

**Robert KORDTs-FREUDINGER¹, Katrin B. KLINGSIECK
(Paderborn) & Eva SEIFRIED (Heidelberg)**

Beforschung einer eigenen Vorlesung: Effekte von Wahlfreiheit auf Leistung und Konfidenz

Zusammenfassung

Basierend auf Forschung zum Lernen in Vorlesungen und der Selbstbestimmungstheorie wurde der Effekt von Wahlfreiheit auf kognitive Testleistung und metakognitive Konfidenz im Rahmen einer eigenen Vorlesung untersucht. $N = 370$ Studierende bearbeiteten semesterbegleitend selbst gewählte sowie vorgegebene schriftliche Reflexionen und absolvierten zu zwei Messzeitpunkten Tests mit konfidenzgewichteten Richtig-Falsch-Items. Testitems zu Inhalten der selbstgewählten Reflexionen wurden zu beiden Messzeitpunkten gleich gut bearbeitet (Leistung; kognitives Maß), während die Sicherheit in korrekte und inkorrekte Lösungen abnahm (Konfidenz; metakognitives Maß). Die Leistung bei Testitems zu Inhalten der vorgegebenen Reflexionen nahm zum zweiten Messzeitpunkt ab, während die Konfidenz in inkorrekte Lösungen zunahm. Mögliche Mechanismen und Konsequenzen der Studienergebnisse, aber auch Grenzen der Studie und somit mögliche Implikationen werden diskutiert.

Schlüsselwörter

Vorlesung, metakognitive Konfidenzurteile, Wahlfreiheit, Forschendes Lehren, Scholarship of Teaching and Learning

¹ E-Mail: robert.kordts.freudinger@upb.de



Researching your own lecture: Effects of choice on performance and confidence

Abstract

Based on research on learning in lectures and on self-determination theory, we investigated the effects of autonomy on cognitive test performance and metacognitive confidence as part of a Scholarship of Teaching and Learning project. During the semester, $N = 370$ university students completed both self-selected and designated written reflections and took tests with confidence-weighted true-false items on two measurement occasions. Items related to content from the self-selected tasks were solved equally well over time, while confidence in correct and incorrect solutions decreased. Performance related to content from the designated tasks increased over time, while confidence in incorrect solutions increased. This paper discusses the possible mechanisms and consequences of the results, as well as the limitations of our study and, thus, the possible implications.

Keywords

Lecture, metacognitive confidence, choice, Scholarship of Teaching and Learning

1 Einführung

Der Artikel beschreibt einen Teil eines Lehrforschungsprojekts zweier Junior-Professor/innen (Erstautor und Zweitautorin), die in ihrer ersten gemeinsamen Vorlesung mit Unterstützung der Ko-Autorin Maßnahmen zur Erhöhung des Lernerfolgs der Studierenden untersuchten.

2 Hintergrund

2.1 Forschendes Lehren

Wenn Hochschullehrende ihre eigene Lehre beforschen und die Ergebnisse publizieren, fällt dies unter das Konzept des Scholarship of Teaching and Learning (BOYER, 1990; HUBER, 2011). Das Konzept des Forschenden Lehrens (SPINATH & SEIFRIED, 2012) ist im Grundsatz ähnlich, allerdings explizit auf die kontinuierliche, iterative Weiterentwicklung einer Lehrveranstaltung fokussiert. Anhand eines Beispiels legen SPINATH & SEIFRIED (2012) den Prozess dar, der von der systematischen Aufarbeitung der Theorie und Empirie über die Generierung empirisch überprüfbarer Hypothesen bis hin zur Implementation der als effektiv gezeigten Maßnahmen sowie der Weiterentwicklung von Theorie und Empirie reicht.

Bei unserem Einstieg in das Forschende Lehren stellte sich uns die Frage nach der Effektivität des Formats Vorlesung (vgl. BLIGH, 2000), vor allem nach dem Verhältnis zwischen kognitiven und metakognitiven Prozessen. In diesem Beitrag konzentrieren wir uns auf die Zunahme des Wissens sowie die Einschätzung des eigenen Wissens. Angesichts des in Vorlesungen üblichen Fehlens des Feedbacks an Studierende zu ihrem Lernprozess interessierte uns insbesondere das metakognitive Monitoring des eigenen Wissens in unserer Veranstaltung.

2.2 Lernerfolg und Metakognition in Vorlesungen

Wenn Studierende in Vorlesungen Wissen hinzugewonnen haben, kann sich auch die Einschätzung des eigenen Wissensstandes, also das metakognitive Wissen, verbessern (vgl. KRUGER & DUNNING, 1999), was wiederum positive Effekte auf den Lernerfolg haben kann (vgl. BANNERT, 2003). Zur gleichzeitigen Erfassung der kognitiven Gedächtnisleistung und des metakognitiven Monitorings schlagen DUTKE & BARENBERG (2009, 2015) einen Leistungstest mit konfidenzgewichteten Richtig-Falsch-Items vor. Dabei wird die kognitive Leistung als

