

Statistik II

Metaanalyse



Troja (2004). Warner Bros. Pictures.

Überblick

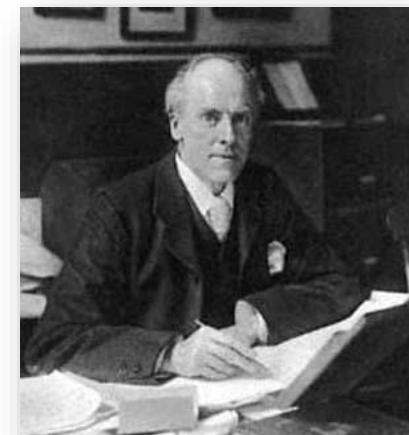
- Einführung
- Vorteile
- Vorgehensweisen
- Probleme

Einführung

- **Metaanalyse:** Zusammenfassung des aktuellen Forschungsstandes zu einer Fragestellung durch statistische Aggregation empirischer Einzelergebnisse inhaltlich homogener Primärstudien
 - **Primärstudien:** Empirische Studien, welche die Untersuchungseinheiten der Metaanalyse darstellen
 - **Homogenität:** Primärstudien untersuchen vergleichbare Fragestellungen
 - **Aggregation:** Zusammenfassung der Primärstudienergebnisse
- **Beispiel:** Metaanalyse zum Signalisierungseffekt: Verbessern Signalisierungen die Lernleistungen und falls ja, wie stark?
- **Zusammenfassung:** Zumeist über die Effektstärken der Studien als AV
- **Populationseffekt:** Einzeleffekte werden zur Schätzung des „wahren“ Populationseffektes zusammengefasst

Einführung (Döring & Bortz, 2015; Sedlmeier & Renkewitz, 2018)

- **Historisch**
 - 1904: Erste bekannte quantitative Ergebniszusammenfassung durch Pearson: Einfache Mittelung von Korrelationen
 - 1976: Erstmalige Verwendung des Begriffs „Metaanalyse“ durch Glass
- **Studentyp:** Metaanalysen kein einheitliches statistisches Verfahren, sondern ein „Studentyp“
- **Metaanalyse vs. narratives Review**
- **Narratives Review:** Verbale Zusammenfassung und Bewertung des aktuellen Forschungsstandes zu einem Thema durch Aufarbeitung und Verdichtung der vorliegenden Literatur



Karl Pearson



Gene Glass

Metaanalyse vs. narratives Review

Um welche Art der Zusammenfassung handelt es sich nach dem unten abgebildeten Abstract?

- A: Metaanalyse
- B: Narratives Review

Research on the use of software programs and tools such as pedagogical agents has peaked over the last decade. Pedagogical agents are on-screen characters that facilitate instruction. [REDACTED] examined the effect of using pedagogical agents on learning by reviewing 43 studies involving 3,088 participants. Analysis of the results indicated that pedagogical agents produced a small but significant effect on learning. The overall mean effect size was moderated by the contextual and methodological features of the studies. The findings revealed that the use of pedagogical agents were more beneficial for K-12 students than post-secondary students. Pedagogical agents that communicated with students using on-screen text facilitated learning more effectively than agents that communicated using narration. The findings of this study have implications for advancing theory and practice, as well as highlighting productive future directions for research.

Quelle: Schroeder, Adesope und Gilbert (2013)

Effektmodelle von Metaanalysen (Döring & Bortz, 2015; Sedlmeier & Renkewitz, 2018)

- **Modell fester Effekte (engl. fixed-effect model):** Annahme, dass sämtliche Studien exakt denselben Populationseffekt erfassen und die Unterschiedlichkeit der Effektgrößen zwischen den Primärstudien auf Stichprobenfehler zurückzuführen ist
- **Modell zufallsbedingter Effekte (engl. random-effects model):** Annahme, dass jeder Primärstudie (aufgrund der Unterschiedlichkeit der Studien) eine andere Populationseffektgröße zugrunde liegen kann
- **Modell gemischter Effekte (engl. mixed-effects model):** Modell kombiniert die beiden Modelle

