

Stefan Krauss & Patrick Wiesner

Ist die Flipped-Classroom-Methode besser als traditioneller Unterricht?

Kritische Bemerkungen zu sieben Metaanalysen für den Schulbereich¹

1. Einleitender Überblick

Flipped-Classroom-Unterricht hat in den letzten Jahren auch aufgrund der COVID-19-Pandemie nochmal an Popularität gewonnen. Auf den ersten Blick könnte der Wissenserwerb zuhause durchaus eine verheißungsvolle Methode sein, um die Unterrichtszeit selbst systematischer zum Einüben beziehungsweise zur Vertiefung von Inhalten verwenden zu können (vgl. 2.1). Wie beinahe jedes schulische Konzept bietet aber natürlich auch der Flipped-Classroom-Unterricht sowohl potenzielle Chancen als auch mögliche Fallstricke (Cevikbas & Kaiser, 2023).

Im vorliegenden Beitrag geben wir einen kurzen Überblick zur Wirksamkeitsforschung bezüglich der Flipped-Classroom-Methode, wobei wir uns auf den schulischen Unterricht beschränken (vgl. 2.2). Hierzu stellen wir zu Beginn einige Metaanalysen vor, die allesamt einen positiven Effekt dieser Unterrichtsform gegenüber traditionellem Unterricht konstatieren (vgl. 2.3). Diese klare empirische Evidenz verunsichert zunächst, würde das doch bedeuten, dass der in Deutschland praktizierte Unterrichtsalltag ernsthaft überdacht werden müsste.

Anschließend werfen wir deshalb einen Blick auf die zugrundeliegenden Originalstudien (für den Primar- und den Sekundarbereich gibt es mittlerweile zahlreiche Einzelstudien). Dabei stellt sich interessanterweise heraus, dass die weit überwiegende Anzahl dieser individuellen Studien lediglich *eine* Klasse im Flipped-Classroom-Unterricht mit *einer* Klasse im traditionellen Unterricht verglichen hat (vgl. 2.4). Um reine Klasseneffekte (z. B. durch Mehrebenenmodellierung) ausschließen zu können, wäre jedoch die empirische Implementation einer substantiellen Anzahl unterschiedlicher Klassen pro Experimentalbedingung erforderlich. Die beiden einzigen Flipped-Classroom-Studien für den Schulbereich, die jeweils

¹ Die theoretischen Überlegungen des vorliegenden Beitrags stammen im Wesentlichen aus Wiesner, Krauss, Stegmüller und Binder (eingereicht). Dort werden Ergebnisse einer Flipped-Classroom-Studie mit $N = 950$ Schüler*innen berichtet ($N = 12$ Klassen pro Experimentalbedingung; siehe Tabelle 2).

eine zweistellige Anzahl von Klassen pro Versuchsbedingung implementiert haben, haben jedoch *keine* positiven Effekte für die Flipped-Classroom-Methode gefunden.

In der empirischen Psychologie bzw. Bildungsforschung ist es keine Seltenheit, dass zahlreiche kleinere Studien wiederholt Effekte finden, die sich in größeren Studien nicht replizieren lassen (vgl. 2.5). Abschließend greifen wir deshalb einige generelle Kritiken an Meta-Studien auf, die eventuell auch alternative Erklärungen für die wiederholte Bestätigung positiver Effekte von Flipped-Classroom-Unterricht liefern könnten (vgl. 2.6). Die im vorliegenden Beitrag angestellten theoretischen Überlegungen sollen dessen Wirksamkeit nicht grundsätzlich widerlegen oder vor einem Einsatz der Flipped-Classroom-Methode warnen (siehe Fazit in Abschnitt 3). Das Ziel des Kapitels ist vielmehr eine Relativierung der scheinbar übermächtigen Evidenz der existierenden Metaanalysen in Bezug auf den Primar- und den Sekundarbereich.

2. Flipped Classroom

2.1 Eine kurze Begriffsklärung

Flipped-Classroom-Unterricht (dt. »umgedrehter« Unterricht) hat nicht nur aufgrund der Corona Pandemie in der Praxis an Bedeutung gewonnen, sondern rückte gerade in den letzten fünf Jahren zunehmend auch in den Fokus empirischer Forschung (Cevikbas & Kaiser, 2020, 2023; Hwang et al., 2019; Kapur et al., 2022). Fast alle Definitionen beinhalten die »invertierte« Reihenfolge des traditionellen Unterrichts als Mindestcharakterisierung. Im Gegensatz zum »traditionellen« Unterricht (d. h., der Wissenserwerb findet in der Schule statt und das Üben überwiegend zuhause) bedeutet Flipped Classroom, dass neues Wissen von den Schüler*innen bereits vor der eigentlichen Unterrichtsstunde selbstständig erworben wird, um im Unterricht dann mehr Zeit für Reflexion, Vertiefung und begleitete Übung zu haben (Bishop & Verleger, 2013; Lage et al., 2000). Oftmals werden für das Lernen zuhause Erklärvideos zur Verfügung gestellt, gelegentlich in Kombination mit zusätzlichen kurzen Quizfragen zur Verständnisüberprüfung (z. B. Hew et al., 2021; Wagner et al., 2020). Ein effektiver Flipped-Classroom-Unterricht setzt jedenfalls voraus, dass a) einerseits den Lernenden Materialien zur Verfügung gestellt werden, die sich für eine eigenständige Erarbeitung des Stoffs eignen, und b) andererseits die Schüler*innen diese Wissensaneignung dann zuhause auch tatsächlich erfolgreich absolvieren.

2.2 Einschränkung auf schulischen Unterricht

Flipped-Classroom-Unterricht ist für alle Fächer und alle Jahrgangsstufen denkbar, und auch Veranstaltungen an der Universität können prinzipiell im Flipped-Classroom-Format stattfinden (in der Tat stammt die Methode sogar ursprünglich aus der universitären Bildung). Es ist wichtig festzuhalten, dass wir uns im Folgenden ausschließlich auf den schulischen Bereich konzentrieren (d. h. Primar- und Sekundarstufe) und das Lernen an der Hochschule (d. h. den tertiären Bereich) unberücksichtigt lassen.

