

Berichte aus der Psychologie

**Michael Krämer, Siegfried Preiser,  
Kerstin Brusdeylins (Hrsg.)**

**Psychologiedidaktik und Evaluation XII**

Dem Wunsch mehrerer Autorinnen und Autoren folgend sei an dieser Stelle darauf hingewiesen, wenn in den folgenden Beiträgen entweder nur die männliche oder nur die weibliche Sprachform genutzt wird, so sind jeweils beide Geschlechter gemeint. Es geschieht ausschließlich der besseren Lesbarkeit halber.

Um den Datenschutz zu wahren, wurde auf die Veröffentlichung der Autorenadressen verzichtet. Wenn Sie Kontakt zu einer Autorin oder einem Autor aufnehmen wollen, schreiben Sie bitte eine e-mail an folgende Adresse. Der Herausgeber leitet Ihren Wunsch gerne weiter: [kraemer@fh-muenster.de](mailto:kraemer@fh-muenster.de)

# **Inhalt**

## **Aus-, Fort- und Weiterbildung in Psychologie – Grundlegende Konzepte**

MICHAEL KRÄMER

Novellierung des Psychotherapeutengesetzes – Anmerkungen zum  
Arbeitsentwurf des Bundesministeriums für Gesundheit 3

ELMAR SOUVIGNIER

„Aktivieren ist besser als Erklären!(?)“ 11  
Einstellungen zum Lehren und deren Bedeutung

JULIA MENDZHERITSKAYA, SABINE FABRIZ, MIRIAM HANSEN,  
NICOLA REIMANN, JAMIE THOMPSON, MALCOLM MURRAY,  
JULIE RATTRAY UND TETI DRAGAS

Förderung der Reflexion über eigene Lehre durch interkulturellen  
kollegialen Austausch im Rahmen von hochschuldidaktischen  
Weiterbildungsprogrammen in Deutschland und England 19

MARIA TULIS

Da ist immer noch Luft drin! Zur Notwendigkeit einer 27  
didaktischen Konzeption kognitiver Umstrukturierungsprozesse  
im Psychologieunterricht

JONATHAN BARENBERG UND STEPHAN DUTKE

Drei Untersuchungen zum Testeffekt – drei Perspektiven 37  
auf das Konzept evidenzbasierter Lehre im Lehramtsstudium

GESA UHDE, BARBARA THIES, HANNAH PERST UND LENA HANNEMANN

Kompetenzorientierte Beratungs- und Begleitstrukturen 45  
im Lehramtsstudium: Selbstreflexionskompetenzen und  
Classroom-Management-Strategien fördern

PAUL HINNERSMANN

Warum mache ich das? Und warum ist diese Frage wichtig? 55  
Ein Workshop zur Motivationspsychologie für Psychologielehrkräfte

ULI SANN UND FRANK UNGER

Konzeption und Weiterentwicklung einer wissenschaftlichen 63  
Weiterbildungsveranstaltung zum „Umgang mit herausfordernden  
Situationen“ in der öffentlichen Verwaltung

## **Psychologie studieren in unterschiedlichen Institutionen**

UTE-REGINA ROEDER UND SARAH-INES MEUDT Psychologie studieren nach Bologna	73
MICHAELA ZUPANIC, JAN P. EHLERS, STEFAN J. TROCHE Auswahlgespräche im Gutachter-Tandem für den Bachelorstudiengang Psychologie der Universität Witten/Herdecke	79
NINA ZEUCH, LENA KEGEL, MARLENE MERTENS UND LEONIE SCHRÖDER Partizipative Elemente bei der Gestaltung von Lehre am Institut für Psychologie der Westfälischen Wilhelms-Universität Münster	89
ROBERT GASCHLER UND STEFAN STÜRMER Erfahren und Üben im Browserfenster – Interaktives Lernen im B.Sc. Psychologie an der FernUniversität	99
SIEGFRIED PREISER UND TIMO STORCK Die Psychologische Hochschule Berlin – ein Ort für die Diversität von Themen, Methoden und Menschen	107

