Testen als evidenzbasierte Lernmethode: Empirische und theoretische Gründe für eine Anwendung im Unterricht

Jonathan Barenberg · Stephan Dutke

Eingegangen: 31. März 2021 / Überarbeitet: 16. November 2021 / Angenommen: 16. November 2021 /
Online publiziert: 4. Dezember 2021
© Der/die Autor(en) 2021

Zusammenfassung In Schule und Hochschule werden Tests meist als Instrument zur Leistungsüberprüfung und -bewertung eingesetzt. Dabei weisen Ansätze des evidenzbasierten Lehrens und Lernens darauf hin, dass Testen als Abrufübung auch zur Lernförderung eingesetzt werden können. Es werden drei Argumente erörtert, die für eine Anwendung von Abrufübungen als Lernmethode in der Unterrichtspraxis sprechen: (1) Entwicklung der Testeffektforschung von Laborkontexten zu Anwendungskontexten sowie direkte Effekte des Testens auf (2) den Transfer von Wissen und (3) auf unterrichtsrelevante, metakognitive Prozesse. Abschließend werden sowohl Schlussfolgerungen für die Umsetzung von Abrufübungen im Unterricht beschrieben als auch Konsequenzen für den Wissenschafts-Praxis-Transfer diskutiert.

Schlüsselwörter Testeffekt · Abrufübung · Transfer · Metakognition · Evidenzbasiertes Lehren und Lernen

Jonathan Barenberg (✉)
Westfälische Wilhelms-Universität Münster, Münster, Deutschland
E-Mail: barenberg@uni-muenster.de

Stephan Dutke
Westfälische Wilhelms-Universität Münster, Münster, Deutschland

Testing, an evidence-based learning method: Empirical and theoretical reasons for application in classroom

Abstract In schools and universities, tests are usually applied to assess learning outcomes. Within the evidence-based teaching framework, however, testing and retrieval practice are also recommended as a strategy which supports learning and retention. Three reasons for applying retrieval practice as a learning method in classroom are discussed: (1) the trend in testing effect research from investigating laboratory settings to applied settings, direct effects of testing on (2) transfer of knowledge, (3) on metacognitive monitoring processes that are relevant for teaching and learning in classroom. Finally, conclusions are drawn for the implementation of retrieval practice in classrooms, and conclusions for the transfer of evidence-based teaching and learning principles from science to practice.

Keywords Testing effect · Retrieval practice · Transfer · Metacognition · Evidence-based teaching

1 Testen als Abrufübung im Kontext evidenzbasierten Lehrens

Versuche, den Erfolg intentionaler Lernprozesse zu erhöhen, setzen häufig bei Enkodierungsprozessen an, also jenen Prozessen, die für eine Verankerung neuer Inhalte im Langzeitgedächtnis sorgen. Bemessen wird der Lernerfolg in vielen Situationen aber daran, inwieweit Gedächtnisinhalte in einer Anforderungssituation tatsächlich zur Verfügung stehen, also inwieweit sie aus dem Langzeitgedächtnis wieder abgerufen werden können. Dies hängt nicht allein von der Enkodierung ab. Die Idee von Abrufübungen geht vielmehr davon aus, dass wiederholtes Abrufen die Verfügbarkeit von gelernten Inhalten erhöht, weitgehend unabhängig von den vorausgegangenen Enkodierungsprozessen. „Testeffekt“ bezeichnet das Phänomen, dass Abrufübungen (z. B. durch einen Test oder ein Quiz) den erneuten Abruf von Lerninhalten begünstigt. Der Effekt des Übens von Abrufprozessen aus dem Langzeitgedächtnis gilt als empirisch gut bestätigt (für Reviews bzw. Metaanalysen siehe z. B. Dunlosky et al. 2013; Roediger und Butler 2011; Rowland 2014; Schwieren et al. 2017; Yang et al. 2021) und kann auch auf theoretischer Ebene erklärt werden (z. B. Karpicke et al. 2014).

Die Erforschung des Testeffekts ist ein gutes Beispiel für den Versuch, Erkenntnisse über grundlegende menschliche Lern-, Gedächtnis- und Motivationsprozesse für unterrichtspraktische Entscheidungen nutzbar zu machen. Dieser Ansatz des evidence-based teaching (z. B. Cranney 2013; Dutke et al. 2017; Schwartz und Gurung 2012) bietet Regelhaftigkeiten menschlichen Lernens als leitend für die förderliche Gestaltung praktischer Lernsituationen an (Graesser 2009). Der Begriff „evidence-based“ verweist darauf, dass die Grundlagen dieses Ansatzes ausführlich empirisch untersucht sind und es auf theoretischer Ebene Erklärungsansätze gibt. Seit 2008 sind im Rahmen des *evidence-based teaching* Ansatzes mehrere Sammlungen solcher begründeten Prinzipien menschlichen Lernens publiziert worden (z. B. Cranney 2013; Dunlosky und Rawson 2015; Dunlosky et al. 2013; Dunn et al. 2013; Graesser

2009; Graesser et al. 2008; Roediger und Pyc 2012). Die inhaltliche Überlappung zwischen diesen Sammlungen ist groß, was nicht verwundert, da sie weitgehend auf den gleichen empirischen Befunden basieren. Deshalb nimmt der Testeffekt in all diesen Sammlungen auch stets eine prominente Position ein.

Die Autor*innen dieser Sammlungen empfehlen explizit die Berücksichtigung lernpsychologischer Prinzipien in der Gestaltung von praktischen Lehr-Lernsituationen. Allerdings sind diese Empfehlungen nicht als normative Vorschriften zu verstehen, da sie auf einer weitestgehend wertfreien Beschreibung von kognitiven Prozessen beruhen. Charakteristisch für diesen Ansatz ist, dass grundlegende Prozesse der menschlichen Informationsverarbeitung als weitgehend unabhängig von spezifischen Lerngegenständen, Lernkulturen, gesellschaftlichen Erwartungen und Normen erachtet werden, da sie eher an Funktionsmerkmalen des menschlichen kognitiven Systems orientiert sind als an fachlich, gesellschaftlich oder kulturell geprägten Unterrichtskontexten. Wenn in diesem Kontext also empfohlen wird, bestimmte theoretisch und empirisch begründete Elemente in Unterrichtsprozesse zu integrieren, ist das als ein wissenschaftlich begründetes Angebot zu verstehen, nicht als die Vorwegnahme einer wertenden Entscheidung, wie Unterricht im Einzelfall gestaltet werden soll.

2 Lernförderliche Effekte von Abrufübungen und Testen

Rowland (2014) unterscheidet zwischen direkten und indirekten Effekten des Testens. Mit indirekten Effekten des Testens sind Veränderungen des Lernverhaltens gemeint. Die Erwartung eines Tests kann beispielsweise motivationale Effekte haben und dem Test *vorausgehendes Lernhandeln* beeinflussen (z.B. durch Anstrengungssteigerung, um ein möglichst gutes Testergebnis zu erzielen). Das Ergebnis eines (Übungs-)Tests liefert Lernenden aber auch Informationen darüber, welche Lerninhalte abrufbar sind und welche (noch) nicht. Diese diagnostische Information können Lernende nutzen, um *nachfolgende Lernaktivitäten* daran auszurichten (z.B. verstärkt Aufmerksamkeit auf Lerninhalte zu richten, die nicht abrufbar sind oder bei deren Abruf Unsicherheit besteht), mit dem Ziel, zukünftige Testergebnisse zu verbessern. Genauso können Lehrende diese Ergebnisinformation zur Ausrichtung ihres *nachfolgenden Lehrhandelns* nutzen (z.B. indem sie die Behandlung eines Unterrichtsthemas intensivieren, zeitlich ausdehnen oder abkürzen).

Mit direkten Effekten des Testens sind dagegen Veränderungen im Langzeitgedächtnis gemeint, die der Abrufprozess selbst bewirkt. So kann beispielsweise der erfolgreiche Abruf einer Information zu einer Stärkung von Verknüpfungen im semantischen Netzwerk durch elaborative Prozesse führen. Diese im engeren Sinne abrufbedingten Effekte sollen im vorliegenden Beitrag näher betrachtet werden, vor allem unter der Perspektive ihrer Relevanz für die Unterrichtspraxis.

In der schulischen wie in der hochschulischen Praxis wird Testen vorrangig als Instrument zur Betrachtung und Bewertung des Ergebnisses von Abrufprozessen verstanden, was auch lernförderlich sein kann, wenn dadurch effektive Lernaktivitäten angeregt werden (siehe auch Endres & Renkl in diesem Band). Weit seltener kommt Testen als Instrument zum Einsatz, um tatsächlich den Abrufprozess von

Informationen aus dem Langzeitgedächtnis zu fördern. Ein wichtiger Grund hierfür scheint die fehlende Kenntnis zu sein, dass der Abrufprozess selbst das langfristige Behalten wirksam fördern kann. Lernende verstehen Testen eher als Instrument zur Beurteilung des Lernstands (Hartwig und Dunlosky 2012; Karpicke et al. 2009; Kornell und Son 2009) und unterschätzen die Wirksamkeit von Abrufübungen, selbst wenn sie dazu angeregt werden, diese einzusetzen (Potts und Shanks 2014).