Der Grund hierfür ist, dass das eigenständige Lernen zuhause ein gewisses Maß an Selbstregulation erfordert, wofür wiederum Motivation und Eigeninitiative der Lernenden notwendig sind. In dieser Hinsicht sind Schule und Universität nicht direkt vergleichbar. Schon die Wahl eines Studienfachs geschieht auf freiwilliger Basis, genauso die Aufrechterhaltung und der erfolgreiche Abschluss eines Studiums. Die Dozent*innen von Vorlesungen und Seminaren sind wiederum nur bedingt für die individuellen Erfolge ihrer Studierenden (sowie deren »permanente Überwachung«) zuständig.

Für Schüler*innen gibt es dagegen generelle Schulpflicht (zumindest bis zum Ende der 9. Klasse) und auch die zu lernenden Fächer können größtenteils nicht selbst gewählt werden. Weiterhin erarbeiten sich Lernende an der Schule neuen Stoff in der Regel nicht selbst. Lehrkräfte an der Schule wiederum haben Aufsichts- und Fürsorgepflichten und stehen für das Lernen der Schüler*innen in einer größeren Verantwortung als Lehrende an der Universität (zumindest was die gängigen unterrichtlichen Qualitätskriterien *kognitive Aktivierung*, *konstruktive Unterstützung* oder *Klassenführung* betrifft; Kunter et al., 2011).

Aus diesen Gründen fokussieren wir im Folgenden den *schulischen* Unterricht. Zum Einsatz der Flipped-Classroom-Methode an der Hochschule gibt es ebenfalls bereits zahlreiche Einzel- sowie Meta-Studien (diesbezügliche empirische Evidenz findet sich z. B. in Güler et al., 2023; Jang & Kim, 2020; Chen et al., 2018).

2.3 Empirische Evidenz aus Meta-Analysen

Tabelle 1 stellt sieben Metaanalysen vor, die die Wirkung der Flipped-Classroom-Methode in der Schule im Vergleich zu traditionellem Unterricht untersucht haben (Cheng et al., 2019; Doğan et al., 2023; Låg & Sæle, 2019; Strelan et al., 2020; van Alten et al., 2019; Wagner et al., 2020; Zhu, 2021).

Alle Metaanalysen kommen zum Schluss, dass Flipped Classroom effektiver ist als traditioneller Unterricht (mit Effektgrößen von $g = 0.21$ bis zu einem Ausreißer mit $d = 1.89$). Es gibt mittlerweile sogar Meta-Studien über Metaanalysen (Hew et al., 2021; Kapur et al., 2022), die sich jedoch überwiegend mit dem Bereich der universitären Bildung befassen.

Tabelle 1: Sieben Metaanalysen zur Wirksamkeit der Flipped Classroom Methode (Effektstärken jeweils im Vergleich zu traditionellem Unterricht)

Metaanalyse Zeitschrift	Altersstufe (Anzahl berücksichtigter Einzelstudien)	Effektstärke (zugunsten von Flipped Classroom) Cohen's d / Hedges g^*
Doğan et al. (2023) ** <i>Education & Information Technologies</i>	Sekundarstufe (11)	$d = 0.58$
	Primarstufe (4)	$d = 1.89$
Zhu (2021) <i>Educational Technology Research and Development</i>	Primar- und Sekundarstufe zusammen (27)	$d = 0.54$
Wagner et al. (2020) <i>Zeitschrift für Pädagogische Psychologie</i>	Sekundarstufe (25)	$d = 0.42$
Strelan et al. (2020) ** <i>Educational Research Review</i>	Sekundarstufe (21)	$g = 0.64$
	Primarstufe (3)	$g = 0.47$
Låg and Sæle (2019) ** <i>AERA Open</i>	Sekundarstufe (16)	$g = 0.45$
	Primarstufe (12)	$g = 0.44$
van Alten et al. (2019) ** <i>Educational Research Review</i>	Sekundarstufe (11)	$g = 0.36$
Cheng et al. (2019) ** <i>Education Tech Research</i>	Sekundarstufe (12)	$g = 0.21$

Bemerkung: * Die Effektstärkemaße g und d unterscheiden sich nur durch eine leicht abweichende Berechnung der gepoolten Standardabweichung:

$|g$ bzw. $d| = 0,2$: kleiner Effekt, $|g$ bzw. $d| = 0,5$: mittlerer Effekt, $|g$ bzw. $d| = 0,8$: großer Effekt

** In diesen Metaanalysen wurde auch der tertiäre Bereich untersucht (nicht in Tab. 1 berücksichtigt)

2.4 Empirische Evidenz aus Einzelstudien

Die Metaanalysen aus Tabelle 1 greifen insgesamt auf etwa 50 Einzelstudien zurück, die die Wirkung von Flipped Classroom im Vergleich zu traditionellem Unterricht empirisch für die Primar- beziehungsweise Sekundarstufe untersucht haben. Tabelle 2 stellt diese verfügbaren Einzelstudien, die in mindestens einer der sieben Metaanalysen aus Tabelle 1 enthalten waren, erstmals im kompletten Überblick dar (inklusive Dissertationen, lediglich unveröffentlichte Studien wurden ausgeschlossen). Die Anfangsbuchstaben der Autor*innen (fett) in Tabelle 1 werden in der linken Spalte in Tabelle 2 verwendet, um anzugeben, welche Studie in welche Metaanalyse einbezogen wurde.