## **Psychologie im Kontext der Berufs- und Allgemeinbildung**

GISLINDE BOVET Psychologische Bildung für Schülerinnen und Schüler	113
PAUL GEORG GEIß <i>Psychological literacy</i> als Ziel des Psychologiestudiums und des Psychologieunterrichts	121
NADJA BADR Beiträge des Psychologieunterrichts zur allgemeinen Bildung auf der Sekundarstufe II	131
HANS-PETER NOLTING Leitlinien psychologischer Allgemeinbildung	139
DAVID FRAISSL Psychologische Bildung: Für eine autonomieorientierte Psychologievermittlung – ein erster Entwurf	145

## Lehren und Lernen - Praxiskonzepte

MARKUS GERTEIS

Aufbau einer wissenschaftlichen Haltung in der Ausbildung  
von angehenden Lehrpersonen 155

BASTIAN HODAPP

„Weniger Referate!“ – Eine empirische Untersuchung zu  
Brainwalking, Thesenpapier, Forschungswerkstatt und Rollenspiel 165

GUIDO BREIDEBACH

Eine Didaktik der differenzierten Individualisierung 173

NATALIE ENDERS

Förderung selbstregulatorischer Fähigkeiten in einem E-Learningseminar  
für Lehramtsstudierende 183

SINA SCHÜRER, BEA BLOH, STEFANIE VAN OPHUYSEN UND  
LARS BEHRMANN

Vermittlungsstrategien zum Forschenden Lernen in der Lehrerbildung –  
Was nutzt der Nutzen? 191

STEFAN ZIMMERMANN, PATRICK MÜLLER UND THOMAS BÄUMER

Projektbasierte Methodenlehre als Plattform zur  
Auseinandersetzung mit gesellschaftsrelevanten Fragestellungen 201

MIRJAM BRÄBLER

Interdisziplinäre Nachhaltigkeitsbildung –  
Eine große Chance für angehende Psycholog\*innen 209

ULRIKE STARKER

Nachhaltigkeitskompetenz fördern durch „Gaming“ –  
ein interdisziplinäres Lehr-Lern-Projekt 217

LARS BEHRMANN

Die Förderung diagnostischer Kompetenzen von Lehrkräften –  
Ein theoriebasiertes Seminarkonzept mit praktischen Übungen 227

LARS HAUTEN UND INGO JUNGCLAUSSEN

Kasuistik im Verfahrensdiallog (KiV) – Neue didaktische Wege in der  
verfahrensdialogischen Psychotherapie-Ausbildung am Institut für  
Psychologische Psychotherapie Berlin (*ppt*) 237

MONIKA SKLORZ-WEINER	
Die Vermittlung von Grundlagen eines professionellen Therapeut-Klient-Verhältnisses im Rahmen des Psychologieunterrichts an einer Berufsfachschule für Atem-, Sprech- und Stimmlehrer/innen	245
KERSTIN BRUSDEYLINS	
Psychoedukation und Schmerzbewältigung in der multimodalen Schmerztherapie	253
AILEEN WOSNIAK	
Entwicklung und Evaluation einer Intervention für Kindergartenfachkräfte zur Förderung des Emotionswissens von Kindern	261
MARKUS KNÖPFEL, TIM KÖHLER UND FRANK MUSOLESI	
Handlungsempfehlungen für die Durchführung von Problem-Based Learning (PBL) auf Basis der Lehrevaluationsergebnisse	265
ANITA KNÖFERLE UND DOROTHEA DORNHEIM	
Entwicklungsförderliche sprachliche Intervention – ein Theorie-Praxis-Seminar für Studierende mit Unterrichtsfach Psychologie	273
BRIGITTE STEINHEIDER, VIVIAN HOFFMEISTER AND JAMES MCKENZIE	
Combining Team-based Learning with Creativity to Increase Learner Engagement in a Non-Traditional Graduate I/O Psychology Program	281
ARISTI BORN	
Mit Psychologie und Praxiserfahrung ins Lernen starten: Die Projektwerkstatt als neues Lernformat im Studiengang Soziale Arbeit	289

