

Eine Analyse der bei Hattie zitierten Klassenfrequenzstudien Forschungsergebnisse zum Thema Klassenfrequenz

Volker Hagemeister

Berliner Landesinstitut für Schule und Medien, Duden-Institut für Lerntherapie

Zusammenfassung

Hatties Aussage, dass sich mit der Einrichtung kleinerer Klassen kaum etwas ändert, ist das Resultat einer einseitigen Auslegung empirischer Studien: Hattie reduziert die Ergebnisse der Studien jeweils auf eine einzige Zahl. Deshalb wird in „Visible Learning« nicht vermittelt, dass in repräsentativ zusammengesetzten kleinen Klassen Kinder aus bildungsfernen Elternhäusern signifikant besser als in großen Klassen gefördert werden, ohne dass dabei die Kinder aus bildungsnahen Elternhäuser vernachlässigt wurden.

Schlüsselbegriffe:

Kritik an Hattie, Klassenfrequenz, Förderung von Minderheiten

An analysis of the class frequency studies cited by Hattie

Volker Hagemeister

Berlin State Institute for School and Media, Duden Institute for Learning Therapy

Abstract

Hattie's statement that the establishment of smaller classes hardly changes anything is the result of a one-sided interpretation of empirical studies: Hattie reduces the results of each study to a single number. Therefore, Hattie's "Visible Learning" does not convey that in representative small classes children from educationally disadvantaged parental homes are supported significantly better than in large classes without neglecting children from educationally advantaged parental homes.

Keywords:

Criticism of Hattie, class size, promotion of minorities

1. Einleitung

1.1 Die Hattie Studie von 2009

Grundlage für Hatties Arbeit »Visible Learning« waren „736 Meta-Analysen, ... die auf 50'000 Studien mit ca. 250 Millionen Lernenden zurückgreifen“ (Hattie, 2015, Vorwort, S. XI). Aus dieser gewaltigen Datenmenge hat Hattie „138 Faktoren extrahiert“, die für das Lernen in der Schule von Bedeutung sind. Die Quintessenz seiner Arbeit sind die unterschiedlichen Effektstärken, die Hattie für diese 138 Faktoren ermittelt hat.

In Anlehnung an Cohen (1988) definiert Hattie (2009, p. 8) die Effektstärke d :

$$\text{Effect size } d = \frac{\text{Mean}_{\text{treatment}} - \text{M}_{\text{control}}}{\text{SD}}$$

Dabei ist $\text{Mean}_{\text{treatment}}$ der Test-Mittelwert der Gruppe, die eine besondere Förderung erhalten hat, $\text{M}_{\text{control}}$ der Mittelwert einer Kontrollgruppe. SD steht für Standardabweichung. Wenn Vor- und Nachtests bei der gleichen Schülergruppe zum Einsatz gekommen sind, so ist

$$\text{Effect size } d = \frac{\text{Mean}_{\text{End of treatment}} - \text{M}_{\text{Beginning of treatment}}}{\text{SD}}$$

Im Abschnitt „Synthesizing the meta-analyses“ führt Hattie aus, dass normaler Unterricht pro Schuljahr Leistungsverbesserungen hervorbringt, die einer Effektstärke von $d = 0.20$ bis $d = 0.40$ entsprechen:

“Teachers typically can attain between $d = 0,20$ to $d = 0,40$ growth per year — and that this is to be considered average. They should be seeking greater than $d = 0.40$ for their achievement gains to be considered above average, and greater than $d = 0,60$ to be considered excellent ... we should judge the success of an innovation relative to $d = 0,40$ (and certainly not $d = 0.0$).” (Hattie, 2009, p. 17)

“Set the bar at $d = 0.40$... This average summarizes the typical effect of all possible influences in education and should be used as the bench-mark to judge effects in education. (Hattie, 2009, P. 16)

Faktoren, die nach Hattie als wenig effektiv eingestuft werden müssen, weil sie deutlich unterhalb seiner *bench-mark* liegen, sind z. B.:

Jahrgangübergreifende Klassen (Hattie, 2015, p. 88, Tabelle 04)	<i>Multi-grade/age classes</i> (Hattie, 2009, p. 74, Table 6.1)	$d = 0,04$
Interne Differenzierung (Hattie, 2015, p. 88, Tabelle 04)	<i>Within-class grouping</i> (Hattie, 2009, p. 74, Table 6.1)	$d = 0,16$
Verringerung der Größe von Klassen (Hattie, 2015, p. 103, Tabelle 05)	<i>Class size</i> (Hattie, 2009, p. 87, Table 6.2)	$d = 0,13$

Besonders wirksame Faktoren sind dagegen nach Hattie z. B.:

Rückmeldungen an die Lernenden (Hattie, 2015, p. 193, Tabelle 08)	<i>Feedback</i> (Hattie, 2009, p. 162, Table 9.1)	$d = 0,73$
<i>Formative Evaluation</i> (Hattie, 2015, p. 193, Tabelle 08)	<i>Providing formative evaluation</i> (Hattie, 2009, p. 162, Table 9.1)	$d = 0,9$
<i>Selbsteinschätzung des eigenen Leistungsvermögens</i> (Hattie, 2015, p. 47, Tab 02)	<i>Self-report grades</i> (Hattie, 2009, p. 39, Table 4.1)	$d = 1,44$

1.2 Weltweite Resonanz

Die in London herausgegebene Zeitschrift »Times Educational Supplement« hat „noch nie ... ein Buch mit den Worten »the holy grail« gepriesen“ (Hattie, 2015, Vorwort, S. VIII).

Weltweit hat John Hatties Buch »Visible Learning«, das 2009 in erster Auflage veröffentlicht wurde, eine ungewöhnlich große Resonanz hervorgerufen: „Alle zitieren Hattie; in den öffentlichen Bildungsblättern, in den wissenschaftlichen Fachjournalen, in den Institutionen der Lehrerbildung, in Lehrerzimmern und auf Elternabenden — überall wird auf die Hattie-Studie Bezug genommen.“ (Terhart, 2014, S. 6).

„Evidenzbasierung im schulischen Kontext wird durch die Arbeiten von John Hattie nochmals verstärkt. Gerade im Bereich von Schule und Unterricht, der enorm von subjektiv-normativen, nicht immer angemessenen, traditionellen und rituellen Verfahren dominiert zu sein scheint, ist dieser Blickwinkel wichtig.“ (Hattie, 2015, Vorwort, S. XXI)

1.3 Kritik an Hatties Studie »Visible Learning«

Hattie habe soziale Einflüsse nicht berücksichtigt: *Snook Social effects are ruled out*

„Social effects are ruled out (in Visible Learning) ... social class background is indeed more important than many of the issues discussed in (Hattie's) book and hence policy decisions cannot be drawn in isolation from the background variables of class, poverty, health in families and nutrition.“ (Snook et al., 2009, p. 94, 95)

Hatties Wirksamkeits-Grenze von $d = 0,4$ gilt nur eingeschränkt:

„Ein ... Effektmaß $d = 0$ bedeutet ... nicht, dass die ... Interventionsmaßnahme (beispielsweise »freie Arbeit«) »null Effekt« habe. Vielmehr zeigt das Maß $d = 0$ nur an, dass es zwischen beiden Gruppen (freie Arbeit und Kontrollgruppe) keinen Unterschied gibt. ... Da der übliche Unterricht einen Effekt erzielt, der zwischen $d = 0,2$ und $d = 0,4$ liegt, sind auch freie Arbeit¹ und offene Lehr- und Lernformen² nicht wirkungslos, sondern dürften ebenfalls in diesem Effektbereich liegen.“ (Steffens und Höfer, 2016, S. 149)

Bei einem Versuchsdesign mit Experimental- und Kontroll-Gruppen kann also auch $d = 0$ als Erfolg gewertet werden. Man kann dann davon ausgehen, dass pädagogische Maßnahmen, durch die z. B. das selbständige Arbeiten oder soziales Lernen gefördert werden sollen, nicht dazu führen, dass der Erwerb des messbaren, fachlichen Wissens vernachlässigt wird.

Werden Effektstärken dagegen aus Vor- und Nach-Tests berechnet, die bei der gleichen Gruppe eingesetzt werden, so ist Hatties Argumentation nachvollziehbar: Ist die gemessene Effektstärke kleiner als 0,4, dann ist zu befürchten, dass die neue Methode den Erwerb des fachlichen Wissens behindert hat, denn auch ohne besondere pädagogische Maßnahmen könnte die Effektstärke, die man aus Vor- und Nachtests ermittelt, bei $d = 0,4$ liegen.

1.4 Ziele dieser Analyse der Hattie-Studie

Angesicht der großen Resonanz und der Akzeptanz, die die Hattie-Studie erfahren hat, ist es für die Bildungspolitik, für die pädagogische Wissenschaft und für die tägliche Arbeit an unseren Schulen von erheblicher Bedeutung, wie verlässlich die Aussagen Hatties sind. Die Auseinandersetzung mit der Hattie-Studie wird im folgenden Text mit dem Abschnitt „Class

¹ »Student control over Learning« $d = 0,04$, Hattie 2009, Table 9.1, p. 162

² »Open vs. traditional« $d = 0,01$, Hattie, 2009, Table 6.1, p. 74

size“ begonnen, weil dieses Thema einerseits die alltägliche Arbeit von Lehrkräften in besonderem Maße berührt und weil andererseits die Reduzierung der Klassenstärke eine sehr kostspielige Reformmaßnahme ist:

„Hatties Befunde besagen: »Kleine Klassen bringen keinen größeren Lernerfolg!« Bei seinem Vortrag an der Universität Oldenburg am 17. April 2013 sagte Hattie, dass dieses Einzelergebnis aus seinen Metaanalysen immer wieder zum Anlass für heftige Attacken auf seine Studie genommen werde. Hatties Antwort: »Die Befunde sind aber leider so.« Es gibt keine nennenswerte Verbesserung der Lernleistungen, wenn die Klassenstärke von 25 auf 15 Schüler reduziert wird.“ (Meyer, 2014, S. 123)

Hat Hattie recht? Sind die Befunde so:

“reducing class sizes ... is ... certainly not worth the billions of dollars that is required to reduce the number of children per classroom.” (Hattie, 2009, p. 86)

1.5 Methoden dieser Analyse

In seiner Tabelle 6.2 hat Hattie eine Reihe bedeutenderer und neuerer Studien zum Thema Klassenfrequenz zusammengefasst (Hattie, 2009, p. 86, 87 / Hattie, 2015, p. 102, 103). Alle Artikel aus Hatties Tabelle 6.2 wurden für diese Analyse noch einmal durchgearbeitet³. Fragen waren dabei z. B.:

- In welcher Weise hat Hattie die Effektstärken berechnet, die er den einzelnen Studien in seiner Tabelle 6.2 zugeordnet hat? Wie hat Hattie soziale Einflüsse und andere Moderatorenvariable berücksichtigt?
- Hattie reduziert die Aussagen der zitierten Studien auf eine einzige Zahl: Auf die Werte für d in Spalte 7 seiner Tabelle 6.2. Wird Hattie damit den einzelnen Studien gerecht oder ist mit dieser Reduktion ein relevanter Verlust an Bedeutung verbunden?