Vorteile und Vorgehensweisen (Rustenbach, 2003)

- **Vorteile von Metaanalysen**
 - Überblick über uneinheitliche Befunde
 - Erhöhung der Power
 - Erhöhung der Validität
 - Entscheidungsgrundlage
- **Ablauf einer Metaanalyse**



- **Ablauf mit einer Primärstudie vergleichbar**



Forschungsfrage (vgl. Rustenbach, 2003)

- **Festlegung relevanter Parameter im Vorfeld**
 - Zusammenhang zwischen UV und AV
 - Auswahlrahmen für Literaturrecherche (z. B. Sprache, Zeitraum)
 - Eher selten: Methodische Mindeststandards (z. B. bezüglich Power, Vollständigkeit der Daten)
 - Potentielle Moderatorvariablen
- **Potentielle Moderatorvariablen:** Inhaltliche oder methodische Merkmale von Studien, welche sich auf die Größe des gefundenen Zusammenhangs auswirken könnten
- **Beispiele für Moderatorvariablen:** Lernervorwissen, räumliches Vorstellungsvermögen, Lernzeit

Sammlung von Primärstudien (vgl. Sedlmeier & Renkewitz, 2018)

- **Vorteile einer umfassenden Literaturrecherche**
 - Je mehr Studien berücksichtigt werden, desto höher die Power
 - Verringerte Gefahr der Ergebnisverzerrung
- **Recherchestrategien**
 - Stichwortsuche im Netz und in Datenbanken (z. B. Google Scholar, Psychinfo, Psycharticles, ERIC)
 - Systematische Durchsicht von Fachzeitschriften & Bibliographien
 - Durchsicht der Literaturverzeichnisse relevanter Artikel
 - Anfragen über Fachgruppenverteiler
 - Kontaktaufnahme mit Experten
 - Möglichst auch Berücksichtigung sog. „grauer Literatur“

Kodierung und Bewertung (Döring & Bortz, 2015)

- **Datenaufbereitung:** Umfasst u. a. Brauchbarkeitsprüfung, Kodierung der Moderatorvariablen und Umkodierung der Kennwerte
- **Brauchbarkeitsprüfung:** Bewertung der Studien hinsichtlich inhaltlicher Angemessenheit und methodischer Standards (ggf. Ausschluss)
- **Kodierung der Moderatorvariablen:** Beurteilung der Studien bezüglich der Ausprägungen in den Moderatorvariablen
- **Umkodierung der Kennwerte:** Zusammenfassung der Effektgrößen und Überführung in eine einheitliche Größe
 - Gruppendifferenzmaße (z. B. Cohens d , Hedges $g+$)
 - Zusammenhangsmaße (z. B. Δ -Maß, Korrelationskoeffizient r)
 - Varianzaufklärungsmaße (z. B. r^2 , η^2)
 - Risikomaße (z. B. Odds Ratio OR oder relatives Risiko RR)

Datenanalyse (Döring & Bortz, 2015)

- **Signifikanztest für den Gesamteffekt**
 - Gewichtung der Effekte der Primärstudien an ihren Stichprobengrößen bzw. Standardfehlern
 - Ermittlung eines Konfidenzintervalls für die geschätzte Effektgröße
- **Homogenitätstest:** Prüft die Wahrscheinlichkeit, dass Effekte der Primärstudien auf denselben zugrundeliegenden Populationseffekt zurückgehen
 - Homogenitätstest mit der Teststatistik Q nach Shadish und Haddock (1994) verbreitetes Verfahren
 - Bei nicht signifikantem Homogenitätstest Annahme, dass ein gemeinsamer Populationseffekt zugrunde liegt

Homogenitätstest

Welche Aussage trifft zur unteren Textpassage zu?