## **Evaluation der Lehre**

STEPHAN DUTKE, UTE-REGINA ROEDER UND JONATHAN BARENBERG	
Findet in Psychologie-Lehrveranstaltungen verteiltes Lernen statt? Eine Untersuchung zu Einflussfaktoren und Effekten auf die metakognitive Lernleistung	299
MICHAEL KRÄMER	
Einflussfaktoren auf den Studienerfolg aus subjektiver Perspektive	307
TIMO HERDEL UND SIEGFRIED PREISER	
Trainingskompetenzen – Validierung eines Trainer-Selbstkonzept- Fragebogens und Evaluation eines Trainings	317

## Evaluation des Lernens

INES DEIBL, JÖRG ZUMBACH UND VIOLA GEIGER

Constructive Alignment im Bereich der Pädagogischen Psychologie –  
Entwicklung und Anwendung eines Fragebogens zur Erfassung von  
Constructive Alignment 327

PETIA GENKOVA UND MANUELA BÖCKENFELD

Generation Y und Stressbewältigungsstrategien: Studiert man heute anders? 335

BASTIAN HODAPP

Entwicklung, Erprobung und Evaluierung eines neuen Prüfungsformates  
beim Forschenden Lernen 343

HEIKE M. BUHL, CARLA BOHNDICK, SABRINA BONANATI, CHRISTIAN GREINER,  
JOHANNA HILKENMEIER UND ROBERT KORDTS-FREUDINGER

Fallbasierte Modulabschlussprüfungen zur Verzahnung von  
Theorie und Praxis im Master des Lehramtsstudiums 351

MICHAELA ZUPANIC, STEFAN J. TROCHE, JAN P. EHLERS

Absolvierendenniveau im formativen Progress Test Psychologie:  
Anspruch oder Wirklichkeit? 359

FLORIAN KLAPPROTH

Die Eignung von Mehrfachwahlaufgaben für die Lernverlaufsdiagnostik 369

ANGELIKA TAETZ-HARRER, MICHAELA ZUPANIC UND  
STEFAN J. TROCHE

Was würden Sie tun? - Generieren von Antwortoptionen in der  
Entwicklung eines *Situational-Judgement-Test* zur Erfassung  
sozialer Kompetenzen 377

# Die Eignung von Mehrfachwahlaufgaben für die Lernverlaufsdiagnostik

**Florian Klapproth**

Mit dem vorliegenden Beitrag wurde geprüft, ob Mehrfachwahlaufgaben änderungssensitiv sind und damit eine geeignete Testform für die Lernverlaufsdiagnostik darstellen. Die Prüfung der Änderungssensitivität erfolgte in einem experimentellen Prä-Post-Kontrollgruppendesign unter Berücksichtigung möglicher Retest-Effekte und der Testschwierigkeit. Es zeigte sich, dass Mehrfachwahlaufgaben änderungssensitiv sind, wenn sie hinreichend schwierig sind. Somit scheinen Mehrfachwahlaufgaben für die Lernverlaufsdiagnostik grundsätzlich geeignet zu sein.

## Theoretischer und empirischer Hintergrund

Das Ziel der Lernverlaufsdiagnostik ist, Lernverläufe von Schülerinnen und Schülern über einen längeren Zeitraum hinweg abzubilden und den Lernfortschritt an die Lehrkräfte zurückzumelden (Deno, 2003). In der Lernverlaufsdiagnostik wird statt einer punktuellen Testerhebung regelmäßig in relativ kurzen Abständen die Entwicklung einer Kompetenz über längere Zeit hinweg durch geeignete Tests verfolgt. Dabei handelt es sich um jeweils andere Tests, die aber stets die Ausprägung derselben Kompetenz messen. In dem vorliegenden Beitrag wird eine experimentelle Untersuchung vorgestellt, in welcher die Eignung eines bestimmten Aufgabenformats (Mehrfachwahlaufgaben) für die Lernverlaufsdiagnostik geprüft wurde.

