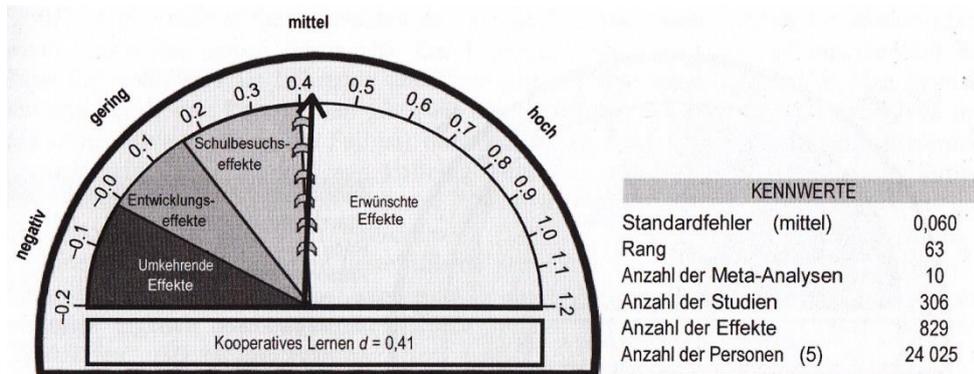


John Hattie: Genie oder Pfuscher?

Ein Vergleich auf drei Ebenen anhand des kooperativen Lernens

John Hattie präsentiert in seinem Buch *Lernen sichtbar machen* (Hattie 2013, S. 251) zum kooperativen Lernen folgendes *Barometer*:



Der Effektwert d von 0.41 ist praktisch deckungsgleich mit dem *Referenzwert* oder *Durchschnittswert aller Studien* von 0.4. Damit ist gesagt, dass Lehrpersonen, die sich um einen exzellenten Unterricht bemühen, diese Methoden eher nicht anwenden sollten oder wenn schon, dann höchst dosiert. Exzellenz beginnt bei Werten im Bereich von 0.60.

Viele im Erziehungsbereich tätige Menschen interpretieren diesen Wert anders: 0.41 sei ja weit weg von Null und damit mit exzellentem Unterricht sehr gut vereinbar. Diese Ansicht wird hier nicht geteilt.

In diesem Artikel soll aber einzig der Frage nachgegangen werden, wie der Wert von 0.41 überhaupt zustande gekommen ist. Vielleicht führt das alleine schon zu einer Klärung, inwieweit kooperatives Lernen wirksam ist.

Metastudie der Metastudien

Wichtig zu wissen ist, dass John Hattie selber keine Studien durchgeführt hat. Er ist ein Studiensammler. Zudem hat er auch nicht einzelne Studien gesammelt, sondern *Metastudien, also Zusammenfassungen von Studien*. Welche Metastudien hat er verwendet, um den *Effektwert von 0.41* zu generieren? Auskunft zu dieser Frage gibt eine Tabelle in seinem Buch auf Seite 422:

Nr.	Domäne	Autor	Jahr	n Studien	n Teilnehmer	n Effekte	d Mittelwert	d SE	CLE	Variablen
650	Unterrichten	Johnson, Maruyama, Johnson, Nelson & Skon	1981	122	—	183	0,73	—	70%	Kooperatives Lernen
651	Unterrichten	Rolheiser-Bennett	1986	23	4 002	78	0,48	—	63%	Kooperatives Lernen
652	Unterrichten	Hall	1988	22	10 022	52	0,31	—	59%	Kooperatives Lernen
653	Unterrichten	Stevens & Slavin	1990	4	—	4	0,48	—	63%	Kooperatives Lernen
654	Unterrichten	Spuler	1993	19	6 137	19	0,54	—	65%	Kooperatives Lernen in Mathematik
655	Unterrichten	Othman	1996	39	—	39	0,27	—	58%	Kooperatives Lernen in Mathematik
656	Unterrichten	Howard	1996	13	—	42	0,37	—	60%	Kooperatives Lernen nach ‚Drehbuch‘
657	Unterrichten	Bowen	2000	37	3 000	49	0,51	0,050	64%	Kooperatives Lernen in Chemie (High School)
658	Unterrichten	Neber, Finsterwald & Urban	2001	12	—	314	0,13	—	54%	Kooperatives Lernen mit Hochbegabten
659	Unterrichten	McMaster & Fuchs	2002	15	864	49	0,30	0,070	58%	Kooperatives Lernen

