacobs

**Hinweis:** Eine Kurzfassung von Jacobs (2009a) findet der interessierte Leser unter:

Jacobs, B. (2009). Leistungssteigerung durch Notendruck? - Die Wirkung der Benotung auf die Studierleistungen in einem Seminar.

URN: urn:nbn:de:bsz:291-psydok-25299

URL: <http://psydok.sulb.uni-saarland.de/volltexte/2009/2529/>