3 Direkte Effekte des Testens und Abrufens in der Unterrichtspraxis

Mit der primären Fokussierung auf Ergebnisse von Abrufprozessen und der Wahrnehmung des Testens als Mittel des Bewertens und Motivierens wird Potenzial zur Verbesserung der Abrufbarkeit von neu erworbenen Lerninhalten vergeben. Für den Einsatz von Abrufübungen in der Unterrichtspraxis spricht, dass direkte Effekte des Testens zunehmend auch in unterrichtlich relevanten Anwendungskontexten gut belegt sind (siehe Abschn. 3.1), diese nicht auf den Informationsabruf begrenzt sind, sondern sich auch auf Leistungen mit Transferanforderungen (siehe 3.2) und auf die Genauigkeit von metakognitivem Monitoring auswirken können (siehe 3.3).

3.1 Von der Grundlagenforschung zur angewandten Forschung

Ursprünglich sollten Studien zu den Effekten wiederholten Abrufs der theoretischen und empirischen Differenzierung von Enkodier- und Abrufprozessen dienen. In den Worten einer der ersten Studien zu diesem Thema bestand das primäre Untersuchungsziel darin, den „relative value of recall and *Einprägung*“ zu bestimmen (Abbott 1909, S. 177, Hervorhebung im Original). Ein zweites wichtiges Erkenntnisziel betraf die Frage, inwieweit der versuchte oder vollendete Abruf eine bereits etablierte Gedächtnisrepräsentation verändern kann (z.B. McDaniel und Masson 1985). Beide Forschungsfragen beziehen sich in erster Linie auf grundlagenwissenschaftliche Erkenntnisziele, (noch) nicht primär auf angewandte Fragestellungen. Die Übertragbarkeit dieser Erkenntnisse in unterrichtliche Anwendungsfelder war sicherlich früh gegeben, der Nachweis eines Nutzens in diesen Feldern fehlte jedoch zunächst.

Dies änderte sich, als Testeffektstudien begannen, die Wirkung von Abrufübungen gezielt in Anwendungskontexten wie Schule und Hochschule zu untersuchen. In den letzten 30 Jahren haben zahlreiche empirische Befunde über unterschiedliche Altersgruppen (von Vorschulkindern bis zu älteren Erwachsenen), über unterschiedliche Lernmaterialien (von einfachem Wortmaterial wie Wortlisten bis zu komplexeren Material wie Texten) und über unterschiedliche Testformate (von geschlossenen bis offenen Antwortformaten) die Generalisierbarkeit des Testeffekts dokumentiert und eine breite empirische Basis für die Anwendung des Testens zur Lernförderung geschaffen (vgl. Dunlosky et al. 2013; Rowland 2014). Darüber hinaus haben insbesondere zwei Befunde die Aufmerksamkeit der Forschung auf die Anwendung des Testens in praktischen Lehr- und Lernkontexten gelenkt und weitere Forschung dazu angeregt: (1) lernförderliche Effekte des Testens (z.B. Butler und Roediger 2007; Kang et al. 2007) mit unterrichtsrelevantem Lernmaterial (z.B. beim Lernen

mit Texten und Unterrichtsvideos), (2) langfristige Effekte des Testens auf die Gedächtnisleistung, auch nach mehreren Wochen (z. B. Butler und Roediger 2007) und Monaten (z. B. McDaniel et al. 2011).

Zwei jüngere Metaanalysen (Adesope et al. 2017; Schwieren et al. 2017) haben sich explizit mit der Frage der Anwendung von Abrufübungen in Lehr-Lernkontexten auseinandergesetzt. Schwieren et al. (2017) fanden über 19 Publikationen mit 72 Effektgrößen hinweg einen mittleren lernförderlichen Effekt des Testens in psychologischen Lehr-Lernkontexten ($d=0,56$), der ähnlich stark war wie Gesamteffekte aus früheren Metanalysen (z. B. Rowland 2014; $g=0,50$), die deutlich mehr Studien unter Laborbedingungen einschlossen. Adesope et al. (2017) verglichen innerhalb ihrer Meta-Analyse auch Studien, die in einem Laborkontext erhoben wurden, mit Studien, die in einen Lehr-Lernkontext eingebettet waren. Auch hier wurden vergleichbare Effektstärken in Labor- ($g=0,62$) und Anwendungskontext ($g=0,67$) berichtet.

Zusammengefasst bestätigen die empirischen Ergebnisse, dass die Wirksamkeit von Abrufübungen für die Verfügbarkeit von Wissen nicht nur in laborexperimentellen Settings gut belegt ist, sondern auch in anwendungsnahen Lehr-Lernkontexten nachgewiesen werden kann. Allerdings weisen die metaanalytischen Befunde (z. B. Rowland 2014; Yang et al. 2021) auch auf einige Moderatoren hin (z. B. Abruf Erfolg in Abrufübungen, Häufigkeit von Abrufübungen, Einsatz von korrektivem Feedback), die die Wirksamkeit von Abrufübungen mitbestimmen. Gerade für die Anwendung in tatsächlichen Lernsettings besteht hier noch weiterer Forschungsbedarf zu potenziellen Moderatoren. Darüber hinaus kann man die Frage, was ein authentisches Lernsetting ausmacht, noch weiter ausdifferenzieren und auch spezifische methodische Kritikpunkte zu Testeffektstudien in authentischen Lernsettings identifizieren, die praxisrelevant und klärungsbedürftig sind (z. B. Barenberg und Dutke 2021). Grundsätzlich hat sich der Forschungsfokus jedoch deutlich und erfolgreich in Richtung auf das Anwendungsfeld Unterricht zubewegt und belegt die Anwendbarkeit von Abrufübungen zur Lernförderung. Hieran anschließend werden zwei weitere direkte Effekte des Testens behandelt, die insofern unterrichtsrelevant sind, als sie über den Informationsabruf aus dem Langzeitgedächtnis hinausgehen: Effekte von Abrufübungen auf den Transfer gelernten Wissens und auf die metakognitive Überwachung des eigenen Lernprozesses.

3.2 Effekte von Abrufübungen auf den Transfer von Wissen

Die Fähigkeit, zuvor enkodiertes Wissen in einer Anforderungssituation wieder abrufen zu können, stellt zwar eine wichtige Voraussetzung für weitergehende Lernprozesse dar, ist aber nicht das alleinige Ziel von Unterricht. Nach Angebot-Nutzungs-Modellen des Unterrichts (z. B. Klieme 2006) wird Unterricht als eine Menge von Angeboten aufgefasst, die von Schüler*innen „im eigenen Handeln genutzt werden“ und zu „Bewußtheiten (sic), Kenntnissen und Fertigkeiten führen“ (Petersen 1937, zitiert nach Prell 1985, S. 27). Auch der in diesem Kontext oft benutzte Kompetenzbegriff (z. B. Rychen und Salganik 2003) geht über die Abrufbarkeit von Wissen hinaus und betont das Ziel, Wissen anwenden zu können. Dies erfordert in der Regel

einen Transfer von Wissen aus einem Aneignungskontext in einen Anwendungskontext.

Im Vergleich zu diesen Zielbegriffen von Unterricht berichten viele Testeffektstudien oft schlichtere Effekte bzw. verwenden Studiendesigns, die Transfereffekte des Testens außer Acht lassen. Experimentelle Testeffektstudien umfassen in der Regel drei Phasen: initiale Lernphase, Übungsphase mit bzw. ohne Abrufübung und Prüfphase. Ein Kritikpunkt an diesem Paradigma ist, dass in vielen Studien die Lerninhalte in der Übungs- und in der Prüfphase unter sehr ähnlichen Testbedingungen abgerufen werden (Carpenter 2012). Im Unterricht unterscheiden sich jedoch häufig die Bedingungen, unter denen sich Lernende einen Lerninhalt aneignen, von den Bedingungen, unter denen sie den gleichen Lerninhalt wieder abrufen sollen (z. B. in einer anderen Unterrichtssituation oder in einer Prüfung). Für die Anwendung von Testen zur Lernförderung in Schule und Hochschule ist es deshalb bedeutsam, ob der förderliche Effekt des Testens auch dann zum Tragen kommt, wenn Übungs- und Prüfbedingungen sich unterscheiden, d. h. ein Transfer in einen neuen Kontext erforderlich ist.

3.2.1 Theoretische Modelle zu Transfereffekten von Abrufübungen

Aus theoretischer Sicht lassen sich hier Modelle der allgemeinen Forschungsliteratur zum Lerntransfer und spezifischere Modelle der Testeffektliteratur unterscheiden (vgl. Pan und Rickard 2018). In der allgemeinen Forschungsliteratur zum Lerntransfer wird zwischen abstraktions- und ähnlichkeitsbasierten Modellen unterschieden (Barnett und Ceci 2002). Bei den abstraktionsbasierten Modellen geht es darum, dass Lernende ein grundlegendes, verallgemeinerbares Prinzip erlernt haben, welches sie auch in Kontexten anwenden können, die sich deutlich vom Aneignungskontext unterscheiden. Das wiederholte Abrufen solch eines verallgemeinerbaren Prinzips durch Testen könnte auch den Abruf in anderen Kontexten erleichtern (z. B. Darabi et al. 2007). Bei den ähnlichkeitsbasierten Modellen, wie dem Ansatz der transferangemessenen Verarbeitung (transfer appropriate processing, Morris et al. 1977), geht es hingegen darum, dass Lernende ähnliche, überlappende Elemente zwischen Aneignungs- und Anwendungskontext erkennen und deshalb Gelerntes von einem Kontext in einen anderen transferieren können. So könnten beispielsweise ähnliche Hinweisreize in Übungs- und Prüftests den Abruf gewünschter Lerninhalte erleichtern, auch wenn sich die Kontexte in anderen Aspekten unterscheiden (z. B. hinsichtlich des Testformats). Folglich weisen sowohl abstraktions- als auch ähnlichkeitsbasierte Modelle in der Theorie auf das Potenzial von Transfereffekten des Testens hin.