Tabelle 2: Übersicht über die Einzelstudien, die in die Metaanalysen von Tabelle 1 eingegangen sind (angeordnet nach der durchschnittlichen Anzahl an Klassen pro Experimentalbedingung); die ersten beiden Studien (Wiesner et al., eingereicht; Wagner & Urhahne, 2021) sind noch in keine Metaanalysen eingegangen

	Autoren (Jahr) Zeitschrift [berücksichtigt in Metaanalyse] ^a	insgesamt: # SuS, (# Klassen)	Fach (Jahrgangsstufe) zusätzliche Info	Flipped Classroom # SuS, (# Klassen)	Kontroll- gruppe # SuS, (# Klassen)	Effektstärke (positiv: zugunsten von Flipped Classroom)	Quelle ^b
1	Wiesner et al. (eingereicht)	950, (37)	Mathematik (6. / 7.)	295, (12)	339, (13)	$d = -0.50^*$	(c)
			Mathematik (6. / 7.) + Unterstützung der Selbstregulation	316, (12)		$d = 0.25$	(c)
2	Wagner & Urhahne (2021) <i>Learning and Instruction</i>	848, (38)	Englisch (9.) + schülerzentriert	215, (10)	169, (8)	$d = -0.14$	(c)
			Englisch (9.) + lehrkraftzentriert	201, (9)	263, (11)	$d = -0.13$	(c)
3	Ramaglia (2015) <i>Dissertation</i> [L]	520, (24)	Mathematik (7.)	137, (6)	137, (6)	$g = 0.13$	(m)
			Mathematik (9.–12.)	129, (5)	49, (3)	$g = 1.09^*$	(m)
			Mathematik (9.–10.)	48, (2)	23, (2)	$g = -0.17$	(m)
4	Bell (2015) <i>Master Thesis</i> [L, Z]	196, (7)	Physik (n.v.)	n.v., (3)	n.v., (4)	$g = 0.07$	(m)
5	Lo et al. (2018) <i>Computers & Education</i> [S, W]	382, (12)	Mathematik (10.–12.)	28, (1)	27, (1)	$d = 0.82^*$	(m)
			Physik (10.–12.)	119, (4)	125, (4)	$d = 0.61^*$	(m)
			Chinesisch (10.–12.)	12, (1)	12, (1)	$d = 0.95$	(m)
			Informatik (10.–12.)	11, (1)	11, (1)	$d = 0.02$	(m)
6	Ripley (2015) <i>Dissertation</i> [C, L]	322, (4°)	Mathematik (6.)	55, (2°)	267, (10°)	$g = 0.01$	(m)
7	Dixon (2017) <i>Dissertation</i> [L, D]	123, (4°)	Biologie (12.)	61, (2°)	56, (2°)	$d = 0.04$	(m)
8	Županec et al. (2018) <i>Journal of Baltic Science Education</i> [D]	112, (4°)	Biologie (7.)	56, (2°)	56, (2°)	$d = 2.15^*$	(c)
9	Atwa et al. (2016) <i>Journal of Personalized Learning</i> [S, D]	109, (4)	Physik (11.)	53, (2)	56, (2)	$d = 0.50^*$	(c)
10	Baki Mohammed Diab, K. M. A. (2016) <i>Dissertation</i> [L]	89, (4)	Mathematik (11.)	39, (2)	40, (2)	$g = 0.97^*$	(m)
11	Bhagat et al. (2016) <i>Educational Technology & Society</i> [L, S, W, Z]	82, (4°)	Mathematik (9.)	41, (2°)	41, (2°)	$g = 0.52^*$	(m)
12	Ceylan & Kesici (2017) <i>Journal of Human Sciences</i> [L]	72, (4°)	Informatik (6.)	37, (2°)	35, (2°)	$g = 0.92^*$	(m)
13	Duffy (2016) <i>Dissertation</i> [D, Z]	87, (4)	Biologie (8.)	44, (2)	43, (2)	$d = 0.28$	(m)
14	Leo & Puzio (2016) <i>Journal of Science Education and Technology</i> [D, L, S]	69, (4)	Biologie (9.)	40, (2)	29, (2)	$g = 0.30$	(m)
15	Salimi & Yousefzadeh (2015) <i>Advances in Language and Literary Studies</i> [C, S, W, Z]	250, (10)	Englisch (7.)	25, (1)	25, (1)	$g = 1.63^*$	(m)
			Arabisch (7.)	25, (1)	25, (1)	$g = 0.51$	(m)
			Mathematik (7.)	25, (1)	25, (1)	$g = 1.83^*$	(m)
			NWT (7.)	25, (1)	25, (1)	$g = 1.57^*$	(m)
			Geographie (7.)	25, (1)	25, (1)	$d = 1.22^*$	(m)

