

## Wenn SchülerInnen sich gegenseitig beurteilen: Wie lernwirksam ist Peer Assessment im Unterricht?

Es gehört zu den Kernaufgaben einer Lehrkraft, SchülerInnen Rückmeldungen auf ihre Lernleistung zu geben – in Form von Noten oder mit mündlichem oder schriftlichem Feedback. Bei der Methode des *Peer Assessment* wird diese Aufgabe von den SchülerInnen selbst übernommen: Sie geben sich wechselseitig Rückmeldung auf ihre erbrachte Leistung. Ob SchülerInnen von *Peer Assessment* profitieren können, untersuchen Double, McGrane und Hopfenbeck (2020) in ihrer Metaanalyse<sup>1</sup> »The impact of peer assessment on academic performance: A meta-analysis of control group studies«. Die AutorInnen gehen auch der Frage nach, in welcher Form *Peer Assessment* wirksam im Unterricht umgesetzt werden kann.

**EINLEITUNG.** *Peer Assessment* taucht in der Literatur unter verschiedenen Begriffen auf, z.B. *peer feedback*, *peer grading* oder *peer evaluation*. In der Metaanalyse von Double und KollegInnen wird *Peer Assessment* als eine Methode verstanden, bei der sich Peers – verstanden als als Mitglieder einer Peer Group, z.B. SchülerInnen der gleichen Schulstufe – wechselseitig Rückmeldung auf ihre erbrachte Leistung geben. Diese Form der Leistungsrückmeldung umfasst den mündlichen, schriftlichen oder digitalen Austausch von aufgabenrelevantem Feedback.

Im Unterricht kann *Peer Assessment* auf unterschiedliche Weise umgesetzt werden.

Lehrkräfte können den SchülerInnen beispielsweise Bewertungskriterien in Form von Leitfäden oder Benotungssystemen zur Verfügung stellen, den digitalen Austausch von Feedback ermöglichen oder *Peer Assessment* als eine regelmäßige Feedbackpraxis in ihrem Unterricht etablieren.

### METAANALYSE IM ÜBERBLICK

|                                       |  |
|---------------------------------------|--|
| <b>Fokus der Studie</b>               | Wirksamkeit von <i>Peer Assessment</i> auf Leistung im Vergleich zu anderen Formen des Assessment (z.B. durch Lehrkraft) |
| <b>Untersuchte Zielgruppe</b>         | SchülerInnen der Primar- und Sekundarstufe und Studierende   |
| <b>Durchschnittliche Effektstärke</b> | Positiver Gesamteffekt von <i>Peer Assessment</i> auf die Leistung ( $g = 0.31$ )  |
| <b>Weitere Befunde</b>                | Effekt von <i>Peer Assessment</i> ist über unterschiedliche Umsetzungsmöglichkeiten und Altersgruppen stabil             |

<sup>1</sup>Die unterstrichenen Fachbegriffe sind auf unser [Online-Glossar](#) verlinkt.

Die bisherige Forschung zu *Peer Assessment* weist auf einen starken Zusammenhang zwischen den Bewertungen von SchülerInnen durch MitschülerInnen im Rahmen von *Peer Assessment* und den Bewertungen der selben SchülerInnen durch die Lehrkraft hin (z.B. Li et al, 2016; Sanchez et al., 2017). Die vorliegende Metaanalyse untersucht nun die Wirksamkeit von *Peer Assessment* auf die schulische und akademische Leistung auch auf der Basis experimenteller Untersuchungen.

**WORUM GEHT ES IN DIESER STUDIE?** Double und KollegInnen untersuchen in ihrer Metaanalyse, wie sich der Einsatz von *Peer Assessment* auf die Leistung von SchülerInnen (Primar- und Sekundarstufe) und Studierenden (tertiäre Stufe) im Vergleich zu Unterricht ohne Assessment, mit Selbstbewertung (engl. *Self Assessment*) oder Bewertung durch die Lehrkraft (engl. *teacher assessment*) auswirkt.

Insgesamt fließen in die Metaanalyse 55 (quasi-)experimentelle Einzelstudien mit insgesamt 143 Effektstärken ein, die im Zeitraum zwischen 1966 und 2018 in internationalen Zeitschriften veröffentlicht wurden oder als »graue Literatur« – d.h. Dissertationen oder Konferenzbeiträge – zur Verfügung standen. In den berücksichtigten Studien wurde *Peer Assessment* auf unterschiedlichste Weise in den Unterricht integriert (siehe *Tabelle 1*) und jeweils mit einer passiven Kontrollgruppe (d.h. kein Assessment) oder einer aktiven Kontrollgruppe (d.h. andere Form des Assessments, etwa durch die Lehrkraft) verglichen.