Dieser Eindruck wird durch spezifischere Modelle der Testeffektliteratur ergänzt (beispielsweise durch das Modell des elaborativen Abrufs, Bjork 1975; Carpenter 2009). Nach diesem Modell wird durch den Abruf von Lerninhalten aus dem Langzeitgedächtnis nicht nur die Gedächtnisspur dieser Inhalte gestärkt, sondern es werden dabei auch Prozesse in Gang gesetzt, die diesen Gedächtnisinhalt elaborieren und mit anderen Inhalten in Verbindung bringen. Es wird angenommen, dass der Abruf einer Information zu einem automatischen Prozess der Aktivierungsausbreitung im semantischen Netzwerk der abgerufenen Information führt. Dadurch werden

nicht nur verschiedene Verknüpfungen der Zielinformation mit anderen Informationen im semantischen Netzwerk gestärkt, sondern auch Verknüpfungen innerhalb des Netzwerks, die nicht die Zielinformation enthalten. Diese Prozesse können also auch den zukünftigen Abruf anderer Informationen als der im Übungskontext abgerufenen begünstigen – ein weiterer Hinweis auf potenzielle Transfereffekte des Testens.

3.2.2 Empirische Befunde zu Transfereffekten von Abrufübungen

Carpenter hat im Jahre 2012 erstmals die bestehende Testeffektliteratur zu dieser Frage gesichtet und empirische Befunde zu Transfereffekten des Testens zusammengefasst. Trotz der damals noch vergleichsweise geringen Zahl von Studien fand sie erste Hinweise auf Transfereffekte des Testens, beispielsweise über verschiedene Testformate und über verschiedene Wissensdomänen hinweg. Beide Transferarten erscheinen auch für die Anwendung in Lehr-Lernkontexten von besonderer Bedeutung. Carpenters Überblicksartikel hat in den darauffolgenden Jahren zu weiterer Forschungsaktivität in dieser Frage angeregt. Pan und Rickard identifizierten (2018) in einer Meta-Analyse zu Transfereffekten des Testens bereits 67 Studien mit 122 Experimenten und 192 Effektstärken. Dieser breitere empirische Zugang umfasste auch weitere Formen des Transfers, die für Unterricht relevant sind (z. B. Transfer auf Anwendungs- und Problemlöseaufgaben, auf schlussfolgerndes Denken und auf den Abruf von zuvor nicht getesteten Inhalten). Über alle Studien hinweg fanden die Autoren einen mittelgroßen Transfereffekt des Testens ($d=0,40$) und bestätigten damit die ersten Hinweise und Schlussfolgerungen von Carpenter (2012) mit meta-analytischen Methoden. Allerdings variierte dieser Effekt deutlich zwischen Studien sowie zwischen verschiedenen Formen von Transfer. Daher wollen wir näher auf die Befundlage für drei Formen von Transfer eingehen, die für den Anwendungskontext in Schule und Hochschule von besonderer Bedeutung sind.

Transfer über Testformate Eine basale, aber häufige Anforderung in der Unterrichtspraxis ist die Fähigkeit von Lernenden, gelerntes Wissen in verschiedenen Situationen abrufen zu können, die sich in der Gestaltung und Anforderung der Abrufsituation unterscheiden (Testformat). So ist es im Fremdsprachenunterricht beispielsweise relevant, dass Lernende Vokabeln nicht nur von der Fremdsprache ins Deutsche abrufen können, sondern auch umgekehrt vom Deutschen in die Fremdsprache. Beim Lernen aus Texten sollen Lernende sowohl unter verschiedenen Alternativen die richtige Information aus dem Text wiedererkennen und identifizieren (Multiple-Choice) als auch bei offenen Fragen die antwortrelevante Information abrufen können. Gerade der letzte Punkt, Informationen in geschlossenen und offenen Testformaten (mit oder ohne Hinweisreizen) abrufen zu können, gilt für eine Vielzahl von Anwendungsfächern und -kontexten. Daher ist die Frage relevant, ob beim Einsatz von Abrufübungen nur der Abruf in einem bestimmten Format trainiert bzw. gefördert wird oder auch der Abruf in anderen Formaten begünstigt wird. Pan und Rickard (2018) haben in ihrer Meta-Analyse 56 Effektstärken aus 29 Studien zu dieser Fragestellung ausgewertet und fanden hier einen mittelgroßen Transfereffekt

des Testens ($d=0,58$). Das ist ein deutlicher Hinweis darauf, dass Abrufübungen die Verfügbarkeit von Wissen in variablen Abrufsituationen unterstützen können.

Transfer auf Anwendungs- und Inferenzaufgaben Eine weitere Anforderung an Lernende ist, dass sie gelerntes Wissen nicht nur abrufen, sondern auch anwenden bzw. Inferenzen (Schlussfolgerungen) aus dem gelernten Wissen ziehen können. Im Fremdsprachenunterricht ist es beispielsweise relevant, nicht nur Vokabeln übersetzen zu können (Fremdsprache-Deutsch und umgekehrt), sondern auch beim Sprechen und Schreiben passende Vokabeln abrufen und im Kontext anwenden zu können; beim Lernen aus Texten sind Lernende häufig gefordert, nicht nur Informationen aus dem Text abrufen und wiedergeben zu können (z.B. den Prozess der Photosynthese zu erklären), sondern auch Informationen aus dem Text abrufen und auf neue Situationen anwenden zu können und Schlussfolgerungen zu ziehen (z.B. zu erklären, warum sich das Wachstum von Pflanzen an verschiedenen Standorten unterscheidet). Daher ist eine weitere relevante Frage, ob der Einsatz von Abrufübungen nur die Leistung in einfachen Abrufsituationen begünstigt oder auch die Leistung in Aufgaben, die schlussfolgerndes Denken und eine Anwendung von Wissen erfordern. Pan und Rickard (2018) haben in ihrer Meta-Analyse 41 Effektstärken aus 17 Studien zu dieser Fragestellung ausgewertet und fanden hier einen kleinen Transfereffekt des Testens ($d=0,32$). Das ist ein Hinweis darauf, dass Abrufübungen auch die Leistung in Anwendungsaufgaben substantiell unterstützen können. Die geringere Effektstärke könnte darauf hinweisen, dass die Verfügbarkeit von Wissen *eine* wichtige Grundlage für die Leistung in Anwendungsaufgaben mit schlussfolgerndem Denken ist, aber weitere Variablen diese Leistung entscheidend mitbeeinflussen.

Transfer auf nicht geübte Inhalte Sowohl Lehrende als auch Lernende sind in der Unterrichtspraxis bei umfangreichen Lerninhalten häufig gefordert zu entscheiden, welche Inhalte sie nach einer ersten Lernphase in welchem Umfang vertiefen, üben oder wiederholen. Im Fremdsprachenunterricht beispielsweise kommen ständig neue Vokabeln dazu, sodass nach dem ersten Kennenlernen neuer Vokabeln diese oft nur noch selektiv erneut aufgegriffen und wiederholt werden; beim Lernen aus Texten werden auch selten alle Informationen eines Textes in gleicher Intensität und Ausführlichkeit im Unterricht behandelt. Daher ist es eine interessante Frage, ob der Einsatz von Abrufübungen neben dem Abruf der getesteten bzw. abgefragten Lerninhalte auch den Abruf nicht getesteter (abgefragter) Lerninhalte begünstigen kann. Pan und Rickard (2018) haben in ihrer Meta-Analyse 17 Effektstärken aus 12 Studien zu dieser Fragestellung ausgewertet und fanden hier keinen signifikanten Transfereffekt des Testens ($d=0,16$). Allerdings waren die zu diesem Zeitpunkt verfügbaren Daten noch sehr überschaubar. In einer aktuellen Meta-Analyse haben Yang et al. (2021) 43 Effektstärken zu dieser Fragestellung ausgewertet und fanden einen kleinen Transfereffekt des Testens ($d=0,32$). Das ist ein Hinweis darauf, dass Abrufübungen auch den Abruf nicht getesteter Lerninhalte substantiell unterstützen können. Entscheidend für den Erfolg dieser Art von Transfer ist der semantische Zusammenhang zwischen getesteten und nicht getesteten Lerninhalten (Little et al. 2011), der zwischen den Studien sehr stark variierte. Für den Fremdsprachenunter-

richt ist es beispielsweise nachvollziehbar, dass der Abruf einzelner Vokabeln aus dem Bereich Tiere auch die Gedächtnisspur nicht abgefragter Vokabeln aus diesem Bereich etwas erhöhen kann (durch automatische Aktivierungsausbreitung im semantischen Netzwerk), aber nicht die Gedächtnisspur anderer nicht assoziierter Vokabeln. Und beim Lernen aus Texten ist es ebenfalls nachvollziehbar, dass das Abrufen einzelner Informationen eines bestimmten Themas (z. B. Römische Verfassung) auch andere damit verknüpfte Informationen aktiviert, aber keinen Einfluss hat auf nicht assoziierte Informationen (z. B. das römische Straßennetz).

