HOW-TO GUIDE #1

Using Data to Drive Change

Understanding Effect Sizes in Programme Evaluations and Research Using the ELOM 4&5 Assessment Tool

June 2023 | Compiled by Andrew Dawes and Junita Henry

WHAT IS AN EFFECT SIZE AND HOW IS IT USED?

Effect sizes are used to quantify the size and direction of performance differences between two groups following an intervention. They also measure the strength of relationships between variables using correlation and regression analyses.

Durlak (2009)[1] explains that effect sizes can be expressed as mean differences, percentages, or correlations. They offer advantages over significance tests by enabling a more comprehensive evaluation of effectiveness across various contexts. Effect sizes serve as an important tool in reporting and interpreting effectiveness (Coe, 2002).[2]

They provide a standardised metric to describe the practical significance of observed effects, independent of sample size. Effect sizes are particularly valuable for comparing study results and interpreting practical implications. Effect sizes can also be unstandardised, which means that they are reported in their original scale.

Researchers and practitioners can use this information to make informed decisions about the practical implications of their findings.

Example 1: Applying effect sizes to quantify strength of variable relationships

If we wanted to know to what extent a set of influences predicted children's numeracy performance, a regression analysis would show the strength of the relationships between each predictor variable and the child's score. Regression analyses provide both standardised (std. β) and unstandardised coefficients (**B**).

Standardised coefficient values are all on the same scale and the regression values range from -1 to + 1.[3] The size of the regression coefficient tells us the amount of influence exerted by each predictor and allows us to compare them.

The first predictor variable chosen for this example is the child's socio-economic status. The second predictor variable is their growth status, measured in height for age. A statistically significant coefficient β= 0.30 for socio-economic status and coefficient β = 0.22 for growth status tells us that both predictors influence performance in numeracy. Socio-economic status has the stronger relationship because its coefficient is higher than the growth status' coefficient.

In the Thrive by Five 2021 Index study, the relative contributions of several predictors to ELOM 4&5 scores were tested and effect sizes derived using standardised coefficients (std. β). For example, the coefficient (effect size) for the difference between children in the poorest group and the wealthiest was std. β **=0.61**; for moderate stunting compared to normal growth it was std. β **=0.37**; and for social-emotional functioning the std. β **=0.33**.

Unstandardised regression coefficients show the amount of increase in the outcome contributed by one unit of increase in the predictor. For example, how many ELOM Total score points (the outcome) are associated with an increase of one month of age

1 Durlak, J. A. (2009). How to select, calculate, and interpret effect sizes. Journal of Pediatric Psychology, 34(9), 917-928.

2 Coe, R. (2002) It's the Effect Size, Stupid. What effect size is and why it is important. Paper presented at the British Educational Research Association annual conference, Exeter, 2002. Available from https://cebma.org/wp-content/uploads/Coe-2002.pdf

3 When there is more than one predictor, and when there is multicollinearity (the predictors are correlated and not independent measures), β can exceed +1 or -1

(the predictor)? A statistically significant unstandardised coefficient of **B** = 1.02 indicates that for every one month of age, we expect a child's ELOM Total score to increase by 1.02 points. Over four months we would expect the score to increase by 4.08 points due to maturation.

Height for age Z scores are measured in SD units, from the World Health Organization reference group median. For example, in the Early Learning Outcomes (ELPO) study, a change of 1 SD in children's height for age produced a 3.17 point change in the ELOM 4&5 Total standard score.

Example 2: Applying effect sizes to compare two groups following an intervention

In an experimental study designed to test the effect of a new numeracy programme, one group of children receive the intervention (the treatment group) while another does not. The group not receiving the intervention is referred to here as the comparison group. The effect size tells us the size of the difference in numeracy scores of the two groups when compared before and after the intervention.

When groups are compared, effect size is measured by the *Standardised Mean Difference* using the following formula:

$$\textbf{Effect size} = \frac{\textbf{[intervention group mean] - [comparison group mean]}}{\textbf{Standard Deviation (pooled or comparison group)}}$$

The effect size is the number of standard deviations (SD) in the difference between the mean of the treatment and the mean of the comparison group.

Classifying effect sizes

Jacob Cohen provides a classification of effect sizes commonly known as Cohen's convention (Coe, 2002):

0.2 SD is a '*small*' effect size; **0.5** SD is a '*medium*' effect size and **0.8** SD or greater, is a '*large*' effect size.

Cohen's conventions are widely used but should be employed with full awareness of the meaning of the findings. Even a small effect may have great practical value. For example, a statistically significant effect size of 0.25 for the relationship between growth status and numeracy has great practical significance: a nutrition programme would likely have a significant impact on the numeracy performance of undernourished children.

Applying effect sizes to ELOM 4&5 Years Assessment Tool Programme Evaluations

The ELOM 4&5 Years Assessment tool measures performance across five key developmental domains for preschool children. The five domains are

- Gross motor development
- Fine motor coordination and visual motor integration
- Emergent numeracy
- Emergent literacy
- Cognition and executive function

When children are assessed using the ELOM 4&5 tool, they receive a score for each of the developmental domains (out of 20 points), as well as a total score that combines all five (100 points).

In Table 1, we present mean ELOM standard score differences and their SD equivalents using Cohen's criteria. These can be used as a quick reference for interpreting the practical significance of findings in research and evaluations when using the ELOM tools in group comparisons.

TABLE 1

Standard Deviation Equivalents (ES) of Statistically Significant Mean Score Differences for ELOM 4&5

	EFFECT SIZE (Cohen's convention)			
	SMALL	**MEDIUM**	**LARGE**	
Difference between Means	**0,20 SD**	**0,50 SD**	**0,80 SD**	**1,00 SD**
ELOM 4&5 Total score	2,81	7,04	11,26	14,07
Gross Motor Development (GMD)	0,84	2,10	3,35	4,19
Fine Motor Coordination & Visual Motor Integration (FMC&VMI)	0,68	1,70	2,71	3,39
Emergent Numeracy & Mathematics (ENM)	0,82	2,05	3,28	4,10
Cognition & Executive Functioning (CEF)	0,85	2,14	3,42	4,27
Emergent Language & Literature (ELL)	0,93	2,32	3,71	4,64

Source: Early Learning Programme Outcomes study

Unstandardised regression coefficients have been used to show the size of the effects of various influences on early learning programme outcomes.

As an example using the formula above, we find a small ES of 0.2SD (a change of 2.81 ELOM Total points) for early learning programme 1. In early learning programme 2, we find a pre-post change of 11.26 in the ELOM Total score (a large ES of 0.8SD). We can compare the two effect sizes. Clearly, we would judge programme 2 to be the more effective in improving children's early learning outcomes. This was the approach taken in the Early Learning Outcomes (ELPO) study. These values can also be used when differences between programme and maturation effects are calculated.