- A: Die Teststatistik Q deutet darauf hin, dass keine Moderatoreffekte vorliegen.
- B: Die Teststatistik Q deutet darauf hin, dass Moderatoreffekte vorliegen.
- C: Weder Antwort A noch Antwort B sind korrekt.

The average effect of conversational style on participants' interest in the topic was small in size ($d=0.15$), and plausible magnitudes for the effect ranged from a small negative effect to a large positive effect (95% CI = -0.13 to 0.44); thus, hypothesis 3—conversational instructional text generates higher levels of interest—was not supported. The homogeneity statistic was not statistically significant ($Q=7.52$, $df=5$, $p=0.185$), but this result should be viewed with caution given the small number of effects available for analysis.

Quelle: Ginns, Martin und Marsh (2013)

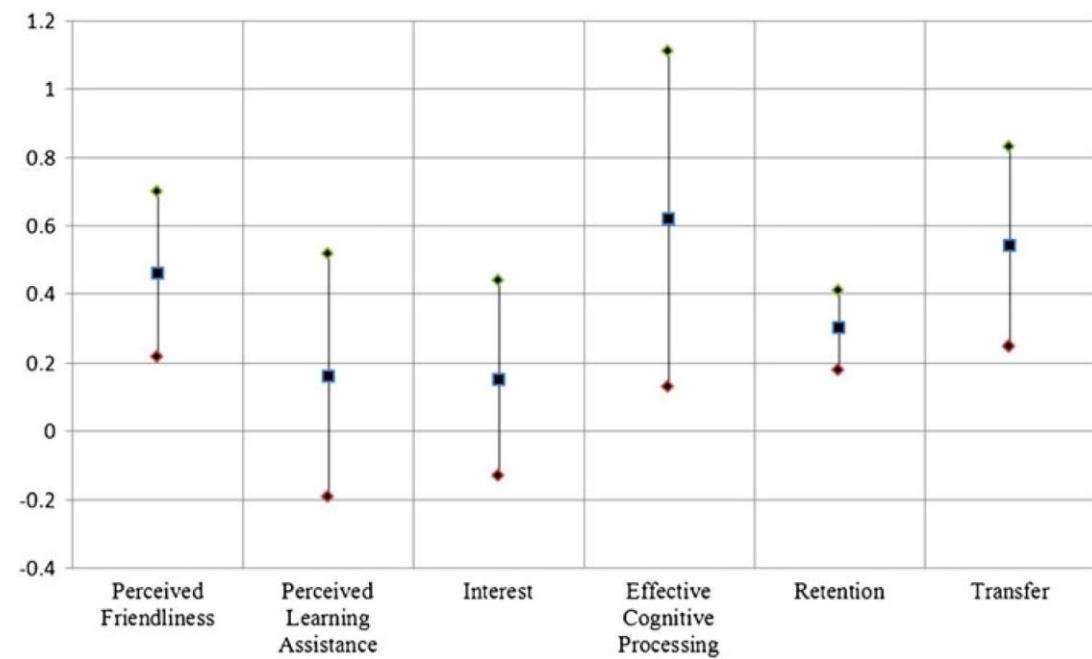
Datenanalyse

- **Moderatorvariablenanalyse**
 - Bei signifikantem Homogenitätstest Annahme von Teilgruppen mit unterschiedlichen Populationseffekten
 - Ermittlung der Bedeutsamkeit von Moderatorvariablen mittels multipler Regression (bzw. mehrfaktoriellen Varianzanalysen) mit den Prädiktoren als Moderatorvariablen und der Effektgröße als Kriteriumsvariable
 - Bei einer Vielzahl von Moderatorvariablen Gruppierung auch clusteranalytisch möglich
- **Maßnahmen zur weiteren Verringerung der Heterogenität:** Z. B. durch Ausschluss von Ausreißern (z. B. Ausreißertest nach Grubbs) oder Ermittlung weiterer Moderatorvariablen

