Effect Sizes in Gifted Education Research

Marcia Gentry
Scott J. Peters
Purdue University

Recent calls for reporting and interpreting effect sizes have been numerous, with the 5th edition of the *Publication Manual of the American Psychological Association* (2001) calling for the inclusion of effect sizes to interpret quantitative findings. Many top journals have required that effect sizes accompany claims of statistical significance. However, too often reports of data analyses in gifted education rely on statistical significance without reporting effect size indices to help interpret quantitative findings. Without a supporting effect size index, erroneous interpretation of results can occur. This Methodological Brief addresses this concern and provides examples and guidance concerning using effect sizes in gifted education quantitative research.

**Keywords:** Effect Size; Quantitative Methods, Statistical Significance, Gifted Education Research, Research Reporting

Numerous authors have addressed the importance of reporting effect sizes (ES; e.g., American Psychological Association [APA], 2001; Cohen, 1994, 1997; Daniel, 1998; Thompson, 1999, 2002). In 1999, the APA Task Force on Statistical Inference called for authors to “always provide some effect-size estimate when reporting a *p* value.” (Wilkinson et al., 1999, p. 103). These authors explained the importance of reporting ES, which include facilitating interpretation of the magnitude of statistical findings, enhancing the ability of researchers to conduct meta-analyses, and improving the quality of the reported research. The APA manual also calls for including ES with quantitative results. However, several authors have noted that, in general, researchers have been slow to heed this call (e.g., Huberty, 2002; Paul & Plucker, 2004).

The quality of research in gifted education has faced criticism by well-known scholars in the field (VanTassel-Baska, Robinson, Coleman, Shore, & Subotnik, 2006). If the research reported in the field of gifted education is to be respected by those within and outside the field, serious consideration concerning the use of ES is warranted. Consistent reporting of ES would facilitate the inclusion of gifted education research in meta-analytic studies (Cooper & Hedges, 1994) as well as improve both the rigor and reputation of the gifted education literature (Matthews et al., in press; VanTassel-Baska et al., 2006). The intent of this Methodological Brief is to provide the *GCQ* readership with information concerning the use, importance of, and interpretation of ES in quantitative studies.

**Effect Sizes in the Gifted Education Research Literature**

Gifted education literature published in the past 25 years contains numerous reports of quantitative research findings as statistically significant without the associated ES (see Matthews et al., 2008; Paul & Plucker, 2004; Plucker, 1997). This practice continues despite frequent calls for the inclusion of ES as best practice in social science research (e.g., APA, 2001). Paul and Plucker suggested that journal editorial policies coupled with authors’ lack of familiarity with ES and their interpretation explain why ES remain uncommon in the gifted education literature. Matthews et al. (2008) examined five major gifted education journals for ES reporting as warranted with significance testing. They found that authors reported ES in conjunction with quantitative findings only 25.6% of the time between 1996 and 2000 and 45.9% of the time between 2001 and 2005.

**Authors’ Note:** Please address correspondence to Marcia Gentry, Gifted Education Resource Institute, Purdue University, 100 North University Street, West Lafayette, IN 47907-2098; e-mail: mgentry@purdue.edu.
The Importance of Effect Size to Researchers and Practitioners

Reporting ES indices enables researchers and practitioners to determine the magnitude of a statistically significant finding using a standardized measure (Grissom & Kim, 2005). Whereas a statistically significant finding indicates that the result is likely not due to chance, an estimate of an ES helps readers interpret the magnitude of the finding within the research context. Additionally, researchers can use an ES to determine sample sizes and plan analyses with sufficient power to detect any real and meaningful findings that may result. Put simply, practitioners and researchers should expect that findings reported in the gifted education literature be accompanied by an ES—a “best practice” in research. When researchers and the journals in which they publish fail to report effect sizes and their interpretation, readers are left with only statistical significance to determine whether the findings have meaningful effects for their particular context.

The following example illustrates the importance and practical usefulness of ES indices. In 2002, Gentry, Gable, and Rizza reported the following concerning whether differences in grade levels and genders existed concerning student attitudes toward school:

There were significant differences for the 2-way MANOVA main effects of grade level ($F_{20, 12322} = 26.64, p < .001$) and gender ($F_{4, 3715} = 20.55, p < .001$) with no significant interaction of grade level and gender. Because a large sample can more easily yield statistical significance, effect size for these results were examined by subtracting Wilks’ Lambda from 1.0 to determine $R^2$ (Cohen, 1988; Tabachink & Fidell, 2006). Although the grade level differences were associated with a medium effect size of $R^2 = .14$, the gender differences were at best a small effect size ($R^2 = .02$). (p. 541)

The ES reported in this case tells the reader the actual magnitude of statistically significant findings. Each main effect was statistically significant at $p < .001$. However, examination of the effect sizes indicates that the grade level effect warrants close attention, whereas gender does not. Without the ES indices the reader would be led to erroneously assume that each result was equally important; and the authors might make a case for the gender differences, which although statistically different, represent little real differences in actual mean scores.

ES Interpretation: Is the Finding Important?

