

## Effektstärken\*

### WAS SIND EFFEKTSTÄRKEN?

Effektstärken (ES) werden in der empirischen Bildungsforschung oft verwendet, um die Wirksamkeit von Bildungsinterventionen zu beschreiben. Es sind statistische Maßzahlen, die den Unterschied zwischen zwei Gruppen beschreiben und so beispielsweise als Indikator für die Effektivität einer bestimmten Unterrichtsmethode dienen können.

### WIE WIRD EINE EFFEKTSTÄRKE BERECHNET?

Die Grundlage einer Effektstärkenberechnung ist eine Differenz zwischen mindestens zwei Messungen: die durchschnittliche Leistung der Experimentalgruppe (nach bestimmter Methode unterrichtet) wird mit der durchschnittlichen Leistung der Kontrollgruppe verglichen (ohne diese Methode unterrichtet).

$$ES = \frac{[Mittelwert_1 - Mittelwert_2]}{Standardabweichung}$$

Abb. 1: Allgemeine Formel Effektstärken.

In der Formel zur Berechnung der Effektstärke steht im Zähler die Differenz der beiden Mittelwerte (siehe Abb. 1). Im Nenner steht ein Maß für die Streuung der Werte in den Gruppen (Standardabweichung). Durch das Teilen durch die Standardabweichung (auch: »Standardisierung«), erhält man ein standardisiertes Maß unabhängig von spezifischen Einheiten wie zum Beispiel der Punktezahl in einem Leistungstest. Dies ermöglicht einen Vergleich verschiedener Effektstärken über unterschiedliche Untersuchungen hinweg. Auch wenn es im Detail verschiedene Verfahren der Berechnung gibt (Cohen's *d*, Hedges' *g*), folgen alle demselben Prinzip »standardisierter Mittelwertsunterschiede«.

### WIE LASSEN SICH EFFEKTSTÄRKEN VERANSCHAULICHEN? EIN FIKTIVES BEISPIEL

Wir betrachten eine Schulklasse, in der eine Intervention im Mathematikunterricht eingesetzt wird. Die SchülerInnen werden in zwei Gruppen aufgeteilt, wobei eine ein Kopfrechenttraining absolviert (Experimentalgruppe) und die andere nicht (Kontrollgruppe). Die Verteilungen der anschließend gemessenen Kopfrechenleistung beider Gruppen sind in Abb. 2 dargestellt. Die beiden Verteilungen zeigen eine hohe Überlappung und die Verteilung der Experimentalgruppe (grün) ist nur leicht nach rechts verschoben. Der kleine Abstand zwischen den Verteilungen deutet auf eine kleine Effektstärke hin.

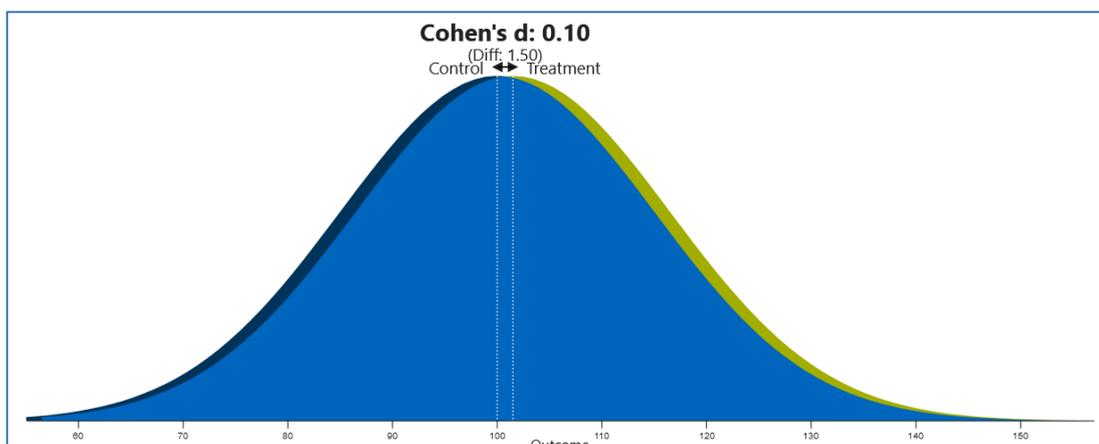


Abb. 2: Grafische Darstellung Cohen's *d* = 0.1

\*Die Ausführungen in diesem Handout beziehen sich auf Effektstärken in experimentellen Studien. Diese bilden die Grundlage für die ausgewählten Metaanalysen und Kurzreviews des Clearing House Unterricht.

Angenommen, statt des Kopfrechentrainings hätte die Experimentalgruppe ein intensives Tutoring in Mathe erhalten. Für dieses Szenario sind die Verteilungen der Mathematiknoten der beiden Gruppen in *Abb. 3* dargestellt.

