



# STATISTICAL VS. SOCIAL SIGNIFICANCE

## AMBASSADOR INSIGHTS

September 8, 2015

*Developed collaboratively by the*  
**LEAP OF REASON  
AMBASSADORS COMMUNITY**

---

# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

“Ambassador Insights” are drawn from online discussions of the [Leap of Reason Ambassadors Community](#), a private community of nonprofit experts, leader practitioners, and instigators. The community’s purpose is to:

- inspire, motivate, and support nonprofit and public sector leaders (and their stakeholders) to build great organizations for greater societal impact
- increase the expectation and adoption of high performance as the path toward that end.

While “Ambassador Insights” do not represent the views of all ambassadors, they have been vetted and approved by each ambassador who contributed to the dialogue and presented to the full Leap of Reason Ambassadors Community for review.

They are intended to help provide context, resources, and information and are by no means an exhaustive review of available resources. Though we strive to keep this information current and accurate, errors can occur. They are therefore provided “as is” with no guarantee of accuracy, completeness, or timeliness.

All publicly available “Ambassador Insights” may be accessed [here](#).

This document, developed collaboratively by the Leap of Reason Ambassadors Community, is licensed under a Creative Commons Attribution-NoDerivatives 4.0



We encourage and grant permission for the distribution and reproduction of copies of this material in its entirety (with original attribution). Permissions beyond the scope of this license are defined as follows:

You may adapt or make derivatives (e.g., remixes, excerpts, or translations) of this document, so long as they **do not**, in the reasonable discretion of the Leap of Reason Ambassadors Community, **alter or misconstrue the document’s meaning or intent**.

Attribution is to be in the following formats:

- For unmodified use of this document, the attribution information already contained in the document is to be maintained intact.
- For adaptations of this document, attribution should substantially follow this format:

“From ‘**Statistical vs. Social Significance**,’ Leap of Reason Ambassadors Community, licensed under [CC BY ND https://creativecommons.org/licenses/by-nd/4.0/](https://creativecommons.org/licenses/by-nd/4.0/)”

The Leap of Reason Ambassadors Community may revoke the additional permissions described above at any time. For questions about copyright issues or special requests for use beyond the scope of this license, please email us at [info@leapambassadors.org](mailto:info@leapambassadors.org).

---



# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

*The error of treating statistical significance as ipso facto equivalent to social significance seems to be standard practice in the program evaluation domain.*

*– David Hunter, Hunter Consulting, LLC*

*Statistical significance and policy relevance are not the same thing. One has to reach a judgment about any project's results by separately considering both forms of significance.*

*– Gordon Berlin, President, MRDC*

## OVERVIEW

The July 2015 exchange among ambassadors about statistically significant impact vs. social (or policy) value in many ways reflects the internal debate over the hot-button issue of external evaluation during the development of “[The Performance Imperative](#).” Ultimately, ambassadors robustly endorsed external evaluation as Pillar 7 of the PI, while making clear that no one type of evaluation is right for every organization at every stage of organizational development. Similarly, ambassadors both questioned and offered thoughtful explanations of the value of statistical significance, while agreeing that it is not the same as social value and that additional context and considerations are necessary to ensure socially valuable outcomes.

The full context of each ambassador's contribution to the exchange is listed in the Appendix.

## BACKGROUND/THE ISSUE AT HAND

In a community post on July 11, Mario Morino solicited ambassadors' thoughts on the Social Innovation Research Center's (SIRC) report, “[Social Innovation Fund: Early Results Are Promising](#),” which described, with reasonable qualification, positive evaluation findings for several SIF-funded programs. David Hunter's concern with the report centered on the way in which SIRC interpreted the results of MRDC's evaluation of Reading Partners, a volunteer tutoring program for kids who are 6-30 months behind their age-mates in reading. While taking no issue with the quality of MRDC's research, which found “statistically significant impacts...equivalent to approximately one-and-a-half to two months of additional progress in reading,” David critiqued SIRC for concluding, “the Reading Partners model is effective.” As David stated, “I want to focus on one issue: **the difference between statistical significance and social significance...and how very frequently (including in this instance) the former is treated as if it were inherently the latter.**”

# STATISTICAL VS. SOCIAL SIGNIFICANCE

David illustrated his point further:

If our child was 30 months behind his or her age cohort in reading, would we be happy if that child caught up 2 months (maximum) after participating in a 28-week reading improvement program? He or she would still be 28 months behind and, I think it is fair to say, still at a great educational disadvantage...In other words, the report does not address whether the statistically significant impacts of Reading Partners has any social value.