2. Ergebnisse

2.1 Bildungspolitisch relevante Aussagen fehlen bei Hattie

In dem Hattie die Wirkung kleiner Klassen auf den Gesamt-Mittelwert $d = 0,13$ reduziert, wird der Blick auf wichtige bildungs- und gesellschaftspolitische Ergebnisse der Klassenfrequenzforschung verstellt:

- In repräsentativ zusammengesetzten kleinen Klassen werden Kinder aus bildungsfernen Elternhäusern signifikant besser als in großen Klassen gefördert, ohne dass dabei die Kinder aus bildungsnahen Elternhäusern vernachlässigt werden (siehe im Folgenden die Abschnitt 3.2, 3.3, 3.4 und 3.6). Der Aufenthalt in den per Zufallsverfahren zusammengestellten kleinen Klassen beim Tennessee STAR-Project führte dazu, dass die Kluft zwischen Schülern europäischer und afrikanischer Herkunft bei den Universitäts-

³ Dabei fehlt nur die Arbeit Urquiolas. Diese unveröffentlichte Dissertation ist bei deutschen Universitätsbibliotheken nicht ausleihbar.

Aufnahmetests um 60% sank, was Hattie keiner Erwähnung wert ist (siehe Abschnitt 3.2).

- Schüler aus kleinen Klassen entscheiden sich häufiger, länger zur Schule zu gehen (siehe Abschnitt 3.5).
- Schüler aus kleinen Klassen müssen seltener einen Kurs wiederholen, sie schwänzen seltener die Schule und es werden seltener Disziplinarstrafen gegen sie verhängt (was vor allem für Kinder aus bildungsfernen Elternhäusern von Bedeutung ist, siehe 3.2, 3.4 und 3.5).

2.2 Falsche Aussagen bei Hattie zum Thema „Erfahrungen im Klassenzimmer“

Berichte über Unterrichtsbeobachtungen gebe es kaum:

In „Visible Learning“ schreibt Hattie:

“It is difficult, however, to find studies that investigate or that demonstrate whether the nature of classroom experiences are different in the smaller than in the larger classes.”
(Hattie, 2009, p. 86; Hattie, 2015, p. 103)

Entgegen dieser Aussage werden allein in 6 der Artikel, die in Hatties Tabelle 6.2 enthalten sind, Unterrichtsbeobachtungen beschrieben. Die Unterschiede im Verhalten von Lehrern und Schülern in großen und kleinen Klassen sind der zentraler Gegenstand in Veröffentlichung von Smith und Glass (1980), von Blatchford (2005) und von Rice (1999). Ferner wird bei Finn (1998 und 2002), Blatchford et al. (2003) und Molnar (1999a und 1999b), über Ergebnisse systematischer Unterrichtsbeobachtungen berichtet (siehe hierzu die Abschnitte 3.2, 3.3, 3.4 und im Anhang 3.7 und 3.12).

2.3 Methodische Mängel bei Hattie

Die Beschreibung von Rechenwegen fehlt:

An keiner Stelle in Abschnitt „Class size“ macht Hattie Angaben dazu, wie er die Effektstärken, die er den einzelnen Studien in seiner Tabelle 6.2 zuordnet, berechnet hat. Solche Angaben sind für den wissenschaftlichen Dialog unverzichtbar. Auch in seinem Zeitschriftenaufsatz von 2005 zum Thema Klassenstärke nennt Hatties keine konkreten Quellen und man findet auch keine Rechenwege. So macht Hattie z. B. zu Molnars Bericht über das SAGE-Projekt nur vage Andeutungen (siehe hier 3.3):

“From the tables in their article, I estimated that the effect-sizes were 0.16 for Reading, 0.20 for Language, and 0.25 for Math” (Hattie, 2005, p. 392)

Auch wenn man alle Artikel, die in Hattie Tabelle 6.2 (2009, p. 87) enthalten sind, durchgearbeitet hat, findet man nur in wenigen Fällen Hinweise darauf, wie Hattie die Effektstärken berechnet haben könnte (siehe 3.2, 3.4 und im Anhang 3.8).

Alle Studien haben bei Hattie das gleiche Gewicht:

Bei Hattie (2009, S. 87) gehen in die Berechnung des Gesamt-Mittelwertes, der in der letzten Zeile seiner Tabelle 6.2 steht, alle Studien mit dem gleichen Gewicht ein, obwohl sich die Studien in Qualität und Quantität erheblich unterscheiden:

Die Anzahl der Testteilnehmer reicht von 3811 bei Dustmann et al. bis nahezu 900'000 bei Glass und Smith (siehe im Folgenden den Abschnitt 3.1).

Beim STAR-Project (hier Abschnitt 3.2) wurden vom Kindergarten bis zum Schulabschluss Daten erhoben, bei Goldstein et al. wird ein Projekt beschrieben, bei dem nach einem halben Jahr die abschließende Bewertung stattfand (siehe im Anhang Abschnitt 3.9)

In der Überschrift auf Seite 87 schreibt Hattie, dass die Tabelle 6.2 Effektstärken enthält, die die Reduzieren der Klassenfrequenz von 25 auf 15 Schüler hervorruft. Tatsächlich wird in den Studien, die in Tabelle 6.2 angeführt werden, mit sehr unterschiedlichen Klassenstärken gearbeitet. Z. B.: Beim STAR-Project sind in den großen Klassen im Mittel 25 Kinder und in den kleinen Klassen 15 Kinder. Bei Angrist und Lavy (1999) reichen die Klassenstärken von 21 bis 38 Kinder (siehe hier die Abschnitte 3,1, 3.2, 3.6 und im Anhang 3.8).

Hinzu kommen weitere Mängel: Hattie geht an keiner Stelle der Frage nach, ob repräsentative Gruppen miteinander verglichen wurden, wie bei den jeweils eingesetzten Tests die Validität zu bewerten ist, ob die Tests normiert waren und die Durchführung standardisiert war. Ebenso macht Hattie keine Angaben zum Studiendesign. Das ist deshalb von Bedeutung, weil Hatties „*bench-mark*“ $d = 0,4$ als untere Grenze bei Schulreformprojekten nur dann seine Berechtigung hat, wenn mit Vor- und Nachtests gearbeitet wurde. Werden dagegen wie beim STAR-Project Experimental-Gruppen mit Kontroll-Gruppen verglichen, liegt die *bench-mark* bei $d = 0,0$ (siehe 3.2).

2.4 Der Abschnitt »Class size« in »Visible Learning« enthält viele Flüchtigkeitsfehler

Literaturzitate sind falsch, z. B.:

- Bei Glass und Smith steht in Hatties Tabelle 6.2 die Jahreszahl „1997“. In Hatties Literaturverzeichnis sucht man jedoch vergeblich nach einer Veröffentlichung von Glass und Smith mit diesem Datum. Zutreffend sind hier wahrscheinlich anstelle von „1997“ die Jahreszahlen 1978 oder 1979 (siehe Abschnitt 3.1).
- Bei Finn ist die Jahreszahl „1988“ falsch. Stattdessen könnte 1998 oder 2002 in Spalte 2 von Tabelle 6.2 stehen (siehe 3.2).
- Bei Blatchford muss man „2005“ vermutlich durch die Jahreszahl 2003 ersetzen (siehe 3.4).

Angaben zur Zahl der Schüler und der Klassen enthalten Unstimmigkeiten:

Wegen Unstimmigkeiten bei Hatties Angaben zur Zahl der Schüler und der Klassen war es teilweise nicht möglich, zweifelsfrei zu klären, welche Studie Hattie jeweils ausgewertet hat, z. B.: Nach Hattie sollen 6'500 Schüler in 79 Klassen am STAR-Project teilgenommen haben. In Finns Veröffentlichungen zum STAR-Projekt sind es dagegen 12'000 Schüler in 329 Klassen (siehe 3.2).

Effektstärken sind mit Fehlern behaftet, z. B.:

- Bei Dustmann, Rajah und van Soest und bei Rice steht in Tabelle 6.2 für d der Wert $-0,04$. Eindeutig falsch ist hier das Vorzeichen (siehe 3.6 und im Anhang 3.12).

- Bei Angrist und Lavy müsste die Effektstärke ein negatives Vorzeichen haben, denn die besseren Testergebnisse wurden hier in den großen Klassen erzielt (siehe 3.6).
- Die Effektstärke $d = 0,09$, die Hattie Glass und Smith zuordnet, ist nicht vereinbar mit dem Wert $d = 0,24$, der eine Zeile darunter in Tabelle 6.2 bei Smith und Glass steht (siehe 3.1 und im Anhang 3.7).

Die deutschsprachigen Auflagen mit dem Titel „Lernen sichtbar machen“ von 2013 und 2015 enthalten im Abschnitt »Klassengröße« die gleichen Flüchtigkeitsfehler wie der Abschnitt »Class size« in »Visible Learning« aus dem Jahr 2009. Offenbar hat bisher niemand Hatties Abschnitt »Klassengröße« einer kritischen Analyse unterzogen, obwohl gerade Hatties Aussagen zum Thema Klassenfrequenz besonders oft zitiert werden.

2.5 Folgerungen

Bezogen auf den Abschnitt »Class size« in Hatties »Visible Learning« von 2009 muss man feststellen, dass es nicht sinnvoll ist, empirische Ergebnisse ganz unterschiedlicher Studien durch Mittelwertbildung ungewichtet zusammen zu fassen, ohne soziale Faktoren und andere Moderatorvariable zu berücksichtigen und ohne das jeweiligen Studien-Designs zu hinterfragen.