Korrektheit der Einschätzung inhaltsbezogener Aussagesätze (*Urteilkorrektheit*) erfasst. Gleichzeitig geben die getesteten Personen an, wie sicher sie sich in ihrem Urteil fühlen: Mit einem Fragezeichen gekennzeichnete Antwortoptionen (Richtig?, Falsch?) bedeuten eine niedrige Konfidenz in die eigene Antwort, mit einem Ausrufezeichen gekennzeichnete Optionen (Richtig!, Falsch!) bedeuten hohe Konfidenz (*Urteilkonfidenz*). DUTKE & BARENBERG (2009) zeigten, dass mit dieser ökonomisch einsetzbaren Methode sowohl Veränderungen des Wissensstandes als auch die Genauigkeit der metakognitiven Konfidenzurteile erfasst werden können. Mit diesem Testformat konnten auch Unterschiede zwischen benoteten und unbenoteten Tests nachgewiesen werden (BARENBERG & DUTKE, 2013): Bei benoteten Tests lagen höhere kognitive Testleistungen und höhere Konfidenzschätzungen vor, allerdings zeigten die Studierenden höhere Konfidenz sowohl in ihre korrekten als auch in ihre inkorrekten Antworten. Dieses Ergebnis könnte durch die Qualität der Motivation bei der Vorbereitung mitverursacht sein: Eine Benotung führt vermutlich v. a. zu einer Steigerung der extrinsischen Motivation oder kontrollierten Regulation.

2.3 Einfluss autonomer und kontrollierter Regulation

Nach der Selbstbestimmungstheorie (vgl. RYAN & DECI, 2000) gibt es qualitative Unterschiede zwischen der autonomen und der kontrollierten Regulation. Als kontrolliert reguliert erleben sich Menschen, wenn sie handeln, um bestimmte Zwecke zu erreichen, und wenn diese Zwecke als von außen determiniert erlebt werden. Autonome Regulation besteht, wenn Menschen entweder von außen kommende Ziele in ihr eigenes Zielsystem integriert oder sich mit diesen Zielen identifiziert haben oder wenn sie intrinsisch motiviert sind – d. h. wenn der Antrieb zum Handeln in der Ausführung des Handelns selbst liegt (vgl. RYAN & DECI, 2000).

Vielfältige Studien haben gezeigt, dass autonome Regulation oder intrinsische Motivation positive Effekte auf akademisches Lernen hat (vgl. GROLNICK & RYAN, 1987; VANSTEENKISTE, LENS & DECI, 2006). Einige Studien haben ferner auch positive Effekte der autonomen Regulation auf das metakognitive Monitoring gezeigt (vgl. LEVESQUE, ZUEHLKE, STANEK & RYAN, 2004). So liegt es

nahe, dass die Qualität der Regulation sich auch auf metakognitive Konfidenzurteile auswirken kann.

In einer Studie von BARENBERG, SEIFRIED, SPINATH & DUTKE (2016) wurde der Effekt der Bearbeitung problembasierter Schreibaufgaben auf Klausurleistungen mit dem oben beschriebenen Testformat untersucht. Studierende bearbeiteten schriftlich zu randomisiert zugewiesenen Vorlesungssitzungen Problemaufgaben, die zu einer tieferen Verarbeitung des Stoffes führen sollten. Ähnlich wie bei BARENBERG & DUTKE (2013) ergab sich für die Problemaufgabenbedingung nicht nur eine höhere kognitive Testleistung, sondern auch höhere Konfidenzschätzungen, allerdings ebenfalls sowohl in korrekte (zutreffende Aussagen korrekt als zutreffend und nicht zutreffende Aussagen korrekt als nicht zutreffend markiert) als auch in inkorrekte Lösungen (nicht zutreffende Aussagen fälschlicherweise als zutreffend bzw. zutreffende Aussagen fälschlicherweise als nicht zutreffend markiert), d. h., die *Urteilssensitivität* veränderte sich nicht. Dieser Effekt könnte mit hoher fremdkontrollierter Regulation erklärt werden, da für die Studierenden keine Wahlfreiheit bestand und damit diesbezüglich eine niedrige autonome Regulation zu vermuten ist (vgl. ASSOR, KAPLAN & ROTH, 2002). Allerdings könnten problembasierte Aufgabenformate grundsätzlich auch autonome Regulation erhöhen (vgl. HMELO-SILVER, 2006). Es lässt sich daher vermuten, dass die Problemaufgaben bei BARENBERG et al. (2016) sowohl autonome als auch kontrollierte Stufen der Regulation angesprochen haben.

3 Fragestellungen

Das aktuelle Lehrforschungsprojekt untersucht daher den Einfluss von Wahlfreiheit bei der Bearbeitung von Aufgaben auf die kognitive Leistung und metakognitive Konfidenzurteile anhand folgender Fragestellungen.

- 1: Welche Effekte hat Wahlfreiheit auf die *Urteilskorrektheit* bei konfidenzgewichteten Richtig-Falsch-Aufgaben in einer Vorlesung?
- 2: Welche Effekte hat Wahlfreiheit auf die *Urteilskonfidenz*?
- 3: Welche Effekte hat Wahlfreiheit auf die *Urteilssensitivität*?

4 Methode

Das Projekt wurde im Wintersemester 2015-16 im Rahmen einer Vorlesung „Entwicklung und Lernen“ in den Studiengängen Master of Education Lehramt für Gymnasien und Gesamtschulen sowie Lehramt für Berufskolleg an der Universität Paderborn durchgeführt. Die Vorlesung bestand aus 13 wöchentlichen Sitzungen zu Themen der Pädagogischen Psychologie. Für insgesamt sechs Sitzungen mussten die Studierenden in kurzen Reflexionstexten die wichtigsten Punkte der Sitzung zusammenfassen.