## Besonderheiten der Lernverlaufsdiagnostik

Für die Lernverlaufsdiagnostik sind zwei Testeigenschaften wesentlich: eine *konstante Testschwierigkeit* und die *Änderungssensitivität* der Tests (Klauer, 2011). Eine konstante Testschwierigkeit stellt sicher, dass gemessene Verbesserungen in den Testwerten nicht auf unterschiedlich schwierige Tests, sondern auf einen tatsächlichen Lernzuwachs zurückgehen. Die Änderungssensitivität gewährleistet, dass ein Lernzuwachs von einem Test auch tatsächlich aufgedeckt werden kann. Um sicher zu gehen, dass die erfassten Veränderungen vom Zufall weitgehend bereinigt sind und daher „wahre“ Veränderungen widerspiegeln, muss auf ein experimentelles Prä-Post-Kontrollgruppendesign zurückgegriffen werden (Klauer, 2011, 2014; Stratford, Binkley



& Riddle, 1996). Durch den Einsatz einer Kontrollgruppe, in der kein Treatment erfolgt, können Retest-Effekte und der Regressionseffekt kontrolliert werden (Klauer, 2011). Der Retesteffekt besteht in der Veränderung der Posttestwerte gegenüber den Prätestwerten, ohne dass das Treatment einen Beitrag zur Veränderung geleistet hat. Dieser Effekt kann zum Beispiel dadurch zustande kommen, dass durch die Darbietung der Testaufgaben ein Trainingseffekt resultiert und dieser sich in den Werten der Post-Messung niederschlägt. Der Regressionseffekt besteht darin, dass Versuchspersonen, die im Prätest extreme Werte aufweisen, im Posttest tendenziell weniger extreme Werte zeigen. Dieser Effekt basiert auf unreliaiblen Messungen, also einem hohen Anteil von Messfehlern. Wenn aufgrund von Messfehlern eine Versuchsperson im Prätest einen extremen Wert erhält, ist die Wahrscheinlichkeit aufgrund der Verteilung der Messfehler höher, in der Postmessung einen weniger extremen Wert zu erhalten.

Es ist sehr wahrscheinlich, dass nicht alle Versuchspersonen sich im gleichen Ausmaß zwischen Prä- und Postmessung verändern, so dass einige sich verbessern, andere auf gleichem Niveau bleiben oder sich sogar verschlechtern. In diesem Fall sollte die Korrelation zwischen Prätest- und Posttestwerten nur mäßig hoch sein. Hohe Korrelationen sollten dagegen ein Indikator für eine geringe Änderungssensitivität sein (Klauer, 2011), da entweder keine Veränderung oder eine Veränderung über alle Versuchspersonen in ähnlichem Ausmaß stattgefunden hat. Letztere wäre ebenfalls ein Hinweis auf einen Retest-Effekt (Guthke & Wiedl, 1996).