In der Einleitung zur deutschsprachigen Ausgabe von »Visible learning« schreiben Beywl und Zierer, dass es „nicht möglich war, die Berechnungen von Hattie zu jedem d -Wert zu überprüfen. Sollten hier Fehler vorliegen, so wurden sie von uns nicht entdeckt. Das Vertrauen in die Arbeit von Hattie war für dieses Vorgehen leitendes Motiv.“ (Beywl und Zierer in Hattie, 2015, S. XX, XXI). Die Analyse des Abschnittes »Class size« bei Hatties zeigt, dass man bei weiteren Auflagen von Hatties Werk nicht darauf vertrauen darf, dass Hattie korrekt gearbeitet hat.

3. Hatties Aussagen zum Thema Klassenfrequenz

In Tabelle 6.2 hat Hattie einige maßgebliche Studien zum Thema Klassenfrequenz aufgelistet. Die Effektstärken, die Hattie diesen Studien in Spalte 7 zuordnet, liegen alle deutlich unter seiner „bench-mark“ von $d = 0,4$:

Table 6.2: Synthesis of meta-analyses and major studies reducing class size from 25 to 15 (Hattie, 2009, p. 87)

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Glass & Smith	1997	77	725	14'358	520'899	0,09	Achievement
Smith & Glass	1980	59	371			0,24	Non-achievement outcomes
Finn	1988	1	1	79	6'500	0,22	Achievement
		1	1	79	6'500	0,12	Achievement (grades 4-6)
		1	1	79	6'500	0,02	Self-concept, Motivation
McGiverin et al.	1989	10	24			0,34	Achievement
Molnar et al.	1999	1	1	411	9'790	0,21	Achievement
Hoxby	2000	1	1	14'593	306'453	0,03	Achievement
Blatchford, P.	2005	1	1	368	9'330	0,23	Achievement

Goldstein et al.	2000	9	36	1'178*	29'440	0,20	Achievement
Dustmann, Rajah, & van Soest	2003	1	1	224	3'811	- 0,04	Achievement
Akerhielm	1995	1	1	1'052*	24'000	0,15	Achievement
Rice	1999	1	1	8'760	24'599	- 0,04	Achievement
Johnson et al.	2003	1	1	168*	3'700	0,00	Achievement
Angrist & Lavy	1999	1	1	1'327	46'455*	0,15	Achievement
Urquiola	2000	1	1	608	10'018	0,20	Achievement
Average		164	1'165	40'728+	948'540+	0,13	

* = estimated. 1989,

Sechs der Artikel, die in Tabelle 6.2 (2009) enthalten sind, werden nun nacheinander daraufhin untersucht, ob sie von Hattie zutreffend ausgewertet wurden. Weitere 7 der 14 Studien aus Tabelle 6.2 werden im Anhang besprochen. Die Artikel, die in den Anhang aufgenommen wurden, sind nicht etwa unwichtig, es ist vielmehr so, dass die im Anhang diskutierten Studien keine wesentlich neuen Aspekte enthalten gegenüber den 6 Studien, die im Folgenden analysiert werden.

3.1 Glass & Smith, Titel Nr. 1 in Hatties Tabelle 6.2

Bei Glass und Smith stehen in Hatties Tabelle 6.2 folgende Zahlen:

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Glass & Smith	1997	77	725	14'358	520'899	0,09	Achievement

Im Literaturverzeichnis zu Hatties »Visible Learning« gibt es keine Veröffentlichung von Glass und Smith aus dem Jahr 1997. Anstelle dieser Jahreszahl könnten in Hatties Tabelle 8.2 bei Glass und Smith die Jahreszahlen 1978 oder 1979 stehen, denn im Jahr 1979 wurde die Meta-Analyse von Glass und Smith zum Thema Klassenfrequenz in der Zeitschrift „Educational Evaluation and Policy Analysis“ veröffentlicht. Eine Vorab-Veröffentlichung gab es 1978. Zu ihrer Datenbasis schreiben Glass und Smith:

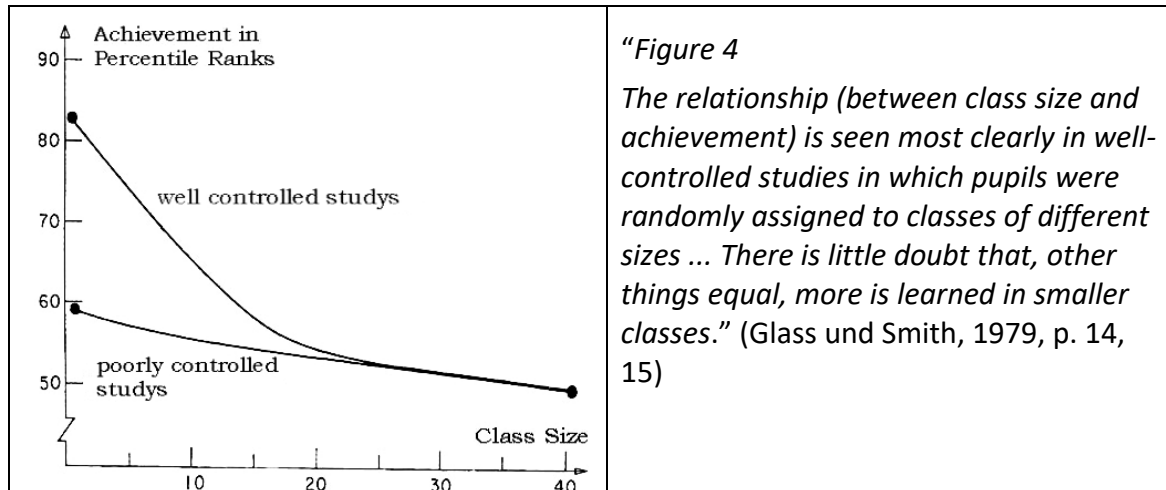
“In all, 77 different studies were read, coded, and analyzed. These studies yielded a total of 725 Δ's. The comparisons are based on data from a total of nearly 900'000 pupils ...” (Glass und Smith, 1979, p. 8)

Während Hattie für die „No. of students“ die Zahl 520'899 nennt (siehe oben), sprechen Glass und Smith von „nearly 900'000 pupils“. Weder in Hatties Zeitschriftenartikel von 2005 zum Thema „Klassenfrequenz“ noch in „Visible Learning“ von 2009 findet man Hinweise dazu, woher Hattie die Zahl 520'899 genommen hat. - Hatte Hattie Zusatzinformationen von Glass und Smith erhalten oder handelt es sich hier einfach nur um Druckfehler oder um Rechenfehler?

Insbesondere stellt sich hier die Frage, ob es sinnvoll ist, dass Hattie die Ergebnisse der Studie von Glass und Smith auf eine einzige Zahl, auf die Effektstärke $d = 0,09$ reduziert. – Ist eine solche Reduktion ein integraler Bestandteil von Meta-Analysen?

Drei Jahre vor der Veröffentlichung der Meta-Analyse von Glass und Smith zum Thema Klassenfrequenz, hatte Glass (1976) den Begriff „Meta-Analyse“ in die sozialwissenschaftliche Forschung eingeführt. Anders als Hattie versteht Glass unter „Meta-Analyse“ nicht, dass man,

ohne Moderatorenvariable zu berücksichtigen, die Ergebnisse einer größeren Anzahl von Studien per Mittelwertbildung auf eine einzige Zahl reduziert. Vielmehr werden bei Glass (1977) Verfahren, wie Moderatoren berücksichtigt werden können, ausführlich diskutiert (z. B. „Tukey's jackknife method“, Glass, 1977, p. 376). Entsprechend wird in der Meta-Analyse zum Thema Klassenfrequenz bei Glass und Smith (1979) der Einfluss verschiedener Moderatoren auf die Testergebnisse bei Klassenfrequenzstudien beschrieben und graphisch dargestellt. Dazu ein Beispiel aus der Meta-Analyse von Glass und Smith, die 1979 veröffentlicht wurde:



Aus der Meta-Analyse von Glass und Smith (1979) erhält man relevante Hinweise, die bei der Planung und Auswertung zukünftiger Klassenfrequenzstudien berücksichtigt werden sollten:

- Klassenfrequenz-Effekte sind bereichsabhängig: Die Reduktion der Frequenz von 40 auf 30 hat andere Ergebnisse zur Folge als die Absenkung der Frequenz von 25 auf 15.
- Bei Klassenfrequenzstudien muss die Zusammensetzung der Schülergruppen kontrolliert werden. Es ist nicht sinnvoll, kleine Schulklassen, die für Problemschüler zusammengestellt wurden, mit großen Klassen zu vergleichen, die nur wenige oder keine Problemschüler enthalten.

Solche wichtigen Aussagen werden von Hattie nicht aufgegriffen. In »Visible Learning« werden die Ergebnisse der umfangreichen Meta-Analyse von Glass und Smiths auf den Effektstärken-Mittelwert $d = 0.09$ reduziert. Wie Hattie diesen Wert berechnet hat, ist nicht nachvollziehbar. Zu dem minimalen Wert $d = 0.09$ passt nicht, was Glass und Smith in der Zusammenfassung der Ergebnisse ihrer Meta-Analyse zum Thema „class size“ schreiben:

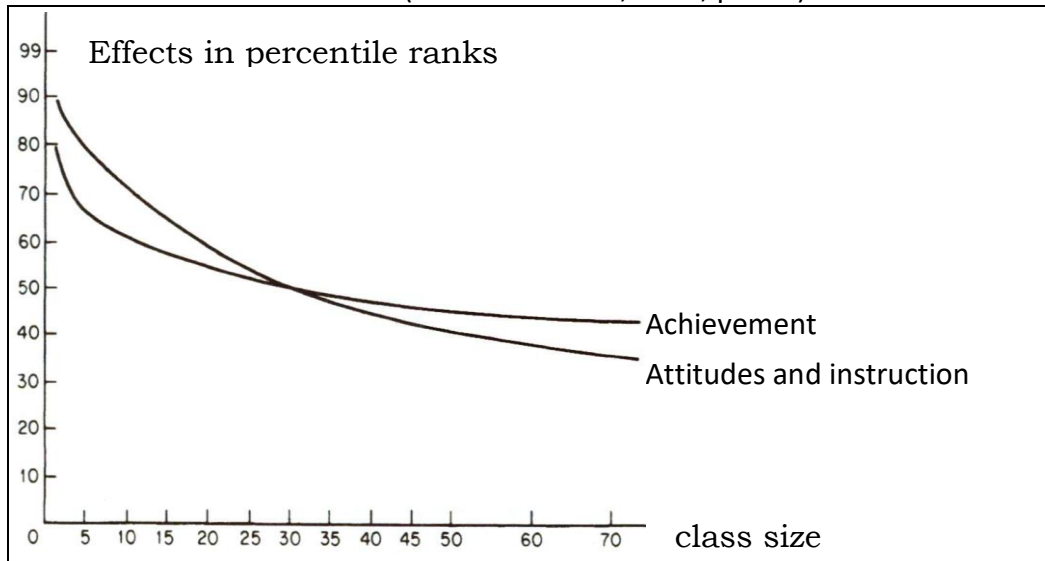
“A clear and strong relationship between class size and achievement has emerged. The relationship seems slightly stronger at the secondary grades than the elementary grades; but it does not differ appreciably ... The relationship is seen most clearly in well-controlled studies in which pupils were randomly assigned to classes of different sizes.”
 (Glass und Smith, 1979, p. 15)

Auch in seinem Aufsatz von 2005 erklärt Hattie nicht, wie er den Effektstärken-Mittelwert $d = 0.09$ berechnet hat, den er Glass und Smith in Tabelle 6.2 zuordnet. Denkbar ist, dass Hattie die Ergebnisse von „well controlled“ und „poorly controlled studys“ zusammengefasst hat. Bei „poorly controlled studys“ kann es durchaus vorkommen, dass in großen Klassen bessere Testergebnisse erzielt werden als in kleinen Klassen (siehe Abschnitt 3.6). Die Effektstärke hat dann ein negatives Vorzeichen, womit sich beim Mitteln über „well controlled“ und „poorly controlled studys“ Effektstärken nahe Null errechnen lassen.

