Exploring the Merits of Adjusting Education Effect Sizes for Program-Outcome Overalignment

Joseph Taylor, UCCS (jtaylo18@uccs.edu)
Joshua Polanin, AIR
Susan Kowalski BSCS

Background/Purpose

When the content foci of interventions and outcome measures are misaligned, a bias can result, affecting the magnitude of the intervention effect estimate (see Slavin, 2008; Institute of Education Sciences, 2018). The purpose of this paper is to discuss a potential process for adjusting effect sizes for program-outcome overalignment.

Specifications for the Adjustment

Program-Outcome Alignment: Concepts and Definitions

Proportional emphasis (PE) is the proportion of a program devoted to a given learning outcome or topic area. Proportional emphasis can be defined as the proportion of total instructional time devoted to a specific topic area (e.g., days per year, or class periods per year).

Test emphasis (TE) is defined as the number of test items that address a given topic area, expressed as a proportion of the total number of test items. A calculation of alignment between a single outcome measure and instruction in the treatment and comparison conditions must account for variation of emphasis by topic.

The alignment (A) between a program and a specific topic within an outcome measure is defined as the difference between the proportional emphasis and test emphasis (PE − TE) for that topic area, with alignment values for treatment condition (A_T) and comparison condition (A_C), respectively. Overalignment of treatment to outcome for topic area *i* can be expressed using a variable called differential alignment (α_i) : the difference between the alignment of the treatment program to the outcome measure and the alignment of the comparison program(s) to the outcome measure. For topic area *i*:

$$\alpha_i = A_{Ti} - A_{Ci}$$
$$\alpha_i = [(PE_{Ti} - TE_i) - (PE_{Ci} - TE_i)]$$
$$\alpha_i = PE_{Ti} - PE_{Ci}$$

where A_{Ti} represents the alignment of the treatment program with the specific topic *i* within the outcome measure, A_{Ci} is the alignment of the comparison program(s) with topic *i* of the outcome measure *i*, PE_{Ti} represents the proportional emphasis of the treatment program on topic area *i*, PE_{Ci} is the proportional emphasis of the comparison program(s) on topic area *i*, and TE_i is the emphasis on topic *i* within the

outcome measure. The concept of differential alignment is illustrated graphically in Figure 1.

[Insert Figure 1 about here]

Toward a more general expression, we extend the case for topic *i* by calculating a weighted sum of differential alignment values across *j* topics addressed by the outcome measure. The weighting ensures that the magnitude of the differential alignment value captures the fact that topic-specific differential alignment introduces more bias for test topics that represent a larger proportion of the total test. Symbolically, for overall differential alignment (α), we have:

$$\alpha = \sum_{i=1}^{j}[(PE_{Ti} - TE_i) - (PE_{Ci} - TE_i)]\,(TE_i) \qquad (1)$$

When interpreting α, values near zero suggest similar outcome measure alignment for both treatment and comparison groups. Positive values suggest that the set of topics within the outcome measure was more heavily emphasized in the treatment program and vice versa. Topic-specific and overall differential alignment values can range between the extremes of 1 and −1. Upon simplifying, differential alignment becomes:

$$\alpha = \sum_{i=1}^{j}(PE_{Ti} - PE_{Ci})(TE_i) = \sum_{i=1}^{j}\alpha_i \qquad (2)$$

Effect Size Adjustment

If an assessment is overaligned with the treatment program, the effect size should be adjusted downward and vice versa. Specifically, an effect size such as Hedges' g (1982), would be adjusted to g':

$$g' = g\,(1 - \alpha) \qquad (3)$$

Because α has a positive value when the outcomes of interest are more heavily emphasized by the treatment condition, equation 3 results in a reduced effect size. With differential alignment ranging from −1.0 to 1.0, the effect sizes adjusted for differential alignment (g') will have the range $2g \geq g' \geq 0$.