Präsentation und Interpretation

- **Präsentation:** Übersichtliche Darstellung der relevanten Parameter der Einzelstudien bzw. Subgruppen in tabellarischer und/oder graphischer Form
- **Interpretation:** Diskussion von Schwächen und Implikationen sowie Ansatzpunkte für weitere Forschung

Abbildung der Effektgrößen (d) und Konfidenzintervalle (95%)



Quelle: Ginns, Martin und Marsh (2013)

Probleme (Döring & Bortz, 2015)

- Notwendige Mindestanzahl an Primärstudien
- Publikationsverzerrung
- Uniformitätsproblem
- „Garbage-in-Garbage-out“-Problem
- Abhängige Untersuchungsergebnisse
- Unvollständige Daten

Notwendige Mindestanzahl an Primärstudien (Döring & Bortz, 2015)

- **Anzahl an Primärstudien:** Aussagekraft der Metaanalyse steigt mit Anzahl an Primärstudien
- **Mindestanzahl**
 - Metaanalyse kann bereits bei zwei Studien sinnvoll sein
 - Forschungspragmatisch allerdings erst ab einer Mindestanzahl an Studien lohnenswert (z. B. ab 12 Studien)
 - Mindestanzahl auch von der Fragestellung abhängig
- **A-priori-Teststärkeanalyse:** Auch für Metaanalysen möglich

Publikationsverzerrung (Döring & Bortz, 2015)

- **Publikationsverzerrung (engl. publication bias):** Erwünschte und signifikante Ergebnisse werden bevorzugt veröffentlicht
- **Gegenmaßnahmen**
 - (Änderung der Herausgeberpolitik der Zeitschriften)
 - Maßnahmen zur Beschaffung unveröffentlichter Studien
 - Verfahren zur Abschätzung der Verzerrung (z. B. Fail-Safe-N-Methode, Funnel-Plot, Rangkorrelationen)
 - Verfahren zur Korrektur der Verzerrung (z. B. Trim-and-Fill-Prozedur)
- **Fail-Safe-N-Methode:** Abschätzung, wie viele nicht signifikante Ergebnisse vorliegen müssten, damit aus einem signifikanten ein nicht signifikanter Gesamteffekt wird

Fail-Safe-N-Methode

Welche Aussage steht an der verdeckten Stelle in der unten abgebildeten Textpassage?

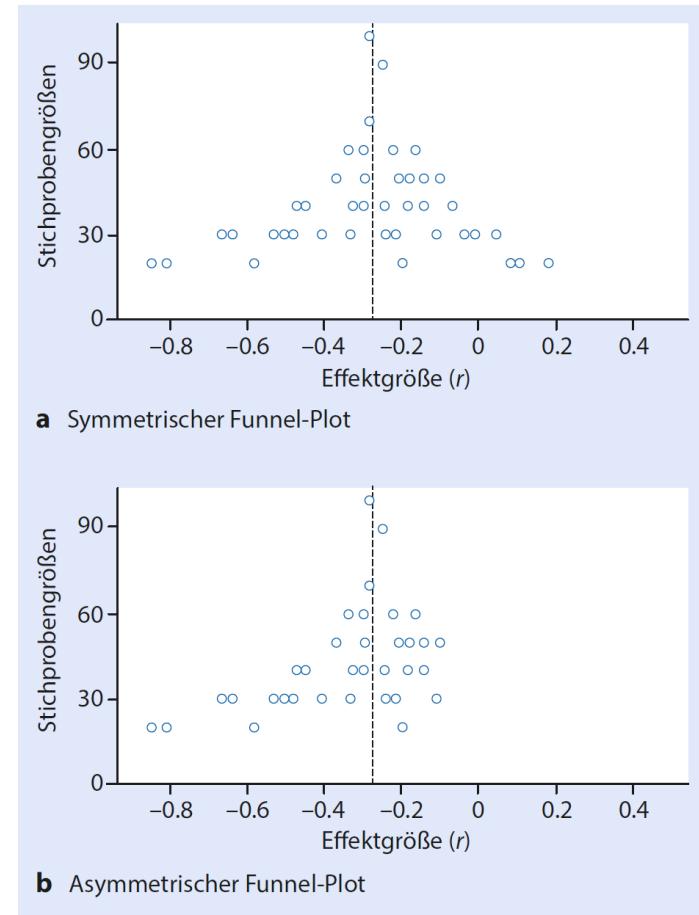
- A: (published) studies with a significant
- B: (unpublished) studies with a significant
- C: (published) studies with a non-significant
- D: (unpublished) studies with a non-significant