4.1 Design und Variablen

Die Studie folgte einem Design mit der unabhängigen Variablen (uV) Wahlfreiheit (selbstgewählte vs. vorgegebene Aufgaben). Um die vor den Reflexionen vorhandenen Wissens- und Konfidenzbestände zu berücksichtigen, gab es zwei Messzeitpunkte; t1 lag in der ersten Sitzung der Vorlesung, t2 in der letzten Sitzung. Die uV Wahlfreiheit wurde über die Vorgabe der mittels Reflexion zu bearbeitenden Sitzung operationalisiert: Drei der Sitzungen wurden für alle Studierenden zur Reflexion vorgegeben; zu welchen drei weiteren Sitzungen sie eine Reflexion schreiben wollten, wurde durch die Studierenden selbst gewählt. Die drei vorgegebenen Sitzungen waren dabei über das Semester hinweg verteilt.

Als abhängige Variablen (aV) wurden zum einen die *Urteilkorrektheit* als die studentische Leistung in den Befragungen zum Wissensstand sowie zum anderen die *Urteilkonfidenz* der Studierenden in ihre Urteile in diesen Tests erfasst. Schließlich wurde die *Urteilssensitivität* durch das Verhältnis der als konfident eingeschätzten korrekten Antworten zu den als konfident eingeschätzten, aber inkorrekten Antworten geschätzt. Die Befragungen, die absolviert werden mussten, deren Ergebnis aber nicht ausschlaggebend für die Studienleistung in der Vorlesung war, wurden nach dem von DUTKE & BARENBERG (2009, 2015) vorgeschlagenen Format mit konfidenzgewichteten Richtig-Falsch-Items von den Lehrenden der Vorlesung konstruiert (Beispiel: „Ein anderes Wort für Bestrafung ist ‚negative Verstärkung‘“). Für t1 wurden insgesamt 13 Items (pro Vorlesungssit-

zung ein Item; alle als falsch zu beurteilen) und für t2 zusätzlich weitere 55 Items (pro Sitzung zwischen vier und fünf zusätzliche Items, insgesamt 68 Items, davon 44 als falsch zu beurteilen) verwendet.

4.2 Durchführung

Für die zu bearbeitenden Reflexionsaufgaben beantworteten die Studierenden schriftlich auf max. 1,5 Seiten pro Sitzung drei Fragen. Um sowohl kognitive als auch metakognitive Prozesse durch die Bearbeitung anzuregen (vgl. BERTHOLD, NÜCKLES & RENKL, 2007; HÜBNER, NÜCKLES & RENKL, 2010), waren jeweils zwei dieser Fragen kognitiv und jeweils eine Frage metakognitiv ausgerichtet: „Was ist für Sie, in eigenen Worten, das Wichtigste aus der Vorlesung dieser Sitzung?“ sowie „Welche Rolle spielt das Thema dieser Sitzung für das schulische Lernen?“ (kognitiv) und „Was ist für Sie zu diesem Thema noch unklar?“ (metakognitiv). Die Studierenden erhielten auf drei ihrer Aufgabenbearbeitungen von geschulten Hilfskräften Rückmeldung zu Inhalt und Tiefe der Reflexionen sowie zu Orthografie und Grammatik.

4.3 Stichprobe

Von den $N = 400$ zur Vorlesung angemeldeten Studierenden lagen für $n = 370$ Studierende (72 % weiblich; 27 % männlich, 1% anderes Geschlecht) Daten zu beiden Messzeitpunkten vor. Die Studierenden waren im Mittel $M = 23.96$ Jahre alt ($SD = 1.95$), im $M = 5.46$ Fachsemester ($SD = 2.88$) und hatten zum Zeitpunkt der Vorlesung im Verlauf ihres bisherigen Studiums bereits $M = 2.00$ Lehrveranstaltungen mit pädagogischem oder psychologischem Inhalt besucht ($SD = 0.93$).

5 Ergebnisse

5.1 Effekte auf die Urteilskorrektheit

Die erste Analyse zur Beantwortung von Fragestellung 1 untersuchte den Effekt der Wahlfreiheit auf die Urteilskorrektheit nur bei den Sitzungsinhalten, zu denen schriftliche Reflexionen abgegeben wurden. Eine Messwiederholungs-ANOVA mit den Faktoren Zeit (t1 vs. t2) X Wahlfreiheit (selbstgewählt vs. vorgegeben) über die Urteilskorrektheit ergab einen Haupteffekt der Zeit, $F(1, 319) = 18.17, p < .01, \eta_p^2 = .05$, wobei die Korrektheit zu t1 höher als die zu t2 war ($M = 0.60, SEM = 0.01$ vs. $M = 0.55, SEM = 0.01$). Der Haupteffekt der Wahlfreiheit war ebenfalls signifikant, $F(1, 319) = 113.79, p < .01, \eta_p^2 = .26$: Items zu Inhalten zu vorgegebenen Aufgaben (im Folgenden: vorgegebene Aufgaben) wurden mit höherer Korrektheit als Items zu Inhalten zu selbstgewählten Aufgaben (im Folgenden: selbstgewählte Aufgaben) bearbeitet ($M = 0.63, SEM = 0.01$ vs. $M = 0.51, SEM = 0.01$). Weiterhin war die Interaktion zwischen Zeit und Wahlfreiheit signifikant, $F(1, 319) = 59.08, p < .01, \eta_p^2 = .16$ (s. Abbildung 1).