Zusammengefasst hat die jüngere Testeffektforschung eine zunehmend breitere empirische Basis hervorgebracht, die auf förderliche Effekte von Abrufübungen auch für den Transfer von Wissen hinweisen. Pan und Rickard (2018) stellten in ihrer Meta-Analyse zwar fest, dass über alle Studien hinweg die Effekte von Abrufübungen kleiner ausfallen, wenn Transfer gefordert ist ($d=0,40$) als wenn kein Transfer notwendig ist ($d=0,68$). Dennoch zeigt die Befundlage, dass Abrufübungen für (hoch)schulrelevante Transferanforderungen förderlich sein können (z. B. Transfer über Testformate, Transfer auf Anwendungs- und Inferenzaufgaben und teilweise sogar Transfer auf nicht geübte Lerninhalte).

3.3 Effekte von Abrufübungen auf die Genauigkeit metakognitiven Monitorings

Neben den abrufbedingten Effekten auf Verfügbarkeit und Transfer von Wissen gibt es auch Belege dafür, dass Abrufübungen die Genauigkeit des metakognitiven Monitorings verbessern können (Rivers 2020). Metakognitives Monitoring bezeichnet die Überwachung des eigenen Lernprozesses und wird oft anhand voraus- oder rückschauender subjektiver Einschätzungen von Lernprozessen und -ergebnissen beurteilt. Die Einschätzung des eigenen Lernprozesses und seiner Ergebnisse stellt eine wichtige Grundlage für die Selbstregulation von Lernen dar (vgl. Nelson und Narens 1990; Winne und Hadwin 1998; Zimmerman 2008). Je genauer das metakognitive Monitoring ist, je genauer es also tatsächliche Lernergebnisse abbildet, umso besser sind Lernende in der Lage, Schwierigkeiten beim Lernen zu erkennen (z. B. mangelnden Fortschritt beim Lernen oder Unsicherheiten beim Abruf oder in der Anwendung von Lerninhalten) und diese Information zu nutzen, um ihr Lernen effektiv zu regulieren (z. B. durch eine Veränderung der Lernstrategie oder die veränderte Zuweisung von Lernzeit). Häufig wird Testen mit metakognitivem Monitoring in Zusammenhang gebracht, weil das Ergebnis von Testen und Abrufen eine diagnostische Information über den Lernstand liefert, der im Sinne metakognitiven Monitorings auch für die Regulation nachfolgender Lernaktivitäten genutzt werden kann (siehe auch Endres & Renkl in diesem Band). Darüber hinaus kann der Prozess des Abrufens selbst aber auch einen unmittelbaren Effekt auf die Genauigkeit des metakognitiven Monitorings haben.

Hinweise darauf, dass Testen und Abrufen die Genauigkeit metakognitiven Monitorings erhöhen kann, stammen u. a. aus der Forschung zu vorausschauenden Einschätzungen der Lernleistung, sogenannten Judgments of Learning (JOL, Kornell und Rhodes 2013). Dabei werden Lernende nach einer ersten Lernphase gebeten einzuschätzen, für wie wahrscheinlich sie es halten, bestimmte Inhalte in einer spä-

teren Überprüfung korrekt abrufen zu können. Nach dieser Überprüfung können die JOL mit der tatsächlich Abrufleistung verglichen werden. So kann festgestellt werden, wie genau die Lernenden das tatsächliche Ergebnis einer Überprüfung des Gelernten vorhergesagt haben. Häufig beruhen JOL auf einem Abrufversuch aus dem Gedächtnis. Daher sind die JOL genauer, wenn sie nicht unmittelbar nach der Lernphase, sondern mit einer Verzögerung gegeben werden (z.B. Dunlosky und Nelson 1997; Schneider et al. 2000). Denn nach einer solchen Verzögerung erfolgt ein Abrufversuch tatsächlich aus dem Langzeitgedächtnis und nicht aus dem Arbeitsgedächtnis. Die erlebte Leichtigkeit bzw. Schwierigkeit des Abrufs ist dann ein recht zuverlässiger Prädiktor für das Gelingen oder Misslingen eines späteren Abrufs (z.B. Koriat 1997). Folgerichtig erhöht sich die Genauigkeit von JOL noch einmal, wenn die betroffenen Inhalte vor dem JOL schon einmal abgerufen wurden (für eine Metaanalyse: Rhodes und Tauber 2011).

Positive Effekte des Abrufs auf die Genauigkeit von JOL konnten für unterschiedliche Lernmaterialien nachgewiesen werden, von einfachen Wortlisten (z.B. King et al. 1980) über das Lernen aus Sachtexten (z.B. Little und McDaniel 2015) bis hin zum Lernen in einem kompletten Studienmodul (z.B. Fernandez und Jamet 2017). Von praktischer Bedeutung ist weiterhin, dass Testen eine Überschätzung der eigenen Leistung verringerte, also eine zu optimistische Prognose des eigenen Lernerfolgs zu vermeiden half (z.B. Cogliano et al. 2019), und den Einfluss von Unterschieden im Vorwissen auf die Leistung im finalen Test neutralisierte. Der positive Effekt des Testens auf die Güte von JOL bleibt auch erhalten, wenn man das Ausmaß des vorherigen Lernens kontrolliert, also beispielsweise die Häufigkeit, mit der Lernende in der Lernphase mit einem zu erlernenden Sachverhalt in Kontakt kommen (z.B. Shaughnessy und Zechmeister 1992).

Die genannten Untersuchungsbeispiele haben die Genauigkeit metakognitiven Monitorings über die Güte der Vorhersage des Lernerfolgs erfasst. Barenberg und Dutke (2019) haben darüber hinaus gezeigt, dass Abrufübungen auch das Potenzial haben, die Genauigkeit von metakognitivem Monitoring in einer späteren Prüfungssituation zu verbessern – also nicht im Sinne einer Vorhersage, sondern als Urteil über das Vertrauen in die Korrektheit der in einer Prüfungssituation gegebenen Antworten. Sie ließen Schüler*innen aus einem Sachtext lernen und testeten Inhalte kurz nach der Lernphase und eine Woche später. Die bereits getesteten Items wurden im zweiten Test nicht nur häufiger korrekt beantwortet als Items, die im ersten Test nicht vorkamen, sondern die Antworten wurden auch mit größerem Vertrauen in ihre Richtigkeit gegeben. Dabei vertrauten die Lernenden aber nicht allen Antworten auf bereits getestete Items in erhöhtem Maße, sondern vor allem jenen Antworten, die tatsächlich korrekt waren. Das Ergebnis zeigt, dass Testen die Genauigkeit metakognitiven Monitorings nicht nur in einer vorausschauenden Perspektive erhöht, sondern auch in einer rückschauenden Perspektive auf bereits gegebene Testantworten. Eine mögliche Erklärung ist auch hier, dass die Erfahrung des leichteren oder schwereren Abrufs als metakognitiver Hinweisreiz für spätere Abrufsituationen dient und die Angemessenheit von Konfidenzurteilen in diesen Situationen erhöhen kann.

3.4 Zusammenfassung der Befunde zu direkten Effekten des Testens und Abrufens

Die Erforschung der direkten Effekte von Abrufübungen zeigt eine kontinuierliche Entwicklung. Die Frage, ob das Abrufen von Gedächtnisinhalten einen eigenen Beitrag zum Lernen liefert, wurde früh in der psychologischen Gedächtnisforschung gestellt und mit labor-experimentellen Methoden untersucht. Der Testeffekt ist empirisch gut bestätigt, auch unter metaanalytischer Perspektive. Ergänzt wurde die grundlagenwissenschaftliche Perspektive durch experimentelle und quasi-experimentelle Feldstudien. Diese zeigten, dass Abrufübungen auch im Kontext realer Lernsettings (z. B. im schulischen Fremdsprachenerwerb oder beim Lernen aus sozial- und naturwissenschaftlichen Texten) und unter deren organisatorischen Rahmenbedingungen (z. B. im Schulunterricht oder in der Hochschullehre) vergleichbare lernförderliche Effekte erzielen können wie mit experimentell konstruiertem Material unter Laborbedingungen. Darüber hinaus wurden direkte Effekte des Testens und Abrufens hervorgehoben, die zwei weitere Anforderungen schulischen Lernens fokussieren. Testen fördert nicht allein die Verfügbarkeit von Wissen, sondern auch den (zumindest nahen) Transfer von Wissen und erfüllt damit eine für die Unterrichtspraxis zentrale Erwartung. Zweitens können Abrufübungen auch die Genauigkeit der metakognitiven Überwachung eigener Lernprozesse fördern und damit eine zentrale Voraussetzung für selbstgesteuertes Lernen verbessern.