16	Jong (2017) <i>Educational Technology & Society</i> [W, Z]	213, (6)	Wirtschaft (11.) high academic	36, (1)	36, (1)	$g = 0.10$	(m)
			Wirtschaft (11.) medium academic	36, (1)	36, (1)	$g = 0.92^*$	(m)
			Wirtschaft (11.) bottom academic	35, (1)	34, (1)	$g = 0.62^*$	(m)
17	Sergis et al. (2018) <i>Computers in Human Behavior</i> [S, Z]	128, (6)	Informatik (8.)	23, (1)	23, (1)	$g = 0.92^*$	(m)
			Mathematik (10.)	20, (1)	20, (1)		
			Geisteswissen- schaft (8.)	20, (1)	20, (1)		
18	Charles-Ogan & Williams (2015) <i>British Journal of Education</i> [S, Z]	100, (2)	Mathematik (n.v.)	55, (1)	45, (1)	$g = 1.69^*$	(m)
19	Tsai et al. (2015) <i>International Journal of Information and Communication Technology Education</i> [C, L]	96, (3)	Informatik (6.)	50, (2)	46, (1)	$g = 0.41^*$	(m)
20	Chao et al. (2015) <i>Engineering Education</i> [A, L, S, W, Z]	91, (2)	Ingenieurwissen- schaft (11.)	46, (1)	45, (1)	$d = 0.78^*$	(m)
21	Esperanza et al. (2016) <i>Lecture Notes in Computer Science</i> [W]	91, (2)	Mathematik (10.–11.)	45, (1)	46, (1)	$d = 0.59^*$	(o)
22	Pengfei & Mingxuan (2015) <i>Proceedings 2015 International Conference of Educational Innovation through Technology (EITT)</i> [L]	90, (2)	Mathematik (5.)	45, (1)	45, (1)	$g = 0.26^*$	(m)
23	Mohanty & Parida (2016) <i>Creative Education</i> [D, S, Z]	90, (2)	Naturwissenschaft & Geschichte (8.)	45, (1)	45, (1)	$d = 1.09^*$	(c)
24	Sezer (2017) <i>Journal of Educational Computing Research</i> [D, W, Z]	68, (2)	Science (6.)	35, (1)	33, (1)	$g = 0.92^*$	(m)
25	Clark (2015) <i>Journal of Educators Online</i> [A, C, L, S, W, Z]	42, (2)	Mathematik (9.)	21, (1)	21, (1)	$g = 0.03$	(m)
26	Denprapat & Chuaychoowong (2016) <i>Journal of Social Academic</i> [L]	67, (2)	Englisch (7.)	n.v., (1)	n.v., (1)	$g = 0.76^*$	(m)
27	Olakanmi (2017) <i>Journal of Science Education and Technology</i> [D, L, W, Z]	66, (2)	Chemie (7.)	33, (1)	33, (1)	$g = 1.16^*$	(m)
28	Wiginton (2013) <i>Dissertation</i> [A, L, Z]	64, (3)	Mathematik (9.)	24, (1)	22, (1)	$g = 1.69^*$	(m)
29	Afrilyasanti et al. (2016) <i>International Journal of English Language and Linguistics Research</i> [S]	62, (2)	Englisch (10.)	30, (1)	32, (1)	$d = 2.77^*$	(c)
30	Schultz et al. (2014) <i>Journal of Chemical Education</i> [A, C, L, S, W]	61, (2)	Chemie (10.–12.)	29, (1)	32, (1)	$g = 0.96^*$	(m)
31	Saunders (2014) <i>Dissertation</i> [A, C, L, Z]	58, (2)	Mathematik (11.)	30, (1)	28, (1)	$g = -0.06$	(m)
32	Špilka & Popper (2014) <i>ICERI 2014 Proceedings</i> [A, L]	56, (2)	Mathematik (8.)	29, (1)	27, (1)	$g = 0.74^*$	(m)
33	Howell (2013) <i>Dissertation</i> [A, D, W]	56, (2)	Physik (9.)	31, (1)	25, (1)	$g = 0.38$	(m)

Kazu & Demirkol (2014)						
34	<i>The Turkish Online Journal of Educational Technology</i> [Z]	54, (2)	Biologie (12.)	27, (1)	27, (1)	$g = 0.57^*$ (m)
35	Kirvan et al. (2015) <i>Computers in the Schools</i> [A, C, L, S, W, Z]	54, (2)	Physik (7. / 8.)	25, (1)	29, (1)	$g = 0.28$ (m)
36	Tugun (2018) <i>TEM Journal</i> [S]	52, (2)	Informatik (9.)	28, (1)	24, (1)	$d = 0.52^*$ (c)
37	Johnson & Renner (2012) <i>Dissertation</i> [L]	52, (2)	Informatik (9.–12.)	26, (1)	26, (1)	$g = - 0.15$ (m)
38	Aidinopoulou & Sampson (2017) <i>Educational Technology & Society</i> [L, W, Z]	49, (2)	Geschichte (5.)	26, (1)	23, (1)	$g = 0.21$ (m)
39	DeSantis et al. (2015) <i>Journal of Interactive Learning Research</i> [A, L, S, Z]	47, (2)	Mathematik (9.)	26, (1)	21, (1)	$g = - 0.12$ (m)
40	Kostaris et al. (2017) <i>Educational Technology & Society</i> [L, S, W, Z]	46, (2)	Informatik (8.)	23, (1)	23, (1)	$g = 1.00^*$ (c)
41	Al-Harbi & Alshumaimeri (2016) <i>English Language Teaching</i> [L, S, W, Z]	43, (2)	Englisch (n.v.)	20, (1)	23, (1)	$g = 0.33$ (m)
42	Katsa M. (2016) <i>International Association for Development of the Information Society</i> [L, W, Z]	40, (2)	Mathematik (10.)	20, (1)	20, (1)	$g = 0.90^*$ (m)
43	Abdelrahman et al. (2017) <i>Online Journal of Educational Technology</i> [S, W]	28, (2)	Englisch (6.)	14, (1)	14, (1)	$d = 1.24^*$ (m)
44	Casem (2016) <i>European Journal of STEM Education</i> [S, W]	24, (2)	Mathematik (12.)	n.v., (1)	n.v., (1)	$d = 0.83$ (c)
45	Gayathri & Vijayarani (2018) <i>International Journal of Research and Analytical Reviews</i> [D]	n.v., (n.v.)	Chemie (11.)	n.v., (n.v.)	n.v., (n.v.)	$d = 1.61^*$ (c)
46	Malto et al. (2018) <i>KnE Social Science</i> [D]	80, (4°)	Biologie (10.)	40, (2°)	40, (2°)	n.v.*
47	Grazia & Hall (2017) <i>Journal of Computers in Maths and Science Teaching</i> [S]	78, (4°)	Mathematik (9.–12.)	39, (2°)	39, (2°)	n.v.
48	Loza et al. (2014) <i>Proceedings of the Latin American Conference</i> [W]	72, (4°)	Mathematik (9.)	37, (2°)	35, (2°)	n.v.*
49	Chen (2016) <i>Journal of Educational Technology Systems</i> [S]	64, (2)	Gesundheitswissenschaft (9.)	33, (1)	31, (1)	n.v.*
50	Yang (2017) <i>Journal of Information Technology Education</i> [W]	57, (2)	Englisch (8.)	31, (1)	27, (1)	n.v.
51	Gross (2014) <i>Dissertation</i> [C]	30, (2)	Englisch (12.)	16, (1)	18, (1)	n.v.
^a	A = van Alten et al. C = Cheng et al. D = Doğan et al. L = Låg and Sæle	[9 eingegangene Studien] [7 eingegangene Studien] [11 eingegangene Studien] [27 eingegangene Studien]		S = Strelan et al. W = Wagner et al. Z = Zhu	[21 eingegangene Studien] [20 eingegangene Studien] [21 eingegangene Studien]	
^b	(c) = Effektstärke berechnet (da sie weder aus der Originalquelle noch aus den Metaanalysen entnommen werden konnte)			(m) = Effektstärke von einer Metaanalyse entnommen (o) = Effektstärke von der Originalpublikation entnommen		
^c	Anzahl der Klassen von den Autoren geschätzt (1–35 SuS entsprechen einer Klasse)					
[*]	Effektstärke: * signifikant (p < 0.05)					