Post-hoc durchgeführte Kontraste ergaben, dass die Korrektheit bei den vorgegebenen Aufgaben von t1 zu t2 abnahm, während sie bei den selbstgewählten Aufgaben gleich blieb ($F(1, 319) = 64.49, p < .01, \eta_p^2 = .17$ (t1: $M = 0.70, SEM = .01$; t2: $M = 0.57, SD = .01$) bzw. $F(1, 319) = 1.84, p = .18$ (t1: $M = 0.50, SEM = 0.15$; t2: $M = 0.52, SEM = 0.01$)). Weitere Kontraste ergaben, dass sich die Korrektheit zwischen selbstgewählten und vorgegebenen Aufgaben zu t1 signifikant unterschied; dies galt auch für t2 ($F(1, 319) = 106.62, p < .01, \eta_p^2 = .25$ bzw. $F(1, 319) = 31.87, p < .01, \eta_p^2 = .09$).

Diese Ergebnisse deuten darauf hin, dass zu beiden Zeitpunkten die vorgegebenen Aufgaben korrekter beantwortet wurden als die selbstgewählten, wobei sich die Quoten aufgrund der Abnahme der Korrektheit bei ersteren Aufgaben zu t2 annäherten.

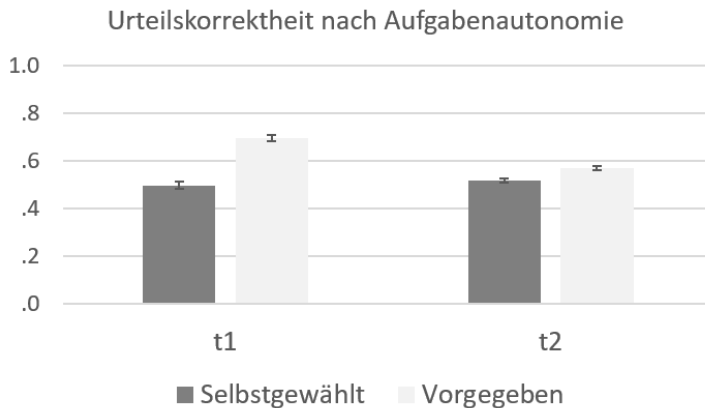


Abb. 1: Mittlere prozentuale Anteile der korrekt abgegebenen Antworten auf Test-items nach Messzeitpunkt und Wahlfreiheit der Aufgabenwahl (Fehlerindikatoren geben den Standardfehler des Mittelwerts wieder).

5.2 Effekte auf die Urteilkonfidenz

Zur Beantwortung von Fragestellung 2 wurde eine analoge Messwiederholungs-ANOVA mit den Konfidenzurteilen der reflektierten Sitzungsinhalte durchgeführt. Der Haupteffekt der Zeit war nicht signifikant, $F(1, 319) < 0.01, p > .99$ (t1: $M = 0.49, SEM = 0.01$; t2: $M = 0.49, SEM = 0.01$), wohl aber der Haupteffekt der Wahlfreiheit, $F(1, 319) = 20.19, p < .01, \eta_p^2 = .06$: Bei den selbstgewählten Aufgaben lag eine höhere Konfidenz vor als bei den vorgegebenen ($M = 0.51, SEM = 0.01$ vs. $M = 0.46, SEM = 0.01$). Dieser Haupteffekt wurde durch die signifikante Interaktion zwischen Zeit und Wahlfreiheit qualifiziert, $F(1, 319) = 55.89, p < .01, \eta_p^2 = .15$ (s. Abbildung 2).

Post-hoc-Kontraste ergaben, dass die Konfidenz bei den vorgegebenen Aufgaben von t1 zu t2 signifikant anstieg, während sie bei den selbstgewählten Aufgaben signifikant abnahm ($F(1, 319) = 15.78, p < .01, \eta_p^2 = .05$ (t1: $M = 0.42, SEM = 0.02$; t2: $M = 0.50, SEM = 0.01$) bzw. $F(1, 319) = 19.45, p < .01, \eta_p^2 = .06$ (t1: $M =$

0.55, $SEM = 0.02$; t_2 : $M = 0.47$, $SEM = 0.01$). Weitere Kontraste ergaben, dass sich die Konfidenz zwischen selbstgewählten und vorgegebenen Aufgaben zu t_1 signifikant unterschied; dies galt auch für t_2 ($F(1, 319) = 43.52$, $p < .01$, $\eta_p^2 = .12$ bzw. $F(1, 319) = 11.35$, $p < .01$, $\eta_p^2 = .03$).

Diese Ergebnisse deuten darauf hin, dass eine Vorgabe zur Bearbeitung zu einer Erhöhung der subjektiven Konfidenz führte und gleichzeitig die Wahlfreiheit bei der Aufgabenwahl die Konfidenz verringerte.

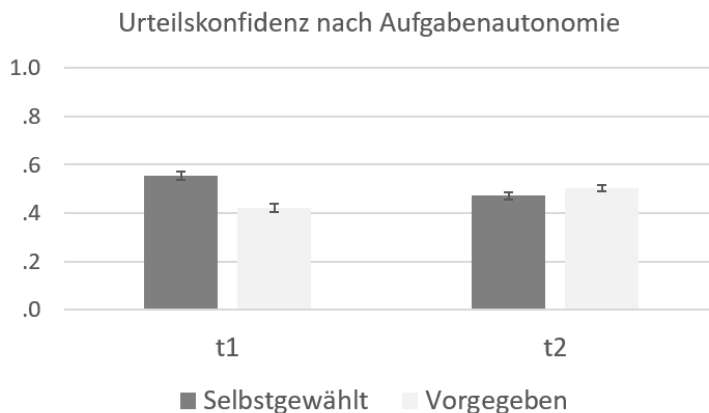


Abb. 2: Mittlere prozentuale Anteile der mit hoher Konfidenz abgegebenen Antworten auf Testitems nach Messzeitpunkt und Wahlfreiheit der Aufgabenwahl (Fehlerindikatoren geben den Standardfehler des Mittelwerts wieder).