Die Leistung der SchülerInnen wird dabei entweder in Form von schriftlichen Leistungen (z.B. Test oder Essays) oder praktischen Leistungen (z.B. Anwendungsaufgaben) erfasst. Um zu untersuchen, unter welchen dieser Bedingungen *Peer Assessment* im Unterricht besonders effektiv ist, nutzen die AutorInnen sogenannte Moderatoranalysen. Dabei untersuchen sie nicht nur den Einfluss verschiedener Umsetzungsmöglichkeiten im Unterricht, sondern auch, in welchem Fach das *Peer Assessment* eingesetzt wurde (siehe Gesamtübersicht).

*Tabelle 1: Umsetzungsmöglichkeiten von Peer Assessment im Unterricht.*

|                       |   |
|-----------------------|---|
| <b>Rolle:</b>         | Die SchülerInnen können innerhalb des Peer Assessments verschiedene Rollen einnehmen: Als Bewertende, zu Bewertende oder beides.  |
| <b>Typ:</b>           | Die häufigsten Umsetzungsformen von Peer Assessment sind Benotung, mündlicher Austausch und schriftliche Beurteilung. Alle Formen können verschieden kombiniert und auch online umgesetzt werden.   |
| <b>Unterstützung:</b> | Mögliche Unterstützungsformen sind z.B. die Vorgabe eines Kriterienkatalogs, eines Bewertungsschemas oder eines Notensystems. Peer Assessment kann aber auch ganz frei und ohne Unterstützung gestaltet werden.   |
| <b>Anonymität:</b>    | Peer Assessment kann offen oder anonym stattfinden. Es kann also transparent gemacht werden, welche/r SchülerIn wen bewertet und wer von wem bewertet wird oder nicht. Außerdem kann die Art der Zuteilung variiert werden: Werden die Rollen offen oder verdeckt zugeteilt oder kann die Zuteilung von den SchülerInnen frei gewählt werden? |
| <b>Häufigkeit:</b>    | Der Einsatz von Peer Assessment kann von der einmaligen Nutzung bis zur regelmäßigen Praxis im Unterricht variieren. Zu vermuten ist, dass mit zunehmender Erfahrung und Übung seitens der SchülerInnen auch die Qualität des Feedbacks gesteigert werden kann.   |

**WAS FAND DIESE STUDIE HERAUS?** Basierend auf den 55 Einzelstudien ergibt die Metaanalyse einen signifikanten positiven Gesamteffekt von *Peer Assessment* auf die Leistung:  $g = 0.31$  (Konfidenzintervall  $g = 0.18$  bis  $g = 0.44$ ). Auch bei gesonderter Betrachtung von SchülerInnen der Sekundarstufe zeigt sich ein signifikant positiver Effekt von *Peer Assessment* auf die schulische Leistung:  $g = 0.44$  (13 Studien).

Mit Blick auf die berücksichtigten Kontrollgruppen zeigt sich, dass SchülerInnen und Studierende mit *Peer Assessment* signifikant bessere Lernleistungen erbrachten als SchülerInnen und Studierende in Kontrollgruppen ohne Assessment ( $g = 0.31$ ) oder mit Bewertung durch die Lehrkraft ( $g = 0.28$ ). Dieser Effekt zeigt sich jedoch nicht im Vergleich zu Kontrollgruppen mit Selbstbewertungen.

In den Moderatoranalysen stellen die AutorInnen fest, dass der positive Effekt von *Peer Assessment* über verschiedene Umsetzungsmöglichkeiten hinweg robust ist. Das heißt, für beinahe alle<sup>2</sup> untersuchten Formen des *Peer Assessment* (Typ, Häufigkeit etc.) lässt sich ein positiver Effekt auf die Leistung der SchülerInnen feststellen. Eine Übersicht aller Moderatoren findet sich im Überblick über alle Einzelbefunde.

**WIE BEWERTET DAS CLEARING HOUSE UNTERRICHT DIESE STUDIE?** Die *Clearing House Unterricht Research Group* bewertet die Metaanalyse anhand der folgenden fünf Fragen und orientiert sich dabei an den Abelson-Kriterien (1995):

**Wie substantiell sind die Effekte?** Insgesamt kommt die Metaanalyse zu einem signifikanten positiven Gesamteffekt von  $g = 0.31$ . Die Größe dieses Effekts zeigt, dass mehr als 60 % der SchülerInnen, die mit *Peer Assessment* gearbeitet haben, eine bessere Lernleistung erzielen als der Durchschnitt der SchülerInnen in den Kontrollbedingungen ohne Assessment oder mit anderen Formen des Assessments. Ergebnisse aus Moderatoranalysen zeigen, dass verschiedene Umsetzungsmöglichkeiten des *Peer Assessment* keine statistisch bedeutsamen Unterschiede in diesem Effekt ausmachen. Somit deuten die Befunde insgesamt darauf hin, dass *Peer Assessment* unabhängig von der jeweiligen Umsetzung positiv auf die Leistung der SchülerInnen wirkt. Erfahren Sie mehr über die Einschätzung von Effektstärken in unserem Handout.