### **Itemformate in der Lernverlaufsdiagnostik**

Die in der Lernverlaufsdiagnostik üblicherweise verwendeten Tests sind Verhaltensstichproben, mit denen prozedurales Wissen erfasst wird, im Wesentlichen die Lesefertigkeit, die Schreibfertigkeit oder die Rechenfertigkeit (Hosp, Hosp & Howell, 2007). Vergleichsweise selten werden in der Lernverlaufsdiagnostik Mehrfachwahlaufgaben verwendet, obwohl diesen eine hohe Objektivität, Zuverlässigkeit und Auswertungsökonomie bescheinigt wird. Mit Mehrfachwahlaufgaben kann vor allem deklaratives Wissen auf unterschiedlichen Ebenen (Reproduktion, Verständnis, Anwendung und Analyse) erfasst werden (Bloom, 1956; Shrock & Coscarelli, 2000). Im Rahmen der Lernverkaufsdiagnostik finden sich Mehrfachwahlaufgaben zum Teil in Lesetests, in denen an festgelegten Stellen im Text von drei Wörtern das korrekte ausgewählt werden muss (Walther, 2014). In einer Testserie zur Erfassung von Kompetenzen in Kindergärten und Schulen (KEKS; May, Bennöhr & Berger, 2014)

wird ebenfalls zum Teil auf Mehrfachwahlaufgaben zurückgegriffen. Beispielsweise werden dort für die Testung von Englisch-Hörverstehen und Deutsch-Grammatik Aufgaben eingesetzt, bei denen die Schülerinnen und Schüler eine korrekte aus insgesamt vier optionalen Antworten auswählen müssen. Auch in dem von Souvignier und Mitarbeitern (z. B. Souvignier, Förster & Salaschek, 2014) entwickelten Ansatz für internetbasierte Lernverlaufsdiagnostik („quop“) wird das Textverständnis mit Mehrfachwahlaufgaben erfasst.

## **Fragestellung und Hypothesen**

Mit dem vorliegenden Beitrag sollte die Frage beantwortet werden, ob Mehrfachwahlaufgaben änderungssensitiv sind und damit eine geeignete Testform für die Lernverlaufsdiagnostik darstellen. Die Prüfung der Änderungssensitivität erfolgte unter Kontrolle möglicher Retest-Effekte und der Testschwierigkeit. Es wurde erwartet, dass Mehrfachwahlaufgaben einen Lernzuwachs erfassen, wenn Versuchspersonen den Inhalt, der durch die Aufgaben erfasst wird, tatsächlich gelernt haben. Im Hinblick auf die Testschwierigkeit wurde angenommen, dass ein schwerer Test Veränderungen besser erfassen sollte und damit änderungssensitiver ist als ein leichter Test, da bei einem schweren Test niedrigere Prätest-Werte und damit ein größeres Veränderungspotential zu erwarten sind als bei einem leichten Test (Grigorenko & Sternberg, 1998; Guthke & Wiedl, 1996). Schließlich wurde angenommen, dass bei einem leichten Test höhere Retest-Reliabilitäten auftreten sollten als bei einem schweren Test (Klauer, 2011).

## **Methode**

### **Versuchspersonen und Versuchsdesign**

Insgesamt nahmen  $N = 180$  Studierende der Psychologie ( $M_{\text{Alter}} = 22.8$  Jahre ( $SD = 4.2$ ), 76 % weiblich) an der Untersuchung teil. Es wurde ein Prä-Post-Kontrollgruppendesign realisiert. Unabhängige Variablen waren das Lernmaterial (eine experimentelle Wortliste und eine Kontroll-Wortliste), die Testschwierigkeit (dreifach gestuft) und der Messzeitpunkt (prä versus post). Abhängige Variable war die Anzahl korrekt gelöster Testitems.

### **Versuchsmaterial**

Das Lernmaterial bestand aus zwei Wortlisten (Experimentalliste und Kontrollliste) mit jeweils 40 Wörtern. Die Wörter stammten aus der „new academic

word list“ (Coxhead, 2000; deutsche Übersetzung durch den Autor). Alle Wörter waren kleingeschrieben (und daher entweder Verben oder Adjektive), um ein Wiedererkennen aufgrund der Schreibweise auszuschließen. Die Wörter der Originalliste sind unterschiedlichen Häufigkeitskategorien zugeordnet (von 1 = sehr häufig bis 10 = sehr selten). Für die Erstellung des Lernmaterials wurden jeweils 8 Wörter zufällig aus jeder Häufigkeitskategorie ausgewählt (sofern sie ein Verb oder Adjektiv waren) und einer der beiden Wortlisten zugeteilt.