In Tabelle 2 haben wir die Flipped-Classroom-Studien als primäres Kriterium nach der *durchschnittlichen Anzahl der Klassen pro Versuchsbedingung* angeordnet (inklusive der Kontrollbedingung mit regulärem Unterricht) und als zweites Kriterium nach der Gesamtanzahl der untersuchten Schüler*innen. Man beachte, dass letztere sich immer bereits in Klassen befinden und sich auch für Interventionsstudien in der Regel nicht ohne Weiteres für eine Randomisierung aus dem Klassenverband lösen lassen. Da deshalb meistens lediglich die Zuweisung von Klassen zu Versuchsbedingungen zufällig erfolgen kann, macht oftmals erst die Implementation mehrerer Klassen eine empirische Studie im Schulkontext zu einem randomisierten Experiment (aber natürlich ist man auch bei Klassen meist auf Gelegenheitsstichproben angewiesen). Da auch die statistische Mehrebenenmodellierung, die potenzielle Klasseneffekte abschwächen kann, nur möglich ist, wenn mehrere Klassen pro Bedingung implementiert werden, haben wir die durchschnittliche Anzahl der Klassen pro Experimentalbedingung als Hauptkriterium für die Beurteilung der »Größe« einer Flipped-Classroom-Studie gewählt. Wenn diese Anzahl der Klassen in den Metaanalysen nicht explizit angegeben war, haben wir diese den Originalstudien entnommen.

Aus Tabelle 2, die insgesamt 51 Artikel enthält, wird deutlich, dass etwa 40 Studien (Nr. 11–41) nur *eine* Klasse pro Versuchsbedingung untersuchten. Weitere vier Studien (Nr. 6–9) haben *zwei* Klassen für jede Bedingung implementiert und lediglich fünf Studien (Nr. 1–5) haben eine größere Anzahl von Klassen pro Experimentalbedingung untersucht. Die bisherige Evidenz ist also zu einem großen Teil quasi-experimentell. Diese Einschränkung wurde auch in den meisten Metaanalysen erwähnt (z. B. Cheng et al., 2019), wenn auch oftmals nicht in dieser Deutlichkeit.

Bislang scheint es demnach lediglich zwei »größere« Flipped-Classroom-Studien für den Schulbereich zu geben (Wiesner et al., eingereicht; Wagner & Urhahne, 2021). Hierzu ist festzuhalten, dass diese beiden Studien noch zu neu waren, um in die Metaanalysen aus Tabelle 1 mit einzugehen.

In Wagner und Urhahne (2021) wurden durchschnittlich zehn Klassen pro Bedingung untersucht (im Englischunterricht). In dieser Studie war jedoch nicht der Flipped Classroom die effektivste Methode, sondern das gemeinsame Anschauen des Erklärvideos in der Klasse, gefolgt von schüler*innenzentriertem Unterricht. Da die Lernmaterialien in allen Bedingungen grundsätzlich identisch waren, hat die Studie von Wagner und Urhahne (2021) eine hohe *interne Validität*, aber gleichzeitig gibt es eine Einschränkung hinsichtlich der *ökologischen Validität*, da Erklärvideos in der Regel nicht im Unterricht gezeigt werden. Interessanterweise liefert diese Studie auch Hinweise auf die entscheidende Rolle, die selbstregulative Fähigkeiten speziell in Bezug auf die Wissensaneignung zuhause und somit für die Wirksamkeit der Flipped-Classroom-Methode im Allgemeinen haben könnten. Von den fünf unterschiedlichen von Wagner und Urhahne (2021) identifi-

zierten »Lerntypen« profitieren »Risikoschüler*innen«, »durchschnittliche Schüler*innen« und »ausgezeichnete Schüler*innen« tendenziell weniger vom Flipped Classroom, während sich diese Unterrichtsform laut dieser großangelegten Studie im Besonderen für »selbstständige« und »selbstbewusste« Schüler*innen eignet.