Since only published studies were included into the meta-analysis, it was necessary to estimate the publication bias. The *fail-safe N* in this meta-analysis was 937, meaning that there would need to be at least 937 [redacted] [redacted] signaling effect to render the overall signaling effect identified in the present meta-analysis non-significant. According to a guideline by [Rosenthal \(1991\)](#) the result of the present meta-analysis seems to be robust to the publication bias since the *fail-safe N* exceeds $5 \times k + 10$ ($5 \times 45 + 10 = 235 < 937$). In addition, the [Begg and Mazumdar \(1994\)](#) rank correlation was non-significant, $\tau(N = 45) = 0.03, p = .79$, indicating that a publication bias was unlikely to be present in the data.

Quelle: Richter, Scheiter und Eitel (2016)

Funnel-Plot (Döring & Bortz, 2015; Sedlmeier & Renkewitz, 2018)

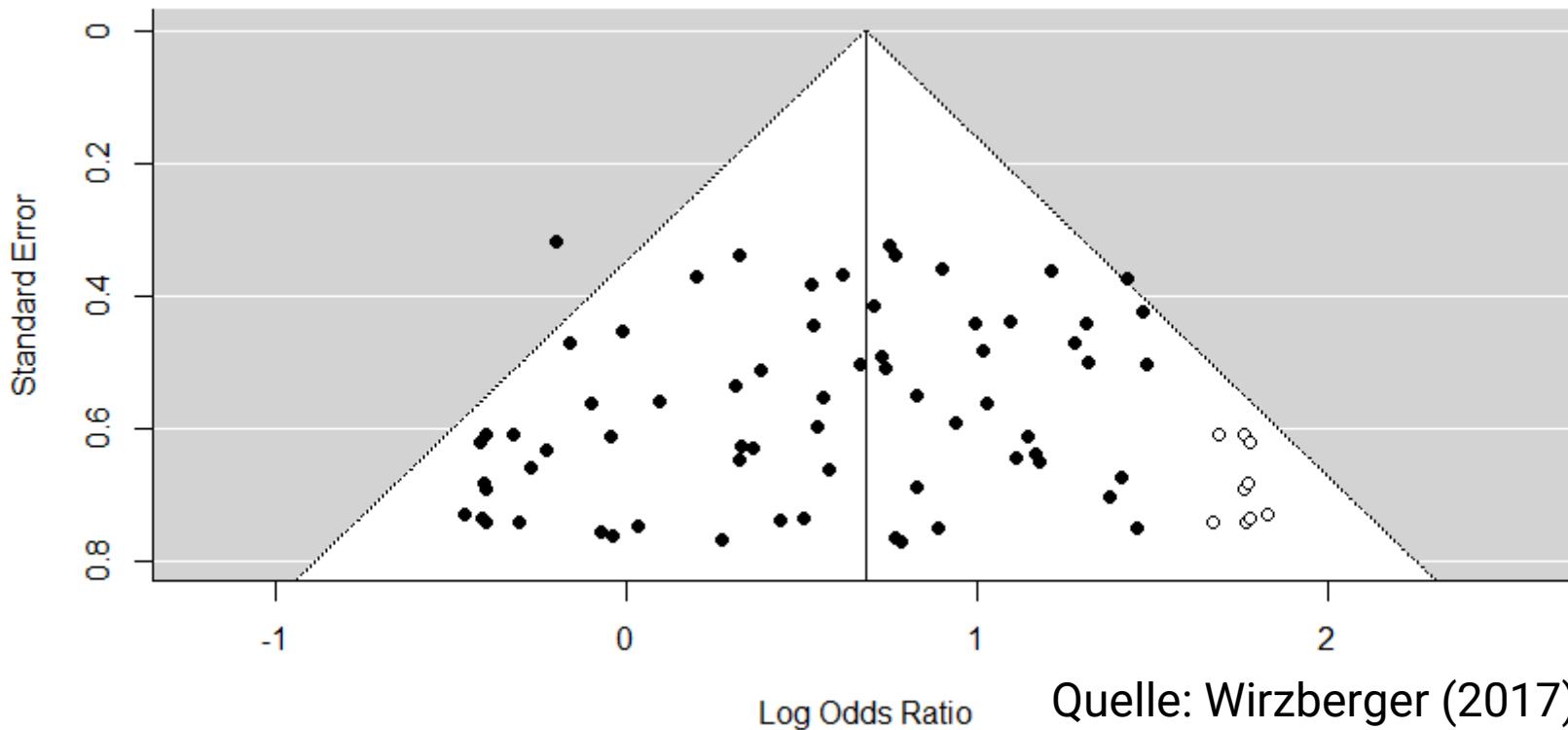
- **Funnel-Plot (auch Trichtergrafik genannt):** Grafisches Verfahren zur Abschätzung der Publikationsverzerrung
- **Vorgehen:** Auf der x-Achse werden die Effektgrößen, auf der y-Achse die Stichprobenumfänge oder Standardfehler der Primärstudien abgetragen
- **Umgedrehter symmetrischer Trichter:** Verteilung sollte einem solchen folgen
- **Signifikanztest** zur Funnel-Plot-Asymmetrie vorhanden
- **Normal-Quantile-Plot** als weiteres graphisches Verfahren