5.3 Effekte auf die Urteilsensitivität

Die vorangegangenen Analysen beantworten noch nicht die Frage, ob die Vorgabe vs. Wahlfreiheit der Aufgabenwahl die Urteilstkonfidenz primär in korrekte oder auch in inkorrekte Antworten förderte, wie es u. a. bei BARENBERG et al. (2016) geschehen war. Um die Urteilsensitivität (Fragestellung 3) abzuschätzen, analysierten wir daher die Urteilstkonfidenz in korrekte und inkorrekte Antworten im Vergleich, jeweils einzeln für selbstgewählte und für vorgegebene Aufgaben.

5.3.1 Analyse für selbstgewählte Aufgaben

Eine Messwiederholungs-ANOVA mit den Faktoren Zeit (t1 vs. t2) X Urteilskorrektheit (korrekt vs. inkorrekt) ergab zwei signifikante Haupteffekte. Der Haupteffekt der Zeit, $F(1,255) = 8.80, p < .01, \eta_p^2 = .03$, gab an, dass die Konfidenz von t1 zu t2 abnahm ($M = 0.54, SEM = 0.02$ vs. $M = 0.48, SEM = 0.02$). Nach dem Haupteffekt der Urteilskorrektheit, $F(1,255) = 31.15, p < .01, \eta_p^2 = .11$, war die Konfidenz in korrekte Urteile höher als diejenige in inkorrekte Urteile ($M = 0.57, SEM = 0.02$ vs. $M = 0.45, SEM = 0.02$). Der Interaktionseffekt Zeit X Urteilskorrektheit war nicht signifikant, $F(1,255) = 0.66, p = .42$ (s. Abbildung 3a).

5.3.2 Analyse für vorgegebene Aufgaben

Eine analoge Messwiederholungs-ANOVA für vorgegebene Aufgaben ergab einen Haupteffekt der Zeit, $F(1,207) = 22.20, p < .01, \eta_p^2 = .10$; demnach nahm die Konfidenz von t1 zu t2 zu ($M = 0.39, SEM = 0.02$ vs. $M = 0.51, SEM = 0.02$). Weiter ergab sich ein Haupteffekt der Urteilskorrektheit, $F(1,207) = 76.28, p < .01, \eta_p^2 = .27$; demnach fiel die Konfidenz in korrekte Urteile höher aus als in inkorrekte Urteile ($M = 0.54, SEM = 0.02$ vs. $M = 0.36, SEM = 0.02$). Beide Haupteffekte wurden durch einen signifikanten Interaktionseffekt qualifiziert, $F(1,207) = 13.85, p < .01, \eta_p^2 = .06$ (s. Abbildung 3b). Post-hoc-Kontraste ergaben, dass die Konfidenz in korrekte Urteile von t1 zu t2 gleich blieb ($M = 0.52, SEM = 0.03$ vs. $M = 0.56, SEM = 0.02$), $F(1,207) = 1.97, p = .16$. Die Konfidenz in inkorrekte Urteile nahm jedoch signifikant von t1 zu t2 zu ($M = 0.26, SEM = 0.03$ vs. $M = 0.47, SEM = 0.02$), $F(1,207) = 32.61, p < .01, \eta_p^2 = .14$. Weitere Kontraste ergaben, dass sich die Konfidenz in korrekte von der in inkorrekte Urteile sowohl zu t1 als auch zu t2 signifikant unterschieden ($F(1,207) = 45.01, p < .01, \eta_p^2 = .18$ bzw. $F(1,207) = 33.39, p < .01, \eta_p^2 = .14$).

Diese Ergebnisse deuten darauf hin, dass die Studierenden sich generell sicherer bei korrekten als bei inkorrekten Urteilen fühlten, sich dieser Unterschied bei vorgegebenen Aufgaben allerdings auf Kosten der Sensitivität reduzierte.

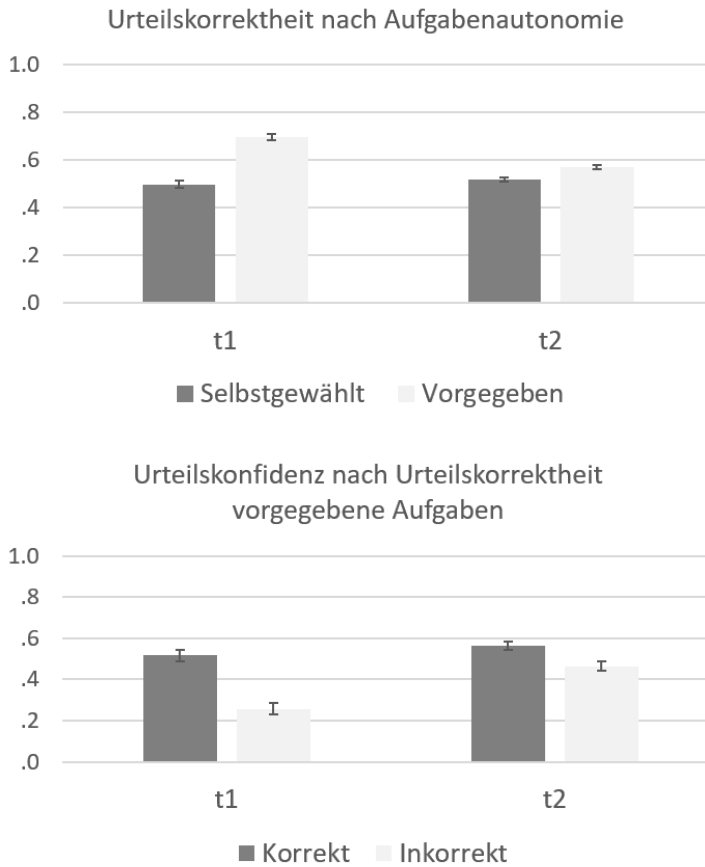


Abb. 3a und 3b: Mittlere prozentuale Anteile der als sicher (konfident) eingeschätzten Antworten auf selbstgewählte Aufgaben (Abb. 3a) sowie bei vorgegebenen Aufgaben (Abb. 3b) nach Messzeitpunkt und Korrektheit der Antwort (Fehlerindikatoren geben den Standardfehler des Mittelwerts wieder).

