Using Effect Size to Assess Impact: Be the 'John Hattie' of Your School

by Chris Birr and Todd Hrenak

In 2009, many of us learned about effect sizes from *Visible Learning* by John Hattie (2009). A simple definition of effect size is "a way of quantifying the size of the difference between two groups" (Coe, 2002). Previously, effect size calculation and interpretation was only seen in research. Now, with a push from John Hattie, effect size evaluations have become more mainstream. Using effect size analyses, school psychologists can provide practical information to address a multitude of school-based research questions. Some examples:

- What was the effect of an individually delivered intervention for first grade students struggling in reading?
- In third grade, what was the effect of a supplemental math program compared to students not provided the program? (if class or school differences exist naturally this allows for program evaluation)
- What is the effect of intervention X when provided to students in grade 3 across the elementary schools in a district?

The purpose here is to provide practicing school psychologists with simple guidance to calculate effect size *and* how to present the data in an objective and unbiased manner. There are great opportunities for school psychologists to help promote and conduct program evaluation and development through the use of data analysis. Calculating and using effect size can be complicated and one must be careful when interpreting this type of data without adequate training on the use of this statistic. In addition, we need to keep seeking improved outcomes for students even with positive effects.

Many of us are experts when reviewing existing research to help shape practice or solve problems to meet needs in our districts. Now, we have another opportunity to conduct our own research and improve outcomes for our own students. We need to remain objective and vigilant to provide high quality, accurate, and unbiased data. Letting the data speak for itself can backfire if the sample size is too small, misinterpreted, or ignored due to lack of understanding or poor explanation. If you calculate it, be ready to carefully and clearly explain the process and results.

This article is intended to give you an introduction and basic understanding of a method to calculate effect size. By the end, it is hoped that the reader would feel comfortable calculating effect sizes according to the method discussed. Keep in mind, there are many ways to calculate and interpret effect size. This article is not sufficient to answer all questions about effect size, but references and resources are included for more information.

How is effect size calculated?

Although there are many types of effect sizes, perhaps the most common definition is the mean difference between two groups, divided by the average standard deviation of the two groups:

$$\text{Effect size} = \frac{\text{[Mean of intervention group] - [mean of class/grade]}}{\text{Pooled standard deviation of grade and intervention group}}$$

In research, the mean of the intervention group would be the mean of the 'experimental' group whereas the mean of the class/grade would be the mean of the 'control' group. For school psychologists, effect size can be calculated using growth scores (fall to spring) from reliable and valid screening assessments. The standard deviation can either be pooled, or an average of grade and intervention group. When pooling or averaging the standard deviations of the groups, this calculation is referred to as Cohen's d. The most reliable method it pooling (using only the standard deviation of the control group ignores part of the population).

Calculating the effect size using only the SD of the control or grade level is called Glass's delta. At times, researchers will use Glass's delta when a large effect was observed from an intervention. The large effect then caused a larger standard deviation skewing the effect size when pooled with the control group.

To calculate effect sizes, one needs:

- Microsoft Excel
- Fall and spring scores from all students in a grade level or class
- Accurate knowledge of students who received the intervention evaluated

Step for calculating effect size:

1. Arrange student data into columns (Fall and Spring for grade and intervention groups)
2. Insert column and calculate difference between spring and fall for each group (Growth)
3. Use the function capability and calculate the SD and mean of each column. (search YouTube to learn how to use functions in MS Excel).
4. Use the obtained numbers and complete the formula. Effect size is typically reported to two decimal places (IES, 2012).
5. Using mean and SD of the national sample, calculate an effect size from intervention to national sample. This adds complexity but allows for a comparison between local and national data. It is possible that the local intervention group will not close the gap with the local grade level yet students are growing more than expected relative to national norms.

How is effect size interpreted?

Hattie's work (2009) provided a definition of effect sizes from 0.4 to 0.7 as moderate and effect sizes of 0.7 and up as in the zone of desired effects. However, this is once again where the research prior to *Visible Learning* was largely based on Cohen's definitions from 1988:

0.2 and below is small
0.5 is moderate
0.8 and up is a large effect.