Die Effektstärke $d = 0,09$, die Hattie Glass und Smith zuordnet, ist nicht vereinbar mit dem Wert $d = 0,24$, der eine Zeile darunter in Tabelle 6.2 bei Smith und Glass steht. Die Effektstärken für „achievement“ und „attitudes and instruction“, die Smith und Glass in Ihren Meta-Analysen zusammengetragen haben, unterscheiden sich nicht um einen Faktor 2,6 (gleich $0,24/0,09$), sondern liegen nahe beieinander (siehe Figure 2):

Figure 2

Graph of the ... relationship between class size and achievement ... and attitudes and instruction (Smith und Glass, 1980, p. 427)



Der nächste Artikel aus Hatties Tabelle 6.2, der hier besprochen wird, ist keine Meta-Analyse: Das STAR-Projekt, über das Finn und andere Autoren (z. B. Achilles et al. 1990) mehrfach berichtet haben, ist eine sehr umfangreiche, „well controlled“ Klassenfrequenzstudie:

3.2 Finn, Titel Nr. 3 in Hatties Tabelle 6.2

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Finn	1988	1	1	79	6'500	0,22	Achievement
		1	1	79	6'500	0,12	Achievement (Kl. 4-6)
		1	1	79	6'500	0,02	Self-concept, Motivation

Finn hat eine ganze Reihe von Artikeln zur Durchführung und Auswertung des Tennessee-STAR-Projects^[4] verfasst. Allerdings ist keine Veröffentlichung aus dem Jahre 1988 darunter. Immerhin findet man im Literaturverzeichnis zu »Visible Learning« (Hattie, 2009) zwei Artikel von Finn, die Hattie zur Berechnung seiner Effektstärken in Tabelle 6.2 herangezogen haben könnte:

- Finn (1998): Class size and students at risk. What is known? What is next?
- Finn (2002): Class-size reduction in grade k-3.

4] STAR: Student Teacher Achievement Ratio

Nicht nur die Jahreszahl, die bei Finn in Tabelle 6.2 steht, ist falsch, sondern auch die die Anzahl der Klassen und der Kinder in Spalte 5 und 6: Würde man 6'500 Kinder auf 79 Klassen verteilen, so läge die Klassenfrequenz im Mittel bei 82 Kindern. – Dagegen findet man zum STAR-Project bei Finn und Achilles (1999, p. 98) folgende Zahlen:

“More than 6'000 students in 329 classrooms (representing 79 schools and 46 districts) participated in the first year, and almost 12'000 students were involved in the course of the 4-year intervention.” (Finn und Achilles, 1999, p. 98)

Das STAR-Project begann im Jahre 1985 mit gut 7000 Kindern. Im letzten Vorschuljahr und im 1. bis 3. Schuljahr wurden Klassen mit 13 bis 17 Kindern und zur Kontrolle Klassen mit 22 bis 26 Kindern nach Zufallsregeln zusammengestellt (Finn, 1998, p. 6-7). - Ebenso wurden die Lehrer im jährlichen Wechsel per Zufallsverfahren den Schulklassen zugeteilt (vgl. Finn, 1998, p. 6-7). Ab Jahrgang 4 wurden alle Kinder auf normal-große Klassen verteilt. Es wurden aber weiterhin Leistungen gemessen und Arbeitshaltung, Schulkarrieren und schließlich die Ergebnisse in den Universitätsaufnahmetests erfasst.

Finns Tabelle 1 enthält die Testergebnisse aus dem Vorschuljahr (K) und den ersten drei Schuljahren beim STAR-Project:

“Table 1 Small-class effect sizes” (Finn, 1998, p. 9)

Scale	Group	K	1	2	3
Total Reading	W	--	.17	.13	.17
	M		.37	.33	.40
	ALL	.18	.24	.23	.26
Total Mathematics	W	.17	.22	.12	.16
	M	.08	.31	.35	.30
	ALL	.15	.27	.20	.23

W: white students, M: minority students (vgl. Finn, 1998, p. 8, 9)

Die Effektstärken in Finns Tabelle 1 spiegeln die Lernfortschritte der kleinen gegenüber den normal großen Klassen wider:

“Each effect size is the mean score for small classes minus the mean score of regular and teacher-aide classes $[S - (R+A)/2]$ in standard deviation units.” (Finn, 1998, p. 8)

Mit Hilfe von Finns Tabelle 1 aus dem Jahr 1998 hat Hattie wahrscheinlich den Wert $d = 0,22$ berechnet, der in Hatties Tabelle 6.2 in Zeile 4 steht. Addiert man in Finns Tabelle 1 alle Zahlen, die bei „All“ stehen und teilt diese Summe durch 8, so erhält man:

$$(0,18 + 0,24 + 0,23 + 0,26 + 0,15 + 0,27 + 0,20 + 0,23) \div 8 = 0,22$$

Da Hattie nur die Zahl 0,22 in seine Tabelle 6.2 aufnimmt, muss man annehmen, dass er die besondere Förderung, die Minderheiten-Kinder in kleinen Klassen erfahren, als unwichtig eingestuft hatte. Hier wird deutlich, wie sehr die „Evidenzbasierung“ in Visible Learning von subjektiven Entscheidungen Hatties beeinflusst wird.

Aus Tabelle 1 bei Finn (1998, S. 9) kann man ablesen, dass die Lernfortschritte der Minderheiten-Schüler, die die kleinen Klassen besucht haben, deutlich größer sind als die Lernfortschritte der weißen Schüler. Hinzu kommt, dass die Lernfortschritte der weißen Schüler in den kleinen Klassen größer sind als die der Kontrollgruppen. Ein Ergebnis der STAR-Projects ist also: In den repräsentativ zusammengesetzten kleinen Klassen werden Kinder aus bildungsfernen Elternhäusern besonders gut gefördert, ohne dass Kinder aus bildungsnahen Elternhäusern dabei benachteiligt werden.

Bei der Bewertung der Zahlen in Finns Tabelle 1 (1998, p. 9) muss man berücksichtigen, dass die Effektstärken bei Finn Aussagen darüber liefern, in wie weit die Lernerfolge in den kleinen Klassen die Ergebnisse der großen Klassen übertreffen. Greift man dazu Hatties Argument auf, dass der übliche Unterricht auch ohne besondere pädagogische Maßnahmen pro Jahr Lernfortschritte erzeugt, die Effektstärken von 0,2 bis 0,4 entsprechen, dann ergibt sich daraus, dass man für „*minority students*“, die repräsentativ zusammengesetzte, kleine Klassen besuchen, Effektstärken in der Größenordnung von $d = 0,6$ bis $d = 0,8$ erwarten kann, wenn mit Vor- und Nachtests und nicht mit Kontrollgruppen bearbeitet wird.

Nun zur nächsten Zeile in Hatties Tabelle 6.2 (2009, p. 87): Hattie ordnet der Wirkung kleiner Klassen beim STAR-Project für die Jahrgänge 4 bis 6 die Effektstärke $d = 0,12$ zu. Wie Hattie diesen Wert berechnet hat, erfährt man weder in »Visible Learning« von 2009, noch in Hatties Zeitschriftenartikel von 2005. Hier schreibt Hattie:

“After the students returned to regular-sized classes, the effects began to decrease ... (Year 4 = 0.12, Year 5 = 0.22, Year 6 = 0.17 and Year 8 = 0.13)” (Hattie, 2005, p. 390).

Da Hattie keine Quelle nennt, ist nicht nachvollziehbar, wie er zu diesen Ergebnissen gekommen ist. Entgegen Hatties Aussage kann man aus Tabelle 4 bei Molnar (1999b) ablesen, dass der Vorsprung der Kinder, die beim STAR-Project bis einschließlich 3. Jahrgang den kleinen Klassen angehört hatten, vom 4. bis zum 8. Jahrgang nicht abnimmt, sondern deutlich zunimmt:

Tabelle 4
(vgl. Molnar, 1999b, p. 31)

	Klasse 4	Klasse 6	Klasse 8
Mathematik	5.9 Monate	8.4 Monate	13 Monate
Muttersprache	9.1 Monate	9.2 Monate	14 Monate
Naturwissenschaften	7.6 Monate	6.7 Monate	13 Monate

Vorsprung der Kinder, die beim STAR-Project bis einschließlich 3. Jahrgang den kleinen Klassen angehört hatten (in „Grade Equivalent“^[5])

Die Kinder, die beim STAR-Project vier Jahr hindurch von der Vorschule bis in den 3. Jahrgang kleine Klassen besucht hatten, sind nachhaltig gefördert worden. Allerdings werden signifikante Effekte der Klassenreduzierung langfristig nur dann erzielt, wenn die Lernenden mindestens 3 Jahre hindurch den kleinen Gruppen angehört hatten:

“Pupils who attended small classes for at least three years had significant sustained benefits through Grade 8” (Finn, 2002, p. 2.13).