In Wiesner et al. (eingereicht) wurden zwölf Klassen pro Bedingung untersucht (im Mathematikunterricht). Neben einer »normalen« Flipped-Classroom-Bedingung, in der die Schüler*innen sich zuhause stochastische Inhalte per Erklärvideos aneignen sollten, gab es eine weitere Flipped Classroom Bedingung, in der die *Selbstregulation* der etwa 10–11-jährigen Kinder explizit unterstützt wurde (vgl. »Unterstützung der Selbstregulation« in Tab. 2). Dies geschah durch eine zusätzliche »mathematikfreie« Unterrichtsstunde, in der den Klassen Strategien zum Umgang mit Lernvideos erläutert und mit ihnen anhand eines nicht-mathematischen Erklärvideos eingeübt wurden (und zwar jeweils für die Phasen »vor dem Video«, »während des Anschauens«, und »nach dem Video«; siehe Wiesner et al., eingereicht). Interessanterweise war der Lernzuwachs im »normalen« Flipped-Classroom-Unterricht geringer als in der Kontrollgruppe mit traditionellem Unterricht. Mit der zusätzlichen Unterstützung durch explizit auf Erklärvideos bezogene Lernstrategien konnte ein kleiner positiver (jedoch nicht-signifikanter) Effekt im Hinblick auf den Lernzuwachs im Vergleich zur Kontrollgruppe und ein signifikanter positiver Effekt gegenüber dem »normalen« Flipped Classroom (d. h. ohne zusätzliche »Strategiestunde«) erzielt werden.

Insgesamt sind die beiden größeren (noch nicht in Metaanalysen berücksichtigten) Studien also weit davon entfernt, die in den vielen kleineren Einzelstudien gefundenen positiven Effekte zugunsten der Flipped-Classroom-Methode zu bestätigen. Das wirft natürlich die Frage auf, ob (und ggfs. welche) alternativen Erklärungen es für die wiederholten positiven Resultate in den kleineren Studien geben könnte.

2.5 Kritische Bemerkungen zu Metastudien im Allgemeinen

In der experimentellen Psychologie ist es nicht ungewöhnlich, dass Effekte, die in kleineren Studien gefunden wurden, in großen und gut kontrollierten Experimenten nicht repliziert werden können. Abgesehen von den bekannten Problemen, die im Rahmen der Replikationskrise wiederholt angeführt wurden (z. B., dass in der Regel nur signifikante Ergebnisse veröffentlicht werden und Studien auch nicht präregistriert werden müssen; vgl. Maxwell et al., 2015; Wiggins & Christopherson, 2019), gibt es gerade in jüngster Zeit vermehrt auch Arbeiten, die insbesondere Metaanalysen kritisch in den Blick nehmen (z. B. Bartoš et al., 2023; Kvarven et al., 2020; Sotola, 2022).

Sotola (2022) beispielsweise untersuchte im *Psychology Bulletin* veröffentlichte Metaanalysen und zeigte, dass das Risiko, dass eine solche Metaanalyse einen posi-

tiven Effekt bestätigt, obwohl in Wirklichkeit ein Nulleffekt vorliegt, immer dann besonders hoch ist, wenn die Anzahl der Versuchspersonen in den berücksichtigten Studien sehr klein ist. Sotola mutmaßt darüber hinaus, dass es für jede in einer Metaanalyse aufgenommene Studie eine weitere Studie geben muss, die nicht veröffentlicht wurde. Der Grund für die Nichtveröffentlichung könnte sein, dass diese geschätzten 50 Prozent aller *tatsächlich durchgeführten* Studien keinen oder einen negativen Effekt gefunden haben.

Während Kvarven et al. (2020) konkret davon ausgehen, dass in Metaanalysen berichtete Effektgrößen um den Faktor drei überschätzt sein könnten, gehen Bartoš et al. (2023) sogar von einer noch größeren Überschätzung aus. Die allgemeine Schlussfolgerung aller drei Beiträge lautet, dass a) die berichteten Effektstärken in der Regel wahrscheinlich überschätzt werden, und dass b) Metaanalysen nicht per se unzuverlässig sind, aber mit Vorsicht interpretiert werden sollten. Metaanalysen erlauben zwar die Synthese einer größeren Anzahl ähnlicher Studien und ermöglichen somit Generalisierungen, sind aber natürlich immer nur so gut wie die eingehenden Einzelstudien (was treffend aus dem Titel des Beitrags von Sotola (2022) deutlich wird: »Garbage in, garbage out«).

2.6 Alternative Erklärungsmöglichkeiten für Flipped-Classroom-Metaanalysen

Es gibt keine spezifischen Besonderheiten, die die Metaanalysen in Tabelle 1 von denen unterscheiden, die von Bartoš et al. (2023), Kvarven et al. (2020) und Sotola (2022) im Allgemeinen untersucht wurden. In Ermangelung gegenteiliger Argumente vermuten wir deshalb, dass die in diesen drei Arbeiten angestellten theoretischen Analysen auch auf die Meta-Ergebnisse in Tabelle 1 zutreffen könnten. Die Hypothese, dass diese Effektstärken ebenfalls überschätzt sein dürften, steht im Einklang mit der Übersichtsarbeit von Kapur et al. (2022), die ebenfalls nahelegt, dass speziell die Effektgrößen in Flipped-Classroom-Metaanalysen aufgrund der oftmals geringen Stichprobengröße der eingeschlossenen Studien und eines Publikationsbias überschätzt sein könnten (siehe auch Hew et al., 2021). Allerdings fokussieren sich Kapur et al. (2022) und Hew et al. (2021) auf das universitäre Lernen (90 % der z. B. von Kapur et al. (2022) berücksichtigten Studien stammen aus dem Hochschulbereich).

3. Fazit

Zunächst haben wir Metaanalysen zur Flipped-Classroom-Methode für den Primar- und Sekundarbereich vorgestellt, die scheinbar ein eindeutiges Bild zugunsten dieser Methode zeichnen (Tab. 1). Anschließend haben wir die individuellen Studien, die diesen Metaanalysen zugrunde liegen, einzeln und der Größe nach

aufgeführt. Dabei kann man feststellen, dass ein Großteil dieser Studien nur ein oder zwei Klassen pro Versuchsbedingung implementiert hat (vgl. Tab. 2).

Demgegenüber stehen mittlerweile zwei größere Flipped-Classroom-Studien mit durchschnittlich 10 bzw. 12 Klassen pro Versuchsbedingung (Wagner et al., 2021; Wiesner et al., eingereicht). Diese beiden Studien, die noch zu neu für die Metaanalysen in Tabelle 1 waren, haben auf der Basis von Mehrebenenmodellierungen *keine* generellen positiven Effekte für die Flipped-Classroom-Methode gefunden.

