A Note on the Use of Cohen's *d* and Common Language Effect

Hans E. Muehsler Naperville, IL Unit District 203
David A. Walker Northern Illinois University
Thomas J. Smith Northern Illinois University

This note reviewed the various variables available in the scholarly literature that have been demonstrated to show a relationship with K-12 student achievement within the categorization system of Hattie (2009). Subsequently, the Hattie framework was used as a heuristic example of applying the effect size Cohen's *d* and also the Common Language Effect.

The present note provides a current framework (Hattie, 2009) for factors that have been shown to be related to K-12 student achievement. By means of this framework, the applied use of a frequently-cited effect size in social science research, Cohen's *d* (Cohen, 1988), as well as the lesser known Common Language Effect (CLE; McGraw & Wong, 1992), which will be used as an aid with the interpretation of *d* effect size results, will be employed to demonstrate the heuristic example.

Framework

Hattie (2009) provides the most up-to-date conceptualization of learning correlates in K-12 with his magnum opus, *Visible Learning: A Synthesis of over 800 Meta-Analyses Relating to Achievement.* This effort refines and expands upon his earlier work (Fraser, Walberg, Welch, & Hattie, 1987). Hattie employs what might be termed "meta-meta-analysis" (i.e., a meta-analysis of previously conducted meta-analyses) to synthesize research findings related to influences on student learning, providing scale of magnitude not before encountered in the research literature. Thus, Hattie, through a synthesis of over 800 meta-analyses using in excess of 50,000 studies predominantly from the international field of educational research, examined potential influences on student learning. Based on the reported sample sizes that were reported in the studies that Hattie included, the number of students examined is estimated to be approximately 236 million.

Domain Categories

Hattie (2009) derived six major domain categories that contributed to student learning, which are predicated on a statistically-based, effect size idea termed the "Zone of Desired Effects." Hattie determined that the use of a singular, common effect size metric of standardized magnitude, Cohen's *d,*

$$d = \frac{M_t - M_c}{s_p},$$

where M_t and M_c are the mean outcomes for the treatment and control (or comparison) groups, respectively, and s_p is the pooled within-groups standard deviation, would be an effective tool to express study outcomes. After the examination of 800 plus meta analyses and many tens of thousands of studies and their associated normal distribution of derived *d* effect sizes, Hattie determined that a $d = 0.40$ should be considered the hinge, or tipping point, between the minimum expected, positive effect of a program or intervention and what he deemed a high, substantial desired effects. Effect sizes between $d = 0.15$ and 0.39, although below the minimum threshold, are considered by Hattie as more "Teacher-Based Effects." Effect sizes between 0.00 and 0.14 are termed "Developmental Effects," while negative effect size values are deemed "Reverse Effects." Thus, given this structure, Hattie's six domain categories comprised of various influences in the "Zone of Desired Effects" that contribute to student learning are, in no order of importance: 1) the student, 2) the teacher, 3) the school, 4) the home, 5) instruction, and 6) curriculum. It is important to note that these six domain categories emerged as result of a trajectory of research, including Hattie's (1987, 1992) initial efforts; Reynolds and Walberg (1998) and Walberg (2006) who posited student aptitude, instruction, and psychological environment as three major factors; and Marzano (2000) who, when positing three factors--student, school, and teacher--suggested that 80% of the variance in achievement could be explained by student variables, 13% by teacher variables, and 7% by school variables.

Limitations of Hattie's Analysis

A number of limitations and delimitations are inherent in Hattie's (2009) work, some of which Hattie himself points out. Only those factors, for example, that can be controlled/influenced by the school are considered as potentially influential by Hattie. Thus, factors such as neighborhood crime rate, for example, are not considered by Hattie as predictors of achievement. Also, the studies included in his meta-meta-analysis (i.e., mega-analysis) only include quantitative studies based on measurable outcomes.

Qualitative studies, or other studies based on observation or non-measurable constructs, are not included. Additionally, there was no attempt in Hattie's work to address more complex mechanisms such as interaction effects that might influence student achievement. A gender × motivation interaction effect, for example, might be evident if the relationship between motivation and achievement was different for boys than it was for girls.

It should be recognized, however, that the mechanisms underlying the effects on student achievement may be much more complex. In their meta-analysis, for example, Nye, Konstantopoulos, and Hedges (2004) show an effect size of $d = 0.32$ for teacher effects. Moreover, the effect of the teacher is stronger in mathematics than it is in reading, and the effect of the teacher is stronger for children from low SES backgrounds than it is for students from higher SES (Terhart, 2011).