Die Verteilung der Experimentalgruppe ist nun im Vergleich zur Kontrollgruppe stark nach rechts verschoben, was auf eine größere Effektstärke hindeutet. Die Berechnung über die Mittelwerte und Standardabweichungen der Verteilungen können diesen Unterschied quantifizieren.

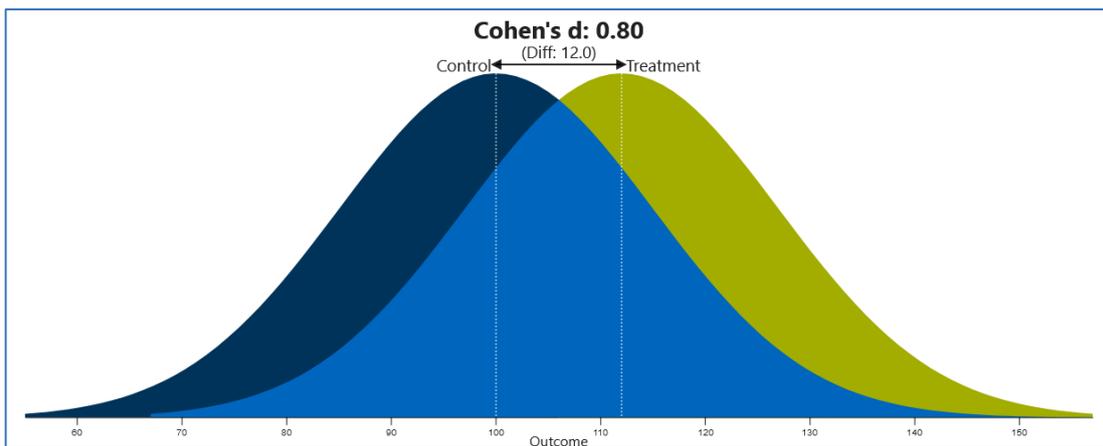


Abb. 3: Grafische Darstellung Cohen's  $d = 0.8$

EINE ÜBERSICHT

Cohen's $d$	Cohen's $U_3$
0	50 %
0.1	54 %
0.2	57,9 %
0.3	61,8 %
0.4	65,5 %
0.5	69,1 %
0.6	72,6 %
0.7	75,8 %
0.8	78,8 %
0.9	81,6 %
1	84,1 %

Tab.1: Umrechnung Cohen's  $d$  zu Cohen's  $U_3$

Zur einfachen Veranschaulichung gebräuchlicher Effektstärkenmaße wie Cohen's  $d$  kann zum Beispiel der Wert Cohen's  $U_3$  als Referenz dienen. Dieser drückt aus, wie viel Prozent der Experimentalgruppe besser als der Durchschnitt der Kontrollgruppe sind. Ein Cohen's  $U_3$  Wert von 69,10% bedeutet zum Beispiel, dass 69,10% der Personen in der Experimentalgruppe über dem Mittelwert der Kontrollgruppe liegen (siehe *Abb.4*, helle Fläche). Für jeden Cohen's  $d$ -Wert kann zur Veranschaulichung ein entsprechender Cohen's  $U_3$ -Wert berechnet werden (siehe *Tabelle 1*).

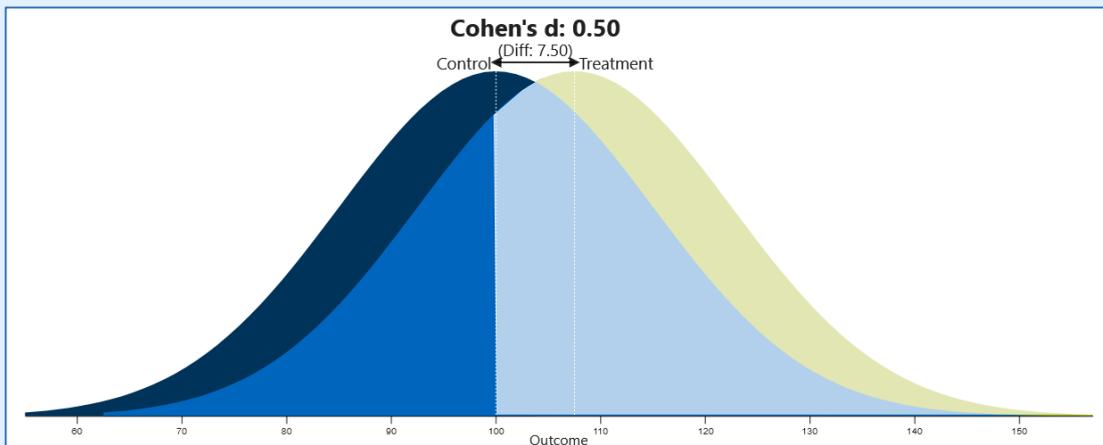


Abb. 4: Grafische Darstellung Cohen's  $U_3 = 69,10\%$

KANN MAN EFFEKTSTÄRKEN VERSCHIEDENER STUDIEN EINFACH MITEINANDER VERGLEICHEN?