6 Diskussion

Nach einer Darstellung und Diskussion der zentralen Befunde der Studie sollen vor allem die Grenzen der Studie und mögliche Implikationen diskutiert werden.

6.1 Effekte der Wahlfreiheit

6.1.1 Urteilskorrektheit

Die Urteilskorrektheit betreffend zeigte sich eine eher negative Entwicklung von t1 zu t2 (Absinken der Leistung). Dafür könnten zum einen die zu t2 neu eingeführten Items (55 von 68 Items) verantwortlich sein, deren Schwierigkeit vorab nicht bekannt war. Zum anderen könnten die schriftlichen Aufgaben – im Vergleich zu den Problemaufgaben bei BARENBERG et al. (2016) – die Verarbeitung der Inhalte zu wenig angeregt haben.

Diese Erklärungen sollten sowohl für die vorgegebenen als auch für die selbstgewählten Aufgaben zutreffen. Bei den selbstgewählten Aufgaben zeigte sich allerdings eine zwar generell niedrigere, aber von t1 zu t2 konstante Leistung. Dies könnte für diese Aufgaben auf eine Verallgemeinerung des Wissens auf die zu t2 neuen Aufgaben hindeuten, was eventuell durch – im Vergleich zu den vorgegebenen Aufgaben – eine stärkere autonome Regulation und dadurch bedingt einen verstärkten Einsatz von Tiefenstrategien erklärt werden könnte (vgl. RYAN & DECI, 2000). Wie alle hier diskutierten Interpretationen ist auch diese aufgrund des vorliegenden Designs spekulativ, da andere Erklärungen aufgrund des Studiendesigns nicht ausgeschlossen sind.

6.1.2 Urteilstkonfidenz

Die metakognitive Konfidenz in die eigenen Antworten zeigte ein zur Urteilskorrektheit konträres Bild. Bei den selbstgewählten Aufgaben lag zu t1 eine höhere Konfidenz vor als bei den vorgegebenen Aufgaben. Die Studierenden haben eventuell die Themen gewählt, bei denen sie sich am sichersten fühlten. Die Abnahme der Konfidenz zu t2 könnte eventuell mit durch die Wahlfreiheit bedingtem stärker-

rem Einsatz von Tiefenstrategien erklärt werden. Die erhöhte Motivation könnte zur stärkeren Anwendung systematischer Urteilsprozesse geführt haben, d. h., die Studierenden könnten sich bei der Konfidenzschätzung der Items zu t2 stärker und systematischer Gedanken gemacht haben, wie genau ihre Urteile waren. Dies könnte mit einer Abnahme heuristischer Urteilsprozesse wie Familiarität und Vertrautheit (vgl. IKEDA, YUE, MURAYAMA & CASTEL, 2016) einhergehen. Diese Prozesse würden darauf hindeuten, dass die Konfidenzschätzungen für die selbstgewählten Aufgaben zu t2 genauer waren als diejenigen für die vorgegebenen Aufgaben.

Die Konfidenz bezüglich der vorgegebenen Aufgaben nahm dagegen von t1 zu t2 zu. Dies könnte eventuell durch die besonderen Inhalte der Sitzungen erklärt werden, da hier v. a. Sitzungen ausgewählt wurden, die als Grundlagenthemen auch in anderen Sitzungen, z. B. in Anwendungskontexten, aufgegriffen und vertieft wurden. Die Studierenden könnten also mit den Themen so vertraut gewesen sein, dass sie heuristisch (vgl. IKEDA et al., 2016) auf ihre höhere Konfidenz schlossen. Auch zur Urteilsconfidenz müssen jedoch aufgrund des vorliegenden Designs alle Interpretationen spekulativ bleiben.

6.1.3 Urteilssensitivität

Wenn Wahlfreiheit bei der Aufgabenwahl dazu führt, dass Studierende ihre Konfidenz systematischer und weniger heuristisch beurteilen, sollte sich auch die Urteilssensitivität erhöhen. In unserer Studie fanden sich bei selbstgewählten und vorgegebenen Aufgaben generell höhere Konfidenz in korrekte als in inkorrekte Lösungen (vgl. BARENBERG et al., 2016).

Die oben genannte Reduktion der Konfidenz für selbstgewählte Aufgaben zeigte sich sowohl in korrekten als auch inkorrekten Lösungen. Hier scheint also generell die Schwelle gestiegen zu sein, eine Antwort als sicher zu beurteilen. Dies könnte mit einer höheren Genauigkeitsmotivation erklärt werden, die durch die oben genannte stärkere Anwendung systematischer Urteilsprozesse verursacht sein könnte. Allerdings zeigte sich für selbstgewählte Aufgaben keine stärkere Konfidenzabnahme für inkorrekte Lösungen, was man bei einer erhöhten Genauigkeitsmotivati-

on erwarten würde. Allerdings ist auch denkbar, dass Studierende in diesem Testformat generell kaum zwischen korrekten und inkorrekten Urteilen unterscheiden können. So fanden BARENBERG & DUTKE (2013, Studie 2) selbst bei Feedback auf die eigene Leistung keinen Effekt auf die Sensitivität der Konfidenzurteile.