Diese besondere Förderung war sogar nach Abschluss der Schule noch wirksam:

Bei den Universitäts-Aufnahmetests (SAT und ACT⁶) waren die STAR-Schüler aus den kleinen Klassen erfolgreicher als die Schüler, die den Kontrollgruppen mit 22 bis 26 Schülern

5 Ein Grade Equivalent (GE) von 13 Monaten besagt nicht, dass ein 8-Klässler auch in der 9. Klasse dem Unterricht folgen könnte. Ein GE von 13 Monaten besagt, dass eine repräsentative Schülerstichprobe, die die Schule 13 Monate länger besucht hat, in dem selben normierten Test im Mittel die gleichen Ergebnisse erzielt wie ein 8-Klässler, dem ein GE von 13 Monaten bescheinigt wird (vgl. Finn und Achilles, 1999, p. 101).

6] SAT und ACT sind Aufnahme-Tests, die an Universitäten in den USA eingesetzt werden:
SAT: Scholastic Aptitude Test; ACT: American College Testing

angehört hatten. Signifikante Verbesserungen wurden wiederum bei den Schülern afrikanischer Herkunft registriert:

„Black students were substantially more likely to take the SAT or ACT if they were assigned to a small rather than regular-size class: 41.3 percent of black students assigned to small classes took at least one of the college entrance exams, compared with 31.8 percent in regular classes ... The chance of such a large difference in test-taking rates between the small and regular class students occurring by chance is less than one in 10'000.“ (Krueger und Whitmore, 2001, p. 19)

“The impact on minority students was particularly strong, thus reducing by 60% the gap in SAT / ACT rates between black students and white students.“ (Finn, 2002, p. 2.4)

Die signifikanten Effekte der Förderung von Minderheiten sind Hattie in „Visible Learning« keiner Erwähnung wert. Stattdessen fragt Hattie, warum seine Meta-Analyse so geringe Effektstärken für kleine Klassen ergeben hat:

“Why are the effect sizes from reducing class size so small?“ (Hattie, 2009, p. 86)

Hatties Antwort ist:

„... teachers of smaller classes adopting the same teaching methods as they were using in larger classes and thus not optimizing the opportunities presented by having fewer students (Finn, 2002).“ (Hattie, 2009, p. 86)

Zu Unrecht verweist Hattie hier auf Finn, denn im Gegensatz zu Hattie geht Finn der Frage nach, warum kleine Klassen beim STAR-Project so wirksam waren:

“Why is it that small classes work as well as they do?“ (Finn, 2002, p. 2.17)

Antworten Finns sind:

- *“The strongest hypothesis about why small classes work concerns students’ classroom behavior. Evidence is mounting that students in small classes are more engaged in learning activities and exhibit less disruptive behavior.“* (Finn, 2002, p. 2.17)
- *“In-grade retentions are reduced (in small classes). Because retained students are disproportionately minority, male, and from low-income homes, the reduction in retentions also reduces the achievement gap in schooling.“* (Finn, 2002, p. 2.5)

Viele der im deutschsprachigen Raum durchgeführten Studien zum Thema Klassenfrequenz haben bisher deshalb keine eindeutigen Ergebnisse hervorgebracht, weil folgende Bedingungen nicht erfüllt waren:

- Die Schülergruppen waren nicht repräsentativ zusammengesetzt (siehe hier Abschnitt 3.1, 3.2, 3.6 und im Anhang 3.9 und 3.11).
- Die kleinen Klassen haben nicht mindestens 3 Jahre hindurch bestanden (siehe Abschnitt 3.2)

Die Arbeit von Molnar et al., die im folgenden Abschnitt besprochen wird, ist wiederum keine Meta-Analyse. Beschrieben werden vielmehr Ergebnisse des „SAGE“ Programms, das in den 1990er Jahren im US-Bundesstaat Wisconsin durchgeführt worden war. Bei diesem Project wurden die Klassen nicht per Zufallsverfahren zusammengestellt (wie beim STAR-Project). Zusätzlich wurden für den Unterricht in den kleinen Klassen extra Curricula und Lehrertrainingsprogramme entwickelt:

3.3 Molnar et al., Titel Nr. 5 in Hatties Tabelle 6.2

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Molnar et al.	1999	1	1	411	9'790	0,21	Achievement

Im Rahmen des SAGE-Programms^[7], über das Molnar et al. (1999a) berichten, wurden an ausgewählten Schulen im US-Bundesstaat Wisconsin Klassen mit 15 Kindern eingerichtet. Die Voraussetzung für die Teilnahme am SAGE-Programm war, dass mindestens 30% der Kinder einer Schulklasse aus finanziell schlecht gestellten Elternhäusern kommen.

Wie beim STAR-Project hat die Analyse der Testergebnisse ergeben, dass von der Einrichtung kleiner Klassen insbesondere Kinder aus bildungsfernen Elternhäusern profitieren. Die Kinder afrikanischer Herkunft, die die kleinen Klassen der zwei SAGE-Jahrgänge besucht haben, erzielen signifikant bessere Nach-Test-Ergebnisse als die Vergleichsgruppen, obwohl die entsprechenden SAGE-Gruppen in den Vortests schlechter abgeschnitten hatten:

“African-American Students' First-Grade Scores, by SAGE or Comparison School”
(aus Tabelle 10 in Molnar et al., 1999a, p. 171)

Scale score	1996-1997		1997-1998	
	SAGE	Comparison	SAGE	Comparison
Mathematics				
Mean pretest	468,89	474,54	472,95	478,72
Mean posttest	524,23	517,03	522,01	506,22

Hattie ordnet in seiner Tabelle 6.2 dem SAGE-Projekt die Effektstärke $d = 0,21$ zu. Wie Hattie diesen Wert berechnet hat, wird nicht erläutert. Wiederum lässt dieser Mittelwert nicht mehr erkennen, dass Kinder afrikanischer Herkunft in den kleinen Klassen des SAGE-Projekts signifikant gefördert wurden. – Auf dieses Ergebnis beim SAGE-Projekt weist Hattie zwar in seinem Zeitschriftenartikel von 2005 zwar hin:

“The effect size was much higher for African-American students.” (Hattie, 2005, p. 392).

Jedoch wird in »Visible Learning« von 2009 im Abschnitt »Class size« (auf den Seiten 85 - 88) die besondere Förderung benachteiligter Kinder nicht erwähnt.

Nicht thematisiert wird ferner bei Hattie, dass im Rahmen des SAGE-Projekts auch Unterricht beobachtet wurde:

“When there are fewer students in a class, teachers develop greater knowledge and understanding of each one, ... students also come to know each other better and are more willing to share their thoughts and problems with the class.” (Molnar et al., 1999a, p. 174)

Blatchford, den Hattie an 7. Stelle anführt, berichtet über ein Klassenfrequenzprojekt, das in England durchgeführt wurde. Die Klassen wurden nicht per Zufallsverfahren zusammengestellt. Aber auch bei diesem Projekt wurde Unterricht in großen und kleinen Klasse systematisch beobachtet:

7] SAGE: Student Achievement Guarantee in Education

3.4 Blatchford, Titel Nr. 7 in Hatties Tabelle 6.2:

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Blatchford, P.	2005	1	1	368	9'330	0,23	Achievement

Wiederum wird in Hatties Tabelle 6.2 eine falsche Jahreszahl genannt: Resultate von Leistungstests sind in Blatchfords Texten aus dem Jahr 2005 nicht enthalten. Im Jahr 2005 berichten Blatchford und Kollegen ausschließlich über Methoden und Ergebnisse der systematischen Unterrichts-Beobachtung (siehe Blatchford, Bassett und Brown, 2005).

Passend wäre in Hatties Tabelle bei Blatchford die Jahreszahl 2003, denn 2003 berichten Blatchford et al. über Ergebnisse von Leistungstests: In ersten bis dritten Klassen wurden über drei Jahre hinweg an Schulen in England Lernfortschritte gemessen. An den Schulen waren Klassen mit 20 und mit 30 Kindern eingerichtet worden.

Bei der Auswertung der Testergebnisse wurden die Schüler in den großen und den kleinen Klassen jeweils in drei Leistungsgruppen unterteilt:

“Pupils were split into three ability groups, based on their pre-reception year literacy scores (bottom 25%, middle 50% and top 25%). There was a strong and statistically significant increase in attainment for all three groups, though there was a larger effect for pupils with lower baseline attainment.” (Blatchford et al., 2003, p. 17)

Die Lernfortschritte der drei Leistungsgruppen aus den kleinen Klassen übertrafen die Ergebnisse der Schüler aus den großen Klassen um folgende Werte (in Standardabweichungen, SD):

untere Leistungsbereich (25 %)	verbessert sich um	0,35 SD
mittlerer Leistungsbereich (50 %)	verbessert sich um	0,20 SD
oberer Leistungsbereich (25 %)	verbessert sich um	0,15 SD

(Daten aus Blatchford et al., 2003, p. 17)

Bildet man den Mittelwert aus den Zahlen in Spalte 3 so erhält man (ungewichtet):

$$(0,35 + 0,2 + 0,15) \div 3 = 0,23$$

Dieser Wert steht in Hatties Tabelle 6.2 bei Blatchford. In dem Hattie die empirischen Resultate Blatchfords et al. (2003) auf *eine einzige Zahl* reduziert, wird wiederum der Blick auf pädagogisch bedeutsame Ergebnisse verstellt, über die auch schon beim STAR- und beim SAGE-Project berichtet wurde: Vor allem Kinder aus bildungsfernen Elternhäusern werden in kleinen Klassen besser gefördert, ohne dass die leistungsstarken Kinder dabei benachteiligt werden.

Ergebnisse von Unterrichtsbeobachtungen bei Blatchford et al. sind:

„Results from the systematic observations showed that children in large classes were twice as likely to be off task ... especially in the passive form of being disengaged from allocated work ... the larger the class the more distracted the children. ...

These results offer support for the view of Finn and Achilles (1999) that one of the main effects of class size is on children’s attentiveness in class.” (Blatchford et al., 2003, p. 13, 14)

“Pupils in small classes were more likely to initiate interactions, respond to the teacher, or sustain dialogue with the teacher into the next time interval. This result is consistent with results on much younger children. Conversely, in larger classes there is more listening ... to the teacher.” (Blatchford et al., 2005, p. 34)

Auch Dustmann, Rajah und van Soest, die Hattie an 9. Stelle anführt, berichten über Klassenfrequenzprojekte, die in England stattgefunden haben. Bei diesem Projekt wurden die Klassen wiederum nicht per Zufallsverfahren zusammengestellt. Trotzdem hat diese Studie bildungspolitisch bedeutsame Ergebnisse hervorgebracht:

3.5 Dustmann, Rajah & van Soest, Titel Nr.9 in Hatties Tabelle 6.2:

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Dustmann, Rajah, & van Soest	2003	1	1	224	3'811	- 0,04	Achievement

Dustmann et al. haben die Daten von ca. 4000 Engländern und Walisern ausgewertet, die über 4 Jahrzehnte hinweg an dem National Child Development Survey (NCDS) teilgenommen hatten.

