Effect sizes

Gary Jones and Deborah Netolicky

What's the idea?

An effect size is a "way of quantifying the difference between two groups" (Coe, 2017, p. 339). It allows us to move beyond the question of whether or not an intervention works and toward an understanding of how well it works in different contexts.

What does it mean?

The formula for an effect size is as follows:

Effect size = (mean of experimental group) - (mean of the control group) divided by standard deviation (SD)

The following is an effect size calculation, showing how time of day affects children's learning (Coe, 2002, referring to Dowson, 2002):

» Half of the children in a class were randomly allocated to listen to a story and respond to questions at 9am

» The other half of the children listened to the same story and responded to questions at 3pm

» The children's understanding of the story was assessed by using a test where the number of correct answers was measured out of 20

» The morning group had an average of score of 15.2, whereas the afternoon group had an average score 17.9 (a difference of 2.7)

» The effect size of the intervention is worked out as: (17.9-15.2)/3.3 = 0.8.

What are the implications for teachers?

There is no generally agreed scale for interpreting the size of an effect size: different researchers put forward different scales.

Coe (2002) suggests that a teaching intervention with an effect size of 0.6 would lead to each pupil improving by approximately one GCSE grade.

John Hattie (2008) reviewed over 800 meta-studies and says the average effect size of a range of educational strategies is 0.4.

The Education Endowment Foundation's Teaching and Learning Toolkit (Higgins et al., 2015) also provides some guidance on effect sizes. It says that: effect sizes between -0.01 to 0.18 are low; between 0.19 and 0.44 are moderate; from 0.45 to 0.69 are high; and 0.7+ are very high.

Wiliam (2016) identifies a number of problems with interpreting effect sizes:

» *Age dependence.* Studies involving younger children tend to have a higher effect size than studies with older pupils, as there tends to be more variation in the performance of older pupils than that of younger pupils.

» *Results tend to vary depending on what outcome is measured.* For example, studies using immediate outcomes, such as classroom tests, tend to have larger effect sizes than studies which use standardised national tests as outcomes.

» *Statistical power.* The greater the statistical power of the experiment (i.e. the likelihood of detecting an effect where there is an effect to be detected), the larger the average effect size of statistically significant studies.

WANT TO KNOW MORE?

» Coe R (2002) It's the Effect Size, Stupid: What Effect Size Is and Why It Is Important. Paper presented at the British Educational Research Association annual conference 12: p.14.

» Coe R, Waring M, Hedges L et al. (2017) Research Methods and Methodologies in Education (2nd Edition). London: SAGE.

» Higgins S, Katsipataki M, Coleman R et al. (2015). The Sutton Trust-Education Endowment Foundation Teaching and Learning Toolkit.

» Simpson A (2017) The Misdirection of Public Policy: Comparing and Combining Standardised Effect Sizes. Journal of Education Policy 32 (4): 450-466.

» Slavin R (2016) What Is a Large Effect Size. Available at: http://www.huffingtonpost.com/robert-e-slavin/what-is-a-large-effect-si_b_9426372.html (accessed 16 May 2019).

» Wiliam D (2016) Leadership for Teacher Learning. West Palm Beach: Learning Sciences International.

CHARTERED COLLEGE OF TEACHING