Für die vorgegebenen Aufgaben fand sich im Unterschied dazu eine Zunahme der Konfidenz nur in inkorrekte Lösungen, während die Konfidenz in korrekte Lösungen gleich blieb. Die in Kapitel 6.1.2 genannte stärkere Nutzung der Vertrautheit für die Konfidenzurteile bei gleichzeitiger Abnahme der Urteilskorrektheit könnte erste Hinweise zur Erklärung dieses Ergebnisses bieten.

Insgesamt zeigt das Ergebnismuster, dass die thematische Wahlfreiheit bei schriftlichen Reflexionsaufgaben keine einfachen, sondern komplexe Effekte auf Urteilskorrektheit, -konfidenz und -sensitivität ausüben kann. Eine generelle Verbesserung der kognitiven und metakognitiven Leistung in einer Vorlesung ist durch diese Intervention offensichtlich nicht zu erwarten. Aufgrund des vorliegenden Studiendesigns lassen sich jedoch nur post-hoc tentative Interpretationen für die erhaltenen Ergebnisse ableiten. Unser Projekt dient somit v. a. einer Ableitung von Hypothesen, die in weiteren Studien geprüft werden sollten.

6.2 Limitationen und Implikationen

Eine Reihe methodischer Aspekte beschränken die Aussagekraft der Studie, von denen die aus unserer Sicht gravierendsten hier dargestellt werden.

Erstens folgte die Datenerhebung keinem experimentellen Design. Die randomisierte Zuteilung von selbstgewählten Aufgaben hätte der Idee der Wahlfreiheit widersprochen. So haben die Teilnehmer/innen die selbstgewählten Aufgaben vermutlich nach bestimmten Aspekten wie inhaltlichem Interesse, Zeitverlauf im Semester oder Aufwandsvermeidungszielen ausgewählt. All diese Faktoren könnten die Ergebnisse beeinflusst haben. Welche Merkmale bei der Auswahl tatsächlich herangezogen wurden, ist nicht bekannt. Systematische verzerrende Effekte durch bestimmte Variablen sind also nicht auszuschließen. Künftige Studien könnten die Autonomie bei der Aufgabenwahl eventuell auf wenige Optionen ein-

schränken. Ferner sollten sich vorgegebene und selbstgewählte Aufgaben nicht bezüglich Schwierigkeit und Interesse der Studierenden unterscheiden. Zudem wäre interessant zu untersuchen, wie die Studierenden das erhaltene Feedback nutzen.

Zweitens ist die uV Wahlfreiheit problematisch. So geht die Selbstbestimmungstheorie (RYAN & DECI, 2000) davon aus, dass die Wahrnehmung der eigenen Autonomie entscheidend für die Motivation ist. Ob diese Wahrnehmung in der Studie wirklich beeinflusst wurde, ist mangels einer Manipulationskontrolle unklar. Da andere Bestandteile von Autonomie, wie Relevanz, nicht berücksichtigt wurden, ist es denkbar, dass die Autonomie nicht als stärker erlebt wurde (vgl. ASSOR et al., 2002). Folgestudien sollten diese Aspekte deshalb direkt erheben, um deren Auswirkungen auf die erlebte Autonomie zu untersuchen.

Drittens wurden, wie oben beschrieben, zu t1 und zu t2 Tests mit unterschiedlich vielen und unterschiedlich bekannten Items miteinander verglichen. Auch wenn die neu hinzugefügten Items aus unserer Sicht für die Fragestellung relevante Informationen enthalten – Verallgemeinerung des Wissens über die bereits getesteten Items (vgl. BARENBERG & DUTKE, 2012) hinaus –, ist die psychometrische Vergleichbarkeit der Tests damit kritisierbar.

Trotz dieser Limitationen liefert die Durchführung des Projekts für unsere eigene Lehre durchaus einige Hinweise. Erstens sind die Ergebnisse nicht so, wie wir es erwartet oder erwünscht hatten: Effekte der Wahlfreiheit auf die Leistung sind kaum vorhanden; die Effekte auf die Urteilsconfidenz sind komplex und post-hoc schwierig zu erklären. Durch das Projekt haben wir jedoch interessante Ansatzpunkte für weitere Studien gewonnen. Zweitens hat uns die Auseinandersetzung mit dem theoretischen Rahmen und den Ergebnissen unserer Studie die Bedeutung einiger psychologischer Konzepte im Kontext von Vorlesungen verdeutlicht, so z. B. den Zusammenhang zwischen Kognition und Metakognition. Generell konnten wir durch unser Projekt interessante Erkenntnisse über unsere eigene Lehre gewinnen, so z. B., dass einfache Effekte augenscheinvalider Interventionen, wie hier der Schreibaufgaben, in der Realität der Lehre eher nicht zu erwarten sind. Umso

spannender ist nun die Frage des Warum, der in nachfolgender Forschung nachgegangen werden kann.