## CONTEXT AND CONSIDERATIONS

Three ambassadors— Gordon Berlin, Kris Moore, and Mari Kuraishi—responded to David’s critique by offering additional considerations for assessing the value of statistical vs. social significance.

### STATISTICAL AND SOCIAL SIGNIFICANCE MUST BE CONSIDERED SEPARATELY

- Gordon Berlin: “Statistical significance and policy relevance are not the same thing. One has to reach a judgment about any project’s results by separately considering both forms of significance.”
  - **Statistical significance represents only an average, and, like any average, there is a lot of dispersion around that mean.** “Some children gained more, some children gained nothing; in fact, some might have lost ground. The analysis and the numbers do not really tell us anything about individual children.”
  - **When the control group also gets extra attention, the social value of statistically significant findings increases.** “In analyzing the results of an experiment one always wants to understand the control group context, keeping in mind that the program’s impact is driven by the treatment difference between those in the program group and those in the control group. In this case, both groups of children got extra attention and resources above and beyond what they would have normally gotten. That RP still made a difference despite the fact that control group children also received school provided special resources, adds to the author’s and MDRC’s confidence that RP is policy relevant.”

### STATISTICALLY SIGNIFICANT READING OUTCOMES MAY SIGNAL GREATER VALUE (IN SCIENCE, MATH, BEHAVIOR) THAN A SINGLE STUDY CAN DETERMINE

- Kris Moore: “While I agree with these cautions, I want to add another consideration, which is that the ability to read, even a little better, makes it more likely that a child will be able to do science and math problems, and perhaps they are more likely to

# STATISTICAL VS. SOCIAL SIGNIFICANCE

read on their own. And, hopefully, if school is less frustrating, a child might even behave a little better. A longitudinal follow-up with data on exploratory (as well as confirmatory) outcomes could answer these kinds of questions.”

## **STATISTICAL SIGNIFICANCE HAS VALUE TO THE INDIVIDUAL STUDENT MAKING SOME GAINS (PRIVATE GOOD), EVEN IF INSIGNIFICANT IN THE BROADER SCHEME OF THINGS (SOCIAL GOOD)**

- Mari Kuraishi: “This also fits the classic economics divide of private good vs. social good...As a parent with a kid with an IEP, I know that ANY gain is sought after and appreciated. But it is frustrating not to know how to get the gains faster and or cheaper.”
  - o Upon review of the exchange, Mari offered this additional explanation: “Markets work in private goods because ‘in theory’ there are multiple suppliers (in non-monopoly situations) that match up to multiple demanders, each with their own utility function reflected in different willingness to pay, and therefore different clearing prices. So in this example that David sets up, ‘If our child was 30 months behind his or her age cohort in reading, would we be happy if that child caught up 2 months (maximum) after participating in a 28 week reading improvement program?’ Different families might be willing to pay different prices for this intervention that results in a net 10-week gain, or 2+ month gain. Positing that repeating this intervention has cumulative results (completely unrealistic, I know) a parent might be willing to pay the price 14 times over to get their kid caught up because the gain to them is a kid that’s caught up, and hypothetically able to participate in public education with no supports etc.

“This equation becomes a lot harder to balance when we have to think about this in the public good context because we are not asking individual households to define their own utility functions and letting them choose their own price point—the decision has to be made for them all as a group (and frequently the households are not consulted in choice of intervention in any event), and moreover the households aren’t paying directly for this intervention. They may be paying indirectly via taxes etc, but they are not at liberty to set price. And in this example we have no idea of alternative interventions, their cost and efficacy, as Gordon points out below.”

---



# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

## A POINT OF CONTENTION: THE COST FACTOR

Gordon and David agreed that whether a program yields statistically significant evaluation results is a different question than whether it has policy relevance. Yet, they offered different perspectives when cost considerations are added to the equation.

**Gordon Berlin: When cost is low, the policy value of modest statistical significance increases.** “In considering policy relevance, cost matters. And because Reading Partners relies on volunteers, the cost is low. If policymakers can get a two month boost on average for all served children for very low cost, in the resource-constrained world we inhabit, policy relevance would rise. Benefits outweigh the modest costs.”

