2020

Issues around Using Academic Return on Investment (A-ROI) for Informing and Improving Decisions

PART I: VALIDITY

YAN, BO

INTRODUCTION

Return on investment (ROI) is a concept that originated in the business world in the early twentieth century. According to the Hagley Museum and Library, the concept and the formula were first developed in 1914 by Donaldson Brown, the Assistant Treasurer of the DuPont Company, for monitoring business performance as the company was grappling with diversifying from explosives to lacquers, Pyralin plastics and dyes[1]. ROI was quickly adopted by the DuPont Company as the primary performance measure for all of its departments and later required for all capital appropriations and projects submitted for approval by the company's senior management.

Because of its simplicity and versatility, ROI became widely adopted by industries. Applying the ROI concept to education, Levenson (2012) coined the term Academic Return on Investment (A-ROI) and called for public school systems to use A-ROI to drive decisions on resource use. Jefferson County Public Schools (JCPS) adopted this concept and started to systematically document and evaluate district investments through a Cycle-based Budgeting model in 2014. Specifically, expected A-ROI, which is defined as academic or academic-related student outcomes achieved per dollar amount invested, is specified for each investment item (e.g., new program, initiative, or position), with an investment cycle assigned for the implementation and evaluation of the program or initiative. At the end of the investment cycle, the actual A-ROI is calculated and provided to leaders as one piece of information for assessing the success of the investment.

The A-ROI calculated by Levenson's formula is essentially a cost-effectiveness ratio (CER). However, the rigor required for conducting a thorough cost-effectiveness analysis is simply beyond the reach of most school districts that have neither the capacity nor the resources necessary to do such work. In addition, the amount of work and time needed for such analysis makes it very difficult to evaluate the large number of investment items typical of large urban school districts. At the same time, there are a number of methodological issues that arise in applying A-ROI for high-stakes budgetary decisions which have not received adequate attention among practitioners and could lead to misguided decisions.

It is our intention to develop an alternative method for calculating A-ROI that strikes a balance between rigor and practicality. Computationally, it should be simple enough so that most practitioners can conduct the analysis on their own. The sacrifice of rigor is compensated by detailed descriptions of when it is acceptable to use the A-ROI results to inform decision making and potential risks, based on both theory and empirical evidence.

In this series, we discuss five basic issues around using A-ROI for informing and improving decision making: 1) validity, 2) uncertainty, 3) commensurability, 4) cost, and 5) unit of analysis. This discussion explains how we conceptualize A-ROI as an alternative method to rigorous cost-effectiveness analysis. It also provides a context for how these issues are approached in our JCPS A-ROI method and how we suggest results produced through this method should be interpreted and properly used in practice, which

[1] The Father of ROI: Donaldson Brown, https://www.hagley.org/librarynews/father-roi-donaldson-brown

will be documented in another report. In this first installment, we focus on the validity of A-ROI as a cost-effectiveness measure.

VALIDITY

Conceptually, the validity of A-ROI for cost-effectiveness comparisons depends on two primary assumptions regarding the time needed to observe an effect and the causal inferences drawn from observed effects, which have implications for the proper time to calculate or recalculate an investment's A-ROI and appropriateness of using the A-ROI result for decision making.

The first assumption is that each investment has a substantive core that remains unchanged in each iteration of its implementation. In education, however, it is widely known that it takes time for investments to be rolled out, take shape, and become institutionalized. As a result, the timing of A-ROI calculation matters because the results could differ, and more importantly, carry different meanings depending on the timing. An early calculation of A-ROI before an investment is fully implemented will likely not be the A-ROI that is of interest after both the effects and costs stabilize[2].

Depending on the situation, the premature calculation of A-ROI could either under- or over-estimate an investment's true effect[3]. For simplicity, Figure 1 shows four hypothetical investment scenarios, with each effect trend line representing a possible path to maturation that an investment could take during the first five years of implementation.