Most of the studies from Hattie's (2009) analysis were conducted in the 1980s and 1990s, although a number come from the 2000s. Their years of publication range from 1980 to 2008. Thus, although the breadth of studies is expansive, a number of the studies now might be considered dated. Terhart (2011) also points out that Hattie is somewhat opaque on the criteria he used to determine whether a particular study would be included in his analysis. Thus, the meta-analyses used in Hattie's analyses likely vary widely in quality. Nonetheless, the sheer breadth of studies included in Hattie's mega-analysis is impressive. To be sure, Hattie's work draws from a number of studies worldwide. Although this global aspect is again impressive in its breadth, and the many factors that influence student learning might certainly be considered universal and global, this approach does not consider more localized effects. That is, effects specific to the United States, for example, are not explicitly addressed. Similarly, factors that influence educational outcomes may differ across countries and cultures, and these country-specific effects may not be apparent in global measures of effect size.

Heuristic Example

As an example, to put the use of the *d* effect size metric in context, we find that Hattie (2009) reviewed seven meta-analyses comprised of 300 studies and 800 effects related to vocabulary programs as an influential aspect of student learning from the Curriculum domain. He found an average *d* effect size = 0.67 or a high, desired effect in this area of reading. Further, Hattie employed the use of the CLE to help with interpretation of effect size results. According to the originators of the CLE, McGraw and Wong (1992), the CLE converts an effect size into a probability (see the Appendix for a Statistical Package for the Social Sciences (SPSS) syntax program to calculate the CLE created by Walker (2014) and based on the original work of McGraw & Wong). Thus, for vocabulary programs, the CLE probability = .47 or 47%, which indicates that 47 times out of 100 the use of a vocabulary program with students would make

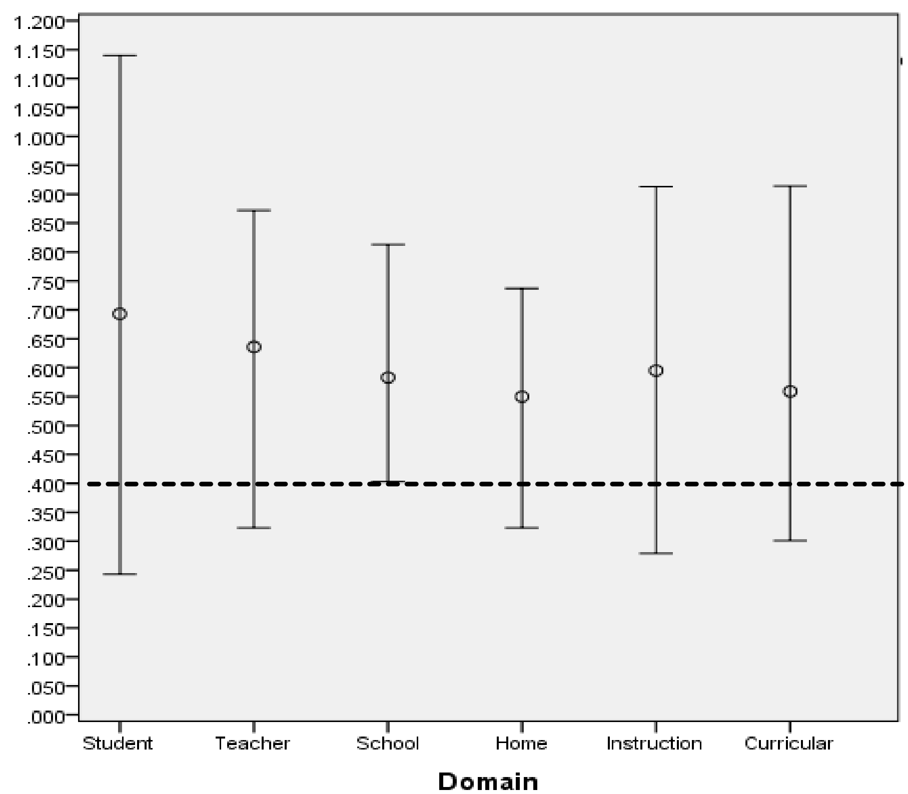

Figure 1. Average Cohen's d with 95% Confidence Intervals across Hattie's six domains.

a positive difference in their reading comprehension. Specifically, if there existed two classes, one implementing the use of a vocabulary program and one without, the vocabulary program class would be effective 47 times out of 100.