In Tabelle 6.2 (Hattie, 2009) wird der Veröffentlichung von Dustmann et al. (2003) die Effektstärke $d = -0,04$ zugeordnet. – Eindeutig falsch ist in diesem Falle das Vorzeichen: Bei Hattie stehen in Tabelle 6.2 Effektstärken mit positivem Vorzeichen, wenn die Leistungen der Kinder bei Abnahme der Klassenfrequenz zunehmen. Bei Dustmann ist es umgekehrt, denn er korreliert die Anzahl der Kinder pro Klasse unmittelbar mit Testpunkten. Deshalb haben bei Dustmann die Effektstärken (bzw. die Korrelationskoeffizienten) ein negatives Vorzeichen, wenn in kleineren Klassen bessere Leistungen erzielt werden:

“class size has a significant and sizeable negative effect on exam results: An increase in class size by one standard deviation decreases the number of O levels achieved by about 0.6” (Dustmann et al., 2002, p. 19).

Dementsprechend hätte in Hatties Tabelle 6.2 bei Dustmann et al. (2003) anstelle von $d = -0,04$ ein deutlich größerer (positiver) Wert für d stehen müssen.

Unabhängig von der Frage, welche Effektstärke dieser Studie zugeordnet wird, muss man feststellen, dass die Analyse der Daten bei Dustmann et al. pädagogisch bedeutsame Ergebnisse hervorgebracht hat:

- “We find a significant ... effect of class size on the probability to stay at school at age 16.” (Dustmann et al., 2002, p. 4)
- “a reduction in class size by 5 pupils would increase life cycle earnings by ... 2930 Pounds.” (Dustmann et al., 2002, p. 30)
- “We find a positive and significant association between class size and truancy boys.” (Dustmann et al., 2002, p. 5)

In folgenden Abschnitt wird die Arbeit von Angrist und Lavy besprochen. Berichtet wird hier über ein umfangreiches Klassenfrequenzprojekt, das in Israel durchgeführt wurde. Die Klassen wurden nicht per Zufallsverfahren zusammengestellt. Jedoch wurde bei der Auswertung der Daten der soziale Hintergrund nachträglich berücksichtigt:

3.6 Angrist & Lavy, Titel Nr. 13 in Hatties Tabelle 6.2:

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Angrist & Lavy	1999	1	1	1'327	46'455	0,15	Achievement

Angrist und Lavy haben Testergebnisse von Schülern aus 3., 4. und 5. Klassen in Israel erfasst. - Wiederum stimmen die Zahlen, die in Hattie Tabelle 6.2 anführt, nicht mit den Angaben der Autoren überein: Bei Hatties sind es in 46'455 Schüler in 1'327 Klassen. Angrist und Lavy (1999, p. 538) nennen dagegen andere Zahlen:

"Table 1 ... reports descriptive statistics, including quantiles, for the population of over 2000 classes in Jewish public schools in each grade (about 62'000 pupils)." (Angrist und Lavy, 1999, p. 538)

Tabelle 1 bei Angrist und Lavy enthält (hier in gekürzter Fassung) Rohwerte, die bei dieser Studie erhoben wurden (1999, p. 539):

5th grade (2019 classes, 1002 schools, tested in 1991)					
Class size	21	26	31	35	38
Average verbal	64,2	69,9	75,4	79,8	83,3
Average math	54,8	61,1	67,8	74,1	79,4
3rd grade (2111 classes, 1011 schools, tested in 1992)					
Class size	22	26	31	35	38
Average verbal	78,4	83,0	87,2	90,7	93,1
Average math	75,0	80,2	84,7	89,0	91,9

Aus diesen Rohwerten kann man unmittelbar ablesen, dass in großen Klassen bessere Testergebnisse erzielt werden als in kleinen Klassen. Dementsprechend hätte in Hatties Tabelle 6.2 bei Angrist und Lavy die Effektstärke ein negatives Vorzeichen haben müssen. Aber vielleicht hat Hattie in diesem Falle den Einfluss sozialer Faktoren berücksichtigt, denn Angrist und Lavy schreiben dazu:

"The raw positive correlation between achievement and class size is clearly an artifact of the association between smaller classes and the proportion of pupils from disadvantaged backgrounds." (Angrist und Lavy, 1999, p. 569)

Entsprechend zeigt sich bei Kontrolle des sozialen Hintergrundes der Schüler, dass benachteiligte Schüler aus kleinen Klassen bessere Testergebnisse erzielen als benachteiligte Schüler aus großen Klassen:

"controlling ... pupil background consistently show a negative association between larger classes and student achievement." (Angrist und Lavy, 1999, p. 569)

Mit dieser Aussage ließe sich eine Effektstärke mit positivem Vorzeichen in Hatties Tabelle durchaus rechtfertigen. Da Hattie nicht erklärt, wie er gerechnet hat, bleibt im Ungewissen, ob der Wert 0,15 für *d* in Tabelle 6.2 ein Flüchtigkeitsfehler ist, oder ob Hattie in diesem Falle ausnahmsweise den Einfluss des sozialen Hintergrundes der Schüler berücksichtigt hat.

Sehr bemerkenswert an Tabelle 1 bei Angrist und Lavy ist ferner, dass die Drittklässler im Jahr 1992 deutlich besser abschneiden als die Fünftklässler im Jahr 1991. Angrist und Lavy erklären dazu, dass 1991 erstmals landesweit in allen 4. und 5. Klassen ein Test eingesetzt

wurde. Die Ergebnisse bei diesem ersten Testeinsatz waren sehr unbefriedigend, was landesweit Testtraining in den Grundschulen zur Folge hatte (Angrist und Lavy, 1999, p. 569).

Hier bestätigt sich, dass nicht ungeprüft vorausgesetzt werden darf, dass Leistungstests valide messen, was bei einer wissenschaftlichen Studie erfasst werden soll (siehe vorne, Abschnitt 2.3). Wenn ein Jahr Testtraining ausreicht, damit Drittklässler deutlich besser abschneiden als untrainierte Viertklässler, dann wird mit dem Test möglicherweise vor allem schnell vermittelbares, abfragbares Wissen erfasst.

Anhang

Die Tabelle 6.2 im Abschnitt „Class size“ von »Visible Learning« (Hattie, 2009, p. 87) enthält 14 Titel. 6 dieser Titel wurden bereits analysiert. Im Folgenden werden 7 weitere Artikel aus Hatties Tabelle 6.2 besprochen:⁸

3.7 Smith & Glass, Titel Nr. 2 in Hatties Tabelle 6.2:

Bei Smith und Glass stehen in Hatties Tabelle 6.2 folgende Zahlen:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
Smith & Glass	1980	59	371	—	—	0,24	Non-Achievement outcomes

In der Klassenfrequenzstudie von Glass und Smith, die 1978 und 1979 veröffentlicht wurde, standen Leistungsmessungen im Zentrum der Meta-Analyse (siehe Abschnitt 3.1). In Ergänzung dazu haben Smith und Glass 59 Artikel erneut ausgewertet, in denen auch Arbeitshaltung und Unterrichtstil erfasst worden waren. Dabei zeigte sich, dass sich das Unterrichtsklima mit abnehmender Klassenfrequenz verbessert:

“Smaller classes are associated with greater attempts to individualize instruction and better classroom climate.” (Smith und Glass, 1980, p. 419)

Die Effektstärke $d = 0,24$, die Hattie in Tabelle 6.2 Smith und Glass zuordnet (siehe oben), ist nicht vereinbar mit dem Wert $d = 0,09$, der bei Glass und Smith in Hatties Tabelle 6.2 steht, denn die Effektstärken für „achievement“ und „attitudes and instruction“, die Smith und Glass in Ihren Meta-Analysen zusammengetragen haben, unterscheiden sich nur geringfügig und nicht, wie bei Hatties um einen Faktor 2,6 (gleich $0,24/0,09$, siehe auch den Abschnitt 3.1).

Da Hattie weder in »Visible Learning« (2009), noch in seinem Aufsatz von 2005 erläutert, wie er die Effektstärken, die er Smith und Glass bzw. Glass und Smith zuordnet, berechnet hat, ist nicht entscheidbar, welcher der beiden Werte falsch ist: $d = 0,24$ oder $d = 0,09$. Möglich ist auch, dass beide Werte falsch sind.

3.8 McGiverin et al., Titel Nr. 4 in Hatties Tabelle 6.2:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
McGiverin et al.	1989	10	24	-	-	0,34	Achievement

Im US-Staat Indiana wurden im Rahmen eines Projektes „Prime Time“ in den Jahren 1984 bis 86 kleine Klasse (im Mittel 19,1 Kinder) eingerichtet. Am Ende des 2. Schuljahres wurden in den kleinen und in normal großen Klassen (im Mittel 26,4 Kinder) standardisierte Tests durchgeführt. Die kleinen Klassen erzielten über einen Zeitraum von 2 Jahren signifikant bessere Ergebnisse:

“The purpose of this investigation was to examine the effects of Indiana's project PRIME TIME on reading and math achievement test scores of second graders who

⁸ Nicht analysiert wird hier die Arbeit Urquiolas aus dem Jahr 2000. Diese unveröffentlichte Dissertation war nicht erhältlich.

had completed 2 years of a state-supported reduced-class-size program. ... The results of this meta-analysis were significant at the .001 level. ... The effect size for the PRIME TIME group was .34 standard deviations” (McGiverin et al., 1989, p. 47).

Die Effektstärke $d = 0,34$, die Hattie hier direkt aus dem Text von McGiverin et al., übernommen hat, hätte er durch 2 teilen müssen, denn bei McGiverin et al. liegt der Berechnung der Effektstärke ein Zeitraum von 2 Jahren zu Grunde, wohingegen Hatties »hinge-point« $d = 0,4$ für Lernfortschritte gilt, die in einem Jahr erzielt werden:

„Teachers typically can attain between $d = 0,20$ to $d = 0,40$ growth per year” (Hattie, 2009, p. 17)

Hattie gibt vor, seine Tabelle 6.2 enthielte Effektstärken für die Reduktion der Klassenfrequenz von 25 auf 15 Kinder. Abweichend davon enthalten bei McGiverin et al. (1989) die großen Klassen im Mittel 26,1 und die kleinen Klassen 19,1 Kinder. Solche Abweichungen findet man auch in anderen Studien: Bei der Studie Blatchfords (2003) wurden Klassen von 30 und 20 Schülern miteinander verglichen (siehe Abschnitt 3.4). Bei Angrist und Lavy (1999) haben die größten Klassen 38 und die kleinsten 21 Schüler (siehe Abschnitt 3.6). Auf diese Abweichungen weist Hattie an keiner Stelle hin, was erforderlich gewesen wäre, da Klassenfrequenz-Effekte bereichsabhängig sind (siehe hier „Figure 4“ in Abschnitt 3.1).