Anschließend wurde auf der Suche nach einer Aufklärung dieses Widerspruchs allgemein-psychologische Literatur zu Rate gezogen, die sich generell kritisch mit Metaanalysen auseinandersetzt (Bartoš et al., 2023; Kvarven et al., 2020; Sotola, 2022). In diesen Arbeiten wird ausgeführt, dass Effekte in Metaanalysen vor allem dann überschätzt sein dürften, wenn die eingehenden Studien auf kleinen Versuchspersonenanzahlen beruhen (was im Besonderen für Tabelle 1 zutrifft).

Wir können im Rahmen des vorliegenden Beitrags keine Aussage darüber machen, ob die Flipped-Classroom-Methode für das Lernen an der Universität geeignet ist (wozu es sogar noch deutlich mehr Metaanalysen als zum schulischen Unterricht gibt; vgl. z. B. Kapur et al., 2022; bzw. Hew et al., 2021), sondern haben universitäre Lehrformate bewusst vom allgemeinbildenden Unterricht an Schulen abgegrenzt. Da für die Organisation eines Hochschulstudiums ohnehin ein gewisses Maß an Selbstregulation erforderlich ist, könnte es durchaus plausibel sein, anzunehmen, dass ein »Flipped Hörsaal« vorteilhaft sein könnte.

Mit Blick auf Schüler*innen kann man dagegen berechtigterweise die Frage stellen, ob es ausreicht, sie zwar mit Erklärvideos zu versorgen, dann damit aber zuhause »alleine zu lassen«. Gerade jüngere Kinder sind vermutlich auf Interaktionen mit der Lehrkraft angewiesen, z. B. wenn sie etwas nicht verstanden haben und Rückfragen stellen möchten. Lehrkräfte können – im Gegensatz zu Erklärvideos – adaptiv auf solche Fragen eingehen, sie können Gestik und Mimik von Kindern interpretieren und sich bei Lern- und Verständnisschwierigkeiten aus ihrem fachdidaktischen Repertoire bedienen.

Interessanterweise unterstützt die Übersicht über Flipped-Classroom-Metaanalysen von Kapur et al. (2022) indirekt die Behauptung einer größeren Relevanz der Unterstützung der Selbstregulation der Lernenden, je jünger diese sind. Kapur et al. (2022) zeigen nämlich, dass auf Hochschulebene die durchschnittliche Effektgröße zugunsten der Flipped-Classroom-Methode (ohne spezifische Unterstützung der Selbstregulation und im Vergleich zum traditionellen Unterricht) am höchsten ist ($g = 0.93$), während der Effekt geringer wird, je jünger die Lernenden werden (high school: $g = 0.63$, elementary school: $g = 0.40$). Auch in der Studie von Wiesner et al. (eingereicht) mit Schüler*innen der sechsten bzw. siebten Jahrgangsstufe, war die Unterstützung der Selbstregulation (in Form von spezifisch auf Erklärvideos zugeschnittenen Lernstrategien) das entscheidende Element. Bei

alleiniger Bereitstellung von Erklärvideos war der Lernzuwachs im Flipped-Classroom-Unterricht unter allen drei implementierten Bedingungen am geringsten.

Solange jedoch Versuche scheitern, scheinbar positive Effekte zugunsten der Flipped-Classroom-Methode auch in kontrollierten »large-scale« Designs mit mehreren Klassen pro Versuchsbedingung nachzuweisen, stellt die scheinbar erdrückende Evidenz aus Tabelle 1 für die Institution Schule noch keinen Anlass zur Besorgnis dar. Insgesamt gewinnt man hierbei eher den Eindruck, dass die Forschung zu Flipped Classroom gerade erst am Anfang steht. In der Tat ist ein weiteres Ergebnis aus Tabelle 2, dass die diesbezügliche Wirksamkeitsforschung erst vor etwa 10 Jahren begonnen hat: Alle 51 Einzelstudien stammen aus dem Zeitraum von 2012–2021 (und die sieben Metaanalysen aus Tab. 1 sind konsequenterweise alle aus den letzten 6 Jahren). Wir stimmen mit Wagner et al. (2020, S. 14) jedenfalls überein, dass »in particular, more randomized controlled trials with larger sample sizes and objective quantitative measures are needed.«

Insbesondere ist noch offen, für welche Fächer und für welche Jahrgangsstufen, aber auch spezifischer für welche Unterrichtsthemen pro Fach, sich Flipped-Classroom-Unterricht im Besonderen anbieten könnte. Gerade in Bezug auf die Unterrichtsfächer kommen Metaanalysen derzeit noch zu unterschiedlichen Schlussfolgerungen. Während Wagner et al. (2020) auf Basis metaanalytischer Verfahren größere Effekte für MINT-Fächer finden, konstatieren Strelan et al. (2020) im Gegenteil gerade gesellschaftswissenschaftlichen Fächern ein größeres Potential für den Flipped-Classroom-Unterricht. Diesbezüglich kann mit Spannung auch auf die Ergebnisse der Studie FALKE-d (Frei et al., 2020) gewartet werden, von der Wiesner et al. (eingereicht) eine Teilstudie war und in der in heterogenen Unterrichtsfächern (Mathematik, Chemie, Musik, Deutsch, Sachunterricht in der Grundschule) auf Grundlage eines parallelen Designs und mit jeweils großen Stichproben ebenfalls die drei Bedingungen (»normaler« Flipped Classroom vs. Flipped Classroom mit zusätzlicher Unterstützung des selbstregulierten Lernens vs. Kontrollgruppe mit traditionellem Unterricht) implementiert wurden.