As can be discerned from Figure 1, the Student domain had the highest average *d* (0.70) or very near Cohen's concept of a "large" effect, but also contained the most error (i.e., the most variability in effect sizes among the sampled studies),, while the Home domain had the smallest average *d* (0.55) or at the benchmark for a moderate effect. Of note, is that all of the domains surpassed Hattie's (2009) threshold of > .040 (i.e., the superimposed line in Figure 1) indicating "zones of desired effect."

Tables 1 through 6 display Hattie's six domains and their influential factors where *d* > 0.40. Comparatively, also presented with each influential factor's average *d* are the *d* range, the average CLE, and the CLE range.

Table 1. Influential Factors from the Student Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Self-Report Grades	1.44	0.47, 3.10	100%	33, 100
Piagetian Programs	1.28	None	91%	None
Prior Achievement	0.67	0.31, 1.19	48%	22, 84
Pre-Term Birth Weight	0.54	0.34, 0.73	38%	24, 52
Concentration/Persistence/Engagement	0.48	0.03, 1.09	34%	2, 77
Motivation	0.48	0.23, 0.73	34%	16, 52
Early Intervention	0.47	0.14, 0.97	33%	10, 69
Preschool Programs	0.45	0.10, 0.56	32%	7, 100
Self-Concept	0.43	0.32, 0.76	30%	23, 54

Table 2. Influential Factors from the Teacher Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Micro Teaching	0.88	0.55, 1.18	62%	39, 83
Teacher Clarity	0.75	None	53%	None
Teacher-Student Relationships	0.72	None	51%	None
Professional Development	0.62	0.37, 0.81	44%	26, 57
Not Labeling Students	0.61	None	43%	None
Quality of Teaching	0.44	0.29, 0.68	31%	20, 48
Expectations	0.43	0.08, 0.82	31%	6, 58

Table 3. Influential Factors from the School Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Acceleration	0.88	None	62%	None
Classroom Behavioral	0.80	0.58, 1.01	56%	41, 78
Classroom Cohesion	0.53	0.17, 0.92	38%	12, 65
Peer Influences	0.53	None	37%	None
Classroom Management	0.52	None	37%	None
Small Group Learning	0.49	0.46, 0.51	35%	33, 36
School Effects	0.48	None	34%	None
School Size	0.43	None	30%	None

Table 4. Influential Factors from the Home Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Home Environment	0.57	0.34, 0.80	40%	24, 56
Socio-Economic Status	0.57	0.50, 0.66	40%	35, 47
Parental Involvement	0.51	0.13, 0.75	36%	9, 53

Table 5. Influential Factors from the Curriculum Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Vocabulary Programs	0.67	0.38, 1.04	47%	27, 74
Repeated Reading Programs	0.67	0.65, 0.68	47%	46, 48
Creativity Programs	0.65	0.37, 1.01	44%	26, 71
Phonics Instruction	0.60	0.24, 1.53	42%	17, 100
Tactile Stimulation Programs	0.58	None	41%	None
Comprehension Programs	0.58	0.10, 1.15	41%	7, 81
Visual-Perception Programs	0.55	0.09, 0.81	39%	6, 57
Outdoor/Adventure Programs	0.52	0.46, 0.61	37%	33, 43
Play Programs	0.50	0.26, 0.74	36%	19, 52
Second/Third Chance Programs	0.50	0.34, 0.66	36%	24, 47
Mathematics	0.45	0.16, 1.01	32%	11, 71
Writing Programs	0.44	0.26, 0.81	31%	18, 57