Hattie haben 10 Metastudien vorgelegen, die ausgewertet wurden. Wie hat er das gemacht? Ganz simpel: Er hat alle Werte in der Spalte „d Mittelwerte“ addiert und durch die Anzahl Metastudien dividiert. Das ergibt diesen vielbeachteten Durchschnittswert von 0.41. Die Metastudien stammen aus den Jahren 1981 – 2002. Damit sind Arbeiten nach 2002 nicht mehr berücksichtigt. Die Metastudien beinhalten 306 Einzelstudien (Summe Spalte „n Studien“) und insgesamt hat Hattie dort 829 Effektvariablen gefunden (Spalte „n Effekte“).

Mit einem gewissen Befremden fällt zuerst einmal auf, dass Hattie überhaupt keine *Gewichtung* vorgenommen hat. So fällt die Metastudie von Stevens & Slavin (1990), die nur 4 Studien mit 4 Effekten beinhaltet, genauso stark ins Gewicht wie die Metastudie von Johnson et al. (1981) mit 122 Studien und 183 Effektwerten. Dieses Vorgehen sollte uns darauf aufmerksam machen, dass auch Hatties Werte nicht sakrosankt sind. Hätten Johnson et al. aus ihrer grossen Metastudie vier kleinere gemacht mit ca. je 30 Studien, dann hätten sie den Wert 0.73 vier Mal generieren können. Damit wäre der Gesamtwert von 0.41 auf 0.49 gestiegen!

Die neueste Metastudie von McMaster & Fuchs (2002)

Wir gehen jetzt eine Ebene nach unten zu den Metastudien und studieren eine davon genauer. Ich habe zu diesem Zwecke die neueste ausgewählt, die von *McMaster & Fuchs* aus dem Jahre 2002. Diese Autoren haben Einzelstudien zwischen 1990 und 2000 gesammelt, die in den publizierten Artikeln genug detailliert beschrieben worden waren, sodass McMaster & Fuchs sich von ihrer Qualität überzeugen konnten. In diesen *qualitative review* wurden 15 Studien aufgenommen.

Das freut das Forscherherz natürlich erst einmal und lässt auf einen hohen Standard bezüglich Studiendesign hoffen. Das scheint sich auch Hattie gedacht zu haben, addiert er doch alle 49 Effekte ohne mit der Wimper zu zucken zusammen und bestimmt den Durchschnittswert, unabhängig ob jetzt in einer Studie nur zwei Effektwerte anführt werden oder gleich deren 12. Die Effektwerte rangieren von -1.20 bis +1.42. Der Durchschnittswert beträgt 0.30. Erwähnenswert ist auch die Tatsache, dass sich alle 15 Studien auf Kinder und Jugendliche mit Lernschwierigkeiten beziehen. Diese Teilnehmer wurden in einigen Fällen sogar aus umfassenderen Studien herausgelöst und für sie gesonderte Werte berechnet. Das erklärt die oft kleinen Versuchsteilnehmerzahlen, die viele der 15 Studien aber leider nahezu unbrauchbar machen. Dazu nachfolgend eine kleine Tabelle:

Studien Nr.	Anzahl Teilnehmer gesamt	Experimentaltgruppen	Kontrollgruppen
11	9	6	3
7	12	(6)*	(6)
5	16	(8)	(8)
15	22	(11)	(11)
4	25	(12)	(13)
6	26	(13)	(13)
3	28	(14)	(14)
1	36	(18)	(18)

* Die in Klammern angegebenen Aufteilungen sind Annahmen meinerseits.

