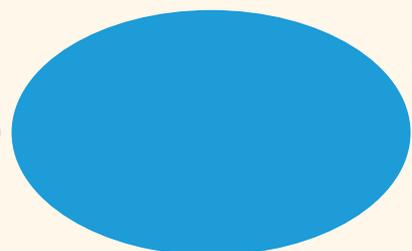
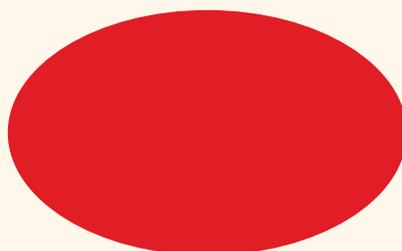
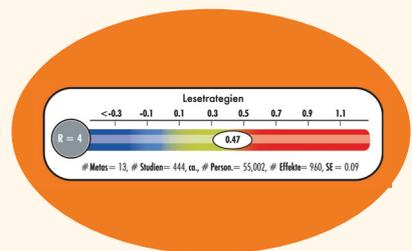
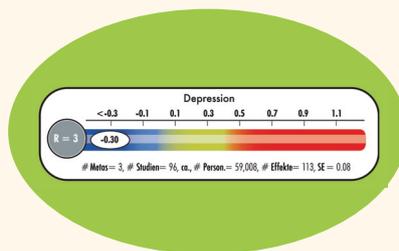
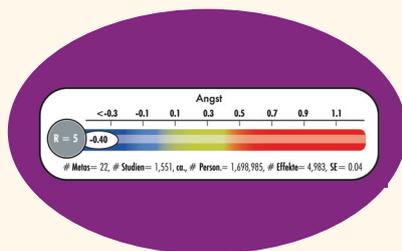
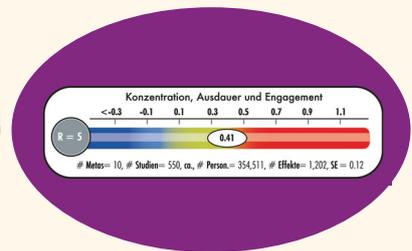
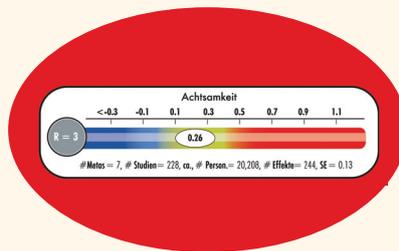
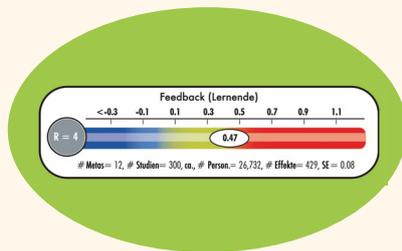
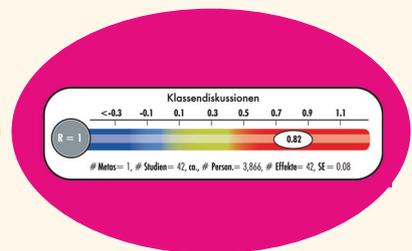
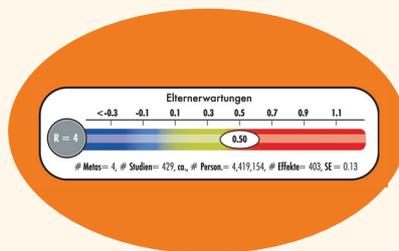
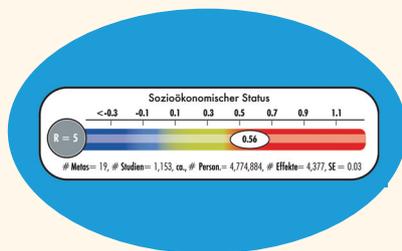


JOHN HATTIE

VISIBLE LEARNING 2.0

Deutschsprachige Ausgabe von „Visible Learning: The Sequel“ besorgt von

Stephan Wernke und **Klaus Zierer**





JOHN HATTIE

VISIBLE LEARNING 2.0

Deutschsprachige Ausgabe von
„Visible Learning: The Sequel“ besorgt von

Stephan Wernke und Klaus Zierer



Schneider Verlag Hohengehren GmbH

Die Originalausgabe ist lieferbar unter dem Titel:

John Hattie

Visible Learning: The Sequel

1st Edition (9781003380542)

All rights reserved.

Authorised translation from the English language edition published by
Routledge, a member of the Taylor & Francis Group, Abingdon, GB

Gedruckt auf umweltfreundlichem Papier (chlor- und säurefrei hergestellt).

Bibliografische Information der Deutschen Nationalbibliothek

Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.dnb.de> abrufbar.

ISBN 978-3-8340-2250-9

Schneider Verlag Hohengehren, D-73666 Baltmannsweiler

Homepage: www.paedagogik.de

Alle Rechte, insbesondere das Recht der Vervielfältigung sowie der Übersetzung, vorbehalten. Kein Teil des Werkes darf in irgendeiner Form (durch Fotokopie, Mikrofilm oder ein anderes Verfahren) ohne schriftliche Genehmigung des Verlages reproduziert werden.

© Schneider Verlag Hohengehren, 73666 Baltmannsweiler 2024.

Printed in Germany. Druck: Appel & Klinger, Schneckenlohe

Visible Learning 2.0

Als das Original von *Visible Learning* im Jahr 2008 veröffentlicht wurde, stellte es sich sofort als eine verlegerische Sensation heraus. Das Interesse an dem Buch war beispiellos und innerhalb weniger Tage war es ausverkauft. Im TES (Times Educational Supplement) wurde es als „Der Heilige Gral des Lehrens“ bezeichnet. Mit dieser Veröffentlichung kehrt John Hattie zu seinem bahnbrechenden Werk zurück. Die Forschungsergebnisse, auf welchen die vorliegende Fortsetzung von *Visible Learning* basiert, stützen sich inzwischen auf mehr als 2.100 Metaanalysen (mehr als doppelt so viele wie in der ursprünglichen Veröffentlichung mit ca. 800 Meta-Analysen), die aus mehr als 130.000 Studien stammen und an denen schätzungsweise mehr als 400 Millionen Lernende aus aller Welt teilgenommen haben.

Dieses Buch ist jedoch mehr als nur eine Neuauflage: Es ist eine Fortsetzung, die das große Ganze beleuchtet, die Umsetzung von Visible Learning in den Schulen reflektiert, wie es verstanden – und manchmal auch missverstanden – wurde und welche Richtung die Forschung in Zukunft einschlagen sollte.

Visible Learning 2.0 bekräftigt den Wunsch des Autors, nicht nur zu behaupten, was funktioniert, sondern auch und vor allem, was *am besten* funktioniert, indem er entscheidende Fragen stellt wie: Warum ist die derzeitige Grammatik des Schulunterrichts in so vielen Klassenzimmern so fest verankert und wie können wir sie verbessern? Warum ist die Lernentwicklungskurve für Lehrpersonen nach den ersten Berufsjahren so flach? Wie können wir die Denkweisen von Lehrpersonen so entwickeln, dass sie sich mehr auf das Lernen und Zuhören konzentrieren (und weniger auf das Lehren und Sprechen)? Wie können wir Forschungsergebnisse in die Diskussionen der Schulen und der Kollegien bringen?

Zu den besprochenen Bereichen gehören:

- Die Forschungsbasis und die Reaktionen auf *Visible Learning*
- Das *Visible Learning* Modell
- Die bewusste Abstimmung von Lern- und Lehrstrategien
- Der Einfluss des Elternhauses, der Lernenden, der Lehrpersonen, der Klassenzimmer, der Schulen, des Lernens und der Lehrpläne auf die Lernleistung
- Der Einfluss von Technologie

Aufbauend auf dem Erfolg des Originals erweitert diese mit Spannung erwartete Fortsetzung Hatties Modell des Lehrens und Lernens auf der Grundlage von Einflussgrößen und ist eine unverzichtbare Lektüre für alle, die im Bildungsbereich tätig sind – sei es als Forschende, Lehrpersonen, Lernende, Schulleitungen, Lehrerausbildende oder politische Entscheidungsträger.

John Hattie ist emeritierter Professor an der Graduate School of Education der Universität von Melbourne, Australien. Er ist einer der weltweit bekanntesten und meistgelesenen Bildungsexperten. Seine Bücher zu *Visible Learning* wurden in 29 Sprachen übersetzt und über zwei Millionen Mal verkauft.

Lob für John Hatties frühere Arbeiten:

„Enthüllt den heiligen Gral des Lehrens“.

Times Educational Supplement

„John Hattie hat einen außerordentlichen Beitrag zur Verbesserung des Lernens geleistet. Visible Learning trat 2009 auf den Plan und erregte unser aller Aufmerksamkeit. Was mir besonders gefällt, ist, wie Hattie und sein Team weiterhin nach innen kritisch und nach außen offen sind, um alle Möglichkeiten in Betracht zu ziehen. Seine neue Reihe auf der Suche nach dem „Goldstandard“ ist typisch für Hatties Engagement und seine Fähigkeit, Kritik anzunehmen, kontinuierlich zu lernen und Verbesserung zu modellieren. Eine Million Bücher! Herzlichen Glückwunsch für eine monumentale Leistung und das Versprechen auf mehr und mehr.“

Michael Fullan, OC, Emeritierter Professor,
OISE/Universität von Toronto

„Ich bin mir nicht sicher, ob John seine eigene Wirkung kennt – ich kenne niemanden im Bildungswesen auf diesem Planeten, der eine höhere Effektstärke hat... Wir Lehrpersonen brauchten ein Sieb, um die Forschung, die den Lehrpersonen hilft, von der Forschung zu trennen, die das nicht tut. John hat es zur Verfügung gestellt, und jetzt wissen wir alle, wie wir unseren eigenen Einfluss steigern können – vielen Dank, John!“

Geoff Petty, Autor von 'Teaching Today' und
„Evidence-Based Teaching“

„Die Bildungsforschung leidet unter der Qual der Wahl – es gibt zu viel davon! John Hattie ist ein entschlossener Vorreiter bei der Zusammenfassung dieser Fülle an Informationen in praktische Begriffe, die verdaulich sind und dennoch zuverlässig das Wesentliche erfassen. Und was noch besser ist: Sie lassen sich leicht in die Unterrichtspraxis übertragen! Hatties Erfolg und Einfluss sind hochverdient.“

Dan Willingham, Professor für Psychologie, Universität von West Virginia,
Autor von „Why Don't Students Like School?“

„Tausende von Studien haben bestätigt, worauf wir uns konzentrieren müssen ... Lehrpersonen auf der ganzen Welt haben nun einen Bezugspunkt für das, worauf es ankommt, um Denken und Handeln zu verbessern“.

Shirley Clarke, Autorin von „Unlocking Formative Assessment“ und
„Thinking Classrooms“

„John Hatties Beiträge für Generationen von Lernenden, Lehrpersonen und Schulleitungen auf der ganzen Welt sind tiefgreifend und nachhaltig. Er besitzt die seltene Kombination aus Mut und Bescheidenheit, die konventionelle Weisheiten furchtlos in Frage stellt und gleichzeitig seine Arbeit kontinuierlich verbessert und erweitert. Er macht unseren Beruf und die Welt der Bildung besser, und Kinder auf allen Kontinenten profitieren von seiner Wissenschaft, seiner Weisheit und seiner praktischen Anleitung.“

Douglas Reeves, Creative Leadership.net

„Hatties Arbeit ermöglicht es uns, die Auswirkungen dessen zu erkennen, was Lehrpersonen im Klassenzimmer tun. Seine Arbeit über die Unterrichtsvorbereitung gehört zu den Pflichtlektüren für alle Lehrpersonen. Beim Unterrichten sind wir oft auf der Suche nach Momenten, in denen uns ein Licht aufgeht. Hattie hat dafür die verborgenen Verdrahtungen sichtbar gemacht.“

Huw Thomas, ehemaliger Schulleiter, College-Dozent und
Diözesandirektor für Bildung

„In einer Welt, in der Pädagogen oft dazu verleitet und ermutigt werden, die neuesten Modeerscheinungen zu nutzen, um das Lernen der Schülerinnen und Schüler zu unterstützen, sind die Bücher von John Hattie zu Visible Learning eine außergewöhnliche Ressource. Jede praktizierende Lehrperson, jede Schulleitung und alle Bildungsforschende sollten diese Bücher in ihrem Regal stehen haben.“

Eric M. Anderman, Professor für Pädagogische Psychologie,
Ohio State University

„John Hattie hat einen Traum, dass eines Tages jedes Kind nicht zufällig, sondern geplant lernt. Er träumt davon, dass eines Tages das Fachwissen, das überall um uns herum vorhanden ist, zusammenkommt und das System verändert. Er träumt davon, dass sich Lehrerinnen und Lehrer eines Tages die wichtige Frage stellen: „Was funktioniert am besten?“, statt nur zu fragen „Was funktioniert?“ Und er träumt davon, dass Lehrerinnen und Lehrer eines Tages immer die maximale Wirkung anstreben, und – das ist das Wichtigste – er lebt seinen Traum. Er hat das Feuer mit seiner Leidenschaft entfacht. Er hat das Lernen entfacht. Und er hat das Lehren entfacht. Herzlichen Glückwunsch John – und: Kenne deinen Einfluss!“

Klaus Zierer, Professor für Schulpädagogik, Universität Augsburg, Deutschland,
Associate Research Fellow des vom ESRC finanzierten Centre on Skills,
Knowledge and Organisational Performance (SKOPE) an der University of Oxford



Inhalt

<i>Vorwort der Herausgeber der deutschen Übersetzung</i>	VII
<i>Danksagungen</i>	IX
1 Die Herausforderung	1
2 Meta-Analysen und die Kritik an <i>Visible Learning</i>	17
3 Das Modell	35
4 Umsetzung des <i>Visible Learning</i> Modells	47
5 Lernende	61
6 Elternhaus und Familie	101
7 Schule und Gesellschaft	119
8 Klassenzimmer	147
9 Lehrperson	173
10 Curriculum	199
11 Zielorientiertes Unterrichten	245
12 Lernstrategien	271
13 Lehrstrategien	285
14 Nutzung von Technologien	313
15 Schulische und außerschulische Einflüsse	327
16 Schlussfolgerungen	339
<u><i>Anhang</i></u>	
<i>Literaturhinweise</i>	351
<i>Literaturverzeichnis</i>	359
<i>Abbildungs-, Tabellen- und Thermometerverzeichnis</i>	383



Vorwort der Herausgeber der deutschen Übersetzung

Bereits drei Werke wurden von John Hattie ins Deutsche übertragen: „Visible Learning“ 2013 unter dem Titel „Lernen sichtbar machen“, „Visible Learning for Teachers“ 2014 unter dem Titel „Lernen sichtbar machen für Lehrpersonen“ und „Visible Learning And The Science Of How We Learn“ 2015 unter dem Titel „Lernen sichtbar machen aus psychologischer Perspektive“. Die Originalausgaben begründeten international ebenso wie die Übersetzungen im deutschsprachigen Raum den Erfolg von John Hattie als einen der renommiertesten Erziehungswissenschaftler weltweit. Insgesamt erreicht die Auflage seiner Werke die drei Millionen Marke.

Mit großer Spannung erwartete die Forschungsgemeinschaft daher die Neuauflage von „Visible Learning“, die als „Visible Learning – The Sequel“ 2023 erschienen ist. Trotz der großen Anstrengungen, die eine Übersetzung nach sich zieht, stand unser Entschluss fest, erneut an die Arbeit für eine deutschsprachige Ausgabe zu gehen. Nach zwölf Monaten freuen wir uns, „Visible Learning 2.0“ an die Leserinnen und Leser zu übergeben.

Wie bei den letzten Übersetzungen auch standen wir diesmal ebenso vor großen Herausforderungen. Mit der Erweiterung des Datensatzes von 800+ Meta-Analysen auf nun über 2.100+ Meta-Analysen hat sich der Datensatz von „Visible Learning“ mehr als verdoppelt. Der Umfang ist folgerichtig gestiegen von ca. 380 Seiten auf ca. 500 Seiten und die Anzahl der Faktoren von 138 auf 357.

Auch inhaltlich hat sich einiges getan: John Hattie hat in den letzten Jahren intensiv an seiner Methode, eine Synthese von Meta-Analysen, gearbeitet und setzt sich im Detail mit der Kritik auseinander, die für ihn immer willkommen war und bis heute ist. Auch hat sich die Anordnung der Faktoren auf nun neun Domänen ausdifferenziert und im Zuge dessen wird dem Einfluss von digitalen Medien eine besondere Aufmerksamkeit geschenkt. Auffallend ist sicherlich auch, dass anstelle der Barometer nun Thermometer das Cover zieren. Der Grund ist ein pragmatischer: Die Thermometer sind platzsparender als die Barometer. Sie folgen aber derselben Logik, nutzen das Effektstärkenmaß als Kriterium und vereinen in sich die wichtigste Statistik, so dass auf einen Blick ersichtlich ist, wie es um die Wirksamkeit eines Faktors aus empirischer Sicht steht.

Aufbauend auf den Erfahrungen aus den vorausgegangenen Übersetzungen wurden die damals getroffenen Konventionen beibehalten: Die Übersetzung bereits existierender Faktoren orientiert sich an den früheren Ausgaben von „Lernen sichtbar machen“ und neue Faktoren wurden nach intensiven Diskussionen und unter Berücksichtigung des internationalen Diskurses ins Deutsche übertragen. Lange Texte, die im Englischen weniger Schwierigkeiten bereiten, haben wir zerlegt und damit zugänglicher gemacht. Auch englischsprachige Originalzitate wurden übersetzt. Nach Möglichkeit wurde eine gendergerechte Sprache verwendet. Unstimmigkeiten, die bei der Durchsicht der Tabellen aufgetaucht sind, haben wir korrigiert. Und schließlich haben wir ein Abbildungs-, Tabellen- und Thermometerverzeichnis ergänzt, um die Suche im Buch zu erleichtern. Nicht mehr enthalten sind die Literaturhinweise zu den Meta-Analysen, die in früheren Ausgaben zu „Visible Learning“ noch mehrere Druckseiten umfassten. Stattdessen verweist John Hattie auf MetaX, eine Online-Plattform unter <https://www.visiblelearningmetax.com/>, auf welcher der aktuelle Datensatz einsehbar ist und im Text vor den Autorennamen mit einem * gekennzeichnet ist.

Anders als im Englischen haben wir für den deutschsprachigen Titel nicht „Die Fortsetzung“ gewählt, sondern sprechen von „Visible Learning 2.0“. Aus unserer Sicht ist nämlich „Visible Learning – The Sequel“ weniger ein Folgeband von „Visible Learning“ als vielmehr eine grundsätzliche Überarbeitung und Erweiterung. Wir hoffen mit dieser Übersetzung das Werk von John Hattie in einer gelungenen Übertragung ins Deutsche vorzulegen, die nun in die großen Fußstapfen tritt und als neues Standardwerk zu „Visible Learning“ viele Leserinnen und Leser findet.

Danksagungen

Dass wir nach zwölf Monaten intensiver Arbeit das vorliegende Buch veröffentlichen können, ist auch das Verdienst einer Reihe von Mitarbeitenden. Es haben mitgewirkt

Michael Poschik & Lucas Jahn an der Universität Augsburg,
Dr. Vanessa Pieper an der Universität Wuppertal und
Dr. Martin J. Meyer an der Universität Vechta.

Ihnen sei an dieser Stelle ebenso herzlich gedankt wie unserem Verleger, *Rainer Schneider*. Schlussendlich möchten wir *John Hattie* unseren Dank aussprechen. Er begleitet uns nun seit über zehn Jahren in enger kollegialer und auch freundschaftlicher Verbundenheit.

Oldenburg und Augsburg, im März 2024

Stephan Wernke und Klaus Zierer



Danksagungen

Ich habe dieses Buch meinem langjährigen Freund Michal Scriven gewidmet. Wie im gesamten Buch sichtbar wird, hat er mein Denken stark beeinflusst. Sein immenses Wissen, seine diesbezügliche Freigebigkeit und seine Fähigkeit, laut und tiefgründig zu denken, haben mich sehr bereichert. Janet und ich haben ihn kürzlich in San Francisco getroffen (er ist 94 Jahre alt) und er inspiriert mich nach wie vor. Wenn ich darüber nachdenke, was mein größter Beitrag zur akademischen Welt gewesen ist, dann ist es, dass ich 20 Jahre lang sein Chef und sein Gutachter gewesen bin. Er ist der Großvater der Evaluation; er hatte Gilbert Ryle (siehe Kapitel 3) als Doktorvater; er hat über Automechanik, wissenschaftliches Arbeiten, Computer und Computertechnik, angewandte Logik, Religions- und Geschichtsphilosophie, Argumentation, Kreativität und kritisches Denken, Kausalität und Erklärung, Psychologie, Psychiatrie und Psychotherapie, Parapsychologie und vieles mehr geschrieben. Er brachte es zu höchsten universitären Ehren im Rudern; sein erstes Buch behandelte die Gasturbine im Autodesign; er wurde als Physiker ausgebildet und wurde Philosoph und Pädagoge; er war der weltweit erste Professor für Evaluation (Janet ist der zweite); er erfand die formative und summative Evaluation, die zielfreie Evaluation und die Metaevaluation, und er gründete den FASTER Forward Fund. Seine Arbeiten zur Evaluation sind grundlegend und er hat nie aufgehört, über Evaluation als Querschnittsdisziplin zu schreiben und zu sprechen. Er ist großzügig, er hat sein außerordentliches Wissen über Messer mit meinen Söhnen geteilt und wir hatten eine großartige Zeit zusammen – danke, Michael und Mary-Ann.

In den vergangenen zwölf Jahren haben so viele Menschen zu diesem Buch beigetragen – als sie Visible Learning (VL+) in ihren Schulen einführten, die Bücher und Artikel über VL von mir und anderen lasen und Kritik übten und über VL bloggten. Das Gegenteil von Liebe ist nicht Hass, sondern Gleichgültigkeit, und ich danke Ihnen allen dafür, dass Sie sich genug Mühe gegeben haben, die Ideen zu lesen, zu kritisieren und zu nutzen.