Quelle: Döring und Bortz (2015)

Trim-and-Fill-Prozedur

- **Trim-and-Fill-Prozedur:** Auf dem Funnel-Plot basierendes Imputationsverfahren („Vervollständigungsverfahren“) liefert eine von der Publikationsverzerrung bereinigte Effektgrößenschätzung



Quelle: Wirzberger (2017)

Uniformitätsproblem (Döring & Bortz, 2015)

- „Äpfel-und-Birnen“-Problem (auch Inkommensurabilitäts-Problem genannt): Wie homogen müssen Studien sein, damit sie „in einen Topf geworfen“ werden können?
- Herausforderung: Hohe Schwierigkeit, inhaltlich und methodisch einander hinreichend ähnliche Primärstudien auszuwählen
- Unterschiedliche Standards für UVs und AVs
 - UVs: Hier eher liberaler Maßstab, da leichte Variationen (z. B. in Operationalisierung) Generalisierung des Effektes erlauben bzw. Differenzierung mittels Moderatorvariablenanalysen möglich ist
 - AVs: Hier eher strenge Kriterien, da ein gemeinsames inhaltliches Konstrukt erfasst werden soll

„Garbage-in-Garbage-out“-(GIGO)-Problem (Döring & Bortz, 2015)

- „Garbage-in-Garbage-out“-Problem: Methodisch schwache Primärstudien („garbage in“) führen zwingend zu einer geringeren Aussagekraft der Metaanalyse („garbage out“)
- Lösungsansätze
 - Ausschluss methodisch schwacher Studien anhand von vorab definierten methodischen Mindeststandards
 - Gewichtung von Primärstudien unterschiedlicher methodischer Qualität
 - Berücksichtigung der methodischen Qualität als Moderatorvariable

Abhängige Untersuchungsergebnisse (Döring & Bortz, 2015)