3.9 Hoxby, Titel Nr. 6 in Hatties Tabelle 6.2:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
Hoxby	2000	1	1	14'593	306'453	0,03	Achievement

Die Studie Hoxbys wurde im US-Bundesstaat Connecticut durchgeführt. Auf der Basis der Daten von Schulämtern und Schulen wurden Vergleiche zwischen Klassen mit 15 bis 30 Kindern vorgenommen (Hoxby, 1998, p. 1). Die unterschiedlich großen Klassen wurden nicht extra zusammengestellt. Große oder kleine Klassen waren das Ergebnis von Schwankungen in der Größe einzelner Jahrgänge:

“Random variation in the population generates exogenous variation in class size.” (Hoxby, 2000, p. 1282)

Hoxby geht davon aus, dass die Anzahl der Kinder eines Jahrgangs an einer Schule eine Zufallsgröße darstellt. Dies wird allerdings durch mehrere Einflüsse verhindert: In Connecticut können Eltern die Schule im zuständigen Bezirk frei wählen. Dabei wird die Auswahl der Schulen durch offizielle Ranglisten erheblich beeinflusst und es gibt Eltern, die umziehen oder ihr Kind ummelden, wenn die Einschulung bevorsteht. Auch die Zusammensetzung von Schulklassen ist in der Regel nicht repräsentativ, denn Schulleitungen neigen dazu, Problemschüler kleineren Klassen zuzuordnen. Ferner muss man bei der Studie von Hoxby berücksichtigen, dass kleine Klassen vor allem in kleinen Schulbezirken Connecticuts gebildet worden waren:

“the data must contain districts that are not large because, in large districts, natural population variation averages out to a great extent within each cohort” (Hoxby, 1998, p. 12)

Bei der Klassenfrequenzstudie von Angrist und Lavy (siehe Abschnitt 3.6) hatte sich gezeigt, dass Schüler aus den kleinen Klassen der kleinen Schulbezirke im Mittel einen ungünstigeren familiären Hintergrund und damit schlechtere Testergebnisse hatten als Schüler aus großen Schulbezirken. Dementsprechend war auch bei Hoxby die Zusammensetzung der kleinen

Klassen nicht repräsentativ. Trotzdem hat diese Studie ein bildungspolitisch sehr bedeutsames Ergebnis hervorgebracht:

“I find no evidence that class size reductions are more efficacious in schools that contain high concentrations of low income students or African-American students.” (Hoxby, 2000, p. 1282)

Dieses Aussage ergänzt die Ergebnisse des STAR-Projects: Beim STAR-Project waren Problemschüler in kleinen, repräsentativ zusammengesetzten Klassen signifikant besser als in großen Klassen gefördert worden. Nach Hoxby werden solche besonderen Fördereffekte nicht beobachtet, wenn die kleinen Klassen Problemschüler in „hoher Konzentration“ enthalten.

3.10 Goldstein et al., Titel Nr. 8 in Hatties Tabelle 6.2:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
Goldstein et al.	2000	9	36	1'178	29'440	0,20	Achievement

In die Meta-Analyse von Goldstein (2000) sollten nur Klassenfrequenz-Studien aufgenommen werden, bei denen normierte Tests zum Einsatz gekommen sind und der Beobachtungszeitraum mindestens ein Schuljahr umfasst (Goldstein, 2000, p. 5). Entsprechend wurden neun US-amerikanische Studien für diese Meta-Analyse ausgewählt: Bei sechs dieser Studien waren die Daten vor 1980 erhoben worden. Eine neuere Studien, die Goldstein et al. für ihre Meta-Analyse ausgewählt hatten, wird im Folgenden kurz angesprochen, weil hier die von Goldstein et al. Selbst gestellten Qualitätsanforderungen offensichtlich nicht erfüllt waren:

San Juan Unified School District (1992): Class Size Reduction Evaluation: Freshman English. Carmichael, USA:

In einem Schulbezirk Kaliforniens waren Klassen im 9. und 10. Jahrgang von 30 auf 20 Schüler verkleinert worden. Im Januar und im Mai 1991 wurde ein Test^{9]} in den auf 20 Schüler reduzierten Klassen eingesetzt. Die Autoren selbst weisen auf die begrenzte Aussagekraft ihrer Studie hin:

“It should be noted that with only one semester experience with this class size reduction program, evaluation findings must be interpreted cautiously.” (San Juan Unified School District, 1992, p. 3).

Hinzu kommt, dass im San Juan Unified School District Kontrollgruppe mit Schülern der gleichen Jahrgänge gefehlt haben. Aus pädagogischer Sicht enthält die Studie aus Carmichael trotzdem interessante Ergebnisse: Auch in Carmichael wird festgestellt, dass sich Schüler in kleinen Klassen durch eigene Beiträge stärker am Unterricht beteiligen (San Juan Unified School District, 1992, p. 6)

3.11 Akerhielm, Titel Nr.10 in Hatties Tabelle 6.2:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
----------------	-------------	-----------------------	-----------------------	-----------------------	------------------------	----------	----------------

9] CTBS/⁴: Comprehensive Tests of Basic Skills, Fourth Edition

Akerhielm, K.	1995	1	1	1'052	24'000	0,15	Achievement
---------------	------	---	---	-------	--------	------	-------------

In ihrer Literaturübersicht kommt Akerhielm (1995) zu folgendem Ergebnis:

“Of a review of 112 studies on class size, only 23 found a statistically significant relationship between class size and student achievement ...” (Akerhielm, 1995, p. 229)

Nur bei 15 der 23 Studien mit signifikanten Ergebnissen wurde festgestellt, dass die Testergebnisse mit abnehmender Klassenfrequenz besser werden. 9 Studien hatten entgegengesetzte Resultate: Hier waren die Testergebnisse in den kleineren Klassen signifikant schlechter. – Den entscheidende Grund für diese widersprüchlichen Ergebnisse sieht Akerhielm darin, dass teilweise nicht berücksichtigt wird, dass Problemschüler in kleinen Klassen oft überrepräsentiert sind:

“Public school students are not randomly allocated to different class sizes, with the lower ability students disproportionately found in the smaller classes. ... Only by using statistically unbiased measures of educational inputs can the debate on educational reform ... be informative, relevant and productive.” (Akerhielm, 1995, p. 239)

Akerhielm kommt zu dem Resümee, dass in Zukunft Studien erforderlich seien, bei denen die Kinder repräsentativ auf große und kleine Klassen verteilt werden. Nur so könne man brauchbare Aussagen zum Thema Klassenfrequenz erhalten. – Bei Hattie werden die Ergebnisse der Studie Akerhielms auf den Wert $d = 0,15$ reduziert, womit ein weiteres Beispiel dafür gegeben ist, dass durch die „Verdichtung der Daten“ die bei Hatties erfassten Studien „weitgehend ihre Bedeutung verlieren“ (Brügelmann, 2013, S. 13).

3.12 Rice, Titel Nr.11 in Hatties Tabelle 6.2:

Authors	Year	No. of studies	No. of effects	No. of classes	No. of students	d	Outcome
Rice	1999	1	1	8'760	24'599	- 0,04	Achievement

Hattie ordnet der Studie von Rice eine Effektstärke von $d = -0,04$ für „Achievement“ zu. Sowohl das Vorzeichen wie auch der Betrag von d sind falsch.

Der zentrale Gegenstand der Studie von Rice waren nicht Resultate von Leistungstests, sondern die Beobachtung von Unterrichtsmethoden, entsprechend dem Titel:

“The impact of class size on instructional strategies and the use of time in high school mathematics and science courses.” (Rice, 1999, p. 215)

Ergebnisse der Studie von Rice, bezogen auf das Fach Mathematik, sind z. B:

“Class size had a negative effect on ... time spent working with small groups, time spent using innovative instructional practices, and time devoted to whole-group discussion. That is, as classes grew larger, less time was spent in these ways.” (Rice, 1999, p. 219)

Wie schon bei Dustmann et al. (siehe Abschnitt 3.5) haben bei Rice (1999) die Effektstärken ein negatives Vorzeichen, wenn die Reduzierung der Klassenfrequenz zu besseren Ergebnissen führt. Bei Hattie ist es umgekehrt. Deshalb hätte auch bei Rice in Hatties Tabelle 6.2 der Wert für d kein negatives Vorzeichen haben dürfen:

„The negative impact of class size on the use of innovative instructional strategies (i.e., smaller classes are related to more innovative strategies)” (Rice, 1999, p. 219, 223).

Bei Rice wurden zwar zusätzlich Ergebnisse von Leistungstests erfasst, aber auch daraus kann Hattie nicht abgeleitet haben, dass kleine Klassen schlechtere Testergebnisse erzielt haben:

“The interaction between class size and average test score was negative and significant for amount of time devoted to innovative instructional strategies“ (Rice, 1999, p. 219).

Mit diesen Angaben lässt sich die Effektstärke $d = 0,04$, die Hatties in Tabelle 6.2 der Arbeit von Rice zuordnet, nicht in Einklang bringen. Möglicherweise hat Hattie aus Tabelle 1 bei Rice (1999, p. 219,) den *Pearson Correlation Coefficient* „-0,04“ in Zeile 3 als Effektstärke (nach Cohen) in seine Tabelle 6.2 aufgenommen.

3.13 Johnson et al., Titel Nr.12 in Hatties Tabelle 6.2:

<i>Authors</i>	<i>Year</i>	<i>No. of studies</i>	<i>No. of effects</i>	<i>No. of classes</i>	<i>No. of students</i>	<i>d</i>	<i>Outcome</i>
Johnson et al.	2003	1	1	168	3'700	0,00	Achievement

Wer nur Visible Learning in der Hand hält, wird keinen Artikel finden, der hierher passt. Der vermutlich zutreffende Artikel von Johnson et al. ist:

Johnson, D. W./Jensen, B./Feeny, S., & Methakullawat, B. (2004). Multivariate analysis of performance of Victorian schools. Paper presented at the Making Schools Better Submit Conference, Melbourne, Australia.