Als ich *VLI* schrieb, sammelte ich alle Artikel über jeden der 150 Einflüsse, fuhr zu unserem Strandhaus in Tairua und las und schrieb. Es hat über 20 Jahre gedauert, bis ich die großen Botschaften herausgearbeitet hatte und der schöne Balkon mit Blick aufs Meer war meine Belohnung für dieses Unterfangen. Für dieses Buch hat die Technik die Führung übernommen, und es wurden keine Bäume mehr gefällt, um die Meta-Analysen oder Studien auszudrucken. Jetzt, da wir von Neuseeland nach Australien umgezogen sind, ist unser Strandhaus in Anglesea weniger vollgestopft mit Papieren und falsch abgelegten Artikeln. Normalerweise bedankt man sich bei seiner Frau, wenn man ein Buch schreibt (und sei es nur dafür, dass sie mir nicht in die Quere kommt, mein einsames Streben ignoriert und schimpft, ich solle mehr Zeit mit ihr und der Familie verbringen). Aber das ist bei mir nicht der Fall. Janet ist meine Partnerin und meine beste Kritikerin, und wir haben so manche Nacht über einzelne Einflussfaktoren, ihre Rolle im Gesamtbild und die zentrale Rolle des evaluativen Denkens diskutiert. Sie hat die Qualität der Botschaften, den Fluss, den Ton und die großen Ideen so verbessert, dass das Buch von einem Kompendium von 350 Einflüssen zu einer großen Geschichte wurde (und sie hat das Barometer in *VLI* und das Thermometer in dieser Fortsetzung entwickelt). Nach 40 gemeinsamen Jahren haben wir gelernt, über Ideen zu streiten, uns zu versöhnen und voranzukommen, nicht nur, um eine Familie zu gründen, sondern auch, um unsere Arbeit und unsere Karrieren voranzubringen.

Die Familie hat es so lange mit *VL* ausgehalten und drei von ihnen sind jetzt Lehrer. Ich bin so stolz darauf, dass sie Lehrer geworden sind – die beste Aufgabe der Welt – und ich wünsche ihnen, dass sie noch lange Freude daran haben, ihre Lernenden zu beeinflussen, während die anderen Verhaltensanalytiker, Evaluatoren und Projektmanager sind. Während wir *VL2* schrieben, wurden wir

Großeltern und haben nun Joel, Kat, Ella und Florence; Kyle, Jess, Emma, Danielle und Charlie; Kieran, Aleisha und Riley (John); sowie Edna und Patterson (die es liebten, wenn ich während der COVID-Pandemie zu Hause festsaß, denn dann gab es viele Spaziergänge und Leckerlis für die Hunde).

Shirley Clarke las und kritisierte jede Seite von *VL2*, schrieb den Kern der Überlegungen zu jedem Kapitel und sie weiß, wie Feedback zu geben ist. Andere, die verschiedene Kapitel kommentiert haben, sind Bill Turner, James Chapman, Steve Graham, Jennifer Buckingham, Korbinian Kiendl, Dan und Lawrence Ingvarson, Lyn Yates, Tim Shanahan, Julie und Ray Smith, Sam Rodgers, John Almarode, Doug Fisher, Nancy Frey, Geoff Petty, Doug Reeves, Greg Ashman, Amy Berry, Peter Blatchford, Lyn Sharratt, Wolfgang Beywl, Klaus Zierer, Dylan Wiliam und Jim Knight. Wir waren nicht immer einer Meinung, aber das ist ja das Schöne an Kritikerinnen und Kritikern. So viele andere haben zu den Ideen beigetragen, darunter alle VL-Berater von Corwin, Osiris, Bazalt und Cognition - sie sind das Herzstück der Umsetzung und ich bin sicher, dass sie ihren Einfluss in diesen Kapiteln erkennen. Ich danke Michael Soules, Chris Devling, Sonja Hollins-Alexander und Vania Tiatto, die diese Arbeit geleitet haben.

Bruce Roberts ist mein Routledge-Redakteur, seit wir uns zum ersten Mal in Budapest getroffen haben. Er begleitete mich bei *VLI*, dieser Fortsetzung und vielen anderen Unternehmungen. Dein Engagement für Qualität, dein sanftes, aber überzeugendes Drängen, die großartigen Mahlzeiten und die lustigen Momente machen das alles lohnenswert – und sein Team in der Produktion hat so viel dazu beigetragen.

Und allen Leserinnen und Lesern unter Ihnen, die die Ideen in ihrer Schule und ihrem Unterricht umgesetzt haben, wünsche ich, dass Sie weiterhin Kritik üben, sie bewerten und ihre Wirkung erkennen.

Die Herausforderung

Seit der Veröffentlichung von *VLI* im Jahr 2008 (Hattie, 2008) hat sich so viel verändert: Das Internet ist jetzt noch allgegenwärtiger, es gibt iPads, der Klimawandel steht auf der Tagesordnung, die gleichgeschlechtliche Ehe wurde in vielen Ländern legalisiert, wir haben die globale Finanzkrise erlebt, das Vereinigte Königreich hat für den Brexit gestimmt, es gab den Arabischen Frühling und es gab 26 Bürgerkriege und internationale Kriege. Darüber hinaus haben wir den Aufstieg populistischer Führer, die MeToo-Bewegung, die Gig-Economy, Angry Birds, Selfie-Sticks, die Entdeckung des Higgs-Bosons, die ersten Bilder eines Schwarzen Lochs, den Aufstieg von Elektroautos, *Fifty Shades of Grey* und die COVID-Pandemie erlebt, die zu erheblichen Störungen in Schulen geführt hat.

In diesem Jahrzehnt trat meine Forschungskarriere als Psychometriker und in der Bildungspsychologie in den Hintergrund und *Visible Learning* übernahm die Führung. Ich bin Großvater geworden (fünf Enkelkinder, ein Enkel); nach 25 gemeinsamen Jahren wurde ich von Elvis Presley mit Janet verheiratet (danke, Peter DeWitt), zog von Neuseeland nach Australien (und folgte Janet, die als Professorin für Evaluation angeworben wurde), erlebte, wie meine Jungs zu Männern wurden; ich ging in den Ruhestand und genieße nun politische Aufgaben in Regierungsbehörden. *Visible Learning* hat meine Art zu denken und zu schreiben wirklich verändert. Seit *VLI* habe ich mehr als 75 Artikel und mehr als 40 Bücher veröffentlicht (Abbildung 1.1) und auf mehr als 400 Konferenzen oder Veranstaltungen weltweit zum Thema *VLI* gesprochen (Anhang A). In Anhang B sind die Kapitel des Buchs *Visible Learning: Guide to Student Achievement – Schools Edition* (Hattie & Anderman, 2022) vermerkt. Diese Kapitel bieten mehr Hintergrund über viele der Einflüsse in dieser Fortsetzung (siehe auch Hattie & Anderman, 2013, für weitere Kapitel). Ein Fehler vieler kritischer Stimmen besteht darin, sich zu sehr auf das eine Buch von 2009 zu konzentrieren, da ich viele der Ideen in *VLI* in diesen anderen Quellen erweitert, geklärt und erforscht habe.

Die Verbesserungen des *VLI*-Modells sind auch eine Folge der Zusammenarbeit mit herausragenden Schul- und Unterrichtsentwicklerinnen und -entwicklern, die das Modell in die Schul- und Unterrichtspraxis implementiert haben. Das *VL+*-Modell wurde von diesem Team in über 10.000 Schulen weltweit umgesetzt. Darüber hinaus haben uns viele Bildungsverantwortliche, Schulleitungen und Lehrpersonen über die Verbesserung dieses Prozesses informiert, was eine weitere Verfeinerung des *VLI*-Modells ermöglichte, eine stärkere Herausarbeitung der Prioritäten, eine genauere Betrachtung der Umsetzungsmodalitäten und der Ausweitung des Erfolgs sowie eine detailliertere Sammlung von Belegen für die Auswirkungen des Modells. Als mich das australische Bundesministerium einlud, den Vorsitz im Vorstand des Australian Institute of Schools and Leaders (www.aitsl.edu.au/) zu übernehmen, hatte ich auch eine stärker politisch ausgerichtete Rolle. Dies bedeutete, dass ich mich regelmäßig mit allen Ministern der Bundesstaaten und Territorien und ihren Generaldirektoren treffen musste (in den neun Jahren als Vorsitzender habe ich nun bereits den 61. Minister bzw. Generaldirektor getroffen – wir wechseln unsere Anführer oft aus).



Abbildung 1.1 Bücher zum Thema VL

Ich stand unter dem Druck, eine zweite Auflage von *Visible Learning* zu schreiben, aber ich widerstand. Ich wusste, dass selbst meine engsten Kolleginnen und Kollegen nicht glücklich sein würden, wenn sich die Wirkung eines Einflusses auch nur geringfügig ändern würde, und damit zu einer neuen Rangliste, die mit den früheren Versionen unvereinbar wäre, und zu verkürzten Diskussionen führen würde.

Die Forschungslandschaft hat sich ebenfalls weiterentwickelt. Es gibt viel mehr Meta-Analysen und nun auch andere Synthesen von Meta-Analysen. Die Debatten über die Interpretation von Effektstärken sind ebenfalls anspruchsvoller geworden. Ich habe viel von meinen Kritikerinnen und Kritikern gelernt: Man kann seine Karriere als Akademiker verbringen, ohne dass es viele bemerken, daher ist es eine Ehre, solche weltberühmten Kritikerinnen und Kritiker zu haben.

Das vorliegende Buch ist also keine zweite Auflage, sondern eine Fortsetzung, die der großen, zugrunde liegenden Geschichte mehr Aufmerksamkeit widmet. Diese Geschichte deckt viel mehr ab als in *VLI*, zumal zusätzliche 1.300 Meta-Analysen hinzugefügt wurden (die Veröffentlichung von

Meta-Analysen wurde 2009 nicht eingestellt). Der Zuwachs ist in Abbildung 1.2 dargestellt, mit etwa 800 in *VLI* und jetzt 2.100 in dieser Fortsetzung.

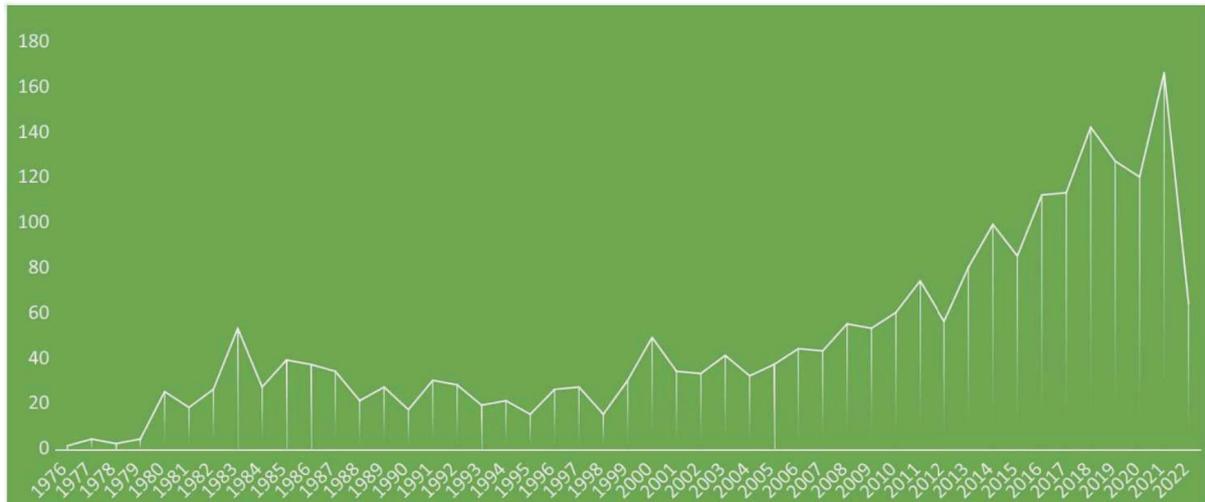


Abbildung 1.2 Anzahl der Meta-Analysen pro Jahr.

Die Details sind wichtig, aber ich habe die Details und Verweise für alle Meta-Analysen in eine Online-Plattform verlagert. MetaX (www.visiblelearningmetax.com) ist eine kostenlose Website mit allen Meta-Analysen, Referenzen, einem Glossar und FAQ, die regelmäßig aktualisiert wird, um die neuesten Meta-Analysen aufzunehmen. Bei dieser Fortsetzung gibt es keine Auflistung aller Meta-Analysen oder Ranglisten, sondern die Betonung liegt auf der wichtigsten Geschichte oder den großen Botschaften von *Visible Learning*. Zu viele haben *VLI* gelesen und kritisiert, als ob jeder Einfluss für sich alleinstünde, obwohl ich mit Nachdruck darauf hingewiesen habe, dass es auf die Kernbotschaften und die Überschneidung ankommt. In *VLI* habe ich argumentiert, dass das Ziel darin bestand, eine Geschichte über die Macht der Lehrpersonen und des Feedbacks zu entwickeln und ein Modell des Lernens und Verstehens zu konstruieren. Die Fortsetzung hebt die Geschichte hervor und bietet zudem eine Reihe von unterstützenden Berichten für diese Geschichte.

Die Herausforderungen

Die gleichen Herausforderungen, die in *VLI* identifiziert wurden, sind immer noch aktuell: Wie kann man den Fokus wechseln von dem, was funktioniert, zu dem, was am besten funktioniert? Warum ist die derzeitige Schulgrammatik, die vielen, aber eben bei Weitem nicht allen dient, in so vielen Klassenzimmern so fest eingebettet, und wie kann man diese verbessern? Warum ist die Lernentwicklungskurve für Lehrpersonen nach den ersten Berufsjahren verhältnismäßig flach? Wie kann man über die generellen Inputs (die Unterrichtsstunden, die Lernziele) hinaus zu den Entscheidungen, die Lehrpersonen ganz aktuell beim Unterrichten treffen, übergehen? Wie kann man sich mehr auf das Lernen konzentrieren? Und was am wichtigsten ist: Wie kann man Forschungsergebnisse als Teil der Diskussionen in die Schulen und die Kollegien bringen?

Der Einsatz erfolgreicher Umsetzungsmodelle in Bildungssystemen, Schulen und Klassenzimmern muss stärker in den Mittelpunkt gerückt werden, ebenso der zunehmende Führungswechsel und seine Auswirkungen, die fehlende Implementation der Evaluierung von Programmen und die überstürzte Suche nach Misserfolgen und deren Behebung, während erstaunliche Erfolge und die Frage, wie wir diese Erfolge ausbauen können, ignoriert werden.

Zu den bedeutendsten Verbesserungen in den letzten zehn Jahren gehört, dass Forschungsergebnisse heute ein regelmäßiges Diskussionsthema in den Schulen sind – dies bedeutet nicht, dass sie privilegiert sind oder ausschlaggebend sind. Die Verbreitung von leichter zugänglichen Ressourcen hat diesen Prozess unterstützt – wie die Education Endowment Foundation, Evidence for Learning, WhatWorks Clearing House, Best Evidence Encyclopedia und andere. Gleeson et al. (2022) befragten 492 Lehrkräfte aus 414 Schulen zu ihrer Nutzung von Forschungsergebnissen. Etwa zwei Drittel gaben an, in den letzten 12 Monaten Forschungsergebnisse genutzt zu haben (91 % der Schulleitungen und 61 % der Lehrpersonen), insbesondere um „eine neue Initiative zu entwerfen und zu planen“ und um „Unterstützung für ein wichtiges Thema oder eine wichtige Entscheidung zu

mobilisieren“, was das Gegenteil dessen ist, was Kirschner et al. (2022) als Pädagogikwahn bezeichnen: eine Reihe von Überzeugungen und Annahmen darüber, was in einem Klassenzimmer passieren sollte, die gekennzeichnet ist durch die Ablehnung von empirischen Befunden, die Akzeptanz des romantischen Denkens, das Zelebrieren des Oberflächlichen in Form von Modeerscheinungen und Mythen, die Behauptung, dass Pädagogik ein Selbstzweck sei, und die Schaffung einer oft toxischen Kultur für Lehrpersonen mit unhaltbarer Arbeitsbelastung.

Der Schwerpunkt muss nun auf der Optimierung der Umsetzung und Mobilisierung dieser Forschung liegen. Unsere Arbeit hat gezeigt, dass dies bedeutet, dass man sich mit den Denkweisen der Lehrpersonen über Evidenz sowie mit ihrer Implementation und Evaluation befassen muss. Dies ist in der vorliegenden Fortsetzung stärker ausgeprägt als in *VLI*.

Die Missverständnisse

Ich habe fünf Versionen von *VLI* geschrieben. Die vierte war die beste, denn sie strotzte nur so vor detaillierten Statistiken, Varianten, Bedingungen, Moderatoren, kurz: ein 500-seitiges Prachtexemplar. Meine beste Kritikerin, Janet, las sie und fragte: „Für welche zwei Menschen auf der Welt hast du das geschrieben?“ Autsch. Ich habe diese Version vernichtet, und die letzte und fünfte Version wurde geschrieben. Diesmal, so sagte sie, hast du das Publikum von zwei auf etwa zehn Personen vergrößert – dann erfand Janet die Barometerskalen, um für mehr Lesbarkeit und auch Interesse zu sorgen. In letzter Minute fügte ich den Anhang hinzu, in dem alle Einflüsse in einer Rangfolge aufgelistet waren – die Rangliste. Die Rangliste führte zu dem häufigsten Missverständnis der Forderungen – zu viele wollten die Einflüsse an der Spitze abhaken und diejenigen, die am Ende der Rangliste stehen, wurden verunglimpft. Andere verkündeten, dass die Einflüsse nicht besonders seien, was zeigte, dass sie keine Seite des Buches gelesen hatten – denn es ging um die Überschneidungen. Genauer gesagt ging es darum, die zugrundeliegenden Ursachen jener Einflüsse zu erklären, die über und unter dem Durchschnitt aller Einflüsse liegen, dem Umschlagpunkt von 0,40. Die Botschaft ist in den Mustern zu finden, nicht in den Details. Es geht um die Interpretation der Belege, nicht um die Belege an sich, um die Kühnheit der großen Botschaften, nicht um das Drumherum der Dezimalstellen.

In *VLI* habe ich darauf hingewiesen, dass die Auseinandersetzungen darüber, was als Evidenz für einen Kausalzusammenhang gilt, so heftig sind wie nie zuvor – einige befürworten nur randomisierte Kontrollstudien (RCT). RCTs sind Studien, bei denen die Probanden nach einem streng zufälligen Verfahren einer Versuchs- oder Kontrollgruppe zugeteilt werden. Es gibt inzwischen viele RCTs in der Bildungsforschung, obwohl viele argumentieren, dass sie streng genommen gar nicht möglich oder sinnvoll sind (Styles & Torgerso, 2018). Ich ziehe es vor, eine Methode nicht als Goldstandard zu verwenden und die Messlatte höher zu legen, indem ich Michael Scrivens (2007) Behauptung vorziehe, dass sich ein höherer Goldstandard auf Studien bezieht, die in der Lage sind, Schlussfolgerungen zu ziehen oder „jenseits begründeter Zweifel“ sind. Vielleicht ist dies eine zu hohe Messlatte, aber zumindest sollten die Beweise „klar und konsistent“ sein und mehr als ein „Beweis des ersten Anscheins“, über einen „wahrscheinlichen Verdacht“ hinausgehend und sicherlich „einen begründeten Verdacht“ übersteigen. Ich habe behauptet, dass ein Hauptziel von *VLI* darin besteht, aus den Studien und ihren Effektstärken eine Geschichte zu weben, die überzeugend, kohärent und verallgemeinerbar ist. Diese Geschichte zu entwickeln, bleibt die Aufgabe dieser Fortsetzung.

Ich habe in *VLI* darauf hingewiesen, worum es in dem Buch *nicht* geht, was zu viel Verwirrung geführt hat. Ich hätte deutlicher sagen sollen, dass es nicht um das Leben selbst in einem Klassenzimmer geht, sondern eher um die Einflüsse auf das Lernen, Denken und Handeln von Schülerinnen und Schülern in Klassenzimmern – auf der Klassen-, Lehrer-, Schul- und Systemebene. Ich habe versucht, die Forschung zur Unterrichtsbeobachtung zusammenzufassen, hatte aber Schwierigkeiten, mich für ein Standard-Effektmaß zu entscheiden. Stattdessen haben wir eine App (VisibleClassroom) entwickelt, die sofortige Transkripte von Lehrgesprächen mit automatisiertem Feedback liefert und in 17.000 Stunden Unterricht eingesetzt wurde (siehe Kapitel 8). Wir haben ein experimentelles Klassenzimmer so aufgebaut, dass wir die intensive Forschung von Graham Nuthall (2007) mit moderner Technologie repliziert haben und viele Daten über das „verborgene Leben der Lernenden“ (Spejkal, 2022) gewannen. Eine große Synthese der Unterrichtsbeobachtungsforschung über das Leben der Schülerinnen und Schüler in den Klassen ist dringend erforderlich.

Ich stellte fest, dass es bei *VLI* nicht um ganz allgemeine Faktoren ging, auf die Lehrpersonen wenig Einfluss haben, wie z. B. soziales Milieu, Armut, familiäre Ressourcen, Gesundheit in Familien

und Ernährung. Lehrpersonen können diese gesellschaftlichen und familiären Probleme nicht ändern. Diese Probleme können jedoch erhebliche Auswirkungen auf ihre Einflüsse innerhalb der Schule haben. Ein wichtiges Ziel der Schulbildung ist es, jungen Menschen Wege aus diesen Verhältnissen zu eröffnen. Kritische Stimmen griffen diese Behauptung auf und bemerkten, dass ich mich nicht um diese Faktoren kümmere und dass Probleme wie Armut nicht am Schultor aufhören. Auch hier zeigt sich, dass sie das Buch nicht gelesen haben, denn es werden viele *Auswirkungen* dieser Probleme innerhalb der Schule erörtert.