- **Problem:** Wie soll mit mehreren Effektgrößen in einer Primärstudie (mit derselben Stichprobe) umgegangen werden?
 - Einerseits: Unabhängigkeit der Messungen Voraussetzung vieler Metaanalysen
 - Andererseits: Informationsverlust durch Ausschluss von Daten
- **Lösungsansätze**
 - Auswahl und Verwendung der wichtigsten Effektgröße pro Studie
 - Zusammenfassung mittels Median- oder Mittelwertsbildung der abhängigen Effektgrößen
 - Berechnung einer multivariaten Metaanalyse

Unvollständige Daten (Döring & Bortz, 2015)

- Probleme durch fehlende statistische Kennwerte („Missing Data“) in den Primärstudien
 - Unvollständige Datenbasis
 - Systematische Verzerrungen
 - Unterschiedliche metaanalytische Ergebnisse
- Umgang mit unvollständigen Daten
 - Nachfrage bei den AutorInnen der Primärstudie
 - Berechnung fehlender Daten aus den berichteten Kennwerten
 - Regressionsanalytische Schätzung fehlender Werte
 - Konservative Schätzung fehlender Daten (z. B. Nulleffekte bei nicht signifikanten Befunden)
 - Kombinierte Signifikanztests

Beispiele für Metaanalysen in Fachzeitschriften

Abstract This article reviews research on the effects of conversational style on learning. Studies of conversational style have variously investigated “personalization” through changing instances of first-person address to second or third person, including sentences that directly address the learner; including more polite forms of address; and making the views and personality of the author more visible. Meta-analyses provided mixed support for a model of learning processes; statistically reliable average effects were found on self-reports of friendliness ($d=0.46$) and effective cognitive processing ($d=0.62$), but not learning assistance ($d=0.16$) and interest ($d=0.15$). Statistically reliable average effects on retention ($d=0.30$) and transfer ($d=0.54$) learning outcomes supported conversational-style redesigns across a range of potential moderators; the clearest apparent boundary condition for learning outcomes across the moderators under analysis was instructional time, with small, non-significant effects being found in studies longer than 35 min. Recommendations for future investigations are discussed.

Quelle: Ginns, Martin und Marsh (2013)

Research on the use of software programs and tools such as pedagogical agents has peaked over the last decade. Pedagogical agents are on-screen characters that facilitate instruction. This meta-analysis examined the effect of using pedagogical agents on learning by reviewing 43 studies involving 3,088 participants. Analysis of the results indicated that pedagogical agents produced a small but significant effect on learning. The overall mean effect size was moderated by the contextual and methodological features of the studies. The findings revealed that the use of pedagogical agents were more beneficial for K-12 students than post-secondary students. Pedagogical agents that communicated with students using on-screen text facilitated learning more effectively than agents that communicated using narration. The findings of this study have implications for advancing theory and practice, as well as highlighting productive future directions for research.

Quelle: Schroeder, Adesope und Gilbert (2013)

Abstract This meta-analytical review focuses on the role of spatial ability when learning with pictorial visualizations. By means of selective theoretical review and meta-analysis (the latter regarding 27 different experiments from 19 studies), several sub-factors of spatial ability are considered as well as dynamic and non-dynamic, interactive and non-interactive visualizations. An overall effect of $r=0.34$ (95%-CI 0.28 to 0.39) demonstrating a medium advantage for high-spatial-ability learners when working with visualizations is calculated. More importantly, two moderators could be identified: Learners with low spatial ability can be significantly supported by a dynamic instead of a non-dynamic visualization as well as by 3d- instead of 2d-illustrations. Results are discussed in consideration of contemporary theories of multimedia learning.