Der Test bestand aus 20 Aufgaben. Jede Aufgabe war eine Mehrfachwahlaufgabe mit vier Antwortoptionen. Die Antwortoptionen waren Wörter, von denen jeweils eines (das Zielwort) aus der Experimentalliste stammte. Keines der Distraktorwörter war in einer der beiden Wortlisten enthalten. Die Schwierigkeit des Tests wurde über die semantische Ähnlichkeit zwischen Zielwort und Distraktorwörtern operationalisiert (Mitkov, Ha, Varga & Rello-Sanchez, 2009). In der leichten Version des Tests war keines der Distraktorwörter dem Zielwort semantisch ähnlich. In der mittelschweren Version waren zwei Distraktorwörter dem Zielwort ähnlich, und in der schweren Version waren alle Wörter innerhalb einer Aufgabe semantisch ähnlich. Es wurde bei der Konstruktion der Testaufgaben darauf geachtet, dass die Distraktorwörter keine phonetische Ähnlichkeit zum Zielwort aufwiesen.

## **Versuchsablauf**

Die Versuchspersonen wurden zufällig den sechs Bedingungen zugewiesen. Die Hälfte der Vpn ( $n_E = 90$ ) sollte die Wörter der Experimentalliste lernen. Die vorgegebene Lernzeit betrug 90 Sekunden. Das Lernergebnis wurde anschließend mit dem Mehrfachwahlaufgaben-Test überprüft. Nach Abschluss der ersten Testphase folgten eine erneute Darbietung der Experimentalliste (für weitere 90 Sekunden) und ein anschließendes Überprüfen des Lernergebnisses mit dem gleichen Test. Zusätzlich wurde eine Kontrollbedingung ( $n_K = 90$ ) realisiert, in der die Vpn zwar den gleichen Test, jedoch eine andere Wortliste (die Kontrollliste) erhielten, um Retest-Effekte zu kontrollieren. Der Versuchsablauf der Kontrollbedingung entsprach dem der Experimentalbedingung.

## **Ergebnisse**

Tabelle 1 zeigt die Mittelwerte und Standardabweichungen der Anzahl korrekt gelöster Aufgaben für alle Bedingungen. Während in der Kontrollbedingung nur ein

geringfügiger mittlerer Zuwachs von Messzeitpunkt 1 zu Messzeitpunkt 2 ( $Diff = 0.20$ ) zu beobachten war, zeigte sich in der Experimentalgruppe ein deutlich höherer mittlerer Zuwachs ( $Diff = 1.44$ ). Eine 2 (Versuchsbedingung) x 2 (Messzeitpunkt) x 3 (Testschwierigkeit) ANOVA mit Messwiederholung über die Summe der korrekten Antworten als abhängiger Variable ergab einen signifikanten Haupteffekt der Versuchsbedingung,  $F(1, 174) = 518.50, p < .001, \eta^2 = .749$ , einen signifikanten Haupteffekt des Messzeitpunkts,  $F(1, 174) = 22.13, p < .001, \eta^2 = .113$ , eine signifikante Messzeitpunkt x Versuchsbedingung-Interaktion,  $F(1, 174) = 12.67, p < .001, \eta^2 = .068$ , eine signifikante Messzeitpunkt x Schwierigkeit-Interaktion,  $F(2, 174) = 3.17, p = .045, \eta^2 = .035$ , und eine signifikante Dreifach-Interaktion,  $F(2, 174) = 3.17, p = .045, \eta^2 = .035$ . Der Haupteffekt der Schwierigkeit,  $F(2, 174) = 1.23, p = .294, \eta^2 = .014$ , und die Versuchsbedingung x Schwierigkeit-Interaktion,  $F(2, 174) = 0.91, p = .405, \eta^2 = .010$ , waren nicht signifikant.