Ein weiteres Missverständnis war die Aussage in *VLI*, es handele sich nicht um ein Buch über Kritik an Meta-Analysen. Ich argumentierte, dass diese Themen an anderer Stelle behandelt werden. Dies führte dazu, dass viele sagten, dass ich mich nicht um die Qualität kümmere und dass ich minderwertige Forschung mit aufgenommen hätte. Auch hier haben einige nicht viel weitergelesen. Slavin (2018) war sehr verärgert darüber, dass ich seine „Best-Evidence“-Methode nicht aufgenommen habe, aber das erkennt, dass ich Meta-Analysen synthetisiert habe, keine Originalstudien. Wie viele argumentiert haben (Glass, 2019), ist die Frage nach der Qualität ein Hauptschwerpunkt von Meta-Analysen, und solche Studien sollten nicht unbedingt verworfen werden. In *VLI* habe ich durchweg bestimmte Studien kommentiert (z. B. im Abschnitt über Lernstile, S. 195-197) und habe einige merkwürdige Ergebnisse in einigen Studien (z. B. zum Thema Ganzwortmethode) in Frage gestellt. Vielleicht hätte ich das Thema Qualität noch stärker in den Vordergrund stellen sollen. In Kapitel 2 dieser Fortsetzung werden nun viele dieser Fragen behandelt und es wird ein Robustheitsfaktor für die Meta-Analysen eingeführt.

Andere haben behauptet, dass ich auf Lernleistung fixiert sei und es doch viele andere Merkmale der Schulbildung gebe. Dies wiederum ignoriert die Aussage im Vorfeld, dass „es natürlich viele Ergebnisse der Schulbildung gibt, wie z. B. Einstellungen, physische Ergebnisse, Zugehörigkeit, Respekt, Bürgersinn und die Liebe zum Lernen. Dieses Buch konzentriert sich auf die Leistung der Lernenden und das ist eine Einschränkung der vorliegenden Synthese“ (S. 6). Es ist erfreulich, Synthesen von Meta-Analysen zur Motivation (Jansen et al., 2022) und zu Lernstrategien (Hattie & Donoghue, 2016) zu sehen. Ich freue mich auf weitere Zusammenfassungen von Einflussfaktoren auf einige der anderen wichtigen Ergebnisse der Schulbildung. Ich entschuldige mich nicht dafür, dass ich mich auf Lernleistung konzentriere und frage mich stattdessen, wie Schulen aussehen würden, wenn Lernleistung kein wichtiges Ergebnis von Schulbildung wäre.

Andere haben meine utopische Sicht der Bildung infrage gestellt oder kritisiert, dass ich Lehrpersonen verunglimpfe oder sie zum Sündenbock für die Kritik mache. Indem ich Lehrpersonen in den Mittelpunkt der stärksten Einflüsse auf die Lernenden stelle, so der Vorwurf, mache ich sie für schulische Misserfolge verantwortlich. Es ist genau das Gegenteil der Fall: Die wunderbaren Erfolge, die wir in den Schulen haben, verdanken wir der Professionalität von Lehrpersonen. Ein anderer Kritiker schrieb, ich sei inkonsequent, wenn ich Lehrpersonen lobte und dann auf ihre Grenzen hinwies – was zeigt, dass dieser Kritiker nicht versteht, was Varianz bedeutet. Die *VL*-Daten sind eindeutig: Exzellenz ist überall um uns herum und es gibt viel davon in unseren Schulen. Das *VL*-Modell verlangt von uns, dass wir den Mut haben, diese hervorragenden Leistungen zuverlässig zu identifizieren, eine Koalition hinsichtlich dieses Erfolgs zu schmieden und andere mit geringeren Einflüssen auf ihre Lernenden einzuladen, sich dieser Koalition anzuschließen. Das größte Hindernis ist oft ein Mangel an Mut.

Zurückschauen, um vorwärtszukommen

Die *VL*-Forschung schaut immer zurück – und das derzeitige System weist eine Fülle von Spitzenleistungen auf. *VL* zielt darauf ab, diese vergangene Forschung zu nutzen, um die gemeinsamen Attribute dieser herausragenden Leistungen zu identifizieren. *The Turning Point* führt aus, dass diese hervorragenden Leistungen auf der Kompetenz des evaluativen Denkens von Lehrpersonen beruht, die das Wesen unseres Berufs ausmacht (Rickards et al., 2021). Diese Form des Denkens wird im vorliegenden Buch näher untersucht. Was mich antreibt, sind die vielen exzellenten Schulleitungen und Lehrpersonen, die ich treffe, wenn ich die Welt bereise. Vielleicht könnte man mich dafür kritisieren, dass ich etwas naiv optimistisch bin, aber ich habe Beweise dafür, dass ich die Welt der Schulen positiv sehe.

In *VLI* habe ich einige entscheidende Teile des Modells ausgelassen. Ich habe zum Beispiel die Macht der Lehrererwartungen betont, aber nicht den noch stärkeren Einfluss der Schülererwartungen;

ich habe mich zu sehr auf die Lehrpersonen konzentriert, die Feedback geben, aber nicht die Bedeutung hervorgehoben, ob oder wie dieses Feedback von den Lernenden gehört, verstanden und umgesetzt wird; ich habe nicht genug Gewicht auf die Qualität der Umsetzung von Lehrstrategien oder schulischen Maßnahmen gelegt; ich habe das Wesen der besonderen Art des Denkens, das effektive Lehrpersonen auszeichnet, nicht ausreichend erforscht; und ich habe die Wichtigkeit übersehen, Lernenden vielfältige Lernstrategien und Kompetenzen für die Teamarbeit zu vermitteln. Es fiel mir schwer, die bemerkenswerten Unterschiede in der Wirksamkeit vieler Lehrstrategien zu verstehen, ebenso die geringe Wirkung der Fachkompetenz der Lehrpersonen und die geringe Wirkung vieler vertiefender Lernstrategien. In der Tat erhalte ich viele E-Mails von Befürwortern von Methoden wie dem problembasierten Lernen, die behaupten, ich hätte Unrecht (sie meinen, die Forschung sei falsch), und sie verstehen nicht, dass sie selbst es sind, die vielleicht brilliant sind. Aber für jede Lehrperson, die gut abschneidet, gibt es eine andere, die weit darunter liegt (das ist es, was ein Durchschnitt bedeutet). Ich habe versucht, in dieser Fortsetzung auf diese Probleme einzugehen.

Lehren und Lernen

Die zentrale Botschaft hat sich nicht geändert und ist ganz einfach: Es kommt darauf an, wie Lehrpersonen, Schulleitungen, Eltern und Lernende *denken*, denn ihre Denkweise, ihre Art zu denken, zu interpretieren und zu bewerten, ist für den Erfolg des Unterrichts entscheidend. Es ist ihre diesbezügliche Kompetenz, die zur Auswahl von Interventionen führt, zur Ausarbeitung und Erklärung von Lernzielen und Erfolgskriterien, ihr Wissen, wann ein Lernender diese Lernziele erreicht hat oder nicht, ihr ausreichendes Verständnis des Lernstandes der Schülerinnen und Schüler, das sie in die Aufgabe einbringen, und ihr ausreichendes Verständnis des Inhalts, um sinnvolle und herausfordernde Aufgaben auf verschiedenen Wegen zum Lernerfolg zu bieten. Das *VL*-Modell erfordert eine Lehrperson, die weiß, wie man eine Reihe von Lehrstrategien einsetzt, um den Lernenden zu helfen, wenn sie etwas *nicht* verstehen, die ihnen Rückmeldung auf den verstandenen oder falsch verstandenen Inhalt gibt und so die Wirkung des Feedbacks zu maximieren, das Lernende erhalten. Dies ist eines der Hauptprinzipien der Montessori-Methode. Das „Sich überflüssig machen“ bezieht sich auf die Vorstellung, dass die Lehrpersonen schrittweise die Verantwortung reduzieren und den Lernenden beibringen, ihre eigenen Lehrpersonen zu werden (zu wissen, was zu tun ist, wenn sie nicht wissen, was zu tun ist). Diese Aussage wird im aktualisierten *Visible Learning*-Modell (Kapitel 3) untersucht.

Ein Faktor, der im *VLI*-Modell erwähnt wurde, war Leidenschaft – die Freude, der Ansporn und die Begeisterung der Lehrperson, die Lernenden zum Lernen zu bringen. Ich habe insbesondere festgestellt, dass Leidenschaft zu den am meisten geschätzten Ergebnissen der Schulbildung gehört. Obwohl sie in den in diesem Buch besprochenen Studien nur selten untersucht wird, ist sie für viele der Einflüsse verantwortlich, die den Unterschied bei den Ergebnissen ausmachen. Es braucht mehr als inhaltliches Wissen, geschickte Lehrtätigkeit oder engagierte Lernende, um einen Unterschied herbeizuführen (obwohl dies hilfreich ist). Es erfordert auch eine Liebe zum Inhalt, eine fürsorgliche Haltung, eine Ethik, um anderen eine Vorliebe oder sogar Liebe für das unterrichtete Fach zu vermitteln. Darüber hinaus ist das Bekenntnis wichtig, dass eine Lehrperson nicht nur lehrt, sondern auch lernt – typischerweise durch die Lernprozesse und -ergebnisse der Lernenden. Dieser Sachverhalt wurde selten zur Kenntnis genommen. Es gibt immer noch zu wenige Studien über die Macht der Leidenschaft, aber sie bleibt ein sichtbares Merkmal vieler Klassenzimmer, insbesondere bei den Lernenden.

In *VLI* gab es viele Bezüge zum Lernen. In dieser Fortsetzung bin ich konkreter geworden und habe ein Kapitel der Frage gewidmet, wie Lehrpersonen die Lernstrategien der Lernenden beeinflussen können (Kapitel 12) und wie Lehr- und Lernstrategien besser aufeinander abgestimmt werden müssen. Im Rahmen des Science of Learning Project (www.slrc.org.au) haben Greg Donoghue und ich ein Team geleitet, das eine Meta-Analyse zu Lernstrategien und ihren Auswirkungen auf die Leistung durchgeführt hat. So entstand ein integratives Modell des Lernens (Donoghue & Hattie, 2021; Hattie & Donoghue, 2016). Wir sahen die Notwendigkeit, dieses Modell zu erstellen, weil einige Strategien an einigen Punkten des Lernprozesses wirksam waren, an anderen Punkten jedoch nicht. Dies führte zu der Entdeckung, dass sich Lernen qualitativ in Phasen des Wissens und Könnens einteilen lässt (Oberflächenverständnis, Tiefenverständnis und Transferverständnis von *VLI*) – eine Entdeckung, die in den eher qualitativen Modellen des Lernens

gut untersucht wurde (z. B. Marton, 2014). Als wir die verschiedenen Lehrstrategien mit diesem Modell abglichen, war bei den meisten Lehrstrategien nicht ersichtlich, wie sie auf die Lernprozesse einwirkten oder sie wirkten rein zufällig. Der intendierten Wirkung von kognitiven Fähigkeiten und Lernfähigkeiten sollte mehr Aufmerksamkeit geschenkt werden.

Das Mantra von *VLI* bleibt bestehen: „Lernen sichtbar machen“ bedeutet, dass Lehrpersonen das Lernen mit den Augen der Lernenden sehen und dass Lernende ihre eigenen Lehrpersonen werden. Es geht um mehr als die Art und Weise, wie wir unterrichten, und viel mehr um die Auswirkungen unseres Unterrichts. Es geht um die Fachkompetenz von Lehrpersonen, aber vor allem darum, wie sie evaluierend denken, wie sie zusammenarbeiten, um Lernprozesse von Lernenden zu interpretieren, damit sie ihre nächsten Unterrichtsentscheidungen treffen können. Es geht um ihre Offenheit, Feedback zu suchen, zu nutzen und zu erhalten und um ihre kollektiven Wirksamkeitserwartungen. Es geht darum sicherzustellen, dass alle (und das bedeutet *alle*) Lernende in einem Schuljahr Input mindestens ein Schuljahr an Lernzuwachs erreichen (und wie ein Jahr Lernzuwachs in dieser Klasse/Schule/Bezirk aussieht). Lehrpersonen müssen Lernende auch für die Herausforderungen des Lernens begeistern, Fähigkeiten und Selbstvertrauen entwickeln, um in Teams und mit den Gleichaltrigen zu arbeiten und den Respekt gegenüber sich selbst und anderen zu fördern.

Ein Jahrzehnt der Wirkung

Das Modell, das in dieser Fortsetzung entwickelt wird, ist stark von der Implementation von *VL* in Schulen auf der ganzen Welt geprägt, die kurz nach der Veröffentlichung des Buches begann. Ich wurde eingeladen, an Schulen zu sprechen, aber die Präsentationen im Nachmittagsprogramm der Schule hatten keine großen Auswirkungen. Zum Zeitpunkt der Veröffentlichung leitete ich ein Team, das ein Beurteilungsmodell für Grundschulen und weiterführende Schulen in Neuseeland entwarf, entwickelte und umsetzte (siehe www.easTTle.org). Ich hatte dem Ministerium mitgeteilt, dass die Entwicklung abgeschlossen sei und sie nun die Pflege des Instruments übernehmen müssten, was bedeutete, dass einige Mitglieder des Teams überflüssig würden.

Sie baten darum, *VL* in Schulen einzuführen, und ich stellte einige Bedingungen. Es musste anpassbar sein (und durfte nicht auf den Fähigkeiten einer oder weniger Personen beruhen), es musste Auswirkungen auf das Lernen der Lernenden haben und die Effekte nachweisen (viel mehr als die üblichen Kriterien von Lehrpersonenfort- und -weiterbildungen, nämlich ob die Lehrpersonen zufrieden waren und viel gelernt hatten), und es durfte nicht bedeuten, dass ich Vorträge halten musste (meine Fähigkeiten liegen im Bereich der Forschung und nicht in der Lehrkräftefort- und -weiterbildung). Debra Masters leitete das Team und schon bald waren alle drei Bedingungen erfüllt und die Nachfrage stieg.

Ursprünglich waren die Programmentwicklung und -durchführung bei Cognition in Neuseeland untergebracht. Cognition ist ein Treuhandunternehmen, das eine Erfolgsbilanz bei der Bereitstellung (in Neuseeland und in Übersee) vorweisen konnte, und ich kannte die Qualität der Mitarbeitenden von Cognition (ich war einige Zeit lang im Vorstand). Nach einigen Jahren erfolgreicher Zusammenarbeit änderte Cognition die Richtung und die Heimat von *VL+* wurde zu Corwin in die USA verlegt (www.visiblelearning.com). Ein neues Team unter der Leitung von Julie Smith, eine umfassende Aktualisierung der Materialien und eine größere Reichweite waren die Folge. Zehn Jahre sind seit dem Beginn der Umsetzung vergangen und Clinton und Clarke (2020) haben eine Evaluierung der umfangreichen Datenbank, der Fallstudien und der qualitativen Daten durchgeführt. Der Titel ihres Berichts lautete „A Decade of Impact“. Diese Umsetzung und Evaluierung haben mein heutiges Denken über das Modell stark beeinflusst.

Der Schwerpunkt von *VL+*

VL+ ist eine Lehrkräftefortbildung, die Professionalität und Schulentwicklung fördert. Zu den Aktivitäten gehören Fortbildungsveranstaltungen, Coaching, Evaluierung und Bewertung sowie eine Reihe von Arbeitsmaterialien. In den letzten Jahrzehnten hat sich das Hauptthema mit dem Aufkommen neuer Forschungsergebnisse gewandelt. Angefangen hat es mit der Macht des Feedbacks und ging dann zu „Kenne deinen Einfluss“ über, wobei der Schwerpunkt auf der Vermittlung von Fähigkeiten lag, jene Strategien zu implementieren und zu evaluieren, die höchstwahrscheinlich einen großen Einfluss auf die Lernleistung haben. Das anhaltende Plädoyer, den Erfolg eher im Hinblick auf

das Lernen der Schülerinnen und Schüler als auf die Zufriedenheit der Lehrkräfte zu sehen, führte zu einem tieferen Dialog über Absichten und Auswirkungen, wobei Martin Luther Kings Appell verwendet wurde, sich zuerst auf den Traum zu konzentrieren (Hattie & Zierer, 2018, S. 166-168).

Die Synthese des Lernens zeigte, wie wichtig es ist, den Unterricht so anzupassen, dass er Lernstrategien und Lernstrategien einbezieht sowie die Kenntnis des richtigen Zeitpunkts und Kontextes, um bestimmte Strategien anzuwenden (die Idee von Kenny Rogers „Wissen, wann man sie halten muss, wissen, wann man sie spielen muss“, siehe Kapitel 12).

Dort, wo *VL+* gut umgesetzt wurde, zeigte sich eine starke Ausprägung kollektiver Wirksamkeitserwartungen in der Schule. Erfolgreiche Innovationen können zu einer verbesserten Zusammenarbeit und Zufriedenheit der Lehrpersonen führen (Blömeke et al., 2021). Wir haben die Umsetzung weiter überwacht, um zu erkennen, wann sich der Schwerpunkt wieder auf Veränderungen im Lehrerzimmer und nicht im Klassenzimmer verlagert. Daher begannen wir mit der Forderung nach kollektiven Wirksamkeitserwartungen, um mindestens ein Jahr Wachstum für ein Jahr Wirkung zu erzielen. Dies führte dazu, dass wir uns darauf fokussierten in Erfahrung zu bringen, wie das Wachstum in diesem Jahr aussieht, zu wissen, was alle Lehrpersonen mit Wirkung meinten, und die Beweise für diesen Begriff der Wirkung in den Lernprozessen der Lernenden zu sehen. Die härteste Lektion aus dem Jahrzehnt der Arbeit in Schulen war weniger das Verständnis und die Nutzung von Forschung, sondern dass es vielen Schulen an einem Konzept für eine umfassende Umsetzung fehlte. Die Schulen führten *VL+* ein und hofften, dass es funktionierte. Also verbrachten wir viel Zeit mit der Entwicklung des DIIE-Modells. In jüngster Zeit zielt die Forschung darauf ab, die spezifischen Fähigkeiten von Lehrpersonen zum evaluativen Denken zu ermitteln, die die tiefgreifende Umsetzung und die daraus resultierenden Auswirkungen auf die Lernenden unterstreichen.

Das DIIE-Modell

Unsere Erfahrungen in den Schulen und die Implementationsforschung haben zur Entwicklung des DIIE-Modells (Diagnose, Intervention, Implementation und Evaluation, siehe Abbildung 1.3) geführt, um eine tiefgreifende Anwendung von *VL* zu gewährleisten. Nachdem man sich darüber geeinigt hat, was Wirkung bedeutet, besteht der Bedarf an einer sehr guten Diagnose, einer Auswahl von wirklich erfolgsversprechenden Maßnahmen bezüglich dieser Diagnose, es muss eine Umsetzungsstrategie gewählt werden (welche die Überwachung der Qualität, des Vertrauens und der Akzeptanz sowie der Dosierung der Umsetzung beinhaltet).

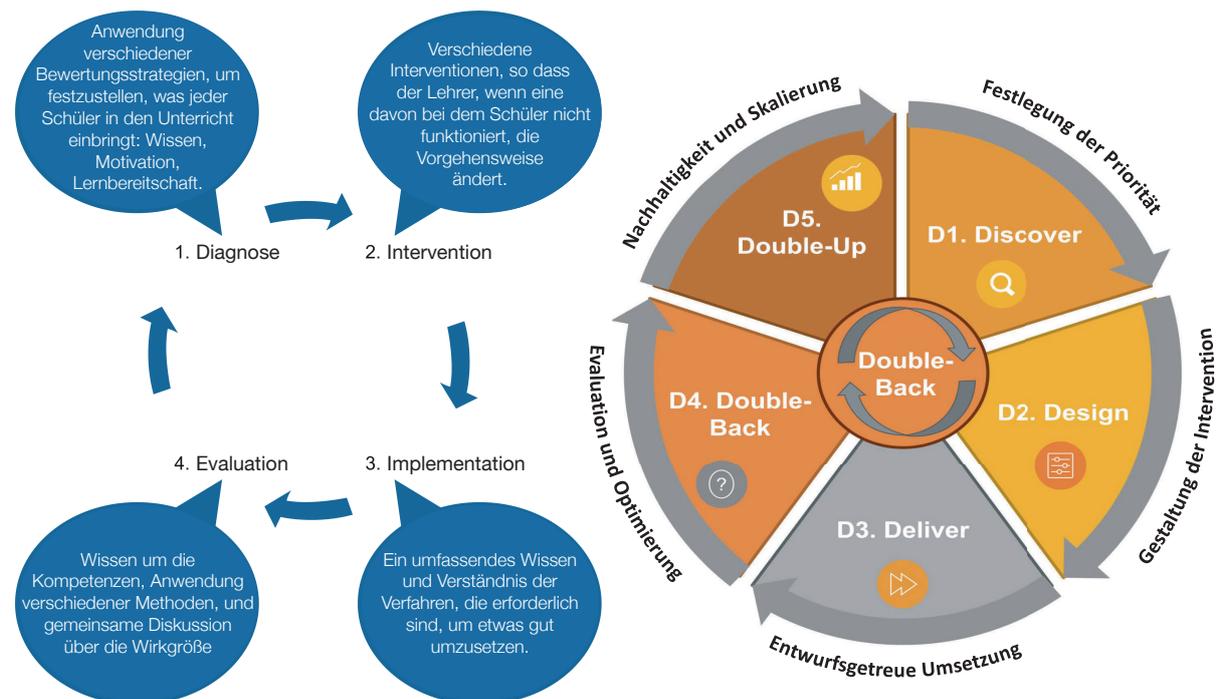


Abbildung 1.3 Das DIIE-Modell und das 5D-Modell

Zudem muss eine durchgängige Evaluierung und auch eine Evaluierung der Auswirkungen auf die Lernenden erfolgen. Allzu oft werden Interventionen ausgewählt, bevor die Probleme diagnostiziert wurden, die mit der Intervention angegangen werden sollen; allzu oft werden Interventionen nicht ausreichend eingeführt und scheitern, weil sie nicht konsequent umgesetzt wurden; allzu oft führt die alltägliche Anpassung dazu, dass dem Modell Tipps und Tricks hinzugefügt werden, so dass die Intervention nicht umgesetzt wird; und allzu oft wird in den Schulen keine Evaluierung vorgenommen, da sie bereits zur nächsten Intervention übergegangen sind. Das DIIE-Modell verfolgt das Ziel, den Verantwortlichen mehr Strenge und klarere Schritte an die Hand geben, um die Auswirkungen der Interventionen auf formative Weise zu überwachen und zu bewerten.