Trotz der positiven Befundlage sind Einschränkungen zu berücksichtigen. *Erstens* gilt es festzuhalten, dass Abrufübungen nur eine unter vielen evidenzbasierten Möglichkeiten der Lernförderung darstellen. Das impliziert, dass übergeordnete Ziele von Unterricht, wie Wissenstransfer und metakognitives Monitoring nicht allein durch Abrufübungen gefördert werden, sondern hierfür auch weitere Strategien und Methoden in Frage kommen. Allerdings können Abrufübungen förderliche Beiträge zur Erreichung dieser Ziele leisten. *Zweitens* ist bei der Beurteilung der Befundlage zu berücksichtigen, dass positive Effekte des Testens immer in Relation zur jeweiligen Kontrollbedingung zu sehen sind. Häufig wird Testen mit dem erneuten Lesen (re-reading) oder dem erneuten Durcharbeiten von Lernmaterialien (restudying) verglichen. Für beide Strategien ist empirisch keine hohe Wirksamkeit nachzuweisen (vgl. Dunlosky et al. 2013). Dennoch sind diese Vergleiche gerechtfertigt, weil re-reading und restudying zu den am häufigsten spontan angewandten Lernstrategien gehören (Karpicke et al. 2009; Dunlosky et al. 2013). Zudem gibt es Befunde, in denen sich Abrufübungen auch im Vergleich mit wirksameren Strategien als überlegen zeigte (z. B. gegenüber Notizen machen: Heitmann et al. 2018). *Drittens* ist zu berücksichtigen, dass die Wirksamkeit von Abrufübungen nur dann nachweisbar ist, wenn der Abruf auch tatsächlich in substantiellem Maße gelingt. Können also in Abrufübungen viele Inhalte nicht abgerufen werden, fällt der Testeffekt geringer aus (z. B. Greving und Richter 2018; Roelle und Berthold 2017; van Gog und Sweller 2015). Die Gründe für das Misslingen des initialen Abrufs sind vielfältig: individuelle Probleme in den vorauslaufenden Lernprozessen oder ein zu hohes Anforderungsniveau im initialen Test sind Beispiele. Auch interindividuelle Unterschiede in Merkmalen, die Lernprozesse ohnehin beeinflussen, wie gering ausgeprägte Intelligenz (Wenzel und Reinhard 2019) oder Arbeitsgedächtniskapazität (Tse und Pu 2012), können die

Wahrscheinlichkeit eines initialen Abrufs verringern und somit auch den Testeffekt verringern. Deshalb könnte es nützlich sein, Tests zum Zweck der Abrufübungen adaptiv zu gestalten, d.h. an den Wissensstand von Lernenden jeweils individuell anzupassen. Erste Befunde (z.B. Heitmann et al. 2018) zeigen eine höhere Wirkung von Abrufübungen, wenn die Auswahl des jeweils folgenden Testitems an der Leistung im vorherigen orientiert wird; bei einer erfolgreichen Bearbeitung folgt ein Item mit komplexeren Anforderungen, bei einer nicht erfolgreichen Bearbeitung dagegen ein leichteres Item. Andere individuelle Merkmale Lernender, beispielsweise in den Bereichen Alter, Vorwissen und kognitive Fähigkeiten, scheinen den Testeffekt weniger zu beeinflussen (für eine Übersicht siehe Dunlosky et al. 2013). Auch Persönlichkeitseigenschaften scheinen den Testeffekt nicht essentiell zu moderieren (z.B. Fellman et al. 2020; Guran et al. 2020; Minear et al. 2018). Die Wirksamkeit von Abrufübungen ist also für unterschiedliche Zielgruppen wahrscheinlich.

4 Schlussfolgerungen für die Anwendung des Testens als Lernmethode

Vor diesem Hintergrund wird deutlich, dass die bisher im Unterricht vorherrschende Sichtweise auf Testen als diagnostisches Instrument zur Bewertung von Leistungen bzw. zur Regulation von Lernaktivitäten zu kurz greift. Die einseitige Fokussierung auf das Ergebnis von Abrufprozessen lässt das Potenzial außer Acht, das sich aus der Übung des Abrufprozesses selbst ergibt (Barenberg und Dutke [im Druck](#)). Die vorliegenden Befunde weisen den Testeffekt als ein evidenzbasiertes Angebot aus, das mit entsprechender Kenntnis von Lehrenden und Lernenden auch gezielt als Strategie einsetzbar ist (z.B. Cranney 2013; Dunlosky et al. 2013; Dunn et al. 2013; Graesser et al. 2008; Pashler et al. 2007). Aus der Forschung zum Testeffekt lassen sich daher sowohl Schlussfolgerungen für die konkrete Anwendung von Abrufübungen im Kontext von Schule und Hochschule ziehen (Abschn. 4.1) als auch übergeordnete Schlussfolgerungen für einen erfolgreichen Wissenschafts-Praxis-Transfer (Abschn. 4.2).

4.1 Schlussfolgerungen für die Anwendung von Abrufübungen

Die Anwendung von Abrufübungen im Unterrichtskontext erfordert nicht allein die Umsetzung einer Methode. Die Herausforderung besteht zunächst darin, die metakognitive Sicht auf das Testen selbst zu verändern. Die einseitige, ergebnisorientierte Sicht auf Testen als Instrument zur Bewertung von Lernständen und zur Regulation von Lernaktivitäten sollte ergänzt werden durch eine prozessorientierte Sicht auf Testen, die auch das lernförderliche Potenzial im Blick hat, das der Ausführung des Abrufs selbst innewohnt.

Dies impliziert nicht nur die Kenntnis der dargelegten Effekte von Testen und Abrufen und des damit verbundenen praktischen Potentials in unterschiedlichen Fachkontexten, sondern es erfordert auch Veränderungen auf der Ebene von Einstellungen und Überzeugungen – sowohl auf Seiten von Lehrkräften als auch auf Seiten von Lernenden. Solange Testen nur Ergebnisse im Blick hat und ein Schema von Leistungsüberprüfung aktiviert, können sich die direkten lernförderlichen

Effekte von Abrufübungen nur unzureichend auswirken. Denn die berichteten Forschungsergebnisse zeigen, dass Abrufübungen nicht nur den kognitiven Lernerfolg, sondern darüber hinaus auch die Genauigkeit von metakognitivem Monitoring erhöht. Dies kann einen Vorteil für die Selbstregulation von Lernaktivitäten darstellen. Zutreffende metakognitive Urteile zu Lerninhalten fördern die effiziente Allokation von Lernzeit, die treffende Auswahl von Lernstrategien und die Nutzung von Lernressourcen (siehe auch Endres & Renkl in diesem Band).

Damit Testen als Lernförderung seine Vorteile im Unterricht entfalten kann, erscheint es sinnvoll, Abrufübungen und Leistungsbeurteilung zu entkoppeln. Für die Anwendung von Abrufübungen könnten folgende Maßnahmen hilfreich sein:

- *Aufklärung über die unterschiedlichen Funktionen des Testens.* Dazu gehört vorrangig, dass Lehrkräfte erklären, dass Testen jenseits von Leistungsüberprüfung einen eigenen Beitrag zur Abrufbarkeit von Gelerntem leistet. Auf alltagspsychologischer Ebene ist Schüler*innen der Unterschied zwischen enkodierungs- und abruforientierten Lernaktivitäten zumindest intuitiv zugänglich. Hieran könnte angeknüpft und erklärt werden, dass Testen der Leistungsfeststellung *oder* dem Lernen dienen kann. In diesem Zusammenhang ist auch zu klären, dass die Ergebnisse eines Tests zur Leistungsüberprüfung eine andere Bedeutung haben als die Ergebnisse eines Tests zur Abrufförderung. Während sie im ersten Fall das Ziel des Testens darstellen, sind sie im zweiten Fall ein Nebenprodukt des fokussierten Prozesses (des Abrufversuchs). Dem Testergebnis gebührt in diesem Zusammenhang weit weniger Aufmerksamkeit als im Falle der Leistungsüberprüfung (allenfalls im Sinne der Orientierung für weitere Lernaktivitäten).
- *Transparenz über die Funktion eines jeweiligen Tests herstellen.* Lernende sollten keine Zweifel haben, welcher der unterschiedlichen Funktionen ein bevorstehender Test dient. Lehrkräfte sollten daher explizit machen, ob ein Test der Leistungsüberprüfung *oder* der Abrufübung dient. Diese Transparenz muss rechtzeitig hergestellt werden, damit Lernende sich darauf einstellen können. Im Falle der Leistungsüberprüfung wird beispielsweise eine gezieltere inhaltliche Vorbereitung erwartet als bei einer Abrufübung. Auch Prozesse der Emotionsregulation werden in Vorbereitung auf Leistungssituationen anders verlaufen als in Erwartung von Abrufübungen, weil die Ergebnisse unterschiedliche Bedeutungen haben.
- *Vertrauen in die unterschiedlichen Funktionen von Tests herstellen.* Im Falle eines Tests als Leistungsüberprüfung hängt das Vertrauen u. a. davon ab, inwieweit Lernenden die Anforderungen transparent erscheinen, sie mit dem Testformat vertraut sind und sie den Test als Grundlage für eine faire Leistungsbewertung ansehen. Im Falle eines Tests als Abrufübung ist maßgeblich, dass Lernende davon überzeugt sind, dass er wirklich nur ihrem Lernen dient. Entsteht der Verdacht, die Ergebnisse könnten entgegen einer Ankündigung doch der Bewertung dienen, kann der Test als Leistungsüberprüfung fehlklassifiziert werden. Lehrkräfte können einer Fehlklassifikation vorbeugen, indem sie beispielsweise die Ergebnisse nicht zur Kenntnis nehmen und nicht festhalten. Sie können auch Schüler*innen Korrektur und Feedback überlassen oder durch andere Maßnahmen verdeutlichen, dass eine Kenntnis des Ergebnisses in diesem Fall vorrangig im Interesse der Lernenden selbst liegt. Die Akzeptanz des Testens als Abrufübung wird zudem gefördert

durch die eigene Erfahrung der Wirksamkeit. Lehrkräfte könnten Lernenden beispielsweise in Demonstrationsversuchen die Wirkung von Abrufübungen vorführen.