**Tab. 1: Mittelwerte und Standardabweichungen der Häufigkeit korrekt gelöster Aufgaben je Versuchsbedingung und Messzeitpunkt**

Messzeitpunkt	Testschwierigkeit	Experimentalgruppe		Kontrollgruppe	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Prätest	gering	15.20	3.17	7.20	2.02
	mittel	13.27	2.64	7.03	2.01
	hoch	13.67	2.40	7.60	2.13
Posttest	gering	15.40	2.42	7.40	2.51
	mittel	15.33	2.54	7.23	2.01
	hoch	15.73	2.98	7.80	2.61

Die signifikante Dreifach-Interaktion zeigt, dass in der experimentellen Bedingung die Änderungssensitivität dann am größten war, wenn der Test eine mittlere oder hohe Schwierigkeit aufwies.

Tabelle 2 zeigt die Koeffizienten der Retest-Reliabilitäten und die zugehörigen Fisher-Z-Werte für die sechs Bedingungen.

**Tab. 2: Koeffizienten der Retest-Reliabilitäten und die zugehörigen Fisher-Z-Werte**

Testschwierigkeit	Experimentalgruppe		Kontrollgruppe	
	r	Z	r	Z
gering	.79	1.07	.17	0.17
mittel	.65	0.78	.14	0.14
hoch	.70	0.87	.62	0.73

Inwieweit sich die Retest-Reliabilitäten voneinander unterscheiden, wurde mit dem  $z$ -Test überprüft. Zuvor wurden die Korrelationskoeffizienten in Fisher-Z-Werte transformiert. In der Experimentalbedingung war der Unterschied der Korrelationskoeffizienten zwischen dem leichten und dem mittelschweren Test signifikant,  $z = 1.95$ ,  $p = .026$ . Weder der Unterschied zwischen dem leichten und dem schweren Test,  $z = 1.35$ ,  $p = .089$ , noch zwischen dem mittelschweren und dem schweren Test,  $z = -0.61$ ,  $p = .271$ , war signifikant. In der Kontrollbedingung sollte kein Lerneffekt auftreten. Folglich sollten die Antworten der Probanden stark vom Zufall abhängig sein und daher nur geringe Zusammenhänge zwischen den beiden Messzeitpunkten bestehen. Tatsächlich zeigten sich für den leichten und den mittelschweren Test geringe Retest-Reliabilitäten, während der Zusammenhang zwischen beiden Messzeitpunkten im schweren Test deutlich höher war und sich signifikant sowohl vom leichten Test,  $z = 3.69$ ,  $p < .001$ , als auch vom mittelschweren Test,  $z = 3.89$ ,  $p < .001$ , unterschied.

## Diskussion

Die Ergebnisse dieser Studie deuten darauf hin, dass Mehrfachwahlaufgaben änderungssensitiv sind, wenn sie hinreichend schwierig sind. Dies zeigte sich sowohl in der Behaltensleistung als auch in der Korrelation zwischen beiden Messzeitpunkten. Somit scheinen Mehrfachwahlaufgaben prinzipiell geeignet zu sein für die Lernverlaufdiagnostik. Auffallend waren die Ergebnisse der Kontrollbedingung. Da in dieser Bedingung die Wörter der Lernliste keine Übereinstimmung mit den Wörtern des Tests aufwiesen, sollte die Häufigkeit korrekter Antworten der Ratewahrscheinlichkeit (25 %) entsprechen, d. h. es wären durchschnittlich 5 korrekte Antworten pro Versuchsperson zu erwarten gewesen. Tatsächlich aber lag die Anzahl korrekter Antworten im Mittel bei knapp 7.4. Denkbar ist, dass manche Zielwörter des Tests aufgrund semantischer oder phonologischer Ähnlichkeit zu Wörtern der Kontrollliste häufiger gewählt wurden als andere Zielwörter. Ferner zeigte sich in der Kontrollbedingung bei