Tabelle 1.1 Ein Rahmen für das 5D-Implementierungsmodell

D1 Discover (Analyse und Diagnose)	D2 Design (Design und Entwicklung)	D3 Deliver (Implementation)	D4 Double-Back (Evaluation)	D5 Double-Up (Weiterführung)
Einigung auf eine Bildungsvision, die es wert ist, vorangetrieben zu werden	Systematische Suche und Einigung auf sehr erfolgsversprechende Interventionen, um zu beginnen.	Implementation und De-Implementation vereinbarter Entwürfe	Überwachung und Evaluation der Implementationen und Einigung auf vorrangige Maßnahmen	Sicherung und Vergrößerung des Einflusses
1.1 Einrichtung einer Steuerungsgruppe 1.2 Sich für eine Bildungsvision entscheiden 1.3 Die Bildungsvision erklären 1.4 Einigungsprozess	2.1 Sondierung von Optionen im Gestaltungsspielraum 2.2 Schulentwicklungsprogramm erstellen 2.3 Stresstest Schulentwicklungsprogramm 2.4 Definition von unwirksamen Strategien 2.5 Beobachtungs- und Evaluationsplan erstellen	3.1 Schulentwicklungsprogramm und Etappenziele 3.2 Umsetzung 3.3 Evaluationsdaten sammeln	4.1 Evaluation überwachen 4.2 Implementation überwachen 4.3 Implementation evaluieren und sich auf weitere Schritte einigen 4.4 Evaluation reflektieren	5.1 Nachhaltigkeit berücksichtigen 5.2 Skalierung berücksichtigen

Das Jahrzehnt der Evaluationsergebnisse

Clinton und Clarke (2020) werteten alle Daten aus dem Jahrzehnt der Implementation von *VL* in über 10.000 Schulen weltweit aus, darunter Daten aus dem School Capability Assessment, der *Visible Learning Matrix*, der Mindframes-Umfrage, der Feedback-Umfrage und einer Meta-Synthese von 47 Fallstudien Schulen.

Die Hälfte der Schulen, die sich für die *VL*-Fortbildungsreihe entschieden haben, verfolgte vorrangig das Ziel, den Unterricht im Allgemeinen zu verbessern, die andere Hälfte wollte sich stärker auf das Lernen der Schülerinnen und Schüler konzentrieren, z. B. darauf, die Effektivität der Lernenden zu steigern, die Selbstständigkeit der Lernenden zu stärken, oder den Lernenden beizubringen, wie sie ihre eigenen Lernziele setzen und ihren eigenen Lernerfolg im Hinblick auf diese Ziele bewerten können. Weitere Beispielzielsetzungen waren sich besser in Gruppenarbeiten zu beteiligen und mit Gleichaltrigen zu kommunizieren, den Unterschied zwischen gutem Verhalten und Lernverhalten zu erkennen, neue Informationen mit bereits bekannten Informationen zu verknüpfen und Tiefenverständnis zu entwickeln.

Wir stellten bald fest, dass viele Schulen noch nicht bereit waren, das *VL*-Modell umzusetzen. Einige wollten nur einzelne Tipps und Tricks, um ihr derzeitiges Repertoire zu ergänzen, andere wollten nur angeben können, dass sie sich mit evidenzbasierten Programmen beschäftigt hatten, einige wollten ausschließlich wissen, wie sie (einige der) Top-Einflüsse aus der *VL*-Rangliste umsetzen konnten, einige hatten ihre Lehrpersonen noch nicht von der Notwendigkeit von Verbesserungen überzeugt und viele hatten sich noch nicht mit der Diagnose ihres aktuellen Verbesserungsbedarfs beschäftigt. Wir verwendeten einen gemeinsamen Refrain: Auf welche Frage ist *Visible Learning* die Antwort? Wir entwickelten ein Maß für die Bereitschaft, das auf die Motivation, die Kapazität in Bezug auf die Intervention, die Verfügbarkeit von Ressourcen (insbesondere Zeit) und auf die

allgemeine Fähigkeit zur Veränderung innerhalb einer Schule fokussierte. Diese Bereitschaftsdiagnose wird zusammen mit dem School Capability Assessment (SCA) durchgeführt. Mit dem SCA werden Schulleitungen und Lehrpersonen aufgefordert, ihre Auswirkungen in 16 Dimensionen zu bewerten, darunter die Lernvorstellungen von Lernenden und Lehrpersonen, die Fortschritte der Lernenden bis hin zu höheren Lernleistungen und das Schulklima.

In den Schulen lag der durchschnittliche Grad der Veränderungsbereitschaft bei 74 %, aber es gibt eine große Variabilität (von 20 % bis 100 %). Dies bestimmte das Ausgangsniveau des Engagements. In einigen Fällen, in denen der Wert sehr niedrig war, wurde die Schule aufgefordert, mehr Zeit und Aufmerksamkeit für die Arbeit mit dem Personal aufzuwenden (z. B. um die SCA-Nachweise zu sammeln und zu interpretieren, die Absicht des Engagements mit VL zu klären, die Fallstudien anderer Schulen zu lesen, Schulleitungen von VL+-Schulen zu besuchen oder sich mit ihnen zu treffen), so dass sie bereitwilliger und bewusster mit den Inhalten des VL-Modells umgehen können – und damit die Wahrscheinlichkeit erhöhen, dass das Modell die gewünschte Wirkung auf ihre Lernenden hat.

Die SCA und die Veränderungsbereitschaft führten dann zu Entscheidungen über spezifische Interventionen und den Aufbau des Schulentwicklungsplans. Innerhalb des Schulentwicklungsplans legen wir großen Wert darauf, die gewünschten kurzfristigen (innerhalb von sechs Wochen), mittelfristigen (innerhalb eines Jahres) und längerfristigen (ein bis drei Jahre) Ziele zu definieren und Ergebnisse zu ermitteln. Der Teil der SCA, der sich auf die Führungsaspekte bezieht, zielt auf die Festlegung der strategischen Richtung ab, die Schulplanung und das Engagement mit VL+. Es geht dabei um die Identifizierung der kritischen Aspekte der Lernleistungsergebnisse, die im Mittelpunkt stehen sollen, und die Ermittlung der eigenen Kenntnisse und Fähigkeiten der Schulleitungen. Diese frühe Arbeit zielt darauf ab, die Lernabsichten und Erfolgskriterien für die VL+-Implementation zu entwickeln, Prioritäten zu setzen, Fähigkeiten zu entwickeln und Coaching anzubieten, um die aktuellen Testdaten der Schule, Einschätzungen der Lehrpersonen, Lernleistungen von Schülerinnen und Schüler sowie deren Aussagen über ihr Lernen zu interpretieren und zu triangulieren.

Die Hauptgründe für die Teilnahme an Fortbildungen waren, dass die Lehrpersonen ihre Einstellung verbessern, sie mehr mit Kolleginnen und Kollegen zusammenarbeiten (z. B. Diskussionen über Dilemmata, schwierige Situationen, Lernende oder Lehrplanthemen) und ein ansprecheres und sicheres Klassenklima schaffen wollten, indem Fehler als Lernchance gesehen werden. Außerdem gab es die Zielsetzung *alle* Lernenden in den Unterrichtsdialog einzubeziehen, mit den Lernenden so zu arbeiten, dass sie ihre Lernabsichten und Erfolgskriterien verstehen, die Wirkung von Feedback zu verbessern und eine Feedback-Kultur in jeder Klasse aufzubauen. Lehrpersonen wollten vom Oberflächenverständnis des Unterrichtens auch zum Tiefenverständnis des Unterrichtens und schließlich zum Transferverständnis des Unterrichtens übergehen. Sie wollten vom bloßen Sammeln von Daten zum wirksamen Interpretieren von Erkenntnissen fortschreiten, die die nächsten Unterrichtsentscheidungen leiten, und Interventionen anwenden, um eine Sprache des Lernens und der Bildung in der gesamten Schule zu identifizieren, zu etablieren, zu fördern und wertzuschätzen (Hattie K., 2021).

Als weiteren Schritt der Diagnose führten wir außerdem Interviews mit den Lernenden. Typischerweise wollten Lernende die Sprache des Lernens und der Bildung beherrschen, um zu formulieren, wie sie effektiver und effizienter lernen können und wie sie erkennen, dass sie gelernt und die Erfolgskriterien erreicht haben. Darüber hinaus wünschten sie sich mehr „Was sind die nächsten Schritte?“-Feedback; Möglichkeiten, das gegebene Feedback zu nutzen; mehr Gelegenheit, um zu hören, wie andere erfolgreich gelernt haben; und engagiertere Aufgaben (Engagement bedeutete für Lernende, dass es sich lohnte, die Aktivitäten auszuführen, da sie dann ihren Lernfortschritt sehen konnten, und nicht, dass sie interessant waren, Spaß machten oder für ihre Zukunft relevant waren).

Das School Capability Assessment wurde nicht nur zu Beginn, sondern noch einmal etwa ein Jahr nach Einführung des Modells als zentraler Bestandteil der Evaluierung eingesetzt. Zu den größten Veränderungen gehörten: Es gibt eine gemeinsame Sprache des Lernens in dieser Schule (0,83); die Schüler werden so unterrichtet, dass sie in der Lage sind, Bewertungen vorzunehmen (0,79); die Lehrpersonen vermitteln den Lernenden klar die intendierten Lernstrategien (0,62); die Lehrpersonen sind mit den verschiedenen Formen effektiven Feedbacks vertraut (0,58); der Schulentwicklungsplan setzt einen Schwerpunkt auf die Entwicklung bewertungsfähiger Lernender (0,57); die Lehrpersonen haben ein klares Bild von der Art der Lernenden, die sie in ihrer Schule haben möchten (0,48); die Lernenden haben die Möglichkeit, den Lehrpersonen Feedback zu geben (0,47); und die Lehrpersonen machen den Lernenden die Lernziele und Erfolgskriterien deutlich (0,42). Nach der Umsetzung der

umfassenden Veränderungen gaben 100 % der Schulleitungen an, dass sich die Teilnahme an dem Programm nachweislich auf das Lernen ausgewirkt hat. 70-80 % stellten eine verstärkte Nutzung von Daten in der Klasse und bei schulischen Entscheidungen fest, zudem eine Verbesserung in der Sprache des Lernens und der Bildung, ein stärkeres Engagement der Lernenden und positive Auswirkungen auf die Tätigkeiten der Lehrpersonen. 84 % berichteten von erhöhten Auswirkungen auf die Lernleistungen.

Die Effekte auf die Lernleistungen waren schwieriger zu erfassen, aber dennoch entscheidend für die Bewertung des Modells. Insgesamt erreichten ein Jahr nach der Umsetzung des Modells 10 % mehr Lernende die jeweiligen nationalen (oder regionalen) Standards bzw. lagen sie darüber, und die Leistungssteigerung bei standardisierten Lese- und Mathematiktests war größer als die Leistungssteigerung bei vergleichbaren Stichproben (unter Verwendung der Testhandbuchnormen). Blewlin und Baldwin (2015) bewerteten beispielsweise die Umsetzung einer zweijährigen *VL+*-Implementierung an 32 Schulen (3.172 Schüler) in Queensland. Die Ergebnisse der Leistungstests (Mathematik und Lesen) zeigten, dass die Lernenden aus den Klassen 1 und 6 höhere Punktzahlen erreichten und signifikant größere Fortschritte machten als die Lernenden, die nicht am *VL+*-Modell teilnahmen.

Eine unabhängige Evaluierung der Umsetzung des *VL+*-Modells an 31 Schulen in Stockholm (Frej et al., 2017) kam zu dem Schluss, dass *VL+* „gut geeignet ist, um zu einer nachhaltigen Entwicklung einer Schule beizutragen und als kohärente Perspektive zu fungieren, die verschiedene Entwicklungsbemühungen zusammenführt und einen gemeinsamen Sprachgebrauch in der Schule in Bezug auf Lehren und Lernen schafft.“ Der Bericht wies auf weitere Effekte hin, wie z. B. die Tatsache, dass sich die Lernenden eines neuen Sprachgebrauchs in Bezug auf das Lernen bedienten, was die Qualität der von den Lernenden geführten Entwicklungsgespräche steigerte und das Gespräch über das Lernen zwischen Lernenden und Lehrpersonen im Allgemeinen verbesserte. Die Schulleitungen sahen *VL+* als das verbindende Element, um wichtige Einzelteile zusammenzuführen. Auf diese Weise wurde es einfacher, Lehrpersonen in gemeinsame Entwicklungsprojekte einzubinden, was zu einem größeren Konsens in der Schule in Bezug auf Lehren und Lernen führte und dazu beitrug, eine Kultur der Abschottung zu durchbrechen. Es entwickelte sich ein neuer gemeinsamer Sprachgebrauch in Bezug auf das Lehren und Lernen, ein höheres Maß an Professionalität und ein gesteigertes Interesse der Lehrpersonen, den Unterricht miteinander zu diskutieren und zu beobachten.

An den *VL+*-Schulen gibt es jetzt einen tief verwurzelten und allumfassenden Ansatz, dass jeder lernen kann und dass alle Schülerinnen und Schüler in ihrem Lernen Fortschritte machen können. Die Lernenden haben sich überdies einen neuen Sprachgebrauch in Bezug auf ihr eigenes Lernen angeeignet, was die Qualität der von den Lernenden geführten Entwicklungsgespräche deutlich erhöht hat. Die Lernenden konnten besser einschätzen, wo sie im Lernprozess stehen, und die Gespräche über das Lernen zwischen Lernenden und Lehrpersonen haben sich allgemein verbessert.

Leeson (2017) untersuchte die Umsetzung in mehr als 150 Schulen im Nordterritorium Australiens. Zu Beginn des *VL+*-Programmes bewerteten die Schulleitungen die Ausgangslage mithilfe des School Capability Assessment (vgl. Abbildung 1.4a; zuerst ihre eigenen Schulen, dann folgt eine Stichprobe von zehn Schulen – jede Schule steht auf einer Linie) als hauptsächlich nicht etabliert (rot), nur in Ansätzen vorhanden (orange), in der Schule offensichtlich (grün) oder in der Schule verankert (blau). Am Ende der drei Jahre hatten sich die meisten Schulen von „nicht etabliert“ zu „offensichtlich“ oder „in der Schule verankert“ entwickelt (Abbildungen 1.4a, 1.4b und 1.4c).

Kapitel 01 - Die Herausforderung

Der sichtbar Lernende				Kenne Deinen Einfluss				Begeistertes und leidenschaftliches Lehren				Feedback			
Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken
1	1	1.5	1	1	1	1	1	1	1	1	1	1	1	1.5	1
2	2.5	2	2.5	2	2.5	1.5	2	3	2	2	3	NA	1	2	1
1	1	1.5	1.5	2	1.5	1	1.5	2	2	2	2	1.5	1	1.5	1
2	1	1.5	1.5	2	1	1	1.5	2	1	1.5	1	1	1	1.5	1.5
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1	2	1.5	1.5	2	2	1	2	2	2	2	1	1	1	1.5	1.5
1	2	2		2	1.5	1	2	2	1	1.5	1.5	1	1	1.5	
1	1	1.5	1	1	1	1	2	2	1	1.5	1	1	1	1.5	1.5
NA	NA	NA	NA	NA	NA	3	1.5	2	2	2	2	NA	1	NA	1
1	1.5	1	1.5	1	1.5	2	2	2	1	2	1.5	2	1	1.5	1
1.22	1.44	1.50	1.44	1.56	1.44	1.35	1.65	1.90	1.40	1.65	1.50	1.19	1.00	1.50	1.17

Abbildung 1.4a Anfangsbewertung der schulischen Leistungsfähigkeit

Der sichtbar Lernende				Kenne Deinen Einfluss				Begeistertes und leidenschaftliches Lehren				Feedback			
Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken
2	2	1.5	1.5	2.5	2.5	2	2.5	2.5	2	2	2	1.5	1	1.5	1
3	2.5	3	2.5	3	3	2.5	2.5	3	3	2.5	3	2	1	2	2
3	2.5	2	2	3	2.5	2	3	3	2.5	2	2.5	3	2	2	2
2	1.5	1.5	1.5	2	2	2	2	3	2	2	2	1.5	1	1	1.5
3	2	2	1.5	3	2.5	2.5	3	3	2	2.5	2	1.5	1	2	2
3	2.5	2	2.5	3	3	2.5	2.5	3	3	2.5	2.5	3	2	2.5	2.5
3	2	2	2	3	2.5	2	2	3	2.5	2.5	2	2	1	2	2
2	2	2	2	3	2	2.5	2.5	3	2	2	2	2	1.5	2	2
3	2	2	2	3	2	2	2.5	3	2	3	2.5	1	1	2	1
2	1.5	2	1.5	2.5	2.5	2.5	2	2	2	2	2	1	1	1.5	1.5
2.60	2.05	2.00	1.90	2.80	2.45	2.25	2.45	2.85	2.40	2.30	2.25	1.85	1.25	1.85	1.70

Abbildung 1.4b Bewertung am Ende des ersten Jahres

Der sichtbar Lernende				Kenne Deinen Einfluss				Begeistertes und leidenschaftliches Lehren				Feedback			
Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken	Leitbild und Werte	Wissen und Verständnis	Persönliche Merkmale	Profession. Praktiken
4	2	2	1	4	3	3	3.5	3	2	2.5	2.5	3	3	3	3
4	3.5	3.5	3	4	4	4	4	4	3	4	4	4	4	3.5	4
3.5	3	2.5	3	4	3.5	3	3.5	4	4	3	3	3	2	2.5	3
3.5	2.5	2.5	3	2	2.5	3	3	4	3	3	2.5	2	2	2.5	2.5
4	2.5	2	2	3	1.5	1	1	4	2	2	2.5	1	1	1.5	1.5
4	4	3.5	3	4	3.5	3	4	4	4	3.5	4	3	2	3.5	3.5
4	4	4	2.5	3	3	2	3.5	4	4	3	3.5	3	2	3	2
4	3	3.5	2.5	1.5	2.5	3	2	3	3	2.5	3	3	3	3	3
3	3.5	4	4	2.5	2.5	4	3	3	4	4	4	3	1	3.5	3.5
2	2	2	1	2	3	2	2	2	2	3	3	3	1	2.5	2
3.70	3.00	2.95	2.50	3.00	2.90	2.80	2.95	3.50	3.10	3.05	3.20	2.80	2.40	2.85	2.80

Abbildung 1.4c Bewertung am Ende des zweiten Jahres. Auszug aus der VL+-Matrix von elf Schulen aus drei Jahren im VL+-Programm

Die Veränderungen bei einigen der wichtigsten Messgrößen in diesen drei Jahren sind in Abbildung 1.5 dargestellt. Die erste Linie in jedem Balken ist der Ausgangswert und die nächsten drei

zeigen, wie sich der Mittelwert aller Lehrpersonen und der Schulleitungen über die drei Jahre verbessert hat.

Die größten Veränderungen zwischen den Beobachtungen in den Klassenräumen vor und nach der Implementation gab es bei der Klärung von Fragen zu den Lernabsichten (von 39 % auf 55 %), bei der Arbeit der Lernenden in Gruppen, und bei wichtigen Gesprächen untereinander (von 19 % auf 32 %). Die Lernenden gaben an, dass sie lernten mit anderen Lernenden und den Lehrpersonen über Aspekte des Lernens zu sprechen (von 31 % auf 63 %); die Lernenden unterstützten sich gegenseitig (von 37 % auf 72 %); die Lernenden suchten nach Feedback und sahen Fehler als Chance (von 39 % auf 81 %); und die Lernenden waren sich ihrer Lernschritte bewusst (von 30 % auf 52 %). Bei den Leistungsmessungen zeigen die Lese- und Mathematikdaten statistisch signifikante Zuwächse zwischen Messzeitpunkt 1 und 2 sowie Effektstärken von über 0,4, die auf das VL+-Modell für die Lernenden der Klassen 1 bis 6 zurückgeführt werden können. Effektstärken von über 0,4 im Lesen sind auch für die Lernenden der Jahrgangsstufen 7-10 erkennbar. Eine Schulleitung schrieb uns, dass „die Einführung von *Visible Learning* im gesamten staatlichen Schulsektor des Nordterritoriums eine messbare Auswirkung auf unsere NAPLAN-Daten (nationale Tests) hat“ (persönliche Mitteilung).

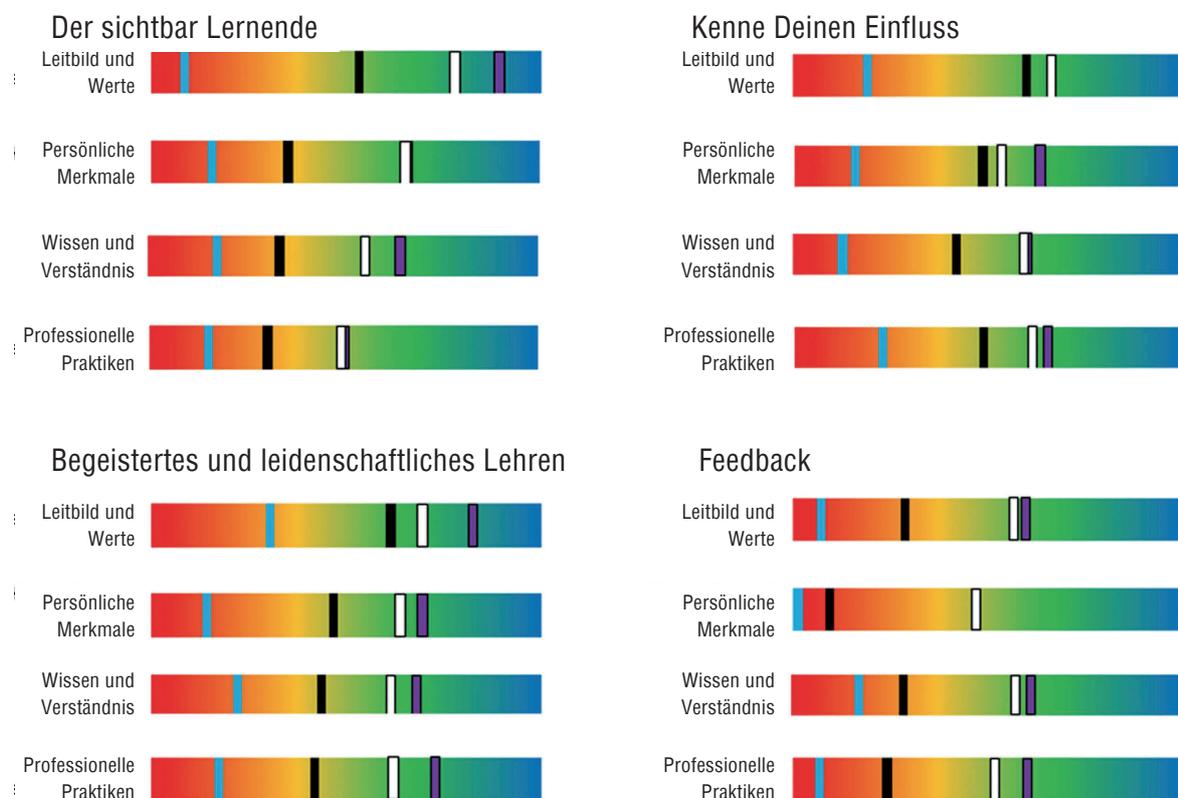


Abbildung 1.5 Durchschnittswerte zu Beginn und im Projektverlauf zu wesentlichen Messgrößen von VL+

In einer anderen Studie wurde ein strukturelles Modell entwickelt, um festzustellen, wie sich die einzelnen Komponenten des VL+-Modells auf die Lernleistungen auswirken (Leeson, 2017). Durch die Modellierung dieser Zusammenhänge ist es möglich, die kausale Wirkung des VL+-Modells auf die Leistung vorherzusagen. Diese Zuwächse waren insbesondere bei Lernenden mit geringen und mittleren Lernleistungen zu beobachten. In ähnlicher Weise waren bei der Mindframes-Erhebung – wenn auch in etwas geringerem Maß – ebenfalls Verbesserungen mit einem Leistungszuwachs bei den Lernenden verbunden (Leeson, 2016). Abbildung 1.6 zeigt das Modell, das am besten zu den empirischen Daten passte, die für jede Komponente des Modells gesammelt wurden (Leeson, 2017).