Quelle: Höffler (2010)

The signaling principle recommends emphasizing relevant aspects of a multimedia message by means of signals (e.g., color coding). We determined the effectiveness of signals that highlight correspondences between text and pictures as well as possible boundary conditions by means of a meta-analysis. To this end, 58 potentially relevant articles were identified in a comprehensive search. After rating the studies based on inclusion criteria and correcting for biases, 27 studies were included in the meta-analysis yielding 45 pairwise comparisons with $N = 2464$ participants. Domain-specific prior knowledge, pacing of the materials, pictorial format, mapping requirements, and distinctiveness of signals were coded as moderators. For transfer and comprehension performance a positive small-to-medium effect size ($r = .17$, 95% CI [0.11, 0.22]) favoring signaled multimedia material was found, which was moderated by prior knowledge. The findings support the effectiveness of the signaling principle in particular for learners with low prior knowledge.

Quelle: Richter, Scheiter und Eitel (2016)

It is assumed that serious games influences learning in 2 ways, by changing cognitive processes and by affecting motivation. However, until now research has shown little evidence for these assumptions. We used meta-analytic techniques to investigate whether serious games are more effective in terms of learning and more motivating than conventional instruction methods (learning: $k = 77$, $N = 5,547$; motivation: $k = 31$, $N = 2,216$). Consistent with our hypotheses, serious games were found to be more effective in terms of learning ($d = 0.29$, $p < .01$) and retention ($d = 0.36$, $p < .01$), but they were not more motivating ($d = 0.26$, $p > .05$) than conventional instruction methods. Additional moderator analyses on the learning effects revealed that learners in serious games learned more, relative to those taught with conventional instruction methods, when the game was supplemented with other instruction methods, when multiple training sessions were involved, and when players worked in groups.

Quelle: Wouters et al. (2013)

Zusammenfassung

- **Metaanalyse:** Zusammenfassung des aktuellen Forschungsstandes zu einer Fragestellung durch statistische Aggregation empirischer Einzelergebnisse inhaltlich homogener Primärstudien
- **Vorteile:** Überblick über uneinheitliche Befunde, Erhöhung der Power & der Validität sowie Nutzung als Entscheidungsgrundlage
- **Vorgehensweise:** Sammlung von Primärstudien, Kodierung und Bewertung, Datenanalyse, Präsentation und Interpretation
- **Probleme:** Notwendige Mindestanzahl an Primärstudien, Publikationsverzerrung, Uniformitätsproblem, „Garbage-in-Garbage-out“-Problem, abhängige Untersuchungsergebnisse und unvollständige Daten

Prüfungsliteratur

- Döring, N., & Bortz, J. (2015). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin: Springer.
 - Metaanalyse (S. 893–943, ohne Rechenbeispiele)

Weiterführende Literatur

- Sedlmeier, P., & Renkewitz, F. (2018). *Forschungsmethoden und Statistik: Ein Lehrbuch für Psychologen und Sozialwissenschaftler* (3. Aufl.). München: Pearson.
 - Metaanalyse (S. 867–904)
- Rustenbach, S. J. (2003). *Metaanalyse: Eine anwendungsorientierte Einführung*. Bern: Huber.
- Shadish, W. R., & Haddock, C. K. (1994). Combining estimates of effect size. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 261–281). New York: Russell Sage Foundation.

Sonstige Literatur

- Ginns, P., Martin, A. J., & Marsh, H. W. (2013). Designing instructional text in a conversational style: A meta-analysis. *Educational Psychology Review*, 25, 445–472.
- Höffler, T. N. (2010). Spatial ability: Its influence on learning with visualizations—a meta-analytic review. *Educational Psychology Review*, 22, 245–269.
- Richter, J., Scheiter, K., & Eitel, A. (2016). Signaling text-picture relations in multimedia learning: A comprehensive meta-analysis. *Educational Research Review*, 17, 19–36.
- Schroeder, N. L., Adesope, O. O., & Gilbert, R. B. (2013). How effective are pedagogical agents for learning? A meta-analytic review. *Journal of Educational Computing Research*, 49, 1–39.
- Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology*, 105, 249–265.