Weiterhin werden in Flipped-Classroom-Studien die Schüler*innen oftmals mit von den Forschenden aufwändig konzipierten Materialien für die Wissensaneignung zu Hause versorgt. Weitestgehend noch offen ist diesbezüglich, welche Effekte ökologisch valider Flipped-Classroom-Unterricht hat, d. h., wenn die Lehrkräfte im Unterrichtsalltag selbständig Materialien entwerfen oder Erklärvideos produzieren müssen.

Literatur

- Bartoš, F., Maier, M., Shanks, D. R., Stanley, T. D., Sladekova, M., & Wagenmakers, E.-J. (2023). Meta-analyses in psychology often overestimate evidence for and size of effects. *Royal Society Open Science*, 10(7), 230224. <https://doi.org/10.1098/rsos.230224>
- Bishop, J., & Verleger, M. (2013). *The Flipped Classroom: A survey of the research*. 2013 ASEE Annual Conference & Exposition, Atlanta, Georgia, United States.
- Cevikbas, M., & Kaiser, G. (2020). Flipped classroom as a reform-oriented approach to teaching mathematics. *ZDM*, 52(7), 1291–1305.
- Cevikbas, M., & Kaiser, G. (2023). Can flipped classroom pedagogy offer promising perspectives for mathematics education on pandemic-related issues? A systematic literature review. *ZDM*, 55(1), 177–191.
- Chen, K. S., Monrouxe, L., Lu, Y. H., Jenq, C. C., Chang, Y. J., Chang, Y. C., & Chai, P. Y. C. (2018). Academic outcomes of flipped classroom learning: A meta-analysis. *Medical education*, 52(9), 910–924.
- Cheng, L., Ritzhaupt, A. D., & Antonenko, P. (2019). Effects of the flipped classroom instructional strategy on students' learning outcomes: A meta-analysis. *Educational Technology Research and Development*, 67(4), 793–824.
- Dogan, Y., Batdi, V., & Yasar, M. D. (2023). Effectiveness of flipped classroom practices in teaching of science: A mixed research synthesis research. *Science & Technological Education*, 41(1), 393–421.
- Frei, M., Asen-Molz, K., Hilbert, S., Schilcher, A., & Krauss, S. (2020). *Die Wirksamkeit von Erklärvideos im Rahmen der Methode Flipped Classroom*. Waxmann.
- Güler, M., Kokoç, M., & Önder Büttüner, S. (2023). Does a flipped classroom model work in mathematics education? A meta-analysis. *Education and Information Technologies*, 28(1), 57–79.
- Hew, K. F., Bai, S., Dawson, P., & Lo, C. K. (2021). Meta-analyses of flipped classroom studies: A review of methodology. *Educational Research Review*, 33, 100393. <https://doi.org/10.1016/j.edurev.2021.100393>
- Hwang, G.-J., Yin, C., & Chu, H.-C. (2019). The era of flipped learning: promoting active learning and higher order thinking with innovative flipped learning strategies and supporting systems. *Interactive Learning Environments*, 27(8), 991–994.
- Jang, H. Y., & Kim, H. J. (2020). A meta-analysis of the cognitive, affective, and interpersonal outcomes of flipped classrooms in higher education. *Education Sciences*, 10(4), 115.
- Kapur, M., Hattie, J., Grossman, I., & Sinha, T. (2022). Fail, flip, fix, and feed – Rethinking flipped learning: A review of meta-analyses and a subsequent meta-analysis. *Frontiers in Education*, 7, Article 956416. <https://doi.org/10.3389/educ.2022.956416>
- Kunter, M., Baumert, J., Blum, W., Klusmann, U., Krauss, S., & Neubrand, M. (2011). *Professionelle Kompetenz von Lehrkräften. Ergebnisse des Forschungsprogramms COACTIV*. Waxmann.
- Kvarven, A., Strömland, E., & Johannesson, M. (2020). Comparing meta-analyses and pre-registered multiple-laboratory replication projects. *Nature Human Behaviour*, 4(4), 423–434.

- Låg, T., & Sæle, R. G. (2019). Does the flipped classroom improve student learning and satisfaction? A systematic review and meta-Analysis. *AERA Open*, 5(3), 233285841987048.
- Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *The Journal of Economic Education*, 31(1), 30–43.
- Maxwell, S. E., Lau, M. Y., & Howard, G. S. (2015). Is psychology suffering from a replication crisis? What does »failure to replicate« really mean? *American Psychologist*, 70(6), 487.
- Sotola, L. K. (2022). Garbage In, Garbage Out? Evaluating the evidentiary value of published meta-analyses using z-curve analysis. *Collabra: Psychology*, 8(1), Article 32571.
- Strelan, P., Osborn, A., & Palmer, E. (2020). The flipped classroom: A meta-analysis of effects on student performance across disciplines and education levels. *Educational Research Review*, 30, 100314.
- van Alten, D. C., Phielix, C., Janssen, J., & Kester, L. (2019). Effects of flipping the classroom on learning outcomes and satisfaction: A meta-analysis. *Educational Research Review*, 28, 100281.
- Wagner, M., Gegenfurtner, A., & Urhahne, D. (2020). Effectiveness of the flipped classroom on student achievement in secondary education: A meta-Analysis. *Zeitschrift Für Pädagogische Psychologie*, 35(1), 11–31.
- Wagner, M., & Urhahne, D. (2021). Disentangling the effects of flipped classroom instruction in EFL secondary education: When is it effective and for whom? *Learning and Instruction*, 75, 101490.
- Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology*, 39(4), 202.
- Wiesner, P., Krauss, S., Stegmüller, N., & Binder, K. (eingereicht). Is Flipped Classroom really superior? Questioning the flip in K-12 teaching.
- Zhu, G. (2021). Is flipping effective? A meta-analysis of the effect of flipped instruction on K-12 students' academic achievement. *Educational Technology Research and Development*, 69(2), 733–761.