Application Example: Undergraduate Physics Education using Time on Task

Consider two semester-long physics courses (Course A and Course B). In this example, time on task (measured in contact hours) is tracked in both courses for each of six topic areas in mechanics. These topic areas coincide with the six subscales of the *Force Concept Inventory* (FCI: Hestenes, Wells, Swackhamer, 1992) by which the efficacy of the courses is measured. Consulting the Hestenes et al., (1992) item alignment, we created a simplified alignment (see Table 1).

[Insert Table 1 about here]

Using the within-topic item totals in Table 1, we calculate the topic-specific test emphases of the FCI included in Table 2.

[Insert Table 2 about here]

Also included in Table 2 are hypothetical data about the proportional emphases of the two courses. For the purposes of presentation, the two courses differ in proportional emphasis for just two FCI topics: *Superposition Principle*, and *Kinds of Force*. We find:

$$\alpha = \sum_{i=1}^{j}(PE_{Ti} - PE_{Ci})(TE_i)$$
$$\alpha = [.07 \text{ } -.20](.10) + [.20 \text{ } -.07](.33)$$
$$\alpha = [-.13](.10) + [.13](.33)$$
$$\alpha = -.013 + .043 = .03$$

The positive value indicates that the net effect of the differential alignment across the two topics was in favor of Course A. That is, Course A put more emphasis (time on task) into *Kinds of Force* and less emphasis into *Superposition Principle*; but because the *Kinds of Force* subscale contributes more to the total FCI score than does *Superposition Principle*, the net effect is a 3% advantage via alignment with the FCI. Using a hypothetical unadjusted effect size of 0.60 in favor of Course A, the adjusted effect size would account for this advantage accordingly. Using equation 5 we have:

$$g' = g\,(1 - \alpha)$$
$$g' = 0.60\,(1 - .03)$$
$$g' = 0.60\,(.97)$$
$$g' = 0.58$$

Discussion and Implications

Biases related to program-outcome overalignment have been noted by the research community and in some cases have prohibited studies from being eligible for syntheses or from being considered in policy/programmatic decisions. We illustrated a process for adjusting effect sizes for biases stemming from differential alignment between the foci of program/courses and those of the outcome measures used to evaluate them. The adjustment described here is relatively small unless the unadjusted effect size is atypically large (in absolute value) or the overalignment between the outcome and one of the treatment conditions is extreme. The goal of this adjustment is to potentially salvage effectiveness information from studies that would otherwise be discounted.

References

Hedges, L. V. (1982). Estimation of effect size from a series of independent experiments. *PsychologicalBbulletin*, *92*(2), 490.

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The physics teacher, 30*(3), 141-158.

Institute of Education Sciences. (2018). *What Works Clearinghouse Procedures and Standards, 4.0*. Retrieved from http://ies.ed.gov/ncee/wwc.

Slavin, R. E. (2008). Perspectives on evidence-based research in education—What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), 5-14.