Die Ergebnisse zeigen, dass die von den Schulen bei ihren Schulbewertungen erzielten Fortschritte (erfasst mithilfe des School Capability Assessment) stark mit den späteren Fortschritten ihrer Lernenden in Bezug auf ihre Lernleistungen zusammenhängen ($r = 0,78$). Darüber hinaus standen die Entwicklung der Ziele der Schulleitungen (d. h. Vision und Werte) ($r = 0,85$) und ihre strategischen Instrumente und Maßnahmen ($r = 0,79$) in einem starken direkten Zusammenhang mit den Leistungssteigerungen der Lernenden im Verlauf eines Projektjahres.

Ein Hauptaugenmerk des VL+-Modells liegt auf der Verbesserung der zehn Geisteshaltungen (Mindframes) der Lehrpersonen. Etwa 30.000 Lehrpersonen haben die Umfrage zu den Geisteshaltungen ausgefüllt (geschätzte Zuverlässigkeit Cronbach's Alpha = 0,75). Abbildung 1.7 zeigt die Mittelwerte von Messzeitpunkt 1 und 2 sowie die Veränderungen der Effektstärke von Messzeitpunkt 1 zu 2. Die größten Verbesserungen sind zu verzeichnen bei „Ich sehe Lernleistungen als Rückmeldung über meine Wirkung und die nächsten Schritte“; „Ich konzentriere mich auf das Lernen von Lernenden und fokussiere auf die Sprache des Lernens und der Bildung“ und „Ich gebe den Lernenden Feedback und helfe ihnen, es zu verstehen, und ich suche, interpretiere und reagiere auf das Feedback, das ich erhalte“.

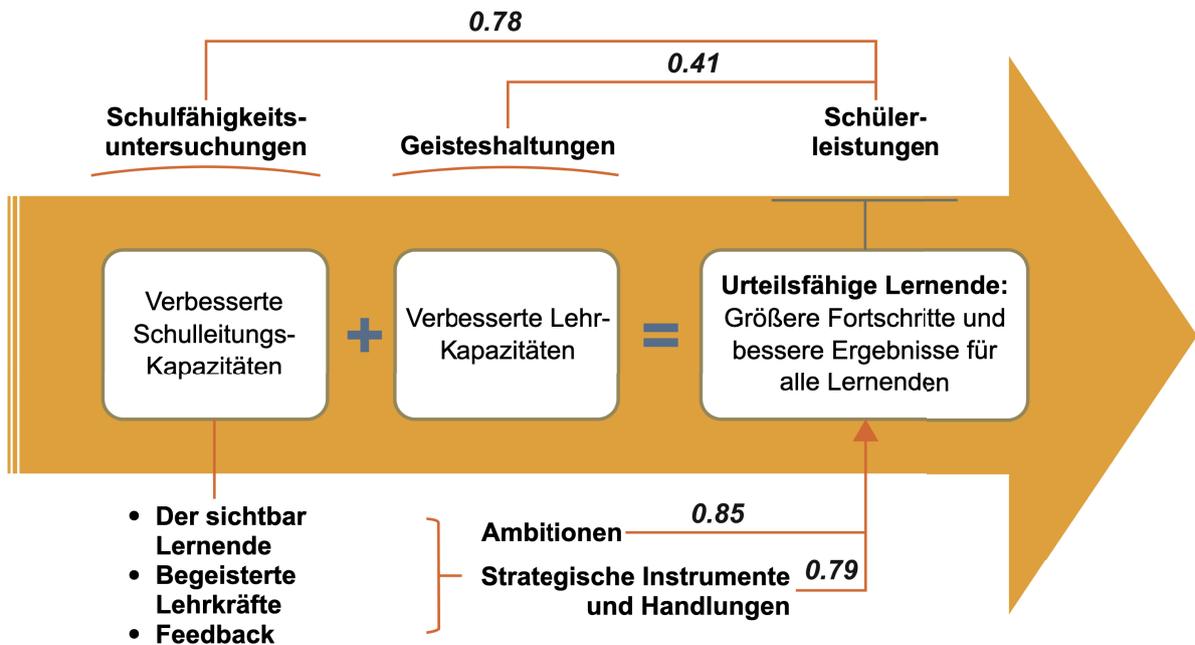


Abbildung 1.6 Modellierung der Wirkung der Evaluierungsinstrumente des VL+-Modells auf die Lernleistungen von Schülerinnen und Schülern

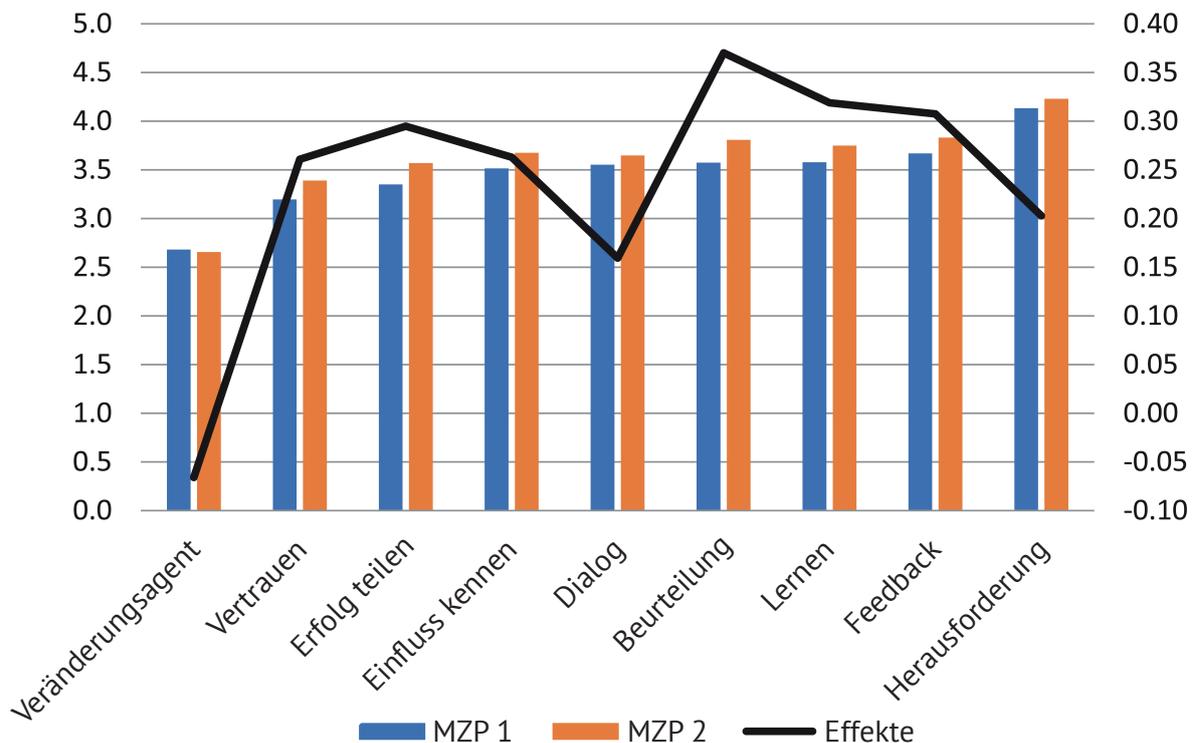


Abbildung 1.7 Effekte und Mittelwertvergleiche der Geisteshaltungen (Mindframes) zu Messzeitpunkt 1 und 2

Geisteshaltungen (Mindframes)

Eine der wichtigsten Botschaften von *VL1* war die Förderung einer Abkehr von Debatten über Schul- und Unterrichtsstrukturen und über die Art und Weise, wie unterrichtet wird, hin zu den Geisteshaltungen, die für den Lernerfolg der Lernenden ausschlaggebend sind. Das Argument lautet, dass es vor allem darauf ankommt, wie Lehrpersonen denken. Diese Geisteshaltungen sind der Wegbereiter für die Auswahl wirkungsvoller Strategien, die Sicherstellung der Umsetzung und die Bewertung der Auswirkungen auf die Lernenden. In unseren neueren Arbeiten haben wir die wichtigsten Grundlagen dieser Geisteshaltungen herausgearbeitet – im Sinne eines evaluativen Denkens (siehe Kapitel 3). Die zehn Geisteshaltungen für Lehrpersonen (Kapitel 9) und Schulleitungen (Kapitel 7) umfassen die Themen Wirkung, Veränderung und Herausforderung sowie Lernen:

Wirkung

1. Ich bin ein Evaluator meines Einflusses auf das Lernen der Schülerinnen und Schüler.
2. Ich sehe Lernleistungen als Rückmeldung über meine Wirkung und die nächsten Schritte.
3. Ich tausche mich mit meinen Kolleginnen und Kollegen sowie Lernenden über meine Vorstellungen von Fortschritt und Wirkung aus.

Veränderung und Herausforderung

4. Ich bin ein Veränderungsagent und bin überzeugt, dass sich alle Lernenden verbessern können.
5. Ich setze konkrete Herausforderungen und folge nicht nur dem Ziel, das Beste zu geben.

Lernen

6. Ich gebe den Lernenden Feedback und helfe ihnen, es zu verstehen, und ich interpretiere und reagiere auf das Feedback, das ich erhalte.
7. Ich benutze nicht nur Monologe, sondern vor allem Dialoge.
8. Ich informiere Lernende von Anfang an darüber, wie Lernerfolg aussieht.
9. Ich entwickle positive Beziehungen und baue Vertrauen auf, so dass das Lernen an einem Ort stattfinden kann, an dem Fehler willkommen sind und von anderen gelernt werden kann und soll.
10. Ich konzentriere mich auf das Lernen von Schülerinnen und Schülern und fokussiere auf die Sprache des Lernens und der Bildung.

Wir haben auch Geisteshaltungen für Lernende (Kapitel 5), Eltern (siehe Kapitel 6) sowie die Schulkultur und das Schulklima (Kapitel 7) entwickelt.

Abschließende Bemerkungen

In dieser Fortsetzung geht es mehr um die großen Botschaften und die Kerngedanken, die den Befunden aus den über 2.100 Metaanalysen zugrunde liegen. Das Modell in Kapitel 3 stützt sich auf die Erkenntnisse aus den Forschungsdaten, ergänzt durch die Erfahrungen aus den *VL+* Fortbildungen in mehr als 10.000 Schulen. In den folgenden Kapiteln finden Sie kurze Zusammenfassungen der wichtigsten Botschaften aus den zahlreichen Einflussfaktoren und die meisten sollten in Verbindung mit den folgenden Kapiteln gelesen werden. Für Details ist es unerlässlich, auf *MetaX* zuzugreifen, um die detaillierten Belege und Verweise auf die verschiedenen Meta-Analysen zu erhalten.

Meta-Analysen und die Kritik an *Visible Learning*

Der Hinweis für das Kreuzworträtsel in der *New York Times* vom 26. November 2019 lautete „Methods of studying pooled data“ (Methoden zur Untersuchung gepoolter Daten) und die gesuchte Antwort war „meta-analyses“ (Meta-Analysen). Es ist ein Meilenstein, wenn ein Wort als hinreichend verbreitet angenommen wird, um im *New York Times Sunday*-Kreuzworträtsel zugelassen zu werden. Einen Anhaltspunkt für die Verbreitung von Meta-Analysen liefert die Anzahl der zitierten Publikationen in Google Scholar: 3,5 Millionen für „Meta-Analyse“, 2,3 Millionen für „Meta-Analyse und Lernleistung“, 4,3 Millionen für „Medizin“, 3,7 Millionen für „Umwelt“, 2,7 Millionen für „Psychologie“ und 2,4 Millionen für „Wirtschaft“.

Die von Gene Glass in den 1970er Jahren erfundene Meta-Analyse ist weit verbreitet und wird heftig kritisiert. Sie ist eine *Methode* und kann, wie alle Methoden, schlecht oder gut umgesetzt werden. In den ersten Jahren waren Meta-Analysen eher Zusammenfassungen vieler Studien; dann wurden sie statistisch ziemlich ausgefeilt (Hedges' g , Heterogenitätsmaße, Zufallsmodelle usw.). Neuere Meta-Analysen verteidigen Interpretationen höherer Ordnung, aber es gibt immer noch zu wenige, die konkurrierende Erklärungen bewerten. Im Bildungsbereich gibt es über 2.100 Meta-Analysen zu Leistungsergebnissen, viele zu anderen schulischen Ergebnissen (Einstellung, Motivation, Lernstrategien) und weitere in der Medizin. Sie haben sich in vielen Domänen etabliert. Die Meta-Analyse ist eine systematische Methode zur Zusammenführung und Überprüfung von Forschungsergebnissen und ist deutlich weniger voreingenommen als manche traditionelle Überprüfungsmethode.

Stellen Sie sich vor, Sie interessieren sich für den reziproken Unterricht, weil Sie gerne die Fähigkeiten des Zusammenfassens, des Fragenstellens, des Klärens und des Vorhersagens unterrichten möchten. Sie führen eine Google (Scholar)-Suche durch und werden bald von den vielen Ressourcen überwältigt – einige behaupten, reziproker Unterricht sei von hohem Wert, andere behaupten das Gegenteil. Was soll man glauben? Also machen Sie sich auf die Suche nach einer Studienübersicht. Eine traditionelle Übersichtsarbeit würde viele Studien heranziehen, die wichtigsten Ergebnisse zusammenfassen und darlegen, wo es Moderatorvariablen gibt (d. h. Faktoren, welche die Auswirkungen des reziproken Unterrichts beeinflussen, wie z. B. das Alter der Lernenden oder das Fach). Der Kritikpunkt an dieser traditionellen Methode ist, dass es bei der Auswahl der Studien zu Verzerrungen kommen kann – wenn der Autor von vornherein geneigt ist, reziproken Unterricht als lohnenswert zu betrachten, spart er möglicherweise negative Artikel aus oder kritisiert alle nicht so positiven Artikel beispielweise als schlecht konzipiert. Dann kam Gene Glass (1976), der eine Möglichkeit erforschte, jeden Artikel nach der Größe des Effekts des reziproken Unterrichts und auch die Moderatoren so zu kodieren, dass es eine statistische Möglichkeit gab, die Größe des Effekts unter verschiedenen Umständen zu betrachten (z. B. welche Moderatoren wichtig sind). Die beiden wichtigsten Statistiken, die von Anfang an von Interesse sind, sind demgemäß der Mittelwert der Effekte und die Streuung um diesen Mittelwert. Die Meta-Analyse ist im Grunde immer noch eine Methode der Überprüfung. Das Überprüfen ist die Grundlage unseres Geschäfts: Wir beginnen oft mit der Untersuchung dessen, was gewesen ist.

Dieses Beispiel zeigt drei wesentliche Vorteile einer Meta-Analyse auf. Sie ist weniger voreingenommen bei der Studienausswahl (obwohl immer eine transparente Auswahl berücksichtigt werden sollte), bestimmt ein Gesamtmaß der Wirkung (Größe des Effekts) und fragt nach den

Moderatoren. Viele Kritiker weisen auf die Schwachstelle des Durchschnitts hin und sie haben Recht: Der Durchschnitt ist nur dann sinnvoll, wenn (a) keine Heterogenität der Effekte vorliegt (was sehr unwahrscheinlich ist) und (b) keine Moderatoren bekannt sind. In der Pädagogik haben wir von Anfang an nach diesen Moderatoren gesucht und selbst wenn wir sie nicht finden können, müssen wir weitersuchen. Ein wichtiger Beitrag der Meta-Analyse ist die Suche und die Methodik zur Entdeckung der Bedeutung verschiedener Moderatoren (oder auch nicht).

Ja, die vielen Meta-Analysen sind in ihrer Qualität unterschiedlich – einige basieren auf weniger als fünf Studien (diese sind nicht in MetaX enthalten), andere enthalten wenige Moderatoren (in der Tat gibt es in allen Meta-Analysen keine besonders große Anzahl an Moderatoren), einige enthalten falsche Schätzungen von Effektgrößen (z. B. viele der Lernstil-Meta-Analysen, siehe *VLI*) und einige haben einen begrenzten Diskussionsbereich und kaum Relevanz für zukünftige Arbeiten. Zu selten werden konkurrierende Einflüsse verglichen, die Kosten (Geld und Zeit) erwähnt und der Grad und die Genauigkeit der Umsetzung einbezogen. Immer noch argumentieren viele, dass die Ergebnisse nur dann ausschlaggebend sind, wenn der Gesamtdurchschnitt statistisch signifikant von Null abweicht, anstatt sie mit der Größe anderer Einflüsse zu vergleichen.

VL basiert nicht auf der Durchführung einer Meta-Analyse von Primärstudien, sondern auf der Meta-Synthese vieler Meta-Analysen. Als *VLI* veröffentlicht wurde, war dies noch etwas Neues, aber jetzt gibt es viel mehr solcher Synthesen, vor allem in der Medizin (wo die Intervention und die Ergebnisse eindeutiger sein können). In der Medizin wird für diese Synthesen von Meta-Analysen der Begriff „umbrella reviews“ verwendet (Papatheodorou, 2019). In der Pädagogik spricht man eher von Meta-Analysen zweiter Ordnung, von Meta-Meta-Analysen oder von der von mir bevorzugten Synthese von Meta-Analysen. Es gibt eine wachsende Zahl dieser Synthesen: zum Englischunterricht (Alsowat, 2020), zur Kreativität (Da Costa et al., 2015), Lehrerbildung (*Dunst et al., 2020), zum sozioökonomischen Hintergrund (Harwell et al., 2017; Korous et al., 2020), zur digitalen Technologie (Higgins et al., 2012; Tamim et al., 2011; Young, 2017), Flipped Classroom (Kapur et al., 2022), Einbeziehung der Eltern (Kim, 2022), zum Mathematik- und Naturwissenschaftsunterricht (Knogler et al., 2022), Fern- und Online-Lernen (Martin et al., 2022), Multimedia-Design (Noetel et al., 2022) und zur Hochschulbildung (Schneider & Preckel, 2017).

Zu den Vorteilen solcher Synthesen von Meta-Analysen gehört, dass die größere Anzahl von Studien einen stärkeren Fokus auf mögliche Ursachen von Heterogenität ermöglicht, wie z. B. das Vorhandensein von Ausreißerstudien, bessere Übersichten über mögliche Moderatoreffekte und mehr Spielraum für tiefergehende Untersuchungen von Macht- und Verzerrungsfragen. Van Aert et al. (2019) untersuchte beispielsweise die möglichen Verzerrungen aufgrund des geringen Stichprobenumfangs, unzuverlässiger Messungen und der Aussagekraft von Studien. Sie verwendeten 83 im *Psychological Bulletin* veröffentlichte Meta-Analysen und 499 systematische Übersichten aus der *Cochrane Database of Systematic Review*. Sie fanden nur minimale Hinweise auf Überschätzung und schwache Hinweise auf Publikationsverzerrungen in der Psychologie und Medizin (siehe jedoch Kvarven et al., 2020). Aber wie bei allen Übersichten liegt der Wert der Synthese von Meta-Analysen in der Qualität der Interpretation, der Erstellung, der Verteidigung oder der Kritik von Modellen zur Maximierung der Interpretationen und den neuen Fragen, die gestellt werden können.

Die Nachweise in diesem Buch basieren auf 2.103 Meta-Analysen. Der Grund für 2.100+ im Titel ist, dass in den meisten Fällen nur der durchschnittliche Effekt je Meta-Analyse enthalten ist, aber wenn es mehr als eine Reihe von Einflüssen innerhalb einer Studie gibt, werden diese auch hinzugefügt (210 Meta-Analysen enthalten mehr als einen Effekt; die Gesamtstichprobe beträgt also 2.313). In Tabelle 2.1 sind die Domänen, Subdomänen und die entsprechenden Kapitel in diesem Buch aufgeführt, in denen jede Domäne ausführlicher behandelt wird. Innerhalb jeder Subdomäne werden in den einzelnen Kapiteln weitere Klassifizierungen der Einflüsse beschrieben.

