

Effect Size for School Psychologists, Part 2

Chris Birr and Todd Hrenak

I can calculate an Effect Size, now what?

In the last article, calculating effect size and interpretation guidelines were discussed. The purpose of effect size is to expand beyond, “Did an intervention work?” to “How well did it work in a range of contexts?” (Coe, 2000). This time, we will explore a few nuances and options available when using effect size. The first caution is that this article is not exhaustive and serves as a continuation to the introduction of effect size interpretation. Effect size should be reported along with other measures and never used in isolation. Some additional scores to report could be:

- mean fall and spring scores
- mean growth from fall to spring from grade and intervention groups
- number of students meeting expected growth
- number of students in the intervention that surpassed the 25th percentile or other set target to indicate growth toward proficiency

Furthermore, comparing data using local and national norms provides a more balanced perspective if your district data tends to skew higher or lower than the national norm. Be sure to clearly indicate when comparing to local or national norms.

As school psychologists, the data available to us are typically limited by the size of the school or district and number of students provided a service or intervention. Unless working in a large district, access to large sample sizes of both control and intervention groups is not often realistic. When Cohen (1988) proposed his initial guidelines for interpretation his interpretations were developed as a percentage of the standard deviation (Lakens, 2013). If an effect size is 0.5, it indicates that the difference equals half a standard deviation.

What if the intervention group has a very large standard deviation?

In the last article, Glass’s delta was mentioned briefly. Glass proposed substituting the control group’s standard deviation for the entire population (Ferguson, 2009). According to Thompson (2007), Glass’s delta is most useful when group sizes are quite large and concerns exist as to whether the intervention may have affected the mean and standard deviation of the experimental group. Thompson continues that Cohen’s d has the advantage of greater precision in estimating the denominator of these standardized effects, because total n is larger if both groups are used to estimate a pooled SD . When using Glass’s delta, the effect of the intervention is tempered but the overall size of the sample is reduced which could impact the effect size obtained. When in doubt, use Cohen’s d to calculate the effect size.

I calculated an effect size for an intervention and it was very small, now what?

Although Cohen’s initial interpretation guidelines are widely accepted, effect sizes in social sciences are often very small (Rosnow & Rosenthal, 2003). Initially, Cohen based these recommendations on an extensive survey of statistics reported in social sciences (Lenth, 2001). Ferguson (2009) indicated that at the time, no agreement had been established as to what magnitude of effect was necessary to establish practical significance. It is also suspected that Cohen never intended for his benchmarks to be accepted as law, he was making an initial recommendation meant for continued re-

search and refinement.

There are researchers who recommend against “canned” effect size interpretation (Lenth, 2001). The caution is against using a target effect size to determine if an intervention or method was successful based only on Cohen’s initial benchmarks. Effect size provides an indication of the degree of effect (positive or negative), and not a simple yes or no whether an intervention worked. Context is important when viewing effect size. Researchers are advised to seek out effects of similar interventions and view within the context of similar studies (Vacha-Haase & Thompson, 2004). One important caveat mentioned by Vacha-Haase and Thompson is that when few or no other studies are available, it is advised to use Cohen’s initial benchmarks. If you are unsure, Cohen’s initial benchmarks are advised:

0.2 and below is small

0.5 is moderate

0.8 and up is a large effect

Where can I find more information on effect size?

Presently, effect size continues to be used primarily in research. Bridging from research to practice will require a thoughtful approach and continual professional development. Listed below are a few frequently cited sources and researchers in the field of effect size use.

The [Institute for Education Sciences \(IES\)](#) published a paper for researchers regarding the use and interpretation of effect size. Dr. Mark Lipsey was the first author of the paper and has contributed significantly to the field of effect size research. Another researcher who has contributed greatly to effect size use is Dr. Bruce Thompson. A quick search of either name and “effect size” on Google Scholar or ERIC will provide more in depth reading and relevant papers.

[What Works Clearinghouse](#) is currently on version 3.0 of their Procedures and Standards Handbook. Again, this handbook is targeted to researchers but it is an excellent source to become a better consumer of research to guide practice. Pages 22-24 of the handbook address effect size. Furthermore, for the WWC, effect sizes of 0.25 standard deviations are considered to be substantively important (p. 23). The WWC also notes that “effect sizes this large will be interpreted as a qualified positive (or negative) effect, even though they may not reach statistical significance in a given study.” This guideline is mentioned as a consideration and not one that should be adopted without careful consideration.