- *Abrufübungen jenseits des Labels Testen.* Grundsätzlich gilt zu berücksichtigen, dass die hier beschriebenen förderlichen Effekte vor allem mit dem Abrufen von Lerninhalten aus dem Langzeitgedächtnis zusammenhängen. Testen impliziert in der Regel den Prozess des Abrufens von Lerninhalten und ist daher eng damit verknüpft. Darüber hinaus sind allerdings aber auch andere Situationen und Formate denkbar, die auch Abrufprozesse anregen, aber nicht mit dem Begriff Testen (und Beurteilen) direkt verknüpft sind. So können Lernende und Lehrende einerseits für sich prüfen, an welchen Stellen sie bereits Abrufübungen nutzen (z. B. mithilfe von Lern- oder Karteikarten); andererseits haben sie auch die Möglichkeit, noch kreative Formen von Abrufübungen im Unterricht einzubauen, die deutlich weniger mit Testen und Leistungsbewertungen verknüpft sind (z. B. spielerische Quiz- und Memoryformate).
- *Berücksichtigung von Abrufbarkeit und Feedback.* Ein weiterer Vorteil von Abrufübungen ist, dass es eine wirksame Strategie für sehr viele Lernende sein kann, also auch sinnvoll ist für Leistungsstärkere wie Leistungsschwächere. Relevant ist, dass der Abruf von Lerninhalten gelingt. Nun ist es selten der Fall, dass bei Abrufübungen alle Lerninhalte abgerufen werden können bzw. von allen Lernenden gleichermaßen. Daher ist es in der Unterrichtspraxis durchaus sinnvoll, Abrufübungen auch mit einer Rückmeldung von richtigen Antworten zu verknüpfen, damit die Lernenden auch aus den Abrufversuchen lernen können, in denen der Abruf nicht gelungen ist. Die empirischen Befunde deuten auch darauf hin, dass der förderliche Effekt von Abrufübungen durch Feedback verstärkt werden kann (z. B. Rowland 2014; Yang et al. 2021).
- *Wiederholung von Abrufübungen und Kombination mit anderen Lernmethoden.* Auch wenn Lerninhalte einmal erfolgreich aus dem Langzeitgedächtnis abgerufen werden konnten, ist das natürlich keine Garantie dafür, dass kein Vergessen mehr stattfindet und diese Inhalte in Zukunft immer abrufbar sind. Viele Befunde aus der Testeffektforschung (vgl. Rowland 2014; Yang et al. 2021) weisen darauf hin, dass wiederholtes Abrufen desselben Lerninhaltes die Verfügbarkeit noch mehr erhöht. Darüber hinaus zeigen einige Studien zum sogenannten sukzessiven Wiedererlernen (z. B. Rawson und Dunlosky 2011; Rawson et al. 2013), dass gerade die Kombination aus wiederholten Abrufübungen und der Möglichkeit zu anderen Lernaktivitäten zwischen diesen Abrufübungen zu nachhaltigem Behalten von Lerninhalten führt.
- *Verfügbarkeit und Transfer von Wissen.* Die angeführten Befunde legen zunächst einmal nahe, dass Abrufübungen vor allem hilfreich sind, um bei Lernenden die Verfügbarkeit von Wissen zu erhöhen. Die Verfügbarkeit von Wissen ist aber auch eine wesentliche Grundlage für andere Unterrichtsziele. So können Abrufübungen auch eingesetzt werden, um die Leistung in Transferaufgaben (z. B. Transfer auf Anwendungs- und Inferenzaufgaben) zu unterstützen. Das schließt allerdings nicht die Bedeutung weiterer Strategien aus, die den Transfer in Anwendungs- und Inferenzaufgaben gezielt vorführen und einüben.

Folglich liegen die Herausforderungen für die Anwendung von Abrufübungen als Strategie im Kontext von Schule und Hochschule vor allem darin, metakognitives Wissen über Strategien und alltagspsychologische Überzeugungen von Lehrkräften und Lernenden zu verändern und günstige Rahmenbedingungen für die Anwendung von Abrufübungen zu schaffen.

4.2 Schlussfolgerungen für den Wissenschafts-Praxis-Transfer

Die Erforschung des Testeffekts entspricht der Grundidee von *evidence-based teaching* als Beispiel für „learning and teaching strategies that have strong empirical evidence (derived from psychological research) for being effective in facilitating student learning“ (Cranney 2013, S. 2). Am Beispiel des Testeffekts wurde ein Teil des Wegs von einem grundlagenwissenschaftlichen Phänomen hin zu einer unterrichtspraktischen Empfehlung nachgezeichnet. Dieser Wissenschafts-Praxis-Transfer war und ist weder einfach noch gradlinig und regt dazu an, im Kontext des *Evidence-based-Teaching*-Programms Bedingungen zu reflektieren, die auch mit Blick auf andere Prinzipien menschlichen Lernens die Übertragung in die Unterrichtspraxis unterstützen können – also nicht auf den Testeffekt beschränkt sind.

- *Vermitteln von Informationen über evidenzbasierte Lernprinzipien.* Viele Merkmale menschlichen Lernens sind Lehramtsstudierenden und ausgebildeten Lehrkräften intuitiv zugänglich. Dennoch ist explizite Information der erste Schritt zur Korrektur unzutreffender oder zur Erweiterung einseitiger alltagspsychologischer Überzeugungen (z. B. Lawson et al. 2019; Menz et al. 2021). Während wissenschaftliche Publikationen in der Regel mit einer präzisen Inhaltsangabe betitelt werden, die für Wissenschaftsanwender*innen oft schwer zu entschlüsseln sind, werben Publikationen im Rahmen des *Evidence-based-Teaching*-Programms auch mit den potenziellen Anwendungsfeldern: „Tips for classroom use and for guiding students’ learning“ (Dunlosky und Rawson 2015, S. 72) oder „Inexpensive techniques to improve education“ (Roediger und Pyc 2012, S. 242) sind Beispiele. Der Zugang des Anwendungsfeldes kann auch durch die Wahl des Publikationsorgans erleichtert werden. Lehrkräfte haben oft keinen Zugang zu psychologischen Fachzeitschriften, teils weil deren Titel keine Schlüsse auf anwendungsfähige, unterrichtsnahe Inhalte zulassen, teils weil Lizenzen erworben werden müssen. Peer-reviewed Open-Science-Publikationen können hier erfolgreicher sein, ebenso wie Fachzeitschriften, die Wissenschaftstransfer programmatisch aufnehmen, wie etwa *The Inquisitive Mind* (<https://de.in-mind.org/>) oder *Psychological Science in the Public Interest* (<https://journals.sagepub.com/home/psi>). Hilfreich sind auch Publikationsformate, die klare Anwendungserwartungen zulassen, wie beispielsweise die „Teacher-Ready Research Reviews“ in *Scholarship of Teaching and Learning in Psychology* (<https://www.apa.org/pubs/journals/stl>).
- *Anwendungsnahe Forschungsdesigns.* Informieren allein reicht aber sicherlich nicht aus. Aus Sicht der Psychologie wird die Transferdistanz in die Unterrichtspraxis oft unterschätzt. Stichproben, die der eigenen Schülerklientel zu wenig ähneln, artifizielles Lernmaterial, schulfremde Umsetzungsbedingungen laden wenig zum Transfer ein. Deshalb sind Studien willkommen, die Praxisvertre-

ter*innen die Funktionsweise von Empfehlungen zumindest partiell vorführen, beispielsweise anhand schulnaher Themen, mit Materialien, wie sie in Schulen bekannt sind, durchgeführt in schulischen oder zumindest schulähnlichen Settings (naher Transfer).