Table 6. Influential Factors from the Instruction Domain

Influence	Average *d*	Range of *d*	Average *CLE*	Range of *CLE* (%)
Provide Formative Evaluation	0.90	0.70, 1.10	64%	49, 78
Interventions for Learning Disabled Students	0.77	0.56, 0.90	55%	40, 64
Reciprocal Teaching	0.74	None	52%	None
Feedback	0.73	0.12, 2.87	52%	8, 100
Spaced vs. Mass Practice	0.71	0.46, 0.96	N/A	N/A
Meta-Cognitive Strategies	0.69	0.67, 0.71	49%	47, 50
Self-Verbalization/Self-Questioning	0.64	0.51, 0.84	45%	36, 59
Problem-Solving Teaching	0.61	0.33, 1.13	43%	23, 80
Teaching Strategies	0.60	0.07, 1.02	43%	5, 72
Cooperative vs. Individualistic Learning	0.59	0.04, 0.45	42%	3, 52
Study Skills	0.59	0.13, 1.62	42%	9, 100
Direct Instruction	0.59	0.21, 0.83	42%	15, 59
Mastery Learning	0.58	0.04, 0.78	41%	3, 59
Worked Examples	0.57	None	40%	None
Concept Mapping	0.57	0.22, 0.81	40%	16, 57
Goals	0.56	0.36, 0.82	40%	25, 58
Peer Tutoring	0.55	0.33, 0.98	39%	23, 69
Cooperative vs. Competitive Learning	0.54	0.28, 0.82	39%	20, 58
Keller's Personalized System of Instruction	0.53	0.49, 0.60	37%	35, 42
Interactive Video Methods	0.52	0.41, 0.65	36%	29, 46
Questioning	0.46	0.13, 0.82	32%	9, 58
Behav. Objectives/Advanced Organizers	0.41	-0.03, 0.89	29%	0, 63
Matching Style of Learning	0.41	-0.03, 0.75	29%	0, 53
Cooperative Learning	0.41	0.13, 0.73	29%	9, 52

References

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

Fraser, B. J., Walberg, H. J., Welch, W. W., & Hattie, J. A. (1987). Synthesis of educational productivity research [Special issue]. *Journal of Educational Research, 11*, 145-252.

Hattie, J. A. C. (1987). Identifying the salient facets of a model of student learning: A synthesis of meta-analyses. *International Journal of Educational Research*, *11*, 187-212.

Hattie, J. A. C. (1992). Measuring the effects of schooling. *Australian Journal of Education, 36*, 5-13.

Hattie, J. A. C. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London: Routledge.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, 111*, 361-365.

Marzano, R. J. (2000). *A new era of school reform: Going where the research takes us*. Aurora, CO: Mid-Continent Research for Education and Learning.

Nye, B., Konstantopoulos, S., & Hedges, L.V. (2004). How large are teacher effects? *Educational Evaluation and Policy Analysis*, *26*, 237–257.

Reynolds, A. J., & Walberg, H. J. (Eds.). (1998). *Evaluation for educational productivity* (Vol. 7). Greenwich, CT: Elsevier Science/JAI Press.

Terhart, E. (2011). Has John Hattie really found the holy grail of research on teaching? An extended review of *Visible Learning. Journal of Curriculum Studies, 43*, 425-438.

Walberg, H. J. (2006). Improving educational productivity: An assessment of extant research. In R. F. Subotnik & H. J. Walberg (Eds.), *The scientific basis of educational productivity* (pp. 103-160). Charlotte, NC: Information Age Publishing.

Walker, D. A. (2014). Common language effect size program [Computer program]. DeKalb, IL: Author.

Send correspondence to: David A. Walker
Northern Illinois University
Email: dawalker@niu.edu

APPENDIX

SPSS Syntax for the CLE

```
DATA LIST LIST /M2 M1 SD2 SD1 (4F9.2).
*******************************************************************************
NOTE: Between BEGIN DATA and END DATA, insert the Mean for the Experimental Group or Post-
Test (M2), the Mean for the Control Group or Pre-Test (M1), the Standard Deviation for the Experimental
Group or Post-Test (SD2), and the Standard Deviation for the Control Group or Pre-Test (SD1),
respectively
*******************************************************************************.
BEGIN DATA
69.70 64.30 2.80 2.60
END DATA.
COMPUTE CLEZ = (M2-M1)/SQRT((SD2**2 + SD1**2)).
COMPUTE CLE = CDFNORM(CLEZ)*100.
EXECUTE.
FORMAT CLE (F8.0).
VARIABLE LABEL CLE 'Common Language Effect: XX Amount of Time (CLE Value) out of 100 the
Use of a Particular Program/Intervention will Make a Positive Difference or XX % of Participants will
Gain Compared to Those Without the Program/Intervention'/.
REPORT FORMAT=LIST AUTOMATIC ALIGN (CENTER)
/VARIABLES= CLE
/TITLE "Common Language Effect Size".
*******************************************************************************
This research is an extension of similar work conducted by:
McGraw, K. O., & Wong, S. P. (1992). A common language effect size
statistic. Psychological Bulletin, 111, 361-365.
*******************************************************************************.
```