Tabelle 2.1 Kapitel, Domänen und Sub-Domänen in diesem Buch

Domäne	Subdomäne	Kapitel
Lernende	Vorwissen und Vorerfahrungen, Einstellungen und Dispositionen, Motivation, Körperliche Voraussetzungen	5
Elternhaus und Familie	Familiäre Ressourcen, Familienstruktur, Häusliches Umfeld	6
Schule und Gesellschaft	Finanzielle Ressourcen und Verantwortungsbereiche, Schularten, Vorschulische Interventionen, Schulkompositionseffekte, Schulleitung	7
Klassenzimmer	Klassenzusammensetzung, Klassenführung, Klassenklima	8
Lehrperson	Eigenschaften der Lehrperson, Lehrerbildung	9
Curriculum	Lesen und Schreiben, Schauspiel, Kunst und Musik, Mathematik und Naturwissenschaften, andere Fächer	10
Zielorientiertes Unterrichten	Lernziele, Erfolgskriterien, Feedback	11
Lernstrategien	Fähigkeiten, Motivierung, Emotion	12
Lehrstrategien	Oberflächen-, Tiefen- und Transferverständnis, Strategien des Vermittelns und Vertiefens	13
Nutzung von Technologien	Soziale Medien, Fernunterricht, Apps, Spiele und Simulationen, Nutzung von Technologie zur Vertiefung	14
Schulische und außerschulische Einflüsse	Schulprogramme, Außerschulische Programme	15

Tabelle 2.2 vermittelt einen Gesamtüberblick über die einzelnen Bereiche und offenbart die größeren Stellschrauben, die unter der Kontrolle der Lehrperson stehen ($d = 0,44$: Curriculum, Lehrstrategien, Lernstrategien, Technologie, Lehrperson), im Vergleich zu den weniger kontrollierbaren Bereichen ($d = 0,21$: Lernende, Elternhaus, Schule, schulische und außerschulische Einflüsse).

Tabelle 2.2 Zusammenfassende Statistiken für jede Domäne

Domäne	Anzahl Meta-Analysen	Anzahl Studien	geschätzte Gesamtanzahl	Anzahl Effekte	Effektstärke	Gewichtete Effektstärke	Standardfehler
Lernende	373	26.245	67.186.805	104.174	0.24	0.23	0.06
Elternhaus und Familie	117	6.676	24.192.643	16.696	0.15	0.15	0.08
Schule und Gesellschaft	146	7.446	10.510.357	26.150	0.19	0.20	0.06
Klassenzimmer	120	4.752	10.686.418	18.689	0.21	0.22	0.06
Lehrperson	81	3.837	7.104.805	8.310	0.53	0.55	0.05
Curriculum	377	17.228	20.639.762	52.289	0.50	0.50	0.08
Zielorientiertes Unterrichten	160	8.674	7.840.346	15.681	0,55	0,56	0,05
Lernstrategien	278	15.821	3.726.064	30.694	0.55	0.53	0.09
Lehrstrategien	263	21.193	3.918.537	41.070	0,49	0,49	0,09
Nutzung von Technologien	350	18.905	7.443.108	32.917	0.36	0.34	0.09
Schulische und außerschulische Einflüsse	48	1.612	43.887.942	6.406	0.25	0.24	0.05
Total	2.313	132.389	207.136.787	353.076	0.42	0.42	0.07

Überraschenderweise enthielt nur die Hälfte der 2.103 Meta-Analysen Informationen zur Stichprobengröße der verschiedenen Studien in einer Meta-Analyse. 200.863.503 Studenten sind das *Gesamt-N* aus diesen 988 Meta-Analysen. Eine Hochrechnung des *Gesamt-N* könnte also $200.863.503 / 988 \times 2.103 = 427.546.505$ betragen. Nehmen wir jedoch an, dass 40 % der Lernenden aus Stichprobenüberschneidungen stammen, so verringert sich die Schätzung auf 256.527.903. Eine andere und vertretbarere Schätzung des Stichprobenumfangs besteht darin, die durchschnittliche Anzahl der Lernenden pro Studie aus den 500 Meta-Analysen mit der kleinsten Stichprobe (also die Anzahl der Studien \times 92 Lernende) zu schätzen. Dies führt zu einer Gesamtschätzung von etwa 207 Millionen. Wie auch immer – es ist eine große Stichprobe.

Bedeutender ist der Robustheitsindex (der später im Buch beschrieben und mit 1 bewertet wird, wenn das *N* nicht vorhanden ist), da er ein vertretbarer Indikator für das Vertrauen in die Ergebnisse ist.

Die Verteilung der Effekte aus allen Meta-Analysen (Anzahl der Effekte) und die zusammenfassenden Effekte sind in Abbildung 2.1 dargestellt. Es gibt kaum Effekte, die sich negativ

auf die Leistung auswirken, und die meisten davon sind erklärbar (da es sich oft um negative Emotionen handelt, wie z. B. Langeweile, Prokrastination, Angst, Mobbing, Krankheit und Depression). Es gibt jedoch einige schulische Auswirkungen, die sich negativ auswirken, und diese sind erwähnenswert (z. B. Mobilität, Sitzenbleiben, Bildschirmzeiten, Unbeliebtheit in der Klasse, Etikettierung). Ansonsten wirkt sich die überwiegende Mehrheit der Einflüsse positiv auf die Leistungen der Lernenden aus. Wie bereits in *VLI* beschrieben, ist es somit trivial, dass eine Lehrperson die Leistungen ihrer Schülerinnen und Schüler steigern kann – fast jede Lehrperson kann das. Die große Frage ist also, was die Einflüsse über dem Durchschnitt von 0.42 (im roten Bereich) von den darunterliegenden (im gelben Bereich) unterscheidet. Es überrascht nicht, dass bei der zusammenfassenden Verteilung einige der größeren und kleineren Effekte wegfallen.

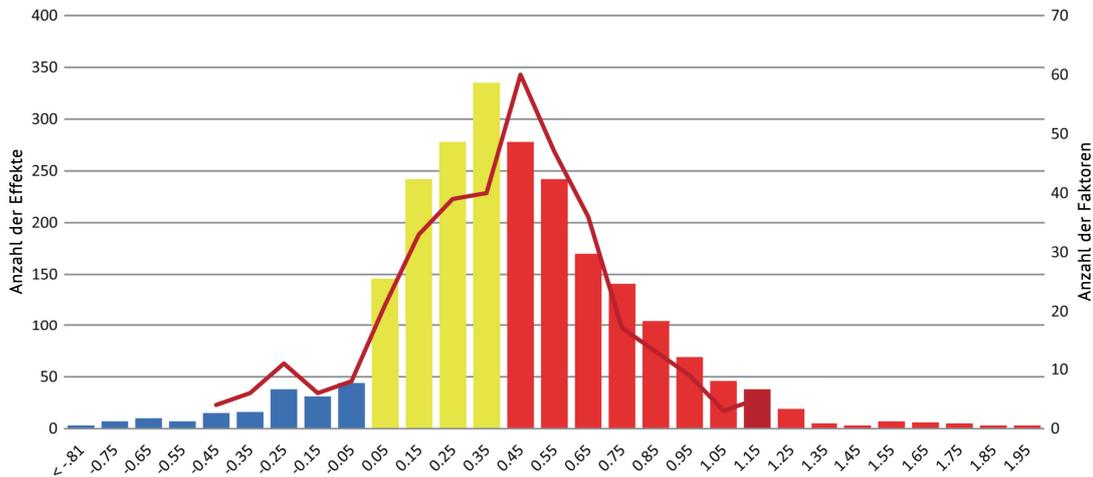


Abbildung 2.1 Verteilung der Effektstärke nach Anzahl der Effekte und nach Anzahl der Faktoren

Die Barometer aus *VLI* sind in dieser Fortsetzung nicht mehr enthalten (sie würden angesichts der über 350 Einflüsse zu viel Platz beanspruchen), daher hat Janet ein neues Diagramm erfunden, um die Einflüsse wieder bildlich darzustellen (Abbildung 2.2). Innerhalb des Thermometers wird der Mittelwert des Einflusses angegeben, wobei das Konfidenzintervall dieses Mittelwertes durch die Enden des Kreises angezeigt wird. Unterhalb des Thermometers befinden sich die zusammengefassten Daten: die Anzahl der Meta-Analysen, die Anzahl der Effekte, die geschätzte Anzahl der Personen und der Robustheitsindex im äußersten linken Kreis (siehe Abbildung 2.2). Das Thermometer verdeutlicht, dass die Temperatur rechts am höchsten ist (sie geht in Richtung rot-heiß), analog zum optimalen Einfluss der Sonne oder hoher Energie, und auf der linken Seite sind die niedrigeren Effekte, die blau dargestellt sind, da wir uns warm anziehen müssen.

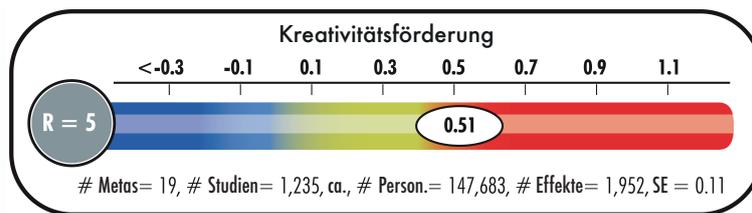


Abbildung 2.2 Das Thermometer, das die zusammenfassenden Informationen, die durchschnittliche Effektgröße, das Konfidenzintervall um diesen Mittelwert und den Robustheitsindex darstellt.

Somit ist die Verwendung des Faktors „Kreativitätsförderung“ mit einer Effektgröße von 0.51 ein Beispiel für ein Programm im roten Bereich, das das Lernen der Schülerinnen und Schüler in ihrem Klassenzimmer also potenziell verbessern kann. Der weiße Kreis zeigt die unteren und oberen Grenzen des Mittelwertes (+2 SE um den Mittelwert), d. h. das Konfidenzintervall um die mittlere Effektstärke. Für die Faktoren, für die es nur eine Meta-Analyse gibt oder für die keine ausreichenden Informationen vorliegen, um einen Standardfehler (SE) zu schätzen, wird der durchschnittliche SE = 0,08 verwendet. Der Faktor „Kreativitätsförderung“ kann bei dieser Schätzung ein hohes Maß an Robustheit aufweisen, was durch eine Robustheit von 5 auf einer Skala von 1 (am niedrigsten) bis 5 (am höchsten) angezeigt wird.

Kritik an der Meta-Analyse

Wichtig ist, dass die Meta-Analyse nur eine Methode neben anderen ist, und wie andere Methoden auch hat sie viel Aufsehen erregt. Ein Blogger hat eine Liste der Kritikpunkte an *VLI* zusammengestellt, darunter die folgenden:

- „Potenziell irreführend“ (Snook et al., 2009)
- „Statistischer Missbrauch, getarnt als statistisches Blendwerk“ (Berk, 2011)
- „Mobbing nach Zahlen“ (Wrigley, 2015)
- „Pseudowissenschaft und ein Kartenhaus“ (Bergeron & Rivard, 2017)
- „Ähneln mehr Schweinehaltung als Wissenschaft“ (Biesta, 2017)
- „Eine Sekte ... eine Tragödie für die australischen Schulleitungen“ (Eacott, 2017)
- „Welche Ungeheuer lauern unter dem Bett der Meta-Meta-Analyse von Hattie?“ (McKnight & Whitburn, 2020)

Diese Kritiken zielen auf zwei unterschiedliche Punkte ab: zum einen handelt es sich um Kritik an der Meta-Analyse bzw. der Synthese von Meta-Analysen und zum anderen ist es Kritik, die sich speziell auf *Visible Learning* bezieht.

Seit Glass die Methode der Meta-Analyse vorstellte, gab es eine Reihe von Weiterentwicklungen und Verbesserungen. Dieser Abschnitt fasst einige dieser Diskussionen zusammen. Es gibt in der Literatur viele Quellen, die jeden einzelnen Punkt näher erläutern. Wiliam (2016) hat die stärkste Kritik geäußert, beginnt aber mit der Feststellung, dass er nicht gegen Meta-Analysen sei und dass die Verwendung von Effektstärken eine „enorme Verbesserung gegenüber früheren Methoden der Berichterstattung über die Ergebnisse von Bildungsexperimenten“ (S. 82) sei. Aber er betont, dass er gegen eine unsachgemäße Verwendung und Interpretation von Meta-Analysen sei. Ich stimme ihm in diesen Punkten zu, aber mit seinen Schlussfolgerungen bin ich nicht einverstanden: Eine Meta-Analyse sei „einfach keine geeignete Technik, um die relativen Auswirkungen verschiedener Ansätze zur Verbesserung des Lernens von Schülerinnen und Schülern zusammenzufassen“. Er empfiehlt stattdessen eine „Überprüfung der besten Forschungsergebnisse“ (die ironischerweise oft viele Meta-Analysen enthalten), die „die Qualität der Studien, die Relevanz der Ergebnisse für die jeweilige Altersgruppe, die berichtet wird, und die Kontexte, in denen Lehrpersonen und Lernende agieren, berücksichtigt“ (S. 98). Wiliam formuliert fünf Hauptaussagen.

Die Intensität und Dauer der Intervention

Bei der Interpretation sollte die Dauer der Intervention, die Größe der von der Intervention betroffenen Gruppen und die Intensität der Intervention berücksichtigt werden. Dies ist seit Langem bekannt und einer der Hauptgründe dafür, dass Glass auf die Bedeutung der Verwendung von Moderatoren in einer Meta-Analyse hinwies. Eine der Stärken der Meta-Analyse sind die Qualität, die Vielfalt und die Informationen, die solche Moderatoren liefern. Häufig haben kürzere Bildungsinterventionen höhere Effekte als längere Bildungsinterventionen, wahrscheinlich weil die Umsetzung des Innovativen den Enthusiasmus und die Aufmerksamkeit der Lehrpersonen und Lernenden wecken kann. Es gibt jedoch zu wenige Meta-Analysen, die die Art und Intensität der Interventionen, die Genauigkeit und Dosierung der Umsetzung kodieren. Dies wird später in der Synthese der Meta-Analysen zum Flipped Classroom (Kapitel 13) besonders hervorgehoben. Wiliam fordert daher zu Recht eine effektivere Kodierung umfangreicherer Moderatoren, insbesondere in Bezug auf die Umsetzung.

Das Schubladenproblem (Publikationsbias)

Ausgehend von dem Argument, dass jeder einen Gewinner liebt, vermuten Kritikerinnen und Kritiker, dass Forschende Studien mit ungünstigen Ergebnissen ungern veröffentlichen. Daher könnten Übersichten, die nur auf veröffentlichten Studien beruhen, die tatsächliche Effektstärke überschätzen. Studien mit ungünstigen Ergebnissen enden dann in der Schublade. Das damit verbundene Schubladenproblem bezieht sich also darauf, wie viele Studien in den Schubladen gelandet sind, die keinen Effekt zeigen. Rosenthal (1979) führte eine Methode zur Schätzung einer ausfallsicheren Anzahl von unveröffentlichten Studien mit einem durchschnittlichen Nulleffekt ein, die erforderlich

wäre, um die gemeldete Effektstärke zu widerlegen. Um die Zusammenhänge zu verdeutlichen, sind in Tabelle 2.3 die Anzahl der Meta-Analysen, die Anzahl der Studien, die Effektstärke und die Anzahl der unveröffentlichten Studien für die Meta-Analysen mit $d > 0,2$ (der Wert, der zur Schätzung der Fail-Safe-Schätzung nach Orwin, 1983, verwendet wurde) dargestellt. Es gibt 39 Meta-Analysen mit weniger als 25 Studien (Mittelwert = 15 Studien), jede mit durchschnittlich 114.404 Probanden, 26 Effektgrößen und einem Mittelwert von 0,21. Es müssen nur 8 Studien je Meta-Analyse in einer Schublade mit $a < 0,2$ oder negativen Effekten liegen, um den aktuellen Mittelwert auf „nicht signifikant“ zu verschieben – und diese 8 Studien finden zu können, erscheint sehr plausibel. Sobald die Zahl der Studien 100 übersteigt, sinkt die Plausibilität dieser Argumentation (obwohl immer noch Vorsicht geboten ist, wenn die Effekte aufgrund einer geringen statistischen Aussagekraft oder aufgrund eines Publikationsbias den wahren Wert überschätzen). Ein Maß für die Robustheit (welches eine Fail-Safe-Schätzung enthält) sollte verwendet werden, um die Aussagekraft und die Zuverlässigkeit der Ergebnisse aus den Meta-Analysen zu ermitteln: Wenn das Maß für die Robustheit bei 3 oder höher liegt, kann man den Ergebnissen vertrauen. Darunter ist Vorsicht geboten, da mehr Studien erforderlich sind, um die Stabilität der Schätzungen zu erhöhen.

Tabelle 2.3 Anzahl der Studien, Meta-Analysen, Effektstärken, Fail-Safe-Schätzung und Robustheitsmaß

Studien-anzahl	Anzahl Meta-Analysen	Anzahl Studien	geschätzte Personen-anzahl	Anzahl Effekte	Effektstärke	Fail-Safe Schätzung	Robustheitsindex
< 25	39	15	114.404	26	0.21	8	1
26–100	91	64	622.515	162	0.38	271	2
101–1.000	196	363	378.497	1.081	0.42	3.777	4
> 1.000	30	1.839	2.559.527	4.191	0.45	167.931	5

Die Altersabhängigkeit der Effektstärke

Wiliam stellte fest, dass die Variabilität bei älteren Lernenden tendenziell größer ist als bei jüngeren und bei Lernenden mit und ohne besondere Bedürfnisse, was insbesondere dann ein Problem darstellt, wenn allgemeinere Maßnahmen (wie Lese-, Schreib- und Rechenfähigkeiten) über eine große Altersspanne hinweg verglichen werden. In solchen Fällen würde man erwarten, dass jüngere Lernende größere Zuwächse (Effektstärken) haben als ältere. Das bedeutet jedoch nicht, dass man bei älteren Lernenden keine höheren Effekte erzielen kann. Im Gegensatz zu Wiliams Behauptungen wurden Alterseffekte in *VLI* und auch in der vorliegenden Überarbeitung erörtert. Außerdem ähneln sich in den Meta-Analysen die durchschnittlichen Effektstärken nach Alter (Tabelle 2.4). Daraus folgt die Erkenntnis, dass man nach alters- oder hintergrundbedingten Moderatoren suchen sollte und falls sie vorhanden sind, sollte man sie beachten und diskutieren.

Tabelle 2.4 Effektstärken moderiert durch das Alter der Teilnehmenden

Alter	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Frühkindliche Bildung	0.40	0.25	117
Elementarbereich	0.36	0.34	130
Sekundarbereich II	0.32	0.33	73
Primar- / Sekundarbereich I	0.36	0.35	56
Tertiärbereich	0.42	0.39	168
Erwachsene	0.40	0.46	16
Alle Altersstufen	0.43	0.37	1.752

Die Sensitivität der Instruktion der verwendeten Ergebnismessung

Die Behauptung lautet, dass eine sinnvolle Kombination von Ergebnissen verschiedener Studien nur dann möglich sei, wenn man davon ausgehe, dass eine Intervention die gleiche Wirkung auf alle unterschiedlichen Ergebnisgrößen habe (z. B. von Lehrpersonen vergebene Noten, andere Testergebnisse, von Forschenden erhobene Ergebnisse) und dass die verwendeten Messungen auf die Interventionsmaßnahmen Einfluss haben. Wiliam argumentierte, dass dieses Problem gelöst werden könne, wenn die Studien nach der Beziehung zwischen der Bewertung und dem Unterricht kodiert würden, z. B. unmittelbar (Klassenzimmertest), nahe, proximal, distal oder entfernt (standardisierte nationale Bewertung). Ich stimme zu, dass es ein Problem ist, wenn dieser Moderator nicht kodiert wird. Ein schwerwiegenderes Problem ist jedoch die Enge oder Weite der Ergebnismessung – mit

höheren Effekten für enge (z. B. Vokabeln) als für weite (z. B. Kreativität) Messungen. Dies wird an verschiedenen Stellen in den nächsten Kapiteln kommentiert.

Die Verallgemeinerbarkeit der Studien

Wiliam argumentiert wie folgt:

„Die Forschung, die tatsächlich durchgeführt wird, ist oft eine seltsame und skurrile Auswahl von Forschung, die durchgeführt werden könnte, und die Verallgemeinerung ... (und) Meta-Analysen sind einfach nicht in der Lage, aussagekräftige Ergebnisse zu liefern, die Führungskräfte nutzen können, um die Aktivitäten der von ihnen geführten Lehrpersonen zu lenken.“ (S. 96-97)

Er verwendet dann einige Auszüge aus der *VLI*-Rangliste, um seine Argumentation zu untermauern. Zum Beispiel argumentiert er, dass es für den Faktor „Beurteilung des eigenen Leistungsniveaus“ ($d = 1,44$) „eindeutig absurd“ sei, dass 9-11-jährige Lernende Fortschritte von 4,6 Jahren machen würden, und dass es „an der Grenze der Glaubwürdigkeit“ liege, dass auch Kindergartenkinder diese Fortschritte machen würden. Ich bin jedoch kein Freund davon, Effektstärken in Alterszuwächse umzurechnen, vor allem, weil dies zu solch absurden Behauptungen führt. Ein $d = 1,44$ bedeutet lediglich, dass es einen Lernzuwachs gibt, der über den Ausgangswert hinausgeht oder mit dem von Lernenden verglichen wird, die nicht an der Intervention teilgenommen haben. Wiliams Behauptung zeigt jedoch, wie vorsichtig man bei der Verallgemeinerung von Interpretationen von Faktoren und Effektstärken sein muss.

Andere Kritikpunkte an der Meta-Analyse

Es überrascht nicht, dass es über die fünf Kritiken, die Dylan nennt, weitere Beanstandungen an der Methode der Meta-Analyse gibt. Darunter die folgenden:

Reductio ad absurdum

Wie ich in *VLI* erwähnte, bemerkte meine Kollegin Alison Jones, wie faszinierend es sei, dass ich Klassenzimmer bis auf die zweite Nachkommastelle verstehen könne. Ihre Bemerkung ist eine ernüchternde Erinnerung daran, wie wichtig es ist, den kulturellen Kontext des Klassenzimmers zu kennen, sich bewusst zu machen, was Lernende und Lehrpersonen aus kultureller und soziologischer Sicht in die Klasse einbringen, und die Erkenntnisse nicht auf einfache Appelle zu reduzieren.

Nicht Äpfel mit Birnen vergleichen

Diese Kritik besagt, dass es unvernünftig ist, verschiedene Arten von Daten zu bündeln. In der Metapher bleibend treffen wir aber Entscheidungen über Obst, und die Moderatoren sind dabei von entscheidender Bedeutung, um zu sehen, wie sich Äpfel von Birnen unterscheiden. Dieses Kombinieren ist allen Formen von Literaturübersichten gemein, und ein Ziel besteht genau darin, Vergleiche zu ziehen und Unterschiede zu entdecken.