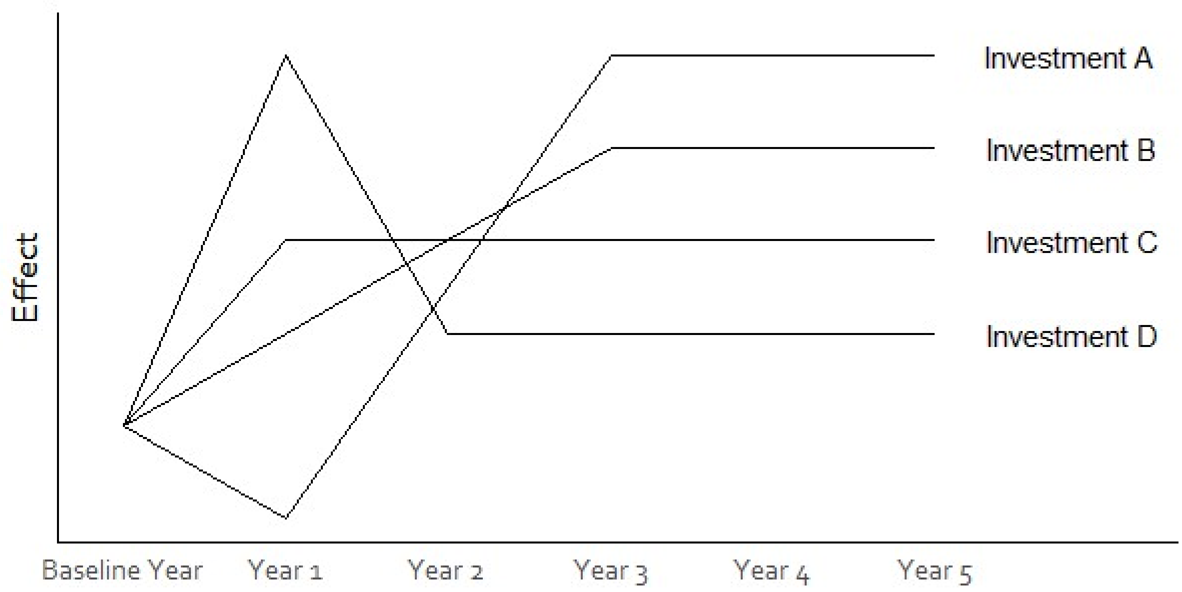

FIGURE 1 CHANGE OF INVESTMENT EFFECT IN THE CONTEXT OF IMPLEMENTATION HISTORY

---

[2] Stabilized effects and costs could still vary randomly. How to detect stabilization of effect and cost is another topic that needs further investigations.

[3] Despite the important implications for program evaluation, the variability of program effects in the first few years of implementation seems to have received little attention. A preliminary investigation of eight literacy program studies and five math program studies reviewed by What Works Clearinghouse found that the majority were conducted at the end of the first year of program implementation.

Among the four investments shown in Figure 1, Investment C has an impact at the end of the first year of implementation and the effect subsequently remains unchanged. For Investment B, the full effect is not seen until the end of three years, which could happen if the effect magnitude is correlated to the level or phase of implementation, or if the effect lags behind the implementation. As a result, the A-ROI result calculated before Year 3 will under-estimate the potential effects.

For Investment D, the effect peaks at the end of the first year of implementation, wanes during the following year, and remains unchanged after that. A typical example of Investment D is technology intervention programs, where the early strong results are largely due to the novelty effect of new technology (Clark, 1983; Tsay et al., 2018). In this case, the A-ROI calculated at the end of the first year of implementation will over-estimate the true effect.

Finally, the A-ROI result for Investment A at the end of the first year would show a negative effect despite the investment having the largest effect after three years of implementation. This is possible when the new program requires an approach that is very different from prevailing conventions, leading to disruptions or confusion in the early stages of implementation[4].

Similarly, costs can be over- or under-estimated. When a program is implemented in phases with different cost elements introduced at each phase, early calculations of A-ROI may under-estimate the total cost. Under-estimation of costs could also arise when implementation necessitates additional cost that is not anticipated during a program's planning process. For example, a program that has received funding to hire more minority teachers as part of a larger literacy program might run into difficulty in recruiting minority teachers and thus need additional funding for that particular aspect. Over-estimation of costs tends to happen with economies of scale. That is, the cost remains unchanged when the number of participants increases or, more generally, when participation growth outpaces cost increases.

Table 1 shows the potential bias in the A-ROI results when over- and/or under-estimation occurs with the effect, cost, or both.

	Cost		
Effect	*Over-estimate*	*Unbiased estimate*	*Under-estimate*
Over-estimate	Unclear	Over-estimate A-ROI	Over-estimate A-ROI
Unbiased estimate	Under-estimate A-ROI	Unbiased A-ROI estimate	Over-estimate A-ROI
Under-estimate	Under-estimate A-ROI	Under-estimate A-ROI	Unclear

TABLE 1 POTENTIAL BIAS IN PREMATURE ESTIMATION OF A-ROI

It is important to point out that the above discussion should not prevent A-ROI from being calculated for an investment during the first few years of its implementation. There are practical reasons for why stakeholders would demand A-ROI information soon after an investment is implemented. For example, district leaders may want to know whether adjustments are needed to improve implementation, or board members may be interested in making sure that an investment they approved is producing results.

[4] It should be noted that estimating mature A-ROI with stabilized effects and costs introduces the possibility of the A-ROI result being confounded by other programs implemented during the maturation period.

However, it does point to the need to communicate to stakeholders how the A-ROI information should be interpreted and used for decision making.

Once A-ROI is calculated for an investment with stable effects and costs, that measure cannot be unquestioningly used year after year, because of the ever-changing nature of educational contexts. For example, in public schools, it is not uncommon to see leadership changes, especially in large urban school districts where the superintendent tenure averages three years (Council of the Great City Schools, 2014).

Following a leadership change, existing programs typically face two potential outcomes. One is that the program no longer has an owner and champion, which then leads to a decreased level of attention and accountability, and eventually deteriorating implementation. The other is that the new leader takes the reins of an inherited program and significantly reshapes it to fit his or her vision, philosophy and approach. In either case, the substantive core is likely to change during the transition although the program often carries the same name. This raises the question of whether the previous measure of A-ROI for such investments should be re-calculated after a transition.