8 Studien weisen also äusserst kleine Teilnehmerzahlen auf. Die anderen 7 Studien sind etwas besser bestückt. Die Teilnehmerzahlen reichen von 58 bis 179. Das grosse Problem waren aber die Lehrpersonen. In vielen Studien kann man ihre Anzahl an einer Hand abzählen. Von einer zufallsbedingten Verteilung der Lehrpersonen auf die Experimental- und Kontrollgruppen kann also sicher nicht ge-

sprochen werden und damit ist natürlich auch in den meisten Fällen unmöglich zu belegen, dass irgendwelche Effekte tatsächlich auf die angewandten Unterrichtsmethoden zurückgeführt werden können. Was wir nämlich aus der erziehungswissenschaftlichen Forschung ganz sicher wissen: Entscheidend für die Unterrichtsqualität ist das Lehrpersonenverhalten und das müsste kontrolliert werden.

In der Folge werden die einzelnen Studien und ihre Problematik kurz besprochen:

1. **Malouf et al. (1990) Studie 1:** 36 Kinder wurden zufällig auf Experimental- und Kontrollgruppe verteilt. Der Versuch dauerte 6 Tage und pro Tag erfolgte das Treatment während 25 Minuten. Über die Lehrpersonen erfährt man kein Wort. Der erreichte Wert von **0.31** zugunsten des kooperativen Unterrichts ist nichtsignifikant, was ja eigentlich nicht erstaunt: Zu wenig Kinder, zu kurze Zeitdauer.
2. **Malouf et al. (1990) Studie 2:** Es handelt sich quasi um die verbesserte Version der ersten Studie. Etwas mehr Teilnehmer (66), aber auch kurze Intervention von 7 Tagen à je 15-20minütiges Treatment. Der Effekt reduzierte sich diesmal auf **0.09**. Wieder keine Angaben zu den Lehrpersonen im Paper von McMaster & Fuchs.
3. **Reiter (1994):** Nur 28 Teilnehmer, die vom Forscher und von zwei Lehrpersonen angeleitet wurden. Die zum Teil hohen Werte **0.39, 0.49, 0.48, 0, 0.53, -0.74** sind aber alle nichtsignifikant. Die Intervention dauerte 8 Wochen. Pro Woche wurden 4 Treatments à 50 Minuten appliziert.
4. **Xin (1996) Studie 1:** Leider nahmen nur 25 Kinder mit Lernbehinderungen an der Studie teil. Sie dauerte aber 20 Wochen und pro Woche wurden 4 Lektionen à 50 Minuten appliziert. 6 Lehrpersonen wurden zufällig auf die Experimental- und Kontrollgruppen verteilt. Das tönt ja mal ganz gut! Die Dauer und Intensität des Treatments sind auch in Ordnung. Allerdings erstaunt etwas doch sehr: die kooperativen Teams wurden Ende jeder Woche mit Zertifikaten belohnt, bei den Kontrollgruppen wurden keine Belohnungen verteilt. Das kann man nun wirklich nicht machen! Das stellt die ganze Studie in Frage. McMaster errechneten selber einen Effektwert für die Kinder mit Lernbehinderungen von **0.34**.
5. **Xin (1996) Studie 2:** Eine fast identische Studie wie die erste von Xin. Es waren aber noch weniger Kinder mit Lernbehinderungen involviert, nämlich 16. Der errechnete Wert von McMaster betrug diesmal aber **-1.20**.
6. **Klingler & Vaughn (1996):** Nur 26 Teilnehmer, die Dauer ist okay: 27 Tage à 35-40 Minuten. Zwei nichtsignifikante Werte von **0.17 und -0.36**.
7. **Klingler et al. (1998):** Nur 12 Kinder mit Lernbehinderungen waren in diese Studie involviert. Das ergibt natürlich auch keine signifikanten Ergebnisse. McMaster führt keine Effektwerte an.
8. **Utay & Utay (1997):** Mit 72 Teilnehmenden zählt diese Arbeit zu den besten der Metastudie. Sie dauerte 12 Wochen lang mit je zwei Treatments. Auch das ist in Ordnung. Alle Kids wurden zuerst instruiert, wie man mit einem Computerprogramm eine Story schreibt. Dann mussten die einen die Story allein schreiben, die anderen kooperativ. Auch die Randomisierung der Kinder ist sehr gut gelöst. Resultat: **0.02, 0.00**. *Die beste Studie soweit produziert aber keine Überlegenheit des kooperativen Lernens!*
9. **O'Melia & Rosenberg (1994):** Mit 179 Kids sehr gut dotiert, 8 wöchige Dauer mit dreimal 30 minütiger Applikation pro Woche ist auch okay. Es ging um kooperatives Hausaufgabenlösen. Allerdings ist der Befund zugunsten des kooperativen Lernens nicht schlüssig, wie McMaster bemerkt, da sich die Gruppen vor Beginn des Versuchs schon unterschieden und zwar zugunsten der Experimentalgruppe. Die Effektwerte betragen **0.39, 0.27 und 0.12**.