Rückwärtsfahren mit dem Auto

Die Meta-Analyse ist eine Rückspiegel-Aktivität: Sie berücksichtigt das, was war, und nicht das, was sein könnte oder sein sollte. Das ist absolut richtig, aber ich würde nicht vorwärtsfahren wollen, ohne darauf zu achten, was hinter mir passiert, und das immer wieder zu prüfen, wenn ich mich auf neue Wege begeben. Das Wesentliche von Forschung besteht darin, stets auf der Suche zu sein. Einige Forschungsbereiche haben trotz überdimensionaler Fürsprache, Kosten und Studien nur geringe Effekte gezeigt, und wenn wir den metaphorischen Rückspiegel ignorieren, könnten wir dazu verurteilt sein, weiterhin falsche Behauptungen aufzustellen, die einen hohen Preis haben. Wenn wir zurückblicken und die Indizien akzeptieren, sind wir viel besser in der Lage, die richtigen Fragen für die Zukunft zu stellen.

Kosten

Die Kritik lautet, dass die Effektstärken die Kosten einer Maßnahme nicht berücksichtigen, was in der Tat der Fall ist, und es wäre wünschenswert, dass es mehr Kosten-Nutzen-Analysen gäbe. Ich bin kein Fan von vielen Umgestaltungen, die Kosten verursachen – der Education Endowment Fund beispielsweise bietet Feedback, Mastery Learning und Phonetik-Unterricht zu sehr geringen Kosten an (bis zu 4.000 Dollar pro Jahr und Klasse). Aber das professionelle Lernen, um diese Fähigkeiten zu verbessern und sie in den Klassen zu implementieren, kann sehr teuer sein – dieses Kostenmodell geht fälschlicherweise davon aus, dass es für Lehrpersonen einfach ist, diese Methoden anzuwenden.

Andere Kritikpunkte

Wir selbst haben auch nach Kritikpunkten an *VL* gesucht und führen diese im Paper „Real Gold vs. Fool’s Gold“ (Hattie & Hamilton, 2019) aus, das parallel zu diesem Kapitel gelesen werden sollte (Tabelle 2.5).

Tabelle 2.5 Liste der Kritikpunkte aus dem „Real Gold“-Paper

A	Belange von Metaanalysen	C	Interpretationen der Effektstärken
1.	Gewichtung	19.	Die Verwendung des Umschlagpunktes von $d = 0,40$ erscheint willkürlich.
2.	Stichprobengröße	20.	Der Umschlagpunkt bietet keine Kontrolle für Moderatoren und Mediatoren.
3.	Stichprobenzusammensetzung	21.	Die durchschnittliche Effektstärke kann durch das Alter moderiert werden.
4.	Reduktionismus	22.	Es gibt zu wenige Moderatoren, die in <i>Visible Learning</i> entdeckt wurden, um den Durchschnitt verwenden zu können.
5.	Qualität und Aggregation	23.	Es ist falsch, sich auf Einflüsse mit hohen Effektstärken zu konzentrieren und die geringen Einflüsse zu vernachlässigen.
6.	Publikationsbias	24.	Korrelation ist nicht gleich Kausalität.
7.	Vergleichbarkeit	25.	Die Nichtberücksichtigung von qualitativen Studien.
8.	Wichtige Studien werden ignoriert	26.	Meta-Analysen berücksichtigen Instruktionen zu wenig.
9.	Die durchschnittlichen Effekte ändern sich.	27.	Meta-Analysen berücksichtigen Kosten zu wenig.
10.	Die Effektgrößen in <i>Visible Learning</i> verändern sich über die Zeit.	28.	Andere widerrufen jetzt ihre eigene Verwendung von Effektstärken.
11.	Historizität der Forschung	29.	<i>Visible Learning</i> ignoriert Debatten darüber, was lernenswert ist.
12.	Ein narrativer Überblick ist besser.	30.	Bei <i>Visible Learning</i> geht es nur um Leistung und das ist eben nicht alles, worum es in der Schule geht.
		31.	<i>Visible Learning</i> ignoriert sozioökonomische Effekte.
		32.	Die Interpretation von Meta-Analysen ist mit vielen Risiken verbunden.
B	Effektstärken	D	Das Modell von <i>Visible Learning</i>
13.	Vermischung verschiedener Effektstärken ist schlechte Wissenschaft.	33.	Es gibt alternative Interpretationen auf der Grundlage der Daten von <i>Visible Learning</i> .
14.	Die Common Language Effect Size (CLE) ist irreführend.	34.	Es ist nur ein Modell.
15.	Die Hälfte aller Statistiken in <i>Visible Learning</i> ist falsch.	35.	Die Hauptaussagen von <i>Visible Learning</i> widersprechen der weit verbreiteten Erfahrung von Lehrpersonen.
16.	Effektstärken werden von Mathematikern nicht verwendet.	36.	<i>Visible Learning</i> kommt zu gegenteiligen Schlussfolgerungen als die eigentlichen Autoren der eingezogenen Studien.
17.	Effektstärken sollten in der Bildungswissenschaft nicht verwendet werden, da sie aus der Medizin stammen.	37.	Die Faktoren sind nicht getrennt.
18.	Die Variabilität der Effekte wird ignoriert.	38.	Verengung auf Lernleistungen

Spezifische Reaktionen auf die Kritik an VL

Die Herausforderungen des Rankings

PISA bewertet Schulen, QS bewertet Universitäten, die englische Premier League bewertet Fußballmannschaften und der Dow Jones bewertet Unternehmen in Ranglisten. Einer der häufigsten Kritikpunkte an VL ist die im Anhang enthaltene Rangfolge der verschiedenen Einflüsse auf die Lernleistungen von Schülerinnen und Schülern. Es wird argumentiert, dass diese Rangliste den Eindruck erweckt, dass jeder Faktor eindeutig und isoliert sei und dass, wenn man einfach mehr von den Faktoren berücksichtige, die ganz oben auf der Liste stehen, und vielleicht die Faktoren ausschaltet, die ganz unten stehen, automatisch (Lern-)Wunder geschehen.

Diese Ansichten sind jedoch falsch. Jeder Faktor ist bei weitem nicht isoliert zu betrachten und der Grund, warum es mehr als 15 Jahre gedauert hat, *Visible Learning* zu schreiben, war die Zeit, die nötig war, um die Beziehung zwischen den vielen Faktoren zu verstehen und ein Modell zu entwickeln, das am besten zwischen den Faktoren oberhalb und unterhalb des Umschlagpunktes unterscheidet. Bei einigen der niedrigeren Einflüsse muss man genau verstehen, warum sie niedrig sind – damit sie verbessert werden können. Niedrigere Einflüsse sollten niemals automatisch als nicht wertvoll abgetan werden (siehe z. B. die Diskussion über problemorientiertes Lernen, Individualisierung und Fachkompetenz der Lehrperson).

Jenseits des Schultores

Viel Kritik stammt von denjenigen, die argumentieren, dass *Visible Learning* nicht genügend Gewicht auf außerschulische Variablen lege (siehe Snook et al., 2009). Diese Kritik ist ein wenig unfair, da *VLI* und die Forschungsdatenbank das Elternhaus einschließlich Familiendynamik, -struktur und deren Ressourcen untersuchen. Es ist jedoch korrekt, dass die Analyse und die Interpretation der Daten viel stärker auf die innerschulischen Einflüsse abzielen, und zwar aus zwei Gründen: Erstens richtet sich die Untersuchung an Lehrpersonen, Schulleitungen sowie bildungspolitische Entscheidungsträgerinnen und -träger, so dass es sinnvoll erscheint, sich auf Elemente zu konzentrieren, auf die sie einen gewissen Einfluss haben. Weder Lehrpersonen noch Schulleitungen können das häusliche Umfeld von Lernenden schnell oder einfach ändern, aber wenn sie zusammenarbeiten, können sie die Schulerfahrung für Lernende aus allen häuslichen Umgebungen erheblich verbessern (Hattie & Hattie, 2022).

Natürlich lassen Lernende ihren häuslichen Hintergrund oder ihre kulturellen Einflüsse nicht am Schultor zurück. Daher ist das Bewusstsein dafür, was Lernende von zu Hause mitbringen, ein wichtiger Faktor und muss von den Schulen berücksichtigt werden, wenn es darum geht, die Bildung aller Lernenden zu verbessern. Es ist eine Binsenweisheit, dass die Verbesserung der Schulerfahrung allein ausreicht, um die Lernergebnisse aller Lernenden zu verbessern und dass die Partnerschaften mit dem Elternhaus hierfür entscheidend sind (dies wird in Kapitel 7 untersucht).

Die Verwendung von $d = 0,40$ als Umschlagpunkt

Scharniere und Gelenke sind großartig: Sie ermöglichen es uns, Schranktüren und Brillenbügel zu schwingen und unsere Knie und Ellbogen zu bewegen. In *VLI* wurde sowohl in der Analyse als auch in den Barometern der Faktoren $d = 0,40$ als Scharnierpunkt angegeben. Ja, in früheren Beiträgen und Präsentationen habe ich wahrscheinlich die Effekte $< 0,40$ verunglimpft und einige sind es nach wie vor wert, verunglimpft zu werden. Aber wie im letzten Abschnitt erwähnt, müssen einige dieser schwachen Effekte besser verstanden werden.

Man sollte sich nicht zu sehr darauf versteifen, $d = 0,40$ in allen Fällen zu verwenden – die Interpretation kann unterschiedlich ausfallen, je nachdem, wie eng (z. B. Wortschatz) oder weit (z. B. Verständnis) das Ergebnis ist, wie hoch die Kosten der Intervention sind, wie anspruchsvoll die Implementation eines Faktors ist, und hinsichtlich vieler weiterer Aspekte. Bei der Umsetzung des VL-Modells lohnt es sich, situatives Wissen darüber zu entwickeln, was im jeweiligen Kontext am besten funktioniert, und sich nicht zu sehr auf 0,40 zu verlassen. Wenn wir beispielsweise in Schulen arbeiten, verwenden wir häufiger den Mittelwert der Klasse oder der Jahrgangsstufe bei einer bestimmten Bewertung, um zwischen über- und unterdurchschnittlichen Leistungen zu unterscheiden.

Darauf aufbauend versuchen wir, die Varianz der Schülerergebnisse zu verstehen. Der Wert von 0,40 ist lediglich der Durchschnitt aller mehr als 2.100 Meta-Analysen und dient als geeigneteres Scharnier als die übliche Null (die es fast allen erlaubt zu behaupten, dass ihre Lieblingsstrategie oder ihr Lieblingsfaktor die Leistung verbessern kann).

Qualität der Studien und Vermischung der Methoden zur Berechnung der Effektstärken

Das Argument ist, dass die Studien in den verschiedenen Meta-Analysen randomisierte, kontrollierte Studien (RCTs), Prä-Post- und Gruppenvergleichsstudien mischen und dass sie eine Vielzahl von Testinstrumenten verwenden, die von standardisierten Schülerleistungstests, IQ-Tests, Erhebungsdaten zur Selbstwahrnehmung, quantifizierten Beobachtungen, von Lehrpersonen durchgeführten Tests bis hin zu korrelativen Daten reichen (alles richtig). Schließlich sind viele der Studien quasi-experimentell und haben keine echte Kontrollgruppe und andere sind viel schwächere Korrelationsstudien (wieder richtig).

Es wird davon ausgegangen, dass die Qualität der Forschung durch die Einbeziehung all dieser Arten von Daten – und nicht nur der RCTs als Goldstandard-Studien – erheblich verbessert wurde. Würden nur perfekte Studien einbezogen, so wären die Daten nicht ausreichend, um Schlussfolgerungen zu ziehen. Im What-Work-Clearinghouse (WWC), das nur RCTs und ähnlich hochwertige Designs zulässt, liegt die durchschnittliche Anzahl der Studien in jeder der 500 Übersichten bei zwei (Lortie-Lortie-Forgues et al., 2021; Stockard & Wood, 2017). Aus zwei Studien kann man nicht viel schließen, außer dass mehr Forschung erforderlich ist.

Wir haben also die Wahl: Entweder wir beschränken uns darauf, die perfekten Studien zu sammeln, oder wir sammeln alle, achten aber sehr darauf, wie wir die Daten interpretieren und welche Schlussfolgerungen wir daraus ziehen. Im Fall von *Visible Learning* wurde der letztere Ansatz gewählt. Es stellt sich also die Frage, ob die Qualität der Meta-Analyse einen Einfluss auf die Gesamtschlussfolgerungen hat.

Es gab zwei Ausnahmen bei diesem Ansatz: Erstens, wenn es sich um schlecht durchgeführte Meta-Analysen handelte, die nicht zu verteidigen waren. Diese wurden nicht mit aufgenommen. Alle Mittelwerte, die größer als $d = 2,00$ waren, wurden nur dann berücksichtigt, wenn die Stichprobengröße groß war, die Studie gut durchgeführt wurde und es andere Studien in dieser Kategorie gab, um Vergleiche anzustellen (es gab $n = 2$ Metas mit $d > 2,0$, eine im Bereich des Flipped Classrooms und eine im Bereich des Lesens). Zweitens, wenn es keine aussagekräftigen Korrelationen gab zwischen der geschätzten Effektgröße und der Anzahl der Studien ($r = 0,01$), der Gesamtzahl oder der geschätzten Anzahl der Teilnehmer ($r = -0,06$) oder der Anzahl der Effekte ($r = 0,01$). Diese Merkmale der Stichprobengröße standen in keinem Zusammenhang und sind ein gewisser Anreiz, Meta-Analysen unabhängig von der Stichprobengröße einzubeziehen.

Qualität der Meta-Analysen

Es gibt eine lebhaft Debatt über die Beurteilung der Qualität von Originalstudien, die in Meta-Analysen verwendet werden (Luchini et al., 2021), aber weit weniger Diskussionen über die Qualität von Meta-Analysen. Bei Meta-Analysen und Synthesen von Meta-Analysen wird insbesondere über einen möglichen Publikationsbias debattiert, d. h. die selektive Veröffentlichung statistisch signifikanter Ergebnisse, die durch den Vergleich der Effekte in veröffentlichten und unveröffentlichten Ergebnissen (z. B. Konferenzen, Dissertationen) überprüft wird. Bei den 2.312 Meta-Analysen waren die berichteten Effekte in der grauen Literatur (Konferenzen und Berichte) geringer und in Büchern am höchsten (Tabelle 2.6). Bei Dissertationen und Zeitschriftenartikeln gab es keine Unterschiede im Durchschnittswert.

Tabelle 2.6 Quelle der Veröffentlichung nach Effektgröße

Quelle	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Konferenz	0.34	0.32	108
Forschungsbericht	0.36	0.30	71
Dissertation	0.41	0.36	285
Fachzeitschriftenartikel	0.42	0.37	1.772
Buch	0.52	0.37	76

Eine weitere Möglichkeit zur Überprüfung des Publikationsbias ist die Frage, ob es Unterschiede in Bezug auf die Qualität der Fachzeitschrift gibt. Mathur und VanderWeele (2021) verglichen die Publikationsquelle (sie klassifizierten sie in Journals niedrigeren und höheren Ranges) in 63 Meta-Analysen auf eine solche Verzerrung und kamen zu dem Schluss, dass „die Publikationsverzerrung vielleicht milder ist als in Meta-Analysen erwartet“ (S. 14). In ähnlicher Weise habe ich die 1.397 Zeitschriften aus den Meta-Analysen nach verschiedenen Maßstäben der Zeitschriftenqualität kodiert (www.scimagojr.com/journalrank.php). Es gibt Hinweise darauf, dass Meta-Analysen mit geringeren berichteten Effekten und größeren Stichproben in Zeitschriften mit höherer Qualität veröffentlicht werden (siehe Tabelle 2.7).

Tabelle 2.7 Korrelationen der Anzahl von Studien, Probanden und Effekten sowie Effektstärke mit zwei Qualitätsindikatoren der Zeitschriften, in denen die Meta-Analysen veröffentlicht wurden

	Anzahl der Studien	Geschätzte Anzahl	Anzahl der Effekte	Effektstärke
SJR – durchschnittliche Anzahl der gewichteten Zitationen in der Zeitschrift in den letzten drei Jahren	0.22	0.05	0.04	-0.11
H – Index – keine Artikel, die mindestens h-Zitationen erhielten	0.19	0.02	0.04	-0.15

Um diesen Hinweis noch weiter zu vertiefen, können die Zeitschriften nach dem Gesamtranking der Journals geordnet und in vier gleich große Gruppen unterteilt werden. Q1 umfasst die Zeitschriften mit dem höchsten Ranking und Q4 die niedrigsten (https://en.wikipedia.org/wiki/SCImago_Journal_Rank). Es ist offensichtlich, dass die höchsten Effektstärken mehr aus den am niedrigsten bewerteten Zeitschriften und die niedrigsten aus den am höchsten bewerteten Zeitschriften stammen. Dies ist ein klarer Hinweis auf eine Verzerrung der Veröffentlichungen, auch wenn der durchschnittliche Effekt der meisten Zeitschriften in den qualitativ hochwertigsten Zeitschriften zu finden ist (was beruhigend ist) und in der Nähe des Umschlagpunktes von 0,40 liegt (Tabelle 2.8).

Tabelle 2.8 Effektstärken in Abhängigkeit von der Qualität der Zeitschrift

Bestes SJR-Quartil	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Q1 – höchste Bewertung	0.38	0.36	1.147
Q2	0.43	0.38	166
Q3	0.52	0.30	55
Q4 – niedrigste Bewertung	0.66	0.46	23

Andere haben empfohlen, die Effektgrößen nach dem Stichprobenumfang (oder dem Kehrwert der Standardabweichung) zu gewichten, um so den Studien mit größeren Stichproben mehr Bedeutung beizumessen. Jeder zusammengefasste Einflussmittelwert wurde nach der Anzahl der Studien gewichtet. Dadurch wird den größeren Studien spürbar mehr Gewicht verliehen. Bei den 358 Zusammenfassungen betrug die durchschnittliche Differenz zwischen dem gewichteten und dem ungewichteten Mittelwert 0,00 (bzw. 0,04 bei den absoluten Differenzen). Sieben Synthesen wiesen Differenzen von mehr als plus/minus 0,20 auf. Es gab sieben Synthesen, bei denen die Unterschiede plus oder minus 0,00 betrugten, und diese werden in den Synthesen der verschiedenen Kapitel aufgeführt: Komplexe Lernzielformulierung (-0,22), Inklusive Beschulung (-0,20), Alternative Bewertungsmethoden (-0,20), Kognitive Aufgabenanalyse (0,20), Erfolgskriterien (0,23), Strategien des Wiederholens und Einprägens (0,25) sowie Bewusstes Üben (0,30). Ist der Effekt positiv, so ist der gewichtete Mittelwert (der wahrscheinlich treffender ist) höher, wenn die größere Stichprobe von Studien berücksichtigt wird; ist er negativ, so führten die größeren Studien zu einer niedrigeren Schätzung.

Land des Erstautors

Mehr als 4.800 Autorinnen und Autoren haben zu den Meta-Analysen in dieser Weiterführung von *Visible Learning* beigetragen. Die Erstautorschaften der Meta-Analysen stammen aus 52 Ländern, fast alle (87 %) aus Ländern mit hohem Einkommen (Tabelle 2.9), die meisten aus den USA (58 %), der Türkei (5,8 %), den Niederlanden (4,7 %), China (4,2 %), Kanada (3,8 %), Australien (3,6 %), Deutschland (3,5 %), Korea (3,2 %), dem Vereinigten Königreich (2,9 %) und Taiwan (1,2 %); sie machen 90 % aller integrierten Meta-Analysen aus (Tabelle 2.10, 2.11). Die höchsten mittleren Effektstärken kommen aus Afrika, Südasien, Mittel- und Osteuropa und insbesondere aus Indonesien

(0,75) und der Türkei (0,73), die niedrigsten aus dem Vereinigten Königreich und Kanada. Dabei ist wahrscheinlich, dass sich die Vergleichsgruppen unterscheiden (in Bezug auf „normalen“ Unterricht), d. h. in den Ländern mit niedrigem Einkommen ist Unterricht eher lehrkraftzentriert, auf Inhalte fokussiert und reglementierter als in den Ländern mit höherem Einkommen, aber dies muss noch weiter untersucht werden.

Es kann davon ausgegangen werden, dass die in den wohlhabenderen Ländern durchgeführte Forschung in hohem Maße auf andere wohlhabende Länder übertragbar ist.