- *Aufgreifen schulischer Probleme und Zielsetzungen.* Allein die Kenntnis, dass kognitionspsychologische Lernprinzipien unterrichtsnah eingesetzt werden *können*, bedeutet nicht, dass dies in der Praxis auch umgesetzt wird. Einen zusätzlichen Impuls kann der Hinweis geben, dass die Umsetzung einer methodischen Empfehlung das Erreichen übergeordneter Ziele von Bildung, (Hoch-)Schule und Unterricht erleichtern oder absichern kann. Wenn beispielsweise nachgewiesen werden kann, dass die Anwendung eines bestimmten evidenzbasierten Lernprinzips die Verfügbarkeit von Wissen über einen längeren Zeitraum verbessert, spricht dies ein übergeordnetes Ziel von Schulunterricht an, nämlich Nachhaltigkeit von Lernen. Sollte sich ein Lernprinzip als nützlich erweisen, auch wenn Lernende große Vorwissensunterschiede aufweisen, könnte dies das übergeordnete Ziel der Inklusion ansprechen. Solche Verknüpfungen von Lernprinzipien mit zentralen, übergeordneten Zielen von schulischer Bildung sollten im Vermittlungsprozess explizit gemacht werden, um die Anwendungsattraktivität zu erhöhen.
- *Evidenzbasierte Lehr-Lernprinzipien in der Aus- und Weiterbildung von Lehrkräften.* Die Aus- und Weiterbildung von Lehrkräften kann von einer reichen Auswahl an Lernprinzipien aus dem Bereich des *evidence-based teaching* profitieren. Diese Themen werden als attraktiv wahrgenommen, weil sie Verbindungen zwischen Theorie, Empirie und praktischer Anwendung nachvollziehbar machen. Zudem eignen sie sich für den selbstreflektierenden Umgang mit Themen der Unterrichtsgestaltung, weil in fast allen Fällen Bezüge zum eigenen Lernverhalten von Studierenden und Lehrkräften hergestellt werden können. Ein besonders geeignetes Erprobungsfeld bieten die Studienprojekte im Praxissemester von Lehramtsstudierenden in der Masterphase (vgl. Souvignier und Dutke 2016). Der Testeffekt ist beispielsweise sehr geeignet, kleine empirische Explorationen im eigenen Unterricht durchzuführen und dabei die Logik von Abrufübungen im Kontext der eigenen Unterrichtsfächer auszuprobieren. Ebenso könnten andere evidenzbasierte Lernprinzipien (z. B. verteiltes Lernen, vgl. Dunlosky et al. 2013) mit überschaubarem Aufwand in Unterrichtsvorhaben oder Studienprojekte der Praxissemesterstudierenden integriert werden. Zusätzlich werden so Themen evidenzbasierten Lehrens in die Schulen getragen, weil sich beispielsweise auch Mentor*innen von Praxissemesterstudierenden mit ihnen auseinandersetzen.

Zusammengefasst wird deutlich, dass erfolgreicher Wissenschafts-Praxis-Transfer im Kontext evidenzbasierten Lehrens kein unidirektionaler Prozess ist, sondern im Idealfall eine Interaktion zwischen grundlagenwissenschaftlicher Ausgangsbasis und Methodik einerseits und praktischen Anforderungen des schulischen, unterrichtlichen Anwendungsfeldes andererseits entsteht. Ein solches Verständnis könnte vielen weiteren evidenzbasierten Prinzipien menschlichen Lernens zu mehr Aufmerksamkeit und Anwendung in der Unterrichtspraxis verhelfen – nicht nur im Fall der Abrufübungen.

Funding Open Access funding enabled and organized by Projekt DEAL.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Abbott, E.E. (1909). On the analysis of the factor of recall in the learning process. *Psychological Monographs*, 11, 159–177. <https://doi.org/10.1037/h0093018>.
- Adesope, O.O., Trevisan, D.A., & Sundararajan, N. (2017). Rethinking the use of tests: a meta-analysis of practice testing. *Review of Educational Research*, 87, 659–701. <https://doi.org/10.3102/0034654316689306>.
- Barenberg, J., & Dutke, S. (2019). Testing and metacognition: retrieval practise affects metacognitive monitoring in learning from text. *Memory*, 27, 269–279. <https://doi.org/10.1080/09658211.2018.1506481>.
- Barenberg, J., & Dutke, S. (2021). Retrieval practice effects in a psychology lecture: Illustrating the relevance of study design, item difficulty, and selection of dependent measures. *Psychology Learning and Teaching*. <https://doi.org/10.1177/14757252211049312>.
- Barenberg, J., & Dutke, S. (in Druck). Das Potential von Testsituationen in der Unterrichtspraxis wird nicht ausgeschöpft: Fehl-Vorstellungen von Testen und Lernen. In G. Steins, B. Spinath, S. Dutke, M. Roth & M. Limbourg (Hrsg.), *Mythen, Fehlvorstellungen, Fehlkonzeppte und Irrtümer in Schule und Unterricht*. Wiesbaden: Springer.
- Barnett, S.M., & Ceci, S.J. (2002). When and where do we apply what we learn? A taxonomy for far transfer. *Psychological Bulletin*, 128, 612–637. <https://doi.org/10.1037/0033-2909.128.4.612>.
- Bjork, R.A. (1975). Retrieval as a memory modifier: An interception of negative recency and related phenomena. In R.L. Solso (Hrsg.), *Information processing and cognition. The Loyola symposium* (S. 123–144). Hillsdale: Erlbaum.
- Butler, A.C., & Roediger, H.L. (2007). Testing improves long-term retention in a simulated classroom setting. *European Journal of Cognitive Psychology*, 19, 514–527. <https://doi.org/10.1080/09541440701326097>.
- Carpenter, S.K. (2009). Cue strength as a moderator of the testing effect: the benefits of elaborative retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 1563–1569. <https://doi.org/10.1037/a0017021>.
- Carpenter, S.K. (2012). Testing enhances the transfer of learning. *Current Directions in Psychological Science*, 21, 279–283. <https://doi.org/10.1177/0963721412452728>.
- Cogliano, M.C., Kardash, C.A.M., & Bernacki, M.L. (2019). The effects of retrieval practice and prior topic knowledge on test performance and confidence judgments. *Contemporary Educational Psychology*, 56, 117–129. <https://doi.org/10.1016/j.cedpsych.2018.12.001>.
- Cranney, J. (2013). Toward psychological literacy: a snapshot of evidence-based learning and teaching. *Australian Journal of Psychology*, 65, 1–4. <https://doi.org/10.1111/ajpy.12013>.
- Darabi, A.A., Nelson, D.W., & Palanki, S. (2007). Acquisition of troubleshooting skills in a computer simulation: worked example vs. conventional problem solving instructional strategies. *Computers in Human Behavior*, 23, 1809–1819.

- Dunlosky, J., & Nelson, T. O. (1997). Similarity between the cue for judgments of learning (JOL) and the cue for test it is not the primary determinant of JOL accuracy. *Journal of Memory and Language*, 36, 34–49.
- Dunlosky, J., & Rawson, K. A. (2015). Practice tests, spaced practice, and successive relearning: tips for classroom use and for guiding students' learning. *Scholarship of Teaching and Learning in Psychology*, 1, 72–78.
- Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., & Willingham, D. T. (2013). Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest*, 14, 4–58.
- Dunn, D. S., Saville, B. K., Baker, S. C., & Marek, P. (2013). Evidence-based teaching: tools and techniques that promote learning in the psychology classroom. *Australian Journal of Psychology*, 65, 5–13. <https://doi.org/10.1111/ajpy.12004>.
- Dutke, S., Bakker, H. E., Papageorgi, I., & Taylor, J. (2017). Editorial: evidence-based teaching—examples from learning and teaching psychology. *Psychology Learning and Teaching*, 16, 175–178. <https://doi.org/10.1177/1475725717701209>.
- Fellman, D., Lincke, L., & Jonsson, B. (2020). Do individual differences in cognition and personality predict retrieval practice activities on MOOCs? *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2020.02076>.
- Fernandez, J., & Jamet, E. (2017). Extending the testing effect to self-regulated learning. *Metacognition and Learning*, 12, 131–156. <https://doi.org/10.1007/s11409-016-9163-9>.
- van Gog, T., & Sweller, J. (2015). Not new, but nearly forgotten: the testing effect decreases or even disappears as the complexity of learning materials increases. *Educational Psychology Review*, 27, 247–264. <https://doi.org/10.1007/s10648-015-9310-x>.
- Graesser, A. C. (2009). Inaugural editorial for Journal of Educational Psychology. *Journal of Educational Psychology*, 101, 259–261. <https://doi.org/10.1037/a0014883>.
- Graesser, A. C., Halpern, D. F., & Hake, M. (2008). *25 principles of learning*. Washington, DC: Task Force on Lifelong Learning at Work and at Home.
- Greving, S., & Richter, T. (2018). Examining the testing effect in university teaching: retrievability and question format matter. *Frontiers in Psychology*, 9, 2412. <https://doi.org/10.3389/fpsyg.2018.02412>.
- Guran, C.-N. A., Lehmann-Grube, J., & Bunzeck, N. (2020). Retrieval practice improves recollection-based memory over a seven-day period in younger and older adults. *Frontiers in Psychology*, 10, 2997. <https://doi.org/10.3389/fpsyg.2019.02997>.
- Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review*, 19, 126–134. <https://doi.org/10.3758/s13423-011-0181-y>.
- Heitmann, S., Grund, A., Berthold, K., Fries, S., & Roelle, J. (2018). Testing is more desirable when it is adaptive and still desirable when compared to note-taking. *Frontiers in Psychology*, 9, 2596. <https://doi.org/10.3389/fpsyg.2018.02596>.
- Kang, S. H. K., McDermott, K. B., & Roediger III, H. L. (2007). Test format and corrective feedback modify the effect of testing on long-term retention. *European Journal of Cognitive Psychology*, 19, 528–558. <https://doi.org/10.1080/09541440601056620>.
- Karpicke, J. D., Butler, A. C., & Roediger III, H. L. (2009). Metacognitive strategies in student learning: do students practise retrieval when they study on their own? *Memory*, 17, 471–479. <https://doi.org/10.1080/09658210802647009>.
- Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-based learning: an episodic context account. In B. H. Ross (Hrsg.), *The psychology of learning and motivation* (S. 237–284). : Elsevier.
- King, J. F., Zechmeister, E. B., & Shaughnessy, J. J. (1980). Judgments of knowing: the influence of retrieval practice. *The American Journal of Psychology*, 93, 329–343. <https://doi.org/10.2307/1422236>.
- Klieme, E. (2006). Empirische Unterrichtsforschung. Aktuelle Entwicklungen, theoretische Grundlagen und fachspezifische Befunde. Einführung in den Thementeil. *Zeitschrift für Pädagogik*, 52, 765–773.
- Koriat, A. (1997). Monitoring one's own knowledge during study: a cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, 126, 349–370. <https://doi.org/10.1037/0096-3445.126.4.349>.
- Kornell, N., & Rhodes, M. G. (2013). Feedback reduces the metacognitive benefit of tests. *Journal of Experimental Psychology: Applied*, 19, 1–13. <https://doi.org/10.1037/a0032147>.
- Kornell, N., & Son, L. K. (2009). Learners' choices and beliefs about self-testing. *Memory*, 17, 493–501. <https://doi.org/10.1080/09658210902832915>.