Conclusion

As mentioned in the first article, effect size can be a powerful indicator when presenting data to stakeholders. Although effect size may be an easier statistic for others to grasp, it is necessary for school psychologists to understand the method of calculation and interpretation. Interested school psychologists are encouraged to review the resources included and seek additional resources to expand their knowledge of effect size. Effect size is a valuable indicator, but when engaged in data based decision making, multiple sources of data are required.

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.
- Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, 40 (5), 532-538.
- 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>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4:863. Retrieved from: <http://journal.frontiersin.org/article/10.3389/fpsyg.2013.00863/abstract>
- Lenth, R. V. (2001). Some practical guidelines for effect sample size determination. *The American Statistician*, 55 (3), 187-193.
- Lipsev, 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>
- Rosnow, R. L., & Rosenthal, R. (2003). Effect sizes for experimenting psychologists. *Canadian Journal of Experimental Psychology*, 57, 221-237. Retrieved from: <http://www.unt.edu/rss/class/mike/5030/articles/rosnowfxsize.pdf>
- Vacha-Haase, T. & Thompson, B. (2004). How to estimate and interpret various effect sizes. *Journal of Consulting Psychology*. 51 (4). 473-481.
- Retrieved from: https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwi7x_eWmKrKAhUMKCYKHcjzDSYQFggkMAE&url=http%3A%2F%2Fwww2.fiu.edu%2F~bliss%2FEffectsize&usq=AFQjCNGnsZElImSSHoeEuWKOWA9MvO3bxRw&bvm=bv.111677986.d.eWE
- Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychology in the Schools*. 44 (5). 423-432.
- 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

Guidance for Progress Monitoring for Specific Learning Disabilities

Guest Editorial By Greg Nyen
 Director of Student Services
 Stevens Point School District

Recently the Stevens Point Area Public School District began the process of reviewing our screening and progress monitoring tools for use in our Multi-leveled Systems of Support (MLSS). Not only were our school psychologists reporting difficulty attaining adequate stability of baseline and progress monitoring data but also in the area of reliability of our trend-line slope. Our teachers were spending inordinate amounts of time probing and re-probing our struggling learners sometimes exhausting all of the probes available to them only to find that our data was unreliable and at times, invalid. Educators and administrators alike were frustrated to say the least.

While interviewing a Computer Adaptive Test program provider who shall remain nameless, they referenced a document that the Department of Public Instruction had provided to them as a reference. Enter the ***Progress Monitoring for Specific Learning Disabilities (SLD) Eligibility Decisions*** authored by members of the special education team at Wisconsin Department of Public Instruction (WDPI).

Co-authored by the school psychology consultant, the specific learning disabilities consultant and a member of the procedural compliance self-assessment/complaint investigation team, this resource provides a much needed level of clarity and reassurance to those of us in education who must make meaning and high stakes decisions from the data sets that are the product of our MLSS efforts. Reassurance and validation of many difficult conversations began to set in at the bottom of page one under the heading of Decision Errors as I read, “Educational, employment and community-based

outcomes for students with special educational needs are poor in comparison to students educated entirely in the general educational system. Therefore, LEAs are keen to ensure they do not make the mistake of identifying a student as needing special educational services when s/he does not.” The section ends with an exclamation point by underscoring how errors may be mitigated by using valid and reliable data.

The document is also a practical tool for the continuous improvement of MLSS. Page 3 outlines the beginning of a very useful table offering a side-by-side comparison of Computer Adaptive Tests and Curriculum Based Measures as they consider the theoretical foundations of probe: brevity; specific skill acquisition; multiple equal, or nearly equal, form availability; sensitivity to small change; reliability and validity; and scientifically based practice. The side-by-side comparison outlines the strengths and weaknesses of two popular and prevalent methods of measurement that don’t require a background in school psychology to understand.

As an administrator responsible for the creation and shared oversight of our own MLSS, I have often struggled with the ‘oh well, the data is what it is’ approach by some of my colleagues who don’t understand the significance of placing a disability label of a student. This document was like gold to me! It validated everything that our student services department members had been espousing for months (since December 1st, 2013 to be exact) and yet also provided a reference point in the form of the table comparison that my colleagues could understand. Job well done by the WDPI!