

Anhang 4: Begründung für die Verwendung eines Drei-Ebenen-Modells

Zunächst wurde ein Random-Effects-Modell ohne Moderatoren und mit vier Analyseebenen (Stichprobenvarianz, Effektstärken, berichtete Stichproben, Studienberichte) berechnet. Die Aufteilung der Varianz wurde nach dem Vorbild von Assink & Wibbelink (2016) ermittelt. Nur etwa 1 % der Varianz der standardisierten Mittelwertsdifferenzen ist demnach auf Stichprobenvarianz der einzelnen Effektstärken zurückzuführen. Fast 7 % der Varianz ist zwischen den Effektstärken innerhalb einzelner Stichproben. Zwischen den Stichproben innerhalb von Berichten gibt es fast keine Varianz und 92 % der Varianz ist auf Unterschiede zwischen den 19 Berichten zurückzuführen. Insgesamt sind damit 99 % der Varianz Heterogenität in den wahren Effekten. Das heißt, nur 1 % der Varianz ist auf Stichprobenfehler zurückzuführen und 99 % können mit Unterschieden zwischen den Studien erklärt werden.

Über die Varianzzerlegung hinaus wurden Likelihood-Ratio-Tests durchgeführt, um zu prüfen, ob die Berücksichtigung der einzelnen Analyseebenen im Mehrebenen-Modell zu einer Verbesserung der Modellgüte führt. Dazu wird das Modell mit allen vier Ebenen verglichen mit einem reduzierten Modell, bei dem jeweils eine Varianzkomponente auf 0 gesetzt wird. Ist das reduzierte Modell signifikant schlechter, sollte die entsprechende Analyseebene berücksichtigt werden.

Die Ebene der einzelnen Stichproben in den Berichten zu berücksichtigen, hat zu keiner Modellverbesserung geführt. Daraus folgt, dass ein Modell mit drei Analyseebenen (Stichprobenvarianz der Effektstärken, Varianz zwischen den Effektstärken, Varianz zwischen den Studienberichten) verwendet werden sollte. Dies ist konform mit den Ergebnissen der Varianzzerlegung und auch plausibel, da in etwa der Hälfte der 19 Studien nur Ergebnisse einer Stichprobe berichtet werden, sodass es nur wenige Studienberichte gibt, in denen es auf dieser Ebene überhaupt Variation geben kann.