Zentraler Bestandteil des Scholarship of Teaching and Learning (vgl. HUBER, 2011) ist dessen Veröffentlichung und damit die Möglichkeit zur Kritik von anderer Seite. Aus eigener Erfahrung mit diesem Projekt können wir sagen, dass sich dieser Schritt lohnt: Durch die Diskussion mit Kolleginnen/Kollegen und das Schreiben dieses Manuskripts haben wir sehr profitiert – sowohl in forschungsmethodischer als auch in inhaltlicher Hinsicht. Uns trieb lange die Frage um, ob wir die hier berichteten „Null-Effekte“ einer Studie mit einigen methodischen Problemen wirklich publizieren sollten, und wenn ja, in welcher Form. Wir denken, dass die klare Darstellung der relevanten Rahmenbedingungen und eine selbstkritische Haltung gegenüber der eigenen Methodik wichtig sind, um die eigene Lehre zur Diskussion zu stellen. Wir hoffen, dies getan zu haben und können mit einer wichtigen Erkenntnis unsererseits schließen: Es lohnt sich, die eigene Lehre zu befor-schen.

7 Literaturverzeichnis

Assor, A., Kaplan, H. & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology*, 72, 261-278.

Bannert, M. (2003). Effekte metakognitiver Lernhilfen auf den Wissenserwerb in vernetzten Lernumgebungen. *Zeitschrift für Pädagogische Psychologie*, 17, 13-25.

Barenberg, J. & Dutke, S. (2012). ‚Testing Effect‘ und metakognitives Monitoring in Psychologie-Lehrveranstaltungen. In M. Krämer, S. Dutke & J. Barenberg (Hrsg.), *Psychologiedidaktik und Evaluation IX* (S. 301-308). Aachen: Shaker.

Barenberg, J. & Dutke, S. (2013). Metacognitive monitoring in university classes: Anticipating a graded vs. a pass-fail test affects monitoring accuracy. *Metacognition and Learning*, 8, 121-143.

Barenberg, J., Seifried, E., Spinath, B. & Dutke, S. (2016). Die Bearbeitung schriftlicher Problemaufgaben erhöht den Lernerfolg in einer Psychologie-Vorlesung. In M. Krämer, S. Preiser & K. Brusdeylins (Hrsg.), *Psychologiedidaktik und Evaluation XI* (S. 323-330). Aachen: Shaker.

Berthold, K., Nückles, M. & Renkl, A. (2007). Do learning protocols support learning strategies and outcomes? The role of cognitive and metacognitive prompts. *Learning & Instruction, 17*, 564-577.

Bligh, D. A. (2000). *What's the use of lectures?* San Francisco: Jossey-Bass.

Boyer, E. L. (1990). *Scholarship reconsidered. Priorities of the professoriate.* Princeton, NJ: The Carnegie Foundation for the Advancement of Teaching.

Dutke, S. & Barenberg, J. (2009). Ökonomische Wissenstests für teilnehmerreiche Psychologie-Lehrveranstaltungen. In M. Krämer, S. Preiser & K. Brusdeylins (Hrsg.), *Psychologiedidaktik und Evaluation VII* (S. 367-376). Aachen: Shaker.

Dutke, S. & Barenberg, J. (2015). Easy and informative: Using confidence-weighted true-false items for knowledge tests in psychology courses. *Psychology Learning & Teaching, 14*, 250-259.

Grolnick, W. S. & Ryan, R. M. (1987). Autonomy in childrens' learning. An experimental and individual difference investigation. *Journal of Personality and Social Psychology, 52*, 890-898.

Hmelo-Silver, C.-E. (2004). Problem-based learning: What and how do students learn? *Educational Psychology Review, 16*, 235-266.

Huber, L. (2011). Forschen über (eigenes) Lehren und studentisches Lernen – Scholarship of Teaching and Learning (SoTL): Ein Thema auch hierzulande? *Das Hochschulwesen, 59*, 118-124.

Hübner, S., Nückles, M. & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning & Instruction, 20*, 18-29.

Ikeda, K., Yue, C. L., Murayama, K. & Castel, A. D. (2016). Achievement goals affect metacognitive judgments. *Motivation Science, 2*, 199-219.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology*, *77*, 1121-1134.

Levesque, C., Zuehlke, A. N., Stanek, L. R. & Ryan, R. M. (2004). Autonomy and competence in German and American university students: A comparative study based on self-determination theory. *Journal of Personality and Social Psychology*, *96*, 68-84.

Ryan, R. M. & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist*, *55*, 68-78.

Spinath, B. & Seifried, E. (2012). Forschendes Lehren: Kontinuierliche Verbesserung einer Vorlesung. In M. Krämer, S. Dutke & J. Barenberg (Hrsg.), *Psychologiedidaktik und Evaluation IX* (S. 171-180). Aachen: Shaker.

Vansteenkiste, M., Lens, W. & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational Psychologist*, *41*, 19-31.

Danksagung

Wir danken Amelie Becker, Lea Grotegut, Christian Janßen, Frederic Küting und Kevin Oergel für ihre wertvolle Hilfe bei der Durchführung der Vorlesung.

Autor/innen



Jun.-Prof. Dr. Robert KORDTS-FREUDINGER || Universität Paderborn, Fach Psychologie || Warburger Str. 100, D-33098 Paderborn

go.upb.de/rkf

robert.kordts.freudinger@upb.de



Prof. Dr. Katrin B. KLINGSIECK || Universität Paderborn, Fach Psychologie || Warburger Str. 100, D-33098 Paderborn

go.upb.de/ppduf

katrin.klingsieck@upb.de



Dr. Eva SEIFRIED || Universität Heidelberg, Psychologisches Institut || Hauptstr. 47-51, D-69117 Heidelberg

www.psychologie.uni-heidelberg.de/ae/paeps/personen/seifried.html

eva.seifried@psychologie.uni-heidelberg.de