Die Autoren weisen selber daraufhin, dass ihre Studie keine gute Grundlage für die Bewertung von Klassenfrequenz-Effekten sei:

“The class size variable for secondary schools was the average class size over the school as a whole. This may not be a good measure of the class size in VCE classes^{10]}.” (Johnson et al., 2004, p. 30)

Dazu wird dann jedoch festgestellt, dass die Klassenfrequenz keine signifikanten Effekte hervorruft und wenn doch signifikante Effekte messbar sind, dann haben diese mal positive, mal negative Vorzeichen. Solche, sich scheinbar widersprechenden Ergebnisse werden dann erzielt, wenn die soziale Zusammensetzung der Klassen nicht berücksichtigt wird (siehe hierzu auch die Abschnitte 3.1,3.2, 3.6 und 3.11).

10] VCE classes: Abschlussklassen im Australischen Bundesstaat Victoria

4. Literatur

- Akerhielm, Karen (1995). Does class size matter? *Economics of Education Review*, 14 (3), 229–241.
- Angrist, Joshua D., und Lavy, Victor (1999). *Using Maimonides' Rule to estimate the effect of class size on scholastic achievement*. *Quarterly Journal of Economics*, 114 (2), 533–575
- Blatchford, Peter, Bassett, Paul, Goldstein, Harvey und Martin, Clare (2003). Are class size differences related to pupils' educational progress and classroom processes? Findings from the institute of education class size study of children aged 5–7 years. *British Educational Research Journal*, 29 (5), 709–730. <http://core.ac.uk/download/pdf/82606.pdf>
- Blatchford, Peter (2005). A multi-method approach to the study of school class size differences. *International Journal of Social Research Methodology*, 8, 195-205.
- Blatchford, Peter, Bassett, Paul und Brown, Penelope (2005). Teachers' and pupils' behaviour in large and small classes: a systematic observation study of pupils aged 10/11 years. *Journal of Educational Psychology*, 97 (3), 454-467.
- Brügelmann, Hans (2011). *Miss Marple neben pisa & co. - Plädoyer für eine Bildungsforschung, die der Praxis nützt*. In Heinz Moser (Hrsg.), *Forschung in der Lehrerbildung. Professionswissen für Lehrerinnen und Lehrer, Band 10* (S. 221-234). Hohengehren: Schneider.
- Brügelmann, Hans (2013). Hat Hatties „Visible Learning“ einen Nutzen für die pädagogische Arbeit vor Ort? BLZ, GEW Bremen. <http://www.gew-hb.de/aktuelles/detailS./neugigkeiten/metaanalysen-der-unterrichtsforschung/>
- Cohen, Jacob (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York: Lawrence Erlbaum Associates.
- Dustmann, Christian, Rajah, Najma und van Soest, Arthur (2002). *Class Size, Education, and Wages*. Institute for the Study of Labor (IZA), Bonn. Discussion paper series, No. 501. <http://ftp.iza.org/dp501.pdf>
- Dustmann, Christian, Rajah, Najma und van Soest, Arthur (2003). Class size, education, and wages. *The Economic Journal*, 113 (485), F99 – F120.
- Finn, Jeremy D. (1998). Class size and students at risk. What is known? What is next? A Commissioned Paper. <http://files.eric.ed.gov/fulltext/ED418208.pdf>
- Finn, Jeremy D. und Achilles, Charles M. (1999). Tennessee's Class Size Study: Findings, Implications, Misconceptions. *Educational Evaluation and Policy Analysis*, 21 (2), 97-109.
- Finn, Jeremy D. (2002). Class-size reduction in grade k-3. In Alex Molnar (Hrsg.), *School reform propols: the research evidence* (pp. 27 – 48) Greenwich, CT.: Information Age Publishing. <http://nepc.colorado.edu/files/Chapter02-Finn-Final.pdf>
- Glass, Gene V. (1976). Primary, secondary and meta-analysis of research. *Educational Researcher*, 10, 3-8. <http://www.jameslindlibrary.org/glass-gv-1976/>
- Glass, Gene V. (1977). Integrating findings: The meta-analysis of research. *Review of Research in Education*, 5, 351-379.
- Glass, Gene V. und Smith, Mary Lee (1978). *Meta-analysis of research on the relationship of class-size and achievement. The class size and instruction project* (Research/Technical No. C8088). San Francisco, CA: Far West Lab. for Educational Research and Development. https://archive.org/details/ERIC_ED168129
- Glass, Gene V. und Smith, Mary Lee (1979). Meta-analysis of research on class size and achievement. *Educational Evaluation and Policy Analysis*, 1 (1), 2-16.
- Goldstein, Harvey, Yang, Min, Omar, Rumana Turner, Rebecca and Thompson, Simon (2000). Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49 (3), 399-412. <http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/meta-analysis-class-size.pdf>

- Hattie, John (2005). The paradox of reducing class size and improved learning outcomes. *International Journal of Educational Research*, 42, 387-425.
- Hattie, John (2009). *Visible Learning - A synthesis of over 800 meta-analyses relating to achievement*. London, New York: Routledge.
- Hattie, John (2013). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von Visible Learning, besorgt von Wolfgang Beywl und Klaus Zierer*. Hohengehren: Schneider. ISBN: 978-3-8340-1190-9
- Hattie, John (2015). *Lernen sichtbar machen. Überarbeitete deutschsprachige Ausgabe von Visible Learning, besorgt von Wolfgang Beywl und Klaus Zierer (3. Aufl.)*. Hohengehren: Schneider. ISBN: 978-3-8340-1450-4
- Hoxby, Caroline M. (1998). *The effects of class size and composition on student achievement: New evidence from natural population variation*. NBER Working Paper No. 6869, December 1998. <http://www.nber.org/papers/w6869.pdf>
- Hoxby, Caroline M. (2000). The effects of class size on student achievement: New evidence from population variation. *Quarterly Journal of Economics*, 115 (4), 1239-1285.
- Johnson, David, Jensen, Ben, Feeny, Simon and Methakullawat, Ben (2004). *Multivariate analysis of performance of Victorian schools*. Melbourne Institute of Applied Economics and Social Research, Australia. https://www.melbourneinstitute.com/downloads/conferences/Papers_presentations/Performance_of_Victorian_Schools.pdf
- Krueger, Alan B. and Whitmore, Diane M. (2001). *Would Smaller Classes Help Close the Black-White Achievement Gap?* Princeton University. <http://dataspace.princeton.edu/jspui/bitstream/88435/dsp01w66343627/1/451.pdf>
- McGiverin, Jenniver, Gilman, David und Tillitski, Chris (1989). A meta-analysis of the relation between class size and achievement. *The Elementary School Journal*, 90 (1), 47-56.
- Meyer, Hilbert (2014). Auf den Unterricht kommt es an! Hatties Daten deuten lernen. In Ewald Terhart (Hrsg.), *Die Hattie-Studie in der Diskussion, Probleme sichtbar machen*. In der Reihe Bildung kontrovers (2. Aufl., S. 117-133). Seelze: Klett/Kallmeyer.
- Molnar, Alex, Smith, Philip, Zahorik, John, Palmer, Amanda, Halbach, Anke und Ehrle, Karen (1999a). Evaluating the SAGE program: A pilot program in targeted pupil-teacher reduction in Wisconsin. *Educational Evaluation and Policy Analysis*, 21 (2), 165-177.
- Molnar, Alex (1999b). *Smaller Classes and Educational Vouchers: A Research Update*. Harrisburg, PA: Keystone Research Center. <http://nepc.colorado.edu/files/keystoneupdate.pdf>
- Pant, Hans Anand (2014). *Visible Evidence? - Eine methodisch orientierte Auseinandersetzung mit John Hatties Meta-Metaanalysen*. In: Terhart, Ewald, Hrsg. (2014): *Die Hattie-Studie in der Diskussion, Probleme sichtbar machen*. In Ewald Terhart (Hrsg.), *Die Hattie-Studie in der Diskussion, Probleme sichtbar machen*. In der Reihe Bildung kontrovers (2. Aufl., S. 134-146). Seelze: Klett/Kallmeyer.
- Pate-Bain, Helen, Boyd-Zaharias, Jane, Cain, Van A., Word, Elizabeth und Binkley, M. Edward (1997). *The student/teacher achievement ratio (STAR) project: STAR follow-up studies*. Lebanon, TN: HEROS. <http://eric.ed.gov/PDFS/ED419593.pdf>
- Rice, Jennifer King (1999). The impact of class size on instructional strategies and the use of time in high school mathematics and science courses. *Educational Evaluation and Policy Analysis*, 21(2), 215-229.
- San Juan Unified School District (1992). *Class Size Reduction Evaluation: Freshman English*. Carmichael, California, USA. <http://files.eric.ed.gov/fulltext/ED344239.pdf>
- Smith, Mary Lee und Glass, Gene V. (1980). Meta-analysis of research on class size and its relationship to attitudes and instruction. *American Educational Research Journal*, 17 (4), 419-433.
- Snook, Ivan, O'Neill, John, Clark, John, O'Neill, Anne-Maree und Openshaw, Roger (2009). Invisible Learnings?: A Commentary on John Hattie's Book – "Visible Learning: A Synthesis of Over 800

- Meta-analyses Relating to Achievement". *New Zealand Journal of Educational Studies*, 44 (1), 93-106.
- Steffens, Ulrich und Höfer, Dieter (2016). *Lernen nach Hattie - Wie gelingt guter Unterricht?* Weinheim und Basel: Beltz. ISBN 978-3-407-25738-3 (Print), ISBN 978-3-407-29472-2 (E-Book)
- Terhart, Ewald (Hrsg.). (2014). *Die Hattie-Studie in der Diskussion, Probleme sichtbar machen*. In der Reihe Bildung kontrovers (2. Aufl.). Seelze: Klett/Kallmeyer.
- Urquiola, Miguel S. (2000). *Essays on educational financing and effectiveness in the United States and Bolivia*. Unpublished Ph.D., University of California, Berkeley, CA.
- Word, Elisabeth R., Johnston, John, Bain, Helen Pate, Fulton, B. Dewayne, Zaharias, Jane Boyd, Achilles, Charles M., Lintz, Martha Nannette, Folger, John and Breda, Carolyn (1990). *The state of Tennessee's student/teacher achievement ratio (STAR) project: Technical report 1985-90*. Nashville: Tennessee State University.