**Wie differenziert sind die Ergebnisse dargestellt?** Die AutorInnen berücksichtigen in ihren Analysen verschiedene Bildungsstufen (Primarstufe, Sekundarstufe und tertiärer Bildungsbereich) und Schulfächer. Zwischen den Bildungsstufen und Schulfächern zeigen sich jedoch keine signifikanten Unterschiede hinsichtlich der Wirksamkeit von *Peer Assessment*. Sowohl SchülerInnen als auch Studierende scheinen in verschiedenen Fächergruppen und bei unterschiedlichen Umsetzungsmöglichkeiten von *Peer Assessment* zu profitieren. Das untersuchte Leistungskriterium wird zwar bei der Studienselektion differenziert in schriftliche und praktische Aufgaben, in den Analysen allerdings nicht getrennt untersucht.

<sup>2</sup>Lediglich beim Einsatz von Noten zeigt sich ein Unterschied: Während der Einsatz von Noten beim *Peer Assessment* bei Studierenden einen signifikanten positiven Effekt auf die Leistung hat ( $g = 0.55$ ), konnte dieser Effekt nicht für SchülerInnen der Primar- und Sekundarstufe gezeigt werden ( $g = 0.002$ , n.s.)

**Wie verallgemeinerbar sind die Befunde?** Die Ergebnisse der Metaanalyse zeigen für nahezu alle Umsetzungsmöglichkeiten von *Peer Assessment* positive Effekte auf die Leistung. Dabei konnten keine statistisch bedeutsamen Unterschiede für verschiedene Fächer oder Bildungsstufen festgestellt werden. Grundsätzlich kann also davon ausgegangen werden, dass der positive Effekt von *Peer Assessment* auf verschiedene Kontexte verallgemeinert werden kann.

Einschränkend ist festzuhalten, dass die geographische Lage nicht als Moderator mitberücksichtigt wurde, weswegen keine Aussage zur Generalisierbarkeit über verschiedene Länder hinweg getroffen werden kann. Auch für differenzierte Aussagen zur Wirksamkeit verschiedener Umsetzungsmöglichkeiten von *Peer Assessment* speziell in der Sekundarstufe und den mathematischen-naturwissenschaftlichen Fächergruppen bedarf es weiterer experimenteller Studien.

**Was macht die Metaanalyse wissenschaftlich relevant?** Die Metaanalyse von Double und KollegInnen ist wissenschaftlich bedeutsam, da sie die Wirksamkeit von *Peer Assessment* für den Lernerfolg systematisch und verlässlich in Ergänzung zu bereits existierenden qualitativen Übersichts- und Literaturarbeiten einschätzt (Dochy et al., 1999; Topping, 1998).

Darüber hinaus zeigt die Metaanalyse zwei wichtige Forschungsdesiderate auf: Sie macht zum einen sichtbar, dass experimentelle Untersuchungen, in denen Effekte des Bewertens und des Bewertetwerdens auf die Leistung differenziert betrachtet werden, bisher fehlen. Die Berücksichtigung dieser jeweiligen Rolle beim *Peer Assessment* ist wichtig, um spezifischere Aussagen über die lernförderlichen Mechanismen treffen zu können. Zudem zeigt die Metaanalyse, dass die SchülerInnen in den berücksichtigten Studien überwiegend als ganze Klassen in Experimental- und Kontrollbedingung, also quasi-experimentell zugeteilt wurden. Für die Zukunft bräuchte es strengere Studiendesigns mit randomisierter Zuweisung der SchülerInnen auf Individualebene (experimentelle Studiendesigns).

**Wie methodisch verlässlich sind die Befunde?** Die Transparenz und Begründung des methodischen Vorgehens entspricht zum Großteil den Kriterien gängiger Anforderungskataloge (z.B. APA Meta-Analysis Reporting Standards). Insbesondere die Studiensuche und -selektion sind vorbildlich dokumentiert. Im Bereich der Kodierung und Analyse der Primärstudien wären für eine bessere Nachvollziehbarkeit genauere Angaben erforderlich – z.B. Details zu den Stichproben, den Designs und zu den Erhebungsinstrumenten der berücksichtigten Studien. Detailliertere Informationen zur methodischen Beurteilung sind dem Rating Sheet zu entnehmen.

**FAZIT FÜR DIE UNTERRICHTSPRAXIS.** Die Befunde der vorliegenden Metaanalyse weisen darauf hin, dass *Peer Assessment* eine lernförderliche Ergänzung für den Unterricht darstellen kann. Der positive Effekt von *Peer Assessment* auf die Leistung ist über verschiedene Altersstufen, Fächer und weiteren Bedingungen hinweg stabil. Dies deutet daraufhin, dass *Peer Assessment* in vielfältiger Form wirksam in den Unterricht implementiert werden kann.

