**Effect Size**

This document is intended for anyone who wants to do a calculation to measure effectiveness. It might be particularly helpful to middle and senior leaders.

**What is an Effect Size (ES)?**

An Effect Size is a measure of the **strength of the relationship** between two variables in a dataset. It helps us to understand the magnitude of a result, using a scale that is independent of the unit of measurement being used. This is particularly useful when comparing change for two different kinds of measure.

For example, it does not matter whether we measure the height of a group of students at age 6 and again at age 10, using inches or centimetres or a specially marked rhubarb stalk each time (note - you can’t mix up the units you use e.g., use inches at age 6 and the rhubarb stalk at age 10), the ES to show how much you have grown will be the same. At the same time, you could do the same with the students’ weight (using pounds/ounces or kilograms or gold ingots), and calculate an effect size for weight. If the effect size for height is greater than the effect size for weight, then you can say that the students have grown taller more than they have put on weight. The effect sizes are comparable, but the differences in heights and weights are not comparable.

It is NOT a measure of the correlation between the two variables nor does it indicate statistical significance (i.e., whether the results could have occurred by chance, or not).

**How big is big?**

The guru on effect sizes, Cohen (1976) said that an effect size of about 0.25 was “small”, about 0.5 was “medium”, and about 0.8 was “large”.

Hattie (2009) found that, on average, an educational intervention would have an effect size of 0.4. Therefore, he suggested that you need to be looking for an ES in excess of 0.4, if you want to demonstrate that your intervention was having an effect over and above the average. All of Hattie’s Visible Learning books and materials are based on maximising the “zone of desired effects” – that is, using strategies that are known to have an effect size in excess of 0.4.

**Calculating an ES**

There are a number of methods of doing this, but the one that is most commonly used is with pre- (T1) and post- (T2) assessment data to show whether some sort of intervention has “made a difference”. The calculation is represented as Cohen’s *d*.

It is desirable, but not essential, that you have two results for each student – a pre-test (T1) and a post-test (T2). This is called a matched dataset.

The formula for Cohen’s *d* is:

 *d* = (mean T1 – mean T2)/std deviation

where *mean* is the arithmetic average of the scores (i.e. sum of the scores divided by the number of scores).

There are a number of different versions of the formula depending on which standard deviation you use, so we will pool the standard deviation for T1 and for T2 (we do not have to do that – the calculator we use will do it for us).

Therefore for your T1 data and for your T2 data, you need to calculate the mean, standard deviation and count how many students there are in each set of data.

Here’s how:

For the data set, set up a table at the bottom of the data file – Mean, SD, N.

Calculate the Mean, SD and N for each column using Excel formulae. For the T1 Stanine columns, my data starts in row 2 and goes down to row 248, so in the appropriate cell type:

|  |  |
| --- | --- |
| Calculation for T1 | Type this formula in the cell |
| Mean | =average(e2:e248) |
| SD | =stdev(e2:e248) |
| N | =count(e2:e248) |

This shows the formula for the SD of the T1 Stanine data



Formula appears here, but you …

… and this is what you see at the bottom where you actually type the formula.



… type in the cell down here

Repeat for column F, T2 Stanine.

|  |  |
| --- | --- |
| Calculation for T2 | Type this formula in the cell |
| Mean | =average(f2:f248) |
| SD | =stdev(f2:f248) |
| N | =count(f2:f248) |

All that is required now is a copy of the ES calculator. In this example we use an online calculator.

Go to <http://www.uccs.edu/~faculty/lbecker/> and enter the values in the boxes and Compute.



Notes:

1. Only the mean and SD are required for each of the two data sets.
2. Use the Post-test (T2) values for Group 1, and Pre-test (T1) for Group 2.
3. If the value of Cohen’s *d* is negative, this means that there was no improvement - the Post-test results were lower than the Pre-tests results.

In the example shown, we would report theES as *d*= 0.36. Two decimal places is enough.

**Interpretation**: Yes, the students have improved (we would hope so over the period of a school year) as the average stanine for the class/school increased from 5.17 to 5.86, but as this ES is (slightly) less than 0.40, we have made less of a difference than an average educational intervention (Hattie, 2009).

However, we should really consult Hattie’s *Visible Learning*, and reference this ES (0.36) against the average ES for interventions of this type. For example, if the intervention was about ‘teaching test-taking and coaching’ to raise achievement, then the average effect size of such interventions if d=0.22 (Hattie, 2009, p. 180), so our ES of 0.36 looks pretty good by comparison with that. If however, the intervention was about ‘using meta-cognitive strategies’ to raise achievement, then the average ES for all studies of this kind was 0.69, so our results are well behind the average for interventions using these strategies. Improved, yes, but some way to go!

NB. There are other online calculators available such as:

<http://www.cognitiveflexibility.org/effectsize/>

<http://www.easycalculation.com/statistics/effect-size.php>

<http://www.cedu.niu.edu/~walker/calculators/effect.asp> (this one uses the N values as well)

All of these gave me the same ES for the above data.