Tabelle 2.9 Anzahl der Länder, Anzahl der Effekte, Mittelwerte und Standardabweichungen für Meta-Analysen moderiert durch Einkommensressourcen des Landes der Erstautorschchaft

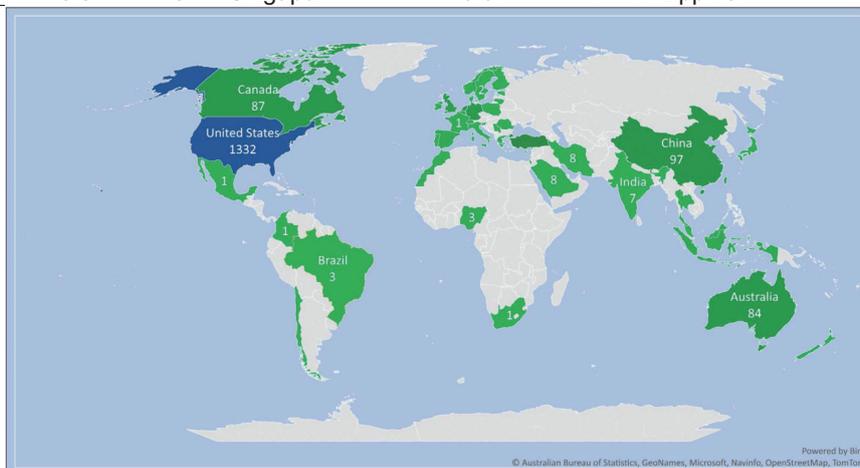
Einkommen	Anzahl Länder	Anzahl Effekte	Mittelwert (M)	Standardabweichung (SD)
Hohes Einkommen	36	2.020	0.43	0.25
Obere Mittelklasse	10	248	0.55	0.22
Untere Mittelklasse	6	43	0.75	0.15

Tabelle 2.10 Anzahl der Länder, Anzahl der Effekte, Mittelwerte und SD für moderierte Meta-Analysen nach Weltregionen des Landes der Erstautorschchaft

Region	Anzahl Länder	Anzahl Effekte	Mittelwert (M)	Standardabweichung (SD)
Latein Amerika & Karibik	5	10	0.29	0.24
Nordamerika	2	1.024	0.34	0.06
Europa & Zentralasien	24	477	0.46	0.15
Mittel-, Ost- & Nordafrika	5	27	0.48	0.20
Ost-Asien & Pazifik	13	362	0.60	0.17
Süd-Asien	1	7	0.62	0.00
Sub-Sahara-Afrika	2	4	0.75	0.06

Tabelle 2.11 Anzahl der Länder, Anzahl der Effekte, Mittelwerte und Standardabweichungen für moderierte Meta-Analysen nach Land der Erstautorschchaft inklusive veranschaulichender Weltkarte

Land	Effektstärke	N	Land	Effektstärke	N	Land	Effektstärke	N
Brasilien	0.04	3	Deutschland	0.37	80	Rumänien	0.61	2
Chile	0.10	4	Belgien	0.37	9	Irland	0.62	1
Israel	0.23	9	USA	0.38	1.332	Indien	0.62	7
Italien	0.23	6	Puerto Rico	0.41	1	Marokko	0.62	1
Mexiko	0.26	1	Polen	0.41	4	Kolumbien	0.64	1
Schweden	0.27	2	Gesamt	0.42	2.312	Estland	0.65	1
Dänemark	0.29	12	Neuseeland	0.43	20	Litauen	0.65	1
Dubai	0.30	1	Luxemburg	0.44	3	Griechenland	0.66	2
UK	0.30	68	Spanien	0.47	20	Saudi-Arabien	0.67	1
Kanada	0.30	87	Zypern	0.49	3	Thailand	0.68	3
Finnland	0.31	4	Portugal	0.51	4	Serbien	0.68	2
Norwegen	0.36	9	Saudi-Arabien	0.52	7	Iran	0.70	8
Frankreich	0.36	1	Taiwan	0.54	28	Macao	0.70	2
Hongkong	0.36	13	Japan	0.56	13	Südafrika	0.71	1
Slowenien	0.37	1	Korea	0.57	73	Türkei	0.73	135
Australien	0.37	84	Malaysia	0.57	6	Indonesien	0.75	20
Schweiz	0.37	4	China	0.58	96	Nigeria	0.79	3
Niederlande	0.37	107	Singapur	0.61	1	Philippinen	1.02	5



Wir müssen jedoch vorsichtig sein, wenn wir die Ergebnisse der meisten Meta-Analysen auf Entwicklungsländer wie Afrika südlich der Sahara oder Südasien – heute als Länder des Globalen Südens bezeichnet – übertragen. Hamilton und Hattie (2022) haben fast 1.000 Metaanalysen und randomisiert kontrollierte Studien aus Ländern des Globalen Südens bewertet und sind zu dem Schluss gekommen, dass die Lernergebnisse in diesen Schulen ganz anders ausfallen. In den Industrieländern konzentriert sich die Bildungspolitik vor allem auf die Qualität des Angebots, auf Bildungsgerechtigkeit und auf die Verbesserung der Lernbedingungen der Schülerinnen und Schüler. In den Entwicklungsländern geht es allzu oft um Zugang zu Bildung und Infrastruktur und daher wird mehr in die Strukturen als in das Lernen investiert.

Zusammenhänge und Kausalität

Bei Kausalaussagen ist immer Vorsicht geboten, und obwohl Korrelationen bei solchen Behauptungen eine wichtige Rolle spielen können (z. B. über Strukturmodelle), ist das bekannte Sprichwort „Eine Korrelation ist keine Ursache“ immer noch ein guter Ausgangspunkt. Der Begriff „Effektstärke“ ist nicht hilfreich, da der Begriff „Effekt“ fälschlicherweise eine Kausalität implizieren kann. Campbell und Stanley (1966) haben Bedingungen für kausale Behauptungen aufgestellt: Es muss ein Zusammenhang zwischen dem Einfluss und der Wirkung bestehen, eine zeitliche Abfolge sowie der Ausschluss von Scheinkorrelationen. Z. B. müssen Forschende Gegenbeispiele zu ihren kausalen Behauptungen untersuchen und „die Anzahl der plausiblen konkurrierenden Hypothesen, die zur Erklärung der Daten zur Verfügung stehen, reduzieren. Je weniger plausible konkurrierende Hypothesen übrigbleiben, desto größer ist der Grad der ‚Bestätigung‘“ (Campbell & Stanley, 1966, S. 206).

Im vorliegenden Buch gibt es viele Beispiele, die mit den Lernleistungen korrelieren (z. B. Geschlecht des Lernenden, Frühgeburtsgewicht, Persönlichkeitsmerkmale), und ich verwende r statt d , wenn ich solche Korrelationen diskutiere. Es gibt aber auch viele Effekte, die auf dem Versuch beruhen, Kausalitäten zu entdecken (Einführung einer Lehrmethode im Vergleich zu anderen Methoden oder Prä-Post-Vergleiche). Es ist jedoch immer darauf zu achten, dass die genannten Bedingungen erfüllt sind. Allzu oft ist die Qualität der Moderatoren nicht hoch genug, um sicherzustellen, dass plausible konkurrierende Hypothesen berücksichtigt werden.

Das Kernstück kausaler Behauptungen ist der Erklärungsansatz in Kapitel 3 und Kapitel 4, um die Mechanismen zwischen den Variablen zu verstehen. Ich versuche diesen Erklärungsansatz bei jeder neu veröffentlichten Meta-Analyse anzuwenden und fordere andere auf, Indizien für die gegenteiligen Tatsachen zu liefern, um die Aussagen möglicherweise zu falsifizieren. Ich habe mich bemüht, in den Behauptungen kühn zu sein, wie es Popper (1992, S. 280) formulierte:

„... kühne Theorien aufzustellen; unser Bestes zu versuchen, um zu zeigen, dass sie falsch sind; und sie vorläufig zu akzeptieren, wenn unsere kritischen Bemühungen erfolglos sind. Unter dem hier entwickelten Gesichtspunkt bleiben alle Gesetze, alle Theorien im Wesentlichen vorläufig, mutmaßlich oder hypothetisch, selbst wenn wir nicht mehr an ihnen zweifeln können.“

Art der Effektstärken

Ein weiterer Kritikpunkt ist die Vermischung von Studien, die auf Korrelations-, Prä-Post- und Experimental-Vergleichsstudien basieren. Dies ist eine berechtigte Kritik, auch wenn es auf der Ebene einer Synthese von Meta-Analysen schwieriger sein kann, dies zu entschlüsseln. Diejenigen, die Meta-Analysen erster Ordnung durchführen, müssen möglicherweise das Studiendesign als Hauptmoderator einbeziehen. Dies könnte das Problem lösen. Die meisten Meta-Analysen sind eindeutig, wenn sie eine Korrelation im Vergleich zu den mittleren Zuwächsen verwenden, obwohl letztere eine Mischung aus Prä-Post-, Post-Gruppen-Vergleich von zwei Gruppen und anderen sein können. In einem Prä-Post-Design ist der Effekt eher ein Maß für den Gewinn, der über die Normalität hinausgeht, aber das gilt auch für den Vergleich der Versuchsgruppe mit der Kontrollgruppe. Eine Übersichtsmethode besteht darin zu fragen, ob es einen Unterschied in den Meta-Analysen gibt, die auf Korrelationen (0,42, $n = 315$) oder mittleren Unterschieden (0,40, $n = 1.443$) als Effektgrößen beruhen. Das ist nicht der Fall.

Alle Meta-Analysen wurden nach einer von Wolfgang Bewyl (persönliche Mitteilung) empfohlenen Methode kodiert. Er argumentierte, dass es mindestens fünf Hauptgruppen von Einflüssen gebe. Diese reichen von Einflüssen, die direkt auf eine kausale Intervention in den Unterricht von Lernenden abzielen, bis hin zu kausalen Studien ohne Intervention, auf die Schulen oft wenig Einfluss haben.

1. Kausal – Intervention – Unterricht: Lehrpersonen können eine solche Wirkung sehr kurz- oder mittelfristig umsetzen (einige Minuten bis zu mehreren Monaten oder wenigen Jahren). Dazu greifen sie absichtlich in die Schule ein (z. B. durch Personalentwicklung) oder verändern den Unterricht. Dabei gibt es zumindest ein implizites Kausalmodell. Beispiele sind die meisten Lehrstrategien, Feedback und neue Programme.
2. Kausal – Intervention: Dies bezieht sich auf die Fälle, in denen ein Einfluss auf eine Veränderung der Lernleistungen zurückgeführt werden kann. Beispiele hierfür beinhalten eine erhöhte Motivation.
3. Korrelation: In diesen Studien wird eine Variable zu einem Leistungsergebnis in Beziehung gesetzt; es gibt nicht notwendigerweise kausale Behauptungen. Beispiele sind Angst, emotionale Intelligenz, Gewissenhaftigkeit und viele emotionale Variablen.
4. Kausal – Intervention – Kontext: Dies bezieht sich auf Interventionen in der Struktur der Schule und des Unterrichts, wie z. B. die Klassengröße und die Gruppierungsmethoden.
5. Kausal – keine Intervention: Hierbei handelt es sich um einen Einfluss, der sich auf die Lernleistung auswirkt, ohne dass es eine (relevante) Rückwirkung gibt. Beispiele hierfür sind physische Merkmale der Lernenden (Krankheiten), der sozioökonomische Status, Schultypen (z. B. Charter Schools) oder auch Merkmale der Lehrperson (z. B. Glaubwürdigkeit).

In einigen Fällen kann es eine Mischung aus Studien geben, die mit der Lernleistung korrelieren, und anderen, die darauf abzielen, die Lernleistung zu verbessern. So kann z. B. die körperliche Fitness oder der BMI mit der Lernleistung korrelieren und mit einem Programm verglichen werden, das die Lernenden bewusst in Bewegung und Sport bringt, um ihr Lernen zu verbessern; oder das Wohlbefinden kann ein Korrelat oder der Schwerpunkt eines gezielten Programms zur Veränderung des Wohlbefindens sein und die Auswirkungen auf die Lernleistung bewerten. Es gibt viele Studien, in denen Lernstrategien mit der Lernleistung in Verbindung gebracht werden, und andere, in denen bewusst Programme zur Verbesserung der Lernstrategien durchgeführt werden, um die Wirkungen auf die Lernleistung zu sehen.

Bei den meisten Einflüssen ist die Klassifizierung eindeutig. Die Auswirkungen sind am höchsten bei den eher kausalen Einflüssen und am niedrigsten bei Korrelationen sowie bei Studien, die sich auf die Strukturen der Schulen beziehen (Kausal – Intervention – Kontext) und bei den kausalen Studien ohne Intervention (Tabelle 2.12).

Tabelle 2.12 Effektgrößen moderiert nach Art des kausalen bis korrelativen Einflusses

Art des Einflusses	Mittelwert (M)	Standardabweichung (SD)	N
Kausal – Intervention – Unterricht	0.51	0.32	1.175
Kausal – Intervention	0.53	0.32	164
Korrelation	0.30	0.44	319
Kausal – Intervention – Kontext	0.30	0.30	369
Kausal – keine Intervention	0.26	0.41	285

Fixed- und Random-Effects-Modelle

In den letzten 15 Jahren sind immer mehr Meta-Analysen dazu übergegangen, die Werte anhand von Random-Effects-Modellen und nicht mehr anhand von Fixed-Effects-Modellen zu schätzen (Hedges & Vevea, 1998). Das Random-Effects-Modell geht davon aus, dass jede Studie aus einer Grundgesamtheit möglicher Effekte stammt und berücksichtigt daher die Breite des Konfidenzintervalls, um den Gesamtmittelwert zu schätzen. Wenn stattdessen die Variabilität der gesamten Effektstärken gleich Null ist oder wenn es gute Gründe für die Annahme gibt, dass die gesamte Effektstärke für alle Studien gleich ist, dann wurde ein Fixed-Effects-Modell verwendet. Trotz der unterschiedlichen Annahmen (und meiner Erwartung, dass die Effektstärken bei der Verwendung des Random-Effects-Modells viel höher sein würden) ist dies nicht der Fall (fixed = 0,40, $n = 1.365$; random = 0,44, $n = 947$).

Bereiche des Lehrplans

Zu den häufigsten Kritiken, die ich per E-Mail erhalte, gehört die Nichtbeachtung der einzigartigen Merkmale von Fachbereichen – Mathematik, Sprachen, Kunst, Musik, Sport (und man könnte fast jeden Lehrplanbereich nennen) sind sicherlich unterschiedlich. Deren Auswirkungen sind in den Sozialwissenschaften (Geisteswissenschaften, Sprachen) prägnanter, aber in den anderen Lehrplanbereichen ähnlich (Tabelle 2.13). Ja, normalerweise sollte dies ein untersuchter Moderator sein, aber wie bereits erwähnt, ist die Enge oder Weite der Zielvariable (unabhängig vom Lehrplanbereich) wahrscheinlich entscheidender. Es wird zwar argumentiert, dass die allgemeinen *VL*-Botschaften unabhängig von der Disziplin gelten, aber es ist zu beachten, dass das Verhältnis zwischen dem Fachwissen und den fachlichen Konzepten innerhalb eines Lehrplanbereichs kritischer ist; daher werden die Lehrpläne nicht ignoriert, sondern sind nur kein wichtiger Moderator für die Interpretation der Lehr- und Lernprozesse.

Tabelle 2.13 Effektstärken, moderiert durch Lehrplandomänen

Fächer	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Human- & Sozialwissenschaften	0.64	0.62	7
Sprachen	0.60	0.29	34
Kreative Kunst	0.55	0.27	29
Naturwissenschaften	0.55	0.37	174
Schreiben	0.53	0.29	26
Lesen	0.48	0.38	238
Mathematik	0.41	0.40	205
Alle	0.38	0.35	1.599

Alter

Fast ebenso häufig wird behauptet, dass die Auswirkungen bei 5-, 15- und 25-Jährigen unterschiedlich seien. Wir alle wissen, dass sich die Grundschulen von den weiterführenden Schulen und den Hochschulen unterscheiden. Dies scheint aber nicht der Fall zu sein. Hier sind die Botschaften also altersübergreifend (Tabelle 2.14). Es wird jedoch darauf hingewiesen, dass bei der Schätzung des Wertes für Erwachsene Vorsicht geboten ist, da diese nicht Teil der Auswahl für die Meta-Analysen in diesem Buch waren und Erwachsene in Englischkursen, Erwachsene in beruflicher Weiterbildung und Erwachsene in Psychologiestudien sehr unterschiedlich sind. Es ist schwierig, repräsentative Stichproben von Erwachsenen zu erhalten.

Tabelle 2.14 Effektgrößen moderiert durch das Alter der Lernenden

Alter	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Frühkindliche Bildung	0.40	0.25	117
Elementarbereich	0.36	0.34	130
Sekundarstufe II	0.32	0.33	73
Primar- & Sekundarstufe I	0.36	0.35	56
Tertiärbereich	0.43	0.40	171
Alle Altersstufen	0.43	0.37	1.752

Lernende mit besonderen Bedürfnissen und Begabungen

Hierbei wird kritisiert, dass die genannten Lernenden sich von durchschnittlichen Lernenden unterscheiden. Tabelle 2.15 zeigt, dass die Meta-Analysen, die sich auf Lernende mit Lernschwierigkeiten beziehen, höhere Effekte aufweisen. Ansonsten gibt es kaum Unterschiede in Bezug auf die Einteilung der Lernenden. Auch bei diesen Meta-Analysen zu Lernschwierigkeiten liegt der Schwerpunkt eher auf spezifischen Inhalten im Bereich des Wissens und bestimmter Fertigkeiten, was wahrscheinlich der Grund für die höheren Effekte ist.

Tabelle 2.15 Effektgrößen moderiert durch die Einteilung der Lernenden

Fächer	Mittelwert (M)	Standardabweichung (SD)	Grundgesamtheit (N)
Alle	0.41	0.35	2.053
Englischsprachige Lernende	0.49	0.37	33
Begabte Lernende	0.45	0.26	20
Lernende mit Lernschwierigkeiten	0.63	0.33	84
Lernende mit Förderbedarf	0.45	0.48	122

Sich überschneidende Daten

Wenn es mehr als eine Meta-Analyse zu einem Faktor gibt, teilen die verschiedenen Meta-Analysen wahrscheinlich viele Einzelstudien, was den durchschnittlichen Effekt verzerren könnte. In unserer Analyse von 28 Meta-Analysen zum Thema „Flipped Classroom“ gibt es zum Beispiel 765 Referenzen (drei lieferten keine Liste der Studien), und 471 der Studien (62 %) waren nur in einer einzigen Meta-Analyse vorhanden (siehe auch Hew et al., 2021). Dies war eine Überraschung, da ich einen geringeren Anteil einmalig rezipierter Artikel erwartet hatte. Einmalig rezipierte Studien zu haben ist wertvoll, um frühere Studienergebnisse zu replizieren (wenn die Studien gleich sind, sollten auch die Mittelwerte und andere Schätzungen gleich sein, obwohl sie für unterschiedliche Moderatoren kodiert werden könnten).

Index der Robustheit

Da ich die Datenbank MetaX immer wieder aktualisiere, ändert sich manchmal der Durchschnitt innerhalb eines Faktors sehr stark, meistens aber nur geringfügig. Dies hat dazu geführt, dass einige die Beständigkeit der Ergebnisse in Frage stellen. Ich werde daher oft gebeten, keine weiteren Meta-Analysen hinzuzufügen, da es verwirrend sei, wenn sich die durchschnittlichen Effekte ändern. Das Interesse an der Erstellung weiterer Meta-Analysen lässt nicht nach, und das ist auch gut so. Das Einfrieren von *VLI*-Daten ergäbe daher keinen Sinn. Ständig finde und lese ich neue Meta-Analysen und füge sie der Datenbank hinzu. Wie es in einigen Teilen dieses Buches ersichtlich ist, können diese Ergänzungen die Schlussfolgerungen aus *VLI* verändern. Es ist nicht verwunderlich, dass sich der durchschnittliche Effekt verändern kann, da die früheren Meta-Analysen teilweise auf wenigen Studien basierten.

Der Robustheitsindex ist ein gewichteter Durchschnitt aus fünf Hauptindikatoren für die Stabilität: die Anzahl der Meta-Analysen, die Anzahl der Studien, die geschätzte Anzahl der Lernenden, die Anzahl der Effekte innerhalb eines Einflusses und die ausfallsichere Anzahl der Studien. Jedes dieser fünf Maße wurde in Quintile eingeteilt, und das durchschnittliche Quintil ist der Robustheitsindex für einen Einfluss (z. B. ist bei der Diskussion des Thermometers der Durchschnitt über die Quintile der fünf Maße für Kreativitätsförderung $R = 5$, was auf eine hohe Robustheit hindeutet, da die Stichprobe am größten war). Für jeden Einfluss wird ein Robustheitsindex angegeben. Liegt dieser zwischen 3 und 5, so kann mehr Vertrauen in diesen bestehen. Die Faktoren mit Robustheitsindizes von 1 und 2 können wahrscheinlich schwanken, wenn im Lauf der Zeit mehr Studien zu diesem Einfluss hinzugefügt werden (Tabelle 2.16).

Schlussfolgerung

Wie in den beiden vorangegangenen Abschnitten dargelegt, konzentriert sich die meiste Kritik an *Visible Learning* auf die Forschungsmethodologie und nicht auf die datenbasierten Interpretationen, Erklärungen oder Implikationen. Ein Hauptgrund für die Veröffentlichung von MetaX war, anderen Forschenden die Jahrzehnte zu ersparen, die es brauchte, um die Datenbank zu erstellen – verbunden mit der Aufforderung, alternative Interpretationen, Erklärungen oder Implikationen vorzuschlagen.

Eine Meta-Analyse ist nur eine Methode und die Kritikerinnen und Kritiker haben Recht, wenn die Aussage von *VL* nur die Präsentation einer Rangfolge von Effekten ist und jeder Faktor als (quasi) unabhängig betrachtet wird. Die Faktoren sind Wahrscheinlichkeitsaussagen, und die Qualität und Genauigkeit bei der Umsetzung von Maßnahmen sind am wichtigsten. Gene Glass' Beitrag bestand darin, dass er sich nicht nur auf die Entwicklung eines Maßes für die Größenordnung konzentrierte, sondern auch nach Moderatoren suchte, da die Antworten so oft in der Variabilität liegen. Mein eigener Doktorvater, Rod McDonald (2014), hat uns eingebläut, dass die Antworten in den Residuen liegen. Marton und Pang (2006) haben eindrucksvoll gezeigt, wie die Fehlervarianz eine Quelle wichtiger Erkenntnisse für Lehrpersonen und Forschende ist. Die Sensibilität einer Lehrperson ist für die unterschiedlichen Reaktionen der Lernenden auf den Unterricht für eine effektive Wirkung unerlässlich. In dieser Weiterführung von *Visible Learning* möchte ich das Modell des *VL* skizzieren, das ich in den letzten 40 Jahren aufgebaut und getestet habe (in Schulen und mit jeder neuen Meta-Analyse).

Tabelle 2.16 Die fünf Quintile der Robustheit und die zu ihrer Berechnung verwendeten Daten

Anzahl Meta- Analysen	Anzahl Studien		Personen- anzahl		Anzahl Effekte		Fail-Safe Schätzung		Robust- heitsindex				
	N	%	N	%	N	%	N	%	N	%			
1	65	18	65	18	69	19	1-72	0	85	24	1	44	12
2	50	14	50	14	73	21	73-176	1-193	57	16	2	79	22
3	85	24	85	24	6.640-23.979	20	177-464	194-972	72	20	3	93	26
4	79	22	79	22	23.980-166.799	20	465-1.260	973-4.063	71	20	4	76	22
5	77	22	77	22	166.800+	20	1.261+	4.064	71	20	5	62	17