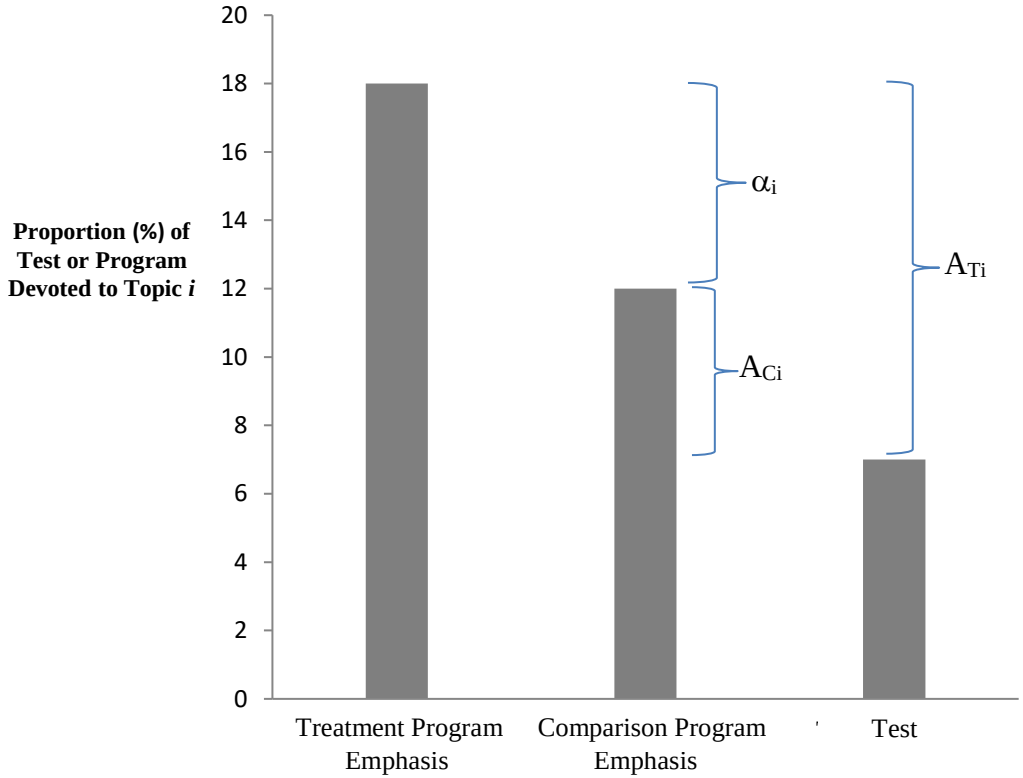

Figure 1. A Graphical Representation of Differential Alignment for a Given Topic (α_i).

Table 1.

Item Alignments: Force Concept Inventory

FCI Subscale	**Specific Topic**	**Item Alignment (Hestenes et al., 1992)**	**Simplified Alignment**	**Total Aligned Items**
Kinematics	Velocity discriminated from position	20E	20E	
	Acceleration discriminated from velocity	21D	21D	6
	Constant acceleration entails parabolic orbit or changing speed	23D, 24E, 25B	23D, 24E, 25B	
	Vector addition of velocities	7E	7E	
First Law	with no force	4B, (6B), 10B	4B, (6B), 10B	8
	velocity direction constant	26B	26B	

	speed constant	8A, 27A	8A, 27A	
	with cancelling forces	18B, 28C	18B, 28C	
Second Law	Impulsive force	(6B), (7E)	(6B), (7E)	4
	Constant force implies constant acceleration	24E, 25B	24E, 25B	
Third Law	for impulsive forces	2E, 11E	2E, 11E	4
	for continuous forces	13A, 14A	13A, 14A	
Superposition	Vector sum	19B	19B	4
Principle	Cancelling forces	(9D), 18B, 28C	(9D), 18B, 28C	
Kinds of Force	Solid contact: passive	(9D), (12 B,D)	(9D), (12 B,D)	13

Solid Contact: Impulsive	15C	15C
Solid Contact: Friction opposes motion	29C	29C
Fluid contact: Air resistance	22D	22D
Fluid contact: buoyant (air pressure)	12D	
Gravitation	5D, 9D, (12B,D), 17C, 18B, 22D	5D, 17C, 18B,
Gravitation: acceleration independent of weight	1C, 3A	1C, 3A
Gravitation: parabolic trajectory	16B, 23D	16B, 23D

Table 2.

Topical Emphases for Two Hypothetical Physics Courses and the Force Concept Inventory

	Hours of Instruction and Topic-Specific Proportional Emphases (45-hour course)					
Course	*Kinematics*	*1st Law*	*2nd Law*	*3rd Law*	*Superposition Principle*	*Kinds of Force*
A	15 (.33)	6 (.13)	9 (.20)	3 (.07)	3 (.07)	9 (.20)
B	15 (.33)	6 (.13)	9 (.20)	3 (.07)	9 (.20)	3 (.07)

	Number of Items and Topic-Specific Test Emphases (Simplified FCI Alignment)					
	Kinematics	*First Law*	*Second Law*	*Third Law*	*Superposition Principle*	*Kinds of Force*
FCI	6 (.15)	8 (.21)	4 (.10)	4 (.10)	4 (.10)	13 (.33)