Standardisierte Effektstärken haben den Vorteil, dass sie keine spezifische Maßeinheit haben und so prinzipiell unterschiedliche Maßnahmen vergleichbar machen. Ein simpler Vergleich der Effektstärken ist dennoch nicht immer sinnvoll. Das liegt zum einen an qualitativen Unterschieden der Interventionen und zum anderen am Zustandekommen der jeweiligen Effektstärke. So können Maßnahmen beispielsweise von einfachen Hervorhebungen in Arbeitstexten bis hin zu mehrwöchigen Unterrichtsprogrammen reichen. Werden ausschließlich die Effektstärken der beiden Maßnahmen verglichen, dann werden zentrale Unterschiede der Maßnahme – wie beispielsweise der Aufwand oder die Komplexität – nicht berücksichtigt.

Zudem bilden Effektstärken nicht ausschließlich die Effektivität einer Intervention ab, sondern sind maßgeblich von einer Reihe weiterer Faktoren einer Studie abhängig. Dazu gehören beispielsweise die Eigenschaften der Stichprobe, die Wahl des Messzeitpunkts, die Güte des Testinstruments und die Behandlung der Kontrollgruppe. Das bedeutet, dass abhängig von der Art der Studie ein und dieselbe Intervention unterschiedlich große Effektstärken produzieren kann.

Aus diesem Grund sollten bei der Interpretation von Effektstärken und dem Vergleich zwischen Interventionen auf Basis von Effektstärken qualitative und kontextuelle Merkmale immer berücksichtigt und diskutiert werden.

### KLEIN, MITTEL, GROSS?

Konventionelle Einteilungen von Effektstärken (Daumenregeln) wie kleine, mittlere oder große Effektstärken sind einfach und daher oft attraktiv. Entscheidungen über die Implementation einer Maßnahme rein auf Basis von Effektstärken und Daumenregeln sind jedoch nicht zu empfehlen. Zum einen sind Effektstärken oft nicht vergleichbar (siehe vorheriger Absatz), zum anderen können sich auch Maßnahmen mit geringen Effektstärken im Einsatz lohnen.

So könnte man sich beispielsweise für eine Intervention mit geringer Effektstärke entscheiden, wenn diese einfach (Aufwand) und günstig (Kosten) zu implementieren ist und viele SchülerInnen erreichen kann (Skalierbarkeit). Grundsätzlich ist zu empfehlen, differenzierte Einschätzungen vorzunehmen und weitere Faktoren (Aufwand, Kosten, Skalierbarkeit) bei der Beurteilung von Bildungsinterventionen zu berücksichtigen.

### Hinweis

#### Statistische Signifikanz

Zusammen mit Effektstärken wird oft die statistische Signifikanz berichtet. Statistische Signifikanz ist von der Effektstärkenangabe zu unterscheiden. Sie zeigt an, ob die gefundene Effektstärke nur auf die untersuchten Personen zutrifft, oder darüber hinaus verallgemeinert werden kann (siehe [Inferenzstatistik](#)).

### REFERENZEN.

- Bakker, A., Cai, J., English, L., Kaiser, G., Mesa, V., & Van Dooren, W. (2019). Beyond small, medium, or large: Points of consideration when interpreting effect sizes. *Educational Studies in Mathematics*, 102(1), 1-8.
- Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher*, 45(5), 283-292.
- Coe, R. (2002). It's the effect size, stupid: What effect size is and why it is important. *Paper presented at the Annual Conference of the British Educational Research Association*, University of Exeter, England, 12-14 September 2002
- Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child development perspectives*, 2(3), 172-177.
- Kraemer, H. C., Morgan, G. A., Leech, N. L., Gliner, J. A., Vaske, J. J., & Harmon, R. J. (2003). Measures of clinical significance. *Journal of the American Academy of Child & Adolescent Psychiatry*, 42(12), 1524-1529.
- Kraft, M. A. (2020). Interpreting effect sizes of education interventions. *Educational Researcher*, 49(4), 241-253.
- Lipsey, M. W., & Wilson, D. B. (2017). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *Quantitative Methods in Criminology*, 65-93.
- Lortie-Forgues, H., Sio, U. N., & Inglis, M. How should educational effects be communicated to teachers? *Educational Researcher*.
- Simpson, A. (2018). Princesses are bigger than elephants: Effect size as a category error in evidence-based education. *British Educational Research Journal*, 44(5), 897-913.
- Simpson, A. (2020). On the misinterpretation of effect size. *Educational Studies in Mathematics*, 103(1), 125-133.