Hattie and Cohen set the bar for effect size upwards of 0.7 when describing a large effect. Keep in mind, Hattie was working with meta analyses and Cohen was providing general guidelines for research consisting of large samples of data. Cohen's guidelines provide an initial framework for interpreting effect sizes, current perspectives in the field suggest that effect sizes be interpreted in context. A future article will discuss interpretation of effect size based on relative context.

For perspective, What Works Clearinghouse uses the following guideline when interpreting effect size, "effect sizes of 0.25 standard deviations or greater are considered substantially important." Most instructional methods or interventions lead to growth to some degree. Using effect size can help one to understand whether the growth observed is practically meaningful compared to a 'control' group.

What other information should be presented?

Effect size is not a "one-size fits all, complete answer" whether an intervention or practice 'worked' or not. When presenting effect size, consider also presenting the following data:

- What happened to students who were provided an intervention? Did they enter another intervention the next year? Did they continue to demonstrate strong growth or meet benchmarks?
- Fall and spring mean scores for grade, intervention, and a national band (i.e. 25th percentile fall and spring) to allow visual analysis. Is the gap closing between groups? A group of students in an intervention may be demonstrating greater growth than the national sample but will never close the gap on district, grade level peers.
- How many students in each group met the tier 1 target or were proficient in spring? Keep in mind that students in interventions may be starting below average and even aggressive growth will not be enough to allow a gap closure.
- What percentage of students in each group met or exceeded the growth target between fall and spring?
- Was the intervention delivered with fidelity? Did students have adequate attendance?

Conclusion

Effect size can be a powerful piece of information to provide when evaluating practices or interventions in a school or district. Now, rather than relying on effect size only to judge existing research, school psychologists with careful application may be able to use the methods to assist with data based decision making ensuring students receive the highest quality services. When working with any data including effect size, objectivity and accuracy must be maintained during calculation, interpretation, and presentation of the data. Effect size is a powerful measurement tool but school psychologists will need to consider more than just effect size when making decisions.

Thanks to Dr. Ed O'Conner, and Dr. David Klingbeil for their suggestions and review of this article.

References:

Coe R. It's the effect size, stupid: what "effect size" is and why it is important. Paper presented at the 2002 Annual Conference of the British Educational Research Association, University of Exeter, Exeter, Devon, England, September 12–14, 2002. Retrieved from http://www.leeds.ac.uk/educol/documents/00002182.htm.

Cohen., J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, *141*(1), 2–18. http://doi.org/10.1037/a0024338

Hattie, J. (2009). V*isible Learning: A synthesis of over 800 meta analyses relating to achievement*. United Kingdom: Routledge.

Lipsey, M.W., Puzio, K., Yun, C., Hebert, M.A., Steinka-Fry, K., Cole, M.W., Roberts, M., Anthony, K.S., Busick, M.D. (2012). Translating the Statistical Representation of the Effects of Education Interventions into More Readily Interpretable Forms. (NCSER 2013-3000). Washington, DC: National Center for Special Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from https://ies.ed.gov/ncser/pubs/20133000/pdf/20133000.pdf

U.S. Department of Education, Institute of Education Sciences, What Works Clearinghouse. (2013, March). *What Works Clearinghouse: Procedures and Standards Handbook (Version 3.0)*. Retrieved from http://ies.ed.gov/ncee/wwc/pdf/reference_resources/wwc_procedures_v3_0_standards_handbook.pdf

Volker, M. (2006). Reporting effect size estimates in school psychology research. Psychology in the Schools, 43(6), 653–672.

Online Resources:

Effect Size Calculator/Information, from the University of Colorado at Colorado Springs: http://www.uccs.edu/lbecker/effect-size.html

Statistics Hell: A statistics website maintained by Dr. Andy Field. The appearance is *interesting* to say the least and even he warns some examples could be considered offensive to some. But, it is full of statistics resources and tutorials. http://www.statisticshell.com/