10. Brandt & Ellsworth (1996): Mit 74 Teilnehmenden und einem 15 Wochen dauerndem Treatment (20 Minuten an drei Tagen pro Woche) zählt diese Studie mit zu den besten des Samples. Die hohen und signifikanten Effektwerte ($p < 0.001$) lassen aufhorchen: **0.80 und 1.02**. Zudem wurden explizit die wichtigen Faktoren für einen kooperativen Unterricht nach Johnson & Johnson angewandt. *Die beste Studie, die signifikante Ergebnisse zugunsten des kooperativen Lernens zeitigt!*

11. Jenkins et al. (1991): Mit nur 9 Teilnehmer ist diese Arbeit zum Vergessen. Die Effektwerte betragen **-1.16, -.23, 1.08** und sind selbstverständlich nichtsignifikant.

12. Jenkins et al. (1994): Die Intervention dauerte ein Jahr, 58 Teilnehmende waren involviert. In der Experimentalschule befanden sich 23, in der Kontrollschule 35 Kinder. Die Krux dieser Arbeit: In der Experimentalgruppe wurde ein ganzes Bündel von Massnahmen implementiert. Kooperatives Lernen war nur einer von vielen Bestandteilen und zudem erfolgte keine Zufallsverteilung der Schüler. McMaster & Fuchs (S.114) schreiben: "Thus, it is difficult to determine the extent to which CL per se played a role in improved student learning". Die Werte von **0.85, 0.77, 1.42, 0.19, 0.55**, obwohl z.T. signifikant, sind kaum zu gebrauchen. Die Studie ist schlicht und einfach irrelevant für unsere Fragestellung.

13. Stevens & Slavin (1995a) und

14. Stevens & Slavin (1995b): Auch in diesen beiden Studien wurden die Experimentalschulen umfassend anders beschult als die Kontrollschulen und kooperatives Lernen war nur eine von vielen Innovationen. Die Untersuchungen dauerten zwei Jahre. Studienteilnehmer waren genügend vorhanden: In Studie 1995a 156 und in Studie 1995b 94. McMaster & Fuchs machen die gleichen Vorbehalte wie in Studie 12: Es sei nicht auszumachen, was das kooperative Lernen an sich bewirkt haben könnte. Beide Studien zusammen generieren 20 Effektwerte, alle im positiven Bereich, aber letztlich vom Studiendesign her betrachtet unbrauchbar: **0.36, 0.35, 0.02, 0.19, 0.37, 0.34, 0.35, 0.24, 0.24, 0.25, 0.33, 0.11, 0.35, 0.17, 0.66, 0.65, 0.65, 0.24, 0.63, 0.34**

15. Gillies & Ashman (2000): Auch diese Studie geht am Thema vorbei, vergleicht sie doch strukturiertes kooperatives Lernen mit unstrukturiertem kooperativem Lernen. Der Vergleich endet mit Vorteilen des strukturierten Ansatzes und es werden zwei Werte angeführt: **1.32 und 0.04**. Warum Hattie diese Werte einrechnet, ist mir schleierhaft.

Diese Kurzübersicht zeigt meiner Meinung nach vor allem eines: Um die erziehungswissenschaftliche Forschung ist es im englischsprachigen Bereich nicht wirklich gut bestellt. Alles in allem sind die Studien eine riesige Enttäuschung.