The second assumption on which the validity of A-ROI rests is that the return is truly due to the investment. Rigorous evaluations collect evidence that rules out alternative explanations for the observed return. This is important because a number of factors may impact investment outcomes, such as natural trends among participants, selection bias whereby gains or losses are really associated with the characteristics of participating students, teachers, or schools rather than the program, or confounding effects from other programs implemented at the same time to improve student achievement.

Table 2 summarizes the nine possible pairs of true and estimated effects. Theoretically, all colored cells in Table 2 represent biased or potentially biased scenarios, all of which should be guarded against by collecting "strong evidence" through well-designed and well-implemented experimental studies, or "moderate evidence" from well-designed and well-implemented quasi-experimental studies as defined by the US Department of Education (2016). Practically, however, they have different implications and significance for gathering and using evidence to facilitate decision making.

		True effect (E)		
		Positive	*Zero*	*Negative*
Estimated effect ($\hat{E}$)	*Positive*	P-P	Z-P	N-P
	Zero	P-Z	Z-Z	N-Z
	Negative	P-N	Z-N	N-N

TABLE 2 POSSIBLE SCENARIOS OF THE OBSERVED RETURN OF AN INVESTMENT

For scenarios represented by red cells, the biased estimates will direct leaders to make incorrect decisions and should definitely be avoided. In P-Z and P-N situations, leaders will be inclined to end effective programs, and in N-P situations, deleterious programs will likely be continued. In these cases, reliance on A-ROI should be avoided as decisions may directly harm students. For yellow cell (Z-P) scenarios, the biased estimate does not harm students directly, but may result in wasting money that could have been spent on effective programs. Here again, decisions relying on the biased estimates are not warranted. For green cells scenarios, the biased estimates will prompt leaders to make good decisions and are thus acceptable despite the bias. The different significance levels of bias and their implications for decision making are summarized in Table 3.

Cell	*Decision Implication*	*Bias Significance*
P-P	Continue effective programs	Acceptable
P-Z, P-N	Discontinue effective programs that help students	Avoid
Z-P	Waste money that could have been spent on effective programs	Avoid
Z-N	Discontinue ineffective programs due to estimated negative results	Acceptable
N-P	Spend money on harmful programs that impact students negatively	Avoid
N-Z, N-N	Discontinue harmful programs due to estimated null or negative results	Acceptable

TABLE 3 SIGNIFICANCE AND IMPLICATION OF BIASED RETURN RESULTS

Undoubtedly, "strong evidence" and "moderate evidence" should be pursued whenever possible and feasible, which gives us the best chance of guarding against the biases represented in the colored cells in Table 2. However, as pointed out earlier, that is not feasible for most school districts, due to the complexity as well as time and effort required to conduct well-designed and well-implemented experimental and quasi-experimental studies.

For practical purpose, our goal is to develop an A-ROI method that provides different levels of protection against various biases that arise as the result of the compromise between rigor and practicality. For example, for any given estimated negative result, the A-ROI method should provide a high-level of confidence that the program does not actually have a positive effect (P-N cell). If the possibility of a true positive effect is ruled out with high confidence, the burden to produce evidence for possible Z-N or N-N results should be less demanding.

For situations represented by cell P-P, there is another level of bias that warrants close attention. That is, true effect E of a program is either under-estimated or over-estimated by $\hat{E}$, which will lead to an unfavorable or favorable assessment, respectively, when it is compared with other programs with unbiased estimates of the investment impact.

This second level of bias seems harder to protect against and requires more rigorous evidence than protection against the biases presented in Table 2. Nevertheless, the point here is that instead of treating all biases as equally unacceptable, for practical purposes, the biases can be differentiated with regard to their implications for decision making and difficulty for producing evidence.

REFERENCES

Clark, R. E. (1983). Reconsidering Research on Learning from Media. *Review of Educational Research*, *53*(4), 445–459.

Council of the Great City Schools. (2014). *Urban School Superintendents: Characteristics, Tenure, and Salary: Eighth Survey and Report* (Urban Indicator). Council of the Great City Schools. https://www.cgcs.org/cms/lib/DC00001581/Centricity/Domain/87/Urban%20Indicator_Superintendent%20Summary%2011514.pdf

Levenson, N. (2012). *Smarter Budgets, Smarter Schools: How to Survive and Thrive in Tight Times*. Harvard Education Press.

Tsay, C. H.-H., Kofinas, A., & Trivedi, S. K. (2018). Novelty Effect and Student Engagement in a Technology-mediated Gamified Learning System. *Academy of Management Proceedings*, *2018*(1), 13030. https://doi.org/10.5465/AMBPP.2018.13030abstract

US Department of Education. (2016). *Non-Regulatory Guidance: Using Evidence to Strengthen Education Investments*. https://ed.gov/policy/elsec/leg/essa/guidanceuseseinvestment.pdf