- Lawson, M. J., Vosniadou, S., Van Deur, P., Wyra, M., & Jeffries, D. (2019). Teachers' and students' belief systems about the self-regulation of learning. *Educational Psychology Review*, 31, 223–251. <https://doi.org/10.1007/s10648-018-9453-7>.
- Little, J. L., & McDaniel, M. A. (2015). Metamemory monitoring and control following retrieval practice for text. *Memory & Cognition*, 43, 85–98. <https://doi.org/10.3758/s13421-014-0453-7>.
- Little, J. L., Storm, B. C., & Bjork, E. L. (2011). The costs and benefits of testing text materials. *Memory*, 19, 346–359. <https://doi.org/10.1080/09658211.2011.569725>.
- McDaniel, M. A., & Masson, M. E. (1985). Altering memory representations through retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11, 371–385. <https://doi.org/10.1037/0278-7393.11.2.371>.
- McDaniel, M. A., Agarwal, P. K., Huelser, B. J., McDermott, K. B., & Roediger III, H. L. (2011). Test-enhanced learning in a middle school science classroom: The effects of quiz frequency and placement. *Journal of Educational Psychology*, 103, 399–414. <https://doi.org/10.1037/a0021782>.
- Menz, C., Spinath, B., & Seifried, E. (2021). Where do pre-service teachers' educational psychological misconceptions come from? The roles of anecdotal versus scientific evidence. *Zeitschrift Für Pädagogische Psychologie*, 35, 143–156. <https://doi.org/10.1024/1010-0652/a000299>.
- Miner, M., Coane, J. H., Boland, S. C., Cooney, L. H., & Albat, M. (2018). The benefits of retrieval practice depend on item difficulty and intelligence. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44, 1474–1486. <https://doi.org/10.1037/xlm0000486>.
- Morris, C. D., Bransford, J. D., & Franks, J. J. (1977). Levels of processing versus transfer appropriate processing. *Journal of Verbal Learning and Verbal Behavior*, 16, 519–533. [https://doi.org/10.1016/S0022-5371\(77\)80016-9](https://doi.org/10.1016/S0022-5371(77)80016-9).
- Nelson, T. O., & Narens, L. (1990). Metamemory: a theoretical framework and some new findings. In G. H. Bower (Hrsg.), *The psychology of learning and motivation* (S. 125–173). New York: Academic Press.
- Pan, S. C., & Rickard, T. C. (2018). Transfer of test-enhanced learning. Meta-analytic review and synthesis. *Psychological Bulletin*, 144, 710–756. <https://doi.org/10.1037/bul0000151>.
- Pashler, H., Bain, P. T., Bottge, B., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning*. Washington, DC: National Center for Education Research, Institute of Education Science, U.S., Department of Education.
- Potts, R., & Shanks, D. R. (2014). The benefit of generating errors during learning. *Journal of Experimental Psychology: General*, 143, 644–667. <https://doi.org/10.1037/a0033194>.
- Prell, S. (1985). Sinn und Zweck empirischer Unterrichtsforschung. *Sonderdruck aus „Gießener Universitätsblätter“*, 18 (2), 27–37.
- Rawson, K. A., & Dunlosky, J. (2011). Optimizing schedules of retrieval practice for durable and efficient learning: how much is enough? *Journal of Experimental Psychology: General*, 140, 283–302. <https://doi.org/10.1037/a0023956>.
- Rawson, K. A., Dunlosky, J., & Sciarrelli, S. M. (2013). The power of successive relearning: Improving performance on course exams and long-term retention. *Educational Psychology Review*, 25, 523–548. <https://doi.org/10.1007/s10648-013-9240-4>.
- Rhodes, M. G., & Tauber, S. K. (2011). The influence of delaying judgments of learning on metacognitive accuracy: a meta-analytic review. *Psychological Bulletin*, 137, 131–148. <https://doi.org/10.1037/a0021705>.
- Rivers, M. L. (2020). Metacognition about practice testing: a review of learners' beliefs, monitoring, and control of test-enhanced learning. *Educational Psychology Review*. <https://doi.org/10.1007/s10648-020-09578-2>.
- Roediger III, H. L., & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. *Trends in Cognitive Sciences*, 15, 20–27. <https://doi.org/10.1016/j.tics.2010.09.003>.
- Roediger III, H. L., & Pyc, M. A. (2012). Inexpensive techniques to improve education: applying cognitive psychology to enhance educational practice. *Journal of Applied Research in Memory and Cognition*, 1, 242–248. <https://doi.org/10.1016/j.jarmac.2012.09.002>.
- Roelle, J., & Berthold, K. (2017). Effects of incorporating retrieval into learning tasks: the complexity of the tasks matters. *Learning and Instruction*, 49, 142–156. <https://doi.org/10.1016/j.learninstruc.2017.01.008>.
- Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological Bulletin*, 140, 1432–1463. <https://doi.org/10.1037/a0037559>.
- Rychen, D. S., & Salganik, L. H. (2003). A holistic model of competence. In D. S. Rychen & L. H. Salganik (Hrsg.), *Key competencies for a successful life and a well-functioning society* (S. 41–62). Göttingen: Hogrefe & Huber.

- Schneider, W., Visé, M., Lockl, K., & Nelson, T.O. (2000). Developmental trends in children's memory monitoring: evidence from a judgment-of-learning task. *Cognitive Development*, 15, 115–134.
- Schwartz, B.M., & Gurung, R.A.R. (Hrsg.). (2012). *Evidence-based teaching for higher education*. Washington, DC: American Psychological Association. <https://doi.org/10.1037/13745-000>.
- Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: a meta-analytic perspective. *Psychology Learning & Teaching*, 16, 179–196. <https://doi.org/10.1177/1475725717695149>.
- Shaughnessy, J.J., & Zechmeister, E.B. (1992). Memory-monitoring accuracy as influenced by the distribution of retrieval practice. *Bulletin of the Psychonomic Society*, 30, 125–128. <https://doi.org/10.3758/BF03330416>.
- Souvignier, E., & Dutke, S. (2016). Studienprojekte im Praxissemester. In E. Jürgens (Hrsg.), *Erfolgreich durch das Praxissemester* (S. 165–176). Berlin: Cornelsen.
- Tse, C.-S., & Pu, X. (2012). The effectiveness of test-enhanced learning depends on trait test anxiety and working-memory capacity. *Journal of Experimental Psychology: Applied*, 18, 253–264. <https://doi.org/10.1037/a0029190>.
- Wenzel, K., & Reinhard, M.-A. (2019). Relatively unintelligent individuals do not benefit from intentionally hindered learning: the role of desirable difficulties. *Intelligence*, 77, 101405. <https://doi.org/10.1016/j.intell.2019.101405>.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker (Ed.), *Metacognition in educational theory and practice* (pp. 227–304). Mahwah, NJ: Erlbaum.
- Yang, C., Luo, L., Vadiello, M.A., Yu, R., & Shanks, D.R. (2021). Testing (quizzing) boosts classroom learning: a systematic and meta-analytic review. *Psychological Bulletin*. <https://doi.org/10.1037/bul0000309>.
- Zimmerman, B.J. (2008). Investigating self-regulation and motivation: historical background, methodological developments, and future prospects. *American Educational Research Journal*, 45, 166–183. <https://doi.org/10.3102/0002831207312909>.