Wie schon gesagt hat das Hattie nicht gehindert, alle Werte zu verwenden und schlicht und einfach den Durchschnittswert von 0.30 auszurechnen.

McMaster & Fuchs haben das übrigens nicht getan! Vermutlich wäre es ihnen nie in den Sinn gekommen, alle aufgeführten Studien über einen Leisten zu schlagen. Ihnen ging es wahrscheinlich eher darum, sich einen Überblick über Studien in diesem Forschungsbereich zu verschaffen.

Ich bin mir auch nicht sicher, ob Hattie überhaupt bemerkt hat, dass zumindest die Studien 12-15 kaum zu seiner Fragestellung passen. Der Verdacht besteht, dass er diese Metastudie doch sehr ungenau rezipiert hat.

Wenn man besagte Studien ausschliesst und den Effektwert ohne sie berechnet, ergibt sich nur ein Wert von 0.12 anstatt 0.30!!

Die Einzelstudie von Brandt & Ellsworth (1996)

Gehen wir jetzt noch eine weitere Ebene nach unten und untersuchen eine *Einzelstudie*. Am geeignetsten ist diejenige von *Brandt & Ellsworth*. Das Studiendesign ist nach der Einschätzung von McMaster & Fuchs in Ordnung, die Werte zugunsten des kooperativen Lernens mit 0.80 und 1.02 beeindruckend. Die zweite gute Studie des Samples von Utay & Utay wäre auch interessant, aber sie spricht überhaupt nicht für kooperatives Lernen.

Es gilt festzuhalten: Aufgrund der bisherigen Erörterungen ist die Arbeit von Brandt & Ellsworth die einzige ernstzunehmende Studie, die für die Effektivität von kooperativem Lernen spricht!

Zuerst schauen wir uns das Studiendesign genau an: Sechs Klassen von 5-9 Jugendlichen mit Lernschwierigkeiten wurden während 15 Wochen mindestens dreimal pro Woche während 20 Minuten kooperativ unterrichtet und zwar nach der Methode der Gründerväter des kooperativen Lernens *Johnson & Johnson*, welche verschiedene Elemente enthält wie etwa *Positive Interdependence* oder *Individual Accountability*. Die involvierten Lehrpersonen waren diesbezüglich während 45 Stunden trainiert worden. 6 andere Klassen mit 5-9 Jugendlichen bildeten die Kontrollgruppe, in denen traditionelle, nicht kooperative Methoden benutzt wurden. Allerdings erfolgte die Zuteilung der Jugendlichen nicht eigentlich randomisiert. Die 12 Klassen waren nur zufällig aus 20 zur Verfügung stehenden Klassen auserwählt worden. Daher waren Pre- und Posttests nötig. Die Pretests ergaben keine Unterschiede zwischen der Experimental- und der Versuchsgruppe.

So weit so gut. Was dann aber über die Versuchsanlage berichtet wird, stimmt einem doch mehr als nachdenklich: Lehrpersonen waren sage und schreibe nur vier involviert! Jede dieser Lehrpersonen unterrichtete einerseits eine oder zwei Experimentalklassen und andererseits auch eine oder zwei Kontrollklassen. Es wurde also davon ausgegangen, dass die in kooperativen Methoden trainierten Lehrkräfte beide Methoden quasi „neutral“ zu applizieren in der Lage seien und daher nur die unterschiedlichen Methoden die Effekte bewirken würden. Das scheint mir doch eine eher abenteuerliche Vermutung zu sein. Zudem finden sich im Paper von Brandt & Ellsworth absolut keinerlei Angaben über die Art des traditionellen Unterrichts.

Damit ist in dieser Studie viel Wünschen und Hoffen drin und leider ziemlich wenig Wissenschaft. Was, wenn diese vier Lehrpersonen alles Kooperationsjunkies waren, die sich in den Experimentalklassen total verausgabten und in den Kontrollklassen einen grottenschlechten traditionellen Unterricht applizierten?

Erstaunlicherweise finden sich in einzelnen Passagen des Papers von Brandt & Ellsworth zwischen den Zeilen Hinweise auf ein eben solches Verhalten!