An important caution regarding ES interpretation involves considering the magnitude of the numerical ES in conjunction with what the researcher deems as contextually significant. For example, Rosenthal, Rosnow, and Rubin (2000) used the example of a study in which researchers found a statistically significant result ($p < .00001$) and an ES of $r^2 = .034$, or 3.4% of variance explained. Use of traditional guidelines (e.g., Cohen, 1988, 1994) might result in the interpretation of this as a trivial effect; however, when put into context of heart attack prevention, the effect becomes much more important with 3.4% fewer heart attacks. If a similar effect was found concerning a national sample of students at-risk for dropping out of school, few would argue 3.4% fewer dropouts as a trivial finding. Thus, arbitrary cut-points for determining “large” or “important” ES should not be used. Instead, the researcher must interpret ES with regard to the context of the study and in light of previous findings in that area of research.

Using Effect Size Indicators: Considerations and Issues

Unlike tests of statistical significance, effect sizes are generally less influenced by sample size and provide the researcher and consumer of research with an indication of the magnitude of the research findings, or in simple terms a measure of difference or association between or among variables. Effect sizes can be expressed as correlational indicators (e.g., correlation coefficients or coefficients of determination including $r^2$, $\eta^2$, $\omega^2$) and as indicators of standardized mean differences (e.g., Cohen’s $d$, Hedges’s $g$). Much has been written on differing types of ES, their strengths, applications, and weaknesses (e.g., Cohen, 1997; Grissom & Kim, 2005), and APA (2001) provides a list of ES indicators. Several issues must be examined when computing ES, including choosing an ES estimate (Grissom & Kim), using confidence intervals (Thompson, 2002), and correcting for positive bias (Grissom & Kim; Thompson, 2006). To review all of these issues is beyond the scope of this brief. However, several points of consideration and interpretation are outlined in Table 1 and suggested readings are referenced.

Estimating ES When They Are Not Reported

When reading an article that does not include ES the reader can calculate or estimate ES using other
Correlational

- Pearson $r$/point-biserial
- Phi ($\phi$)
- $r^2$
- $\eta^2$

**Mean difference**

- Cohen’s $d$
- Hedges’s $g$

Table 1

<table>
<thead>
<tr>
<th>Type</th>
<th>Common Use</th>
<th>Interpretation</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlational</td>
<td>Comparing two $m_i$ or more groups with $\chi^2$ tests</td>
<td>Directional, report with sign ($+/-$) As a correlation Variance explained Variance explained</td>
<td>Can be computed from $t$ or $F$ statistic Only for truly discrete/dichotomous variables Partial $\eta^2$ can be computed which includes error</td>
</tr>
<tr>
<td>Mean difference</td>
<td>Comparing two $m_i$, or more groups $t$ tests or analysis of variance (ANOVA)</td>
<td>As a $Z$ score As a $Z$ score</td>
<td>Two groups have same sample size Similar to Cohen’s $d$, except controls for different sample sizes</td>
</tr>
</tbody>
</table>

Note: This is only a partial list of some of the most common effect size measures.

information normally provided such as results from a $t$ test or analysis of variance (ANOVA). Most quantitative studies comparing groups also report descriptive statistics such as the means and standard deviations. Consider, for example, a study reporting means and standard deviations for Group 1 ($M = 56$, $SD = 7$) and Group 2 ($M = 44$, $SD = 6$) and reporting that each group contains an equal sample size. Cohen’s $d$ can be calculated as follows:

$$d = \frac{\text{mean}_1 - \text{mean}_2}{\sqrt{\frac{(SD_1^2 + SD_2^2)}{2}}} = \frac{56 - 44}{\sqrt{7^2 + 6^2}} = 1.841$$

The ES of 1.841, resulting from the secondary calculation, is analogous to a $Z$ score; in this case, the mean of Group 1 was 1.841 standard deviations above that of Group 2. Meta-analytic researchers often estimate ES in this manner. However, it is always best for researchers to report ES in their primary analyses to promote more accurate secondary analyses. The category of ES calculations known as percentage of variance accounted for (e.g., $\eta^2$), can also be computed from almost any study making this type of ES commonly used in secondary-analysis (Thompson, 1999). It is also worth noting that virtually any one measure of ES can be converted into any other measure of ES, and most textbooks on the topic contain tables related to such conversions (e.g., Grissom & Kim, 2005, p. 109).

In summary, reporting ES is clearly important and doing so would improve authors’ familiarity with ES, improve the quality of quantitative research in the field, and provide consumers of the research with assurances that researchers follow recommended practices in reporting social science research (e.g., Wilkinson et al., 1999). By placing value on, attending to, and reporting ES, researchers in the field of gifted education can develop a more robust, meaningful, respected, and defensible research base.

To view a 1-hour lecture on effect sizes presented by Dr. Bruce Thompson at Texas A&M University, visit the following Web site: [http://realvideo.coe.tamu.edu:8080/ramgen/distance/stats/Thompson.rm](http://realvideo.coe.tamu.edu:8080/ramgen/distance/stats/Thompson.rm)

**References**


Cohen, J. (1997). The earth is round ($p < .05$). In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 21–35). Mahwah, NJ: Lawrence Erlbaum.


Marcia Gentry directs the Gifted Education Resource Institute at Purdue University where she enjoys working with doctoral students and engaging in research and gifted education professional development. She remains active in the field through service to NAGC and AERA and by writing, reviewing, and presenting research aimed to improve education for children, youth, and teachers.

Scott J. Peters is a doctoral candidate in gifted education at Purdue University. His research interests include the application of advanced quantitative methods to gifted education research questions and to the development of rigorous instrumentation for the purpose of student identification.