Für die SchülerInnen bietet *Peer Assessment* die Möglichkeit, sich kritisch mit dem Lerninhalt auseinanderzusetzen und die eigenen Leistungen zu reflektieren. Anhand der berück-

sichtigten Einzelstudien lässt sich allerdings auch feststellen, dass die bisherige Forschung zum *Peer Assessment* noch recht unspezifisch ist. So fehlen Untersuchungen, die die verschiedenen Rollen beim *Peer Assessment* oder die konkreten Kontexte mit einbeziehen, um hier evidenzbasierte Empfehlungen zur konkreten Umsetzung aussprechen zu können. Um die Effektivität unterschiedlicher Umsetzungsmöglichkeiten und lernförderliche Mechanismen noch besser zu verstehen, sind daher weitere Forschungsbemühungen notwendig.

### STUDIENBEISPIEL

Wang und KollegInnen (2017) untersuchen in ihrer Studie die Wirksamkeit von *Peer Assessment* im Informatikunterricht der 9. Klasse. In den zehn Wochen der Untersuchung sollten die SchülerInnen die Programmiersprache »Scratch« lernen. Die Stichprobe bestand aus vier Schulklassen mit insgesamt 166 SchülerInnen. Alle vier Klassen erhielten eine Einführung in »Scratch« und die Aufgabe, ein eigenes Projekt zu programmieren.

In zwei Klassen sollten sich die SchülerInnen anschließend gegenseitig Feedback auf ihre individuellen Projekte geben (Experimentalgruppe mit 80 SchülerInnen, *Peer Assessment*). Dafür standen ihnen klare Bewertungskriterien (z.B. zu Thema und Inhalt oder Gestaltung) und Bewertungsstufen (von 1 bis 4) zur Verfügung. In den anderen beiden Klassen erhielten die SchülerInnen Feedback von der Lehrkraft (Kontrollgruppe mit 86 SchülerInnen).

Um die Wirksamkeit von *Peer Assessment* im Vergleich zur Bewertung durch die Lehrkraft zu überprüfen, nahmen alle SchülerInnen an einem Test zu Programmierkenntnissen und -fähigkeiten sowie Fragebögen zum kritischen Denken und ihren Einstellungen teil – einmal vor Beginn (Prätest, zur Kontrolle der Eingangsvoraussetzungen) und zum Abschluss des Kurses (Posttest). Zusätzlich erhielten die SchülerInnen zu beiden Zeitpunkten Programmieraufgaben, um ihre praktischen Programmierfertigkeiten zu testen. Im Ergebnis erzielte die Experimentalgruppe in allen Bereichen des Posttests signifikant bessere Ergebnisse als die Kontrollgruppe.

## REFERENZEN.

- Abelson, R. P. (1995). *Statistics as principled argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Ahn, S., Ames, A. J. & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research*, 82 (4), 436–476.
- Dochy, F., Segers, M. & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher Education*, 24 (3), 331–350.
- Double, K. S., McGrane, J. A. & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32 (2), 481–509.
- Li, H., Xiong, Y., Zang, X., L. Kornhaber, M., Lyu, Y., Chung, K. S. & K. Suen, H. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment & Evaluation in Higher Education*, 41 (2), 245–264.
- Sanchez, C. E., Atkinson, K. M., Koenka, A. C., Moshontz, H. & Cooper, H. (2017). Self-grading and peer-grading for formative and summative assessments in 3rd through 12th grade classrooms: A meta-analysis. *Journal of Educational Psychology*, 109 (8), 1049.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68 (3), 249–276.
- Wang, X.-M., Hwang, G.-J., Liang, Z.-Y. & Wang, H.-Y. (2017). Enhancing students' computer programming performances, critical thinking awareness and attitudes towards programming: An online peer-assessment attempt. *Journal of Educational Technology & Society*, 20 (4), 58–68.

## LINKS.

Zur Metaanalyse von [Double et al, 2020](#)  
Zum Studienbeispiel von [Wang et al, 2017](#)

## ZITIEREN ALS.

Diery, A., Knogler, M. & CHU Research Group (2022). Wenn SchülerInnen sich gegenseitig beurteilen: Wie lernwirksam ist Peer Assessment im Unterricht? [www.clearinghouse-unterricht.de](http://www.clearinghouse-unterricht.de), *Kurzreview* 28.



Dieses Kurzreview ist lizenziert unter einer [Creative Common Namensnennung - Keine Bearbeitungen 4.0 International Lizenz](#). Verwendung und Verbreitung unter Namensnennung erlaubt, keine Veränderungen gestattet.