So gab es in den traditionell unterrichteten Klassen 20% mehr Absenzen. Es traten mehr Disziplinprobleme auf und es ging auch weniger kameradschaftlich her und zu. Nach Meinung der Lehrer hätten sich die Kids in diesen Klassen auch generell apathisch verhalten. Ethnische Reibereien seien ebenfalls häufiger aufgetreten und etwa 70% der Studenten hätten den traditionellen Unterricht als irrelevant oder langweilig empfunden.

Brandt & Ellsworth sind natürlich der Ansicht, dass die diesbezüglich aufgetretenen Unterschiede auf das Konto der verschiedenen Methoden zurückzuführen seien. Aber das ist reines Wunschdenken. Darüber lassen sich aufgrund der Studie letztlich keine Aussagen machen.

Diese Anmerkungen bezüglich der Auswirkungen des traditionellen Unterrichts stehen auch völlig in Widerspruch zu Hatties Einschätzung. Er betont an vielen Stellen in seinem Buch doch sehr häufig, dass traditioneller Unterricht alles andere als langweilig sein müsse. Sein Favorit ist z.B. *Direkte Instruktion*, ein effektives Verfahren, das zwar eher traditionell daherkommt, aber mit Sicherheit nicht zu obigen negativen Erscheinungen führen muss.

Wie dem auch sei: Die Studie von Brandt & Ellsworth enttäuscht vom Studiendesign her betrachtet und die hohen Effektwerte sind alles andere als zwingend.

Fazit

Wenn man erziehungswissenschaftliche Studien etwas genauer unter die Lupe nimmt, kommt man nicht umhin festzustellen, dass praktisch alle Arbeiten Lichtjahre vom Goldstandard empirischer Forschung, der randomisierten Doppelblindstudie, entfernt sind. Daher sind die gewonnenen Resultate mit äusserster Vorsicht zu geniessen und der Verdacht besteht, dass da nur allzu oft geschönt wird. Hattie hat ja feststellen müssen, dass mehr als 95% seiner 50'000 Studien positive Resultate zeitigen. Alles scheint zu wirken. Das aber kann nicht sein. Sein Referenzwert von 0.40 ist daher mehr als gerechtfertigt!

Ich bin der Meinung, dass man die 0.4 ohne mit der Wimper zu zucken subtrahieren kann. Dann ergibt sich für kooperatives Lernen ein Wert von Null. Aufgrund der hier referierten Studien würde es mich nicht erstaunen, wenn der tatsächliche Wert aber noch etwas tiefer läge.

Was soll man nach diesen Erörterungen überhaupt von Hatties Herkulesarbeit halten? Kann man aus Stroh Gold spinnen wie im Grimm'schen Märchen von Rumpelstilzchen? Ich glaube schon! Das geniale an Hattie ist, dass er den Durchschnittswert von 0.4 zum Referenzwert erkürt hat und so ermöglicht, die Spreu vom Weizen zu trennen. Aber das muss man auch wirklich tun! Und wenn man sich dann die Praktiken und Methoden anschaut, die bei Hattie Werte zwischen - sagen wir einmal, - 0.55 und 0.9 ergeben haben, beginnt man zu erahnen, wie exzellenter Unterricht ausschauen könnte. Kooperative Lernmethoden haben damit aber kaum etwas zu tun!

Allerdings scheint Hattie teilweise doch unseriös gearbeitet zu haben, wie der Umgang mit der oben besprochenen Metastudie von McMaster und Fuchs nahelegt!

Literatur

Brandt, F.J. & Ellsworth, N.J. (1996). *Effects of cooperative learning on the academic achievement and self-esteem of urban adolescents with learning disabilities*. Learning Disabilities, 7, S. 9-13

Hattie, J. (2013). *Lernen sichtbar machen*. Baltmannsweiler: Schneider

McMaster, K.N. & Fuchs, D. (2002). *Effects of cooperative learning on the academic achievement of students with learning disabilities: An update of Tateyama-Sniezek's review*. Learning Disabilities Research and Practice, 17(2), S. 107-117.