hoher Testschwierigkeit eine vergleichsweise hohe Korrelation zwischen beiden Messzeitpunkten. Eine positive Korrelation zwischen beiden Messzeitpunkten kann auf einen Retest-Effekt zurückgeführt werden. Dieser war offenbar stärker ausgeprägt bei dem schwierigen Test als bei den beiden leichteren Tests. Möglicherweise führte die semantische Ähnlichkeit der Optionen zu einem besseren Wiedererkennen derjenigen Optionen, die im ersten Durchgang gewählt worden waren.

Inwieweit die Ergebnisse dieser Studie auf reale Situationen übertragbar sind, müsste in weiteren Untersuchungen geprüft werden. Es lässt sich jedoch vorsichtig optimistisch prognostizieren, dass das Potential von Mehrfachwahlaufgaben – die Erfassung deklarativen Wissens auf unterschiedlichen Ebenen (vgl. Bloom, 1956; Fink, 2003; Shrock & Coscarelli, 2000) – im Rahmen von Lernverlaufsdiagnostik ausgeschöpft werden kann.

## **Literatur**

- Bloom, B. S. (1956). *Taxonomy of educational objectives. The classification of educational goals. Handbook I: Cognitive Domain*. New York: McKay.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112, 155-159.
- Coxhead, A. (2000). A new academic word list. *TESOL Quarterly*, 34, 213-238.
- Deno, S. L. (2003). Developments in curriculum-based measurement. *The Journal of Special Education*, 37, 184-192.
- Fink, L. D. (2003). *Creating significant learning experiences. An integrated approach to designing college courses*. San Francisco: Wiley.
- Grigorenko, E. L. & Sternberg, R. J. (1998). Dynamic testing. *Psychological Bulletin*, 124, 75-111.
- Guthke, J. & Wiedl, K. H. (1996). *Dynamisches Testen. Zur Psychodiagnostik der intraindividuellen Variabilität*. Göttingen: Hogrefe.
- Hosp, M. K., Hosp, J. L. & Howell, K. W. (2007). *The ABCs of CBM: A practical guide to curriculum-based measurement*. New York: Guilford Press.
- Klauer, K. J. (2011). Lernverlaufsdiagnostik – Konzept, Schwierigkeiten und Möglichkeiten. *Empirische Sonderpädagogik*, 3, 207-224.
- Klauer, K. J. (2014). Formative Leistungsdiagnostik: Historischer Hintergrund und Weiterentwicklung zur Lernverlaufsdiagnostik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 1-17). Göttingen: Hogrefe.

- May, P., Bennöhr, J. & Berger, C. (2014). Lernentwicklungsmonitoring mit KEKS. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 258-280). Göttingen: Hogrefe.
- Mitkov, R., Ha, L. A., Varga, A. & Rello-Sanchez, L. (2009). Semantic similarity of distractors in multiple-choice tests: Extrinsic evaluation. In R. Basili & M. Pennacchiotti (Eds.), *Proceedings of the EACL 2009 Workshop* (pp. 49-56). Athens, Greece: Association for Computational Linguistics.
- Shrock, S. & Coscarelli, W. (2000). *Criterion-referenced test development* (2<sup>nd</sup> edition). Silver Spring: International Society for Performance Improvement.
- Souvignier, E., Förster, N. & Salaschek, M. (2014). Quop: Ein Ansatz internetbasierter Lernverlaufsdiagnostik mit Testkonzepten für Lesen und Mathematik. In M. Hasselhorn, W. Schneider & U. Trautwein (Hrsg.), *Lernverlaufsdiagnostik* (S. 239-256). Göttingen: Hogrefe.
- Stratford, P. W., Binkley, J. M. & Riddle, D. L. (1996). Health status measures: Strategies and analytic methods for assessing change scores. *Physical Therapy*, 76, 1109-1123.