**David Hunter: Is scaling inexpensive programs that get only modest results ultimately good for society?** “Is it really good for our society and its people to have relatively inexpensive programs that produce weak results at best? Results that are unlikely to exert a meaningful influence over the life prospects of the people who most need a helping hand? Wouldn't we want our social policy analysts to take this wider view? And might it not well be true that paying for inexpensive but weak programs will, in the end, cost our society more (in lost opportunities for people who needed more robust help than the weak programs provide—and the loss in contributions they might have made had they received such help)?

“Consider the example of home visitation programs for single mothers. If there were someone in your family who qualified for such a program, would you want her to receive services from the Nurse-Family Partnership or Healthy Families America? In the former there are proven and significant (life-changing) impacts both for the mother and sixteen years later for their children. In the latter, there also are proven impacts—but the effects are very modest in comparison and don't seem to be strong enough to change life trajectories. The Nurse-Family Partnership is more expensive. Is that a good reason for policy analysts and makers to favor Healthy Families America when looking to spend public moneys?”

---



# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

## RESOURCES FOR UNDERSTANDING SOCIAL VALUE

Dean Fixsen responded by offering several references from social science that go beyond RCTs/statistical significance to help “describe and assess social validity.” As he stated, “Present day views of social sciences have wed themselves to randomized group designs as the only way of knowing and in that mode statistical significance is the only test of what matters. Science is broader than that, and any science aimed at improving human service outcomes definitely goes far beyond that narrow view.” References recommended by Dean (with links added):

- Wolf, M. M. (1978). [Social Validity: The case for subjective measurement or how applied analysis is finding its heart](#). *Journal of Applied Behavior Analysis*, 11, 203-214.
- Schwartz, I. S., & Baer, D. M. (1991). [Social Validity Assessments: Is Current Practice State-of-the- Art](#). *Journal of Applied Behavior Analysis*, 24(2), 189-204.
- Wolf, M. M. (1991). [Why did social validity become a classic?](#) *Current Contents*, 23, 8.
- Sudsawad, P. (2005). Concepts in clinical scholarship - [A conceptual framework to increase usability of outcome research for evidence-based practice](#). *American Journal of Occupational Therapy*, 59(3), 351-355.
- Hurley, J. J. (2012). [Social Validity Assessment in Social Competence Interventions for Preschool Children: A Review](#). *Topics in Early Childhood Special Education*, First published online April 6, 2012.

## CONCLUSION

Our thanks to David Hunter, Gordon Berlin, Dean Fixsen, Kris Moore, and Mari Kuraishi for sharing their insights.

---



# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

## APPENDIX

### FULL CONTEXT OF EACH AMBASSADOR'S INDIVIDUAL CONTRIBUTION TO THE EXCHANGE

#### REQUEST FOR INPUT FROM THE COMMUNITY

##### FROM MARIO MORINO, JULY 11, 2015

Although you probably didn't miss this one (it's gotten some good play in the blog-/twitter-sphere), we're interested in your thoughts on SIRC's ["Social Innovation Fund: Early Results Are Promising"](#) report.

##### RESPONSE FROM DAVID HUNTER, JULY 11, 2015

I am responding to the evaluation of Reading Partners by MDRC, which shows participant impacts at a level of statistical significance. Kids who are 6-30 months behind their age-mates in reading benefit from the program at a statistically significant level of confidence (95% or better) via comparison to a control group. Without going too far into the weeds, I want to focus on one issue: the difference between statistical significance and social significance...and how very frequently (including in this instance) the former is treated as if it were inherently the latter. I have been thinking about this quite a bit lately, and thought that perhaps other ambassadors might be interested in this matter.

In its evaluation report on Reading Partners MDRC states in its summary that "Reading Partners had a positive and statistically significant impact on three different measures of student reading proficiency" (ES-5). Yet if we read on, we find that "These statistically significant impacts are equivalent to approximately one-and-a-half to two months of additional progress in reading among the program group members relative to students in the control group, who did not participate in the Reading Partners program" (ES-5). Now we need to ask ourselves: If our child was 30 months behind his or her age cohort in reading, would we be happy if that child caught up 2 months (maximum) after participating in a 28 week reading improvement program? He or she would still be 28 months behind and, I think it is fair to say, still at a great educational disadvantage. Indeed, if we take a linear approach to this, after five (!!!) years of participation our child would have caught up a maximum of 20 months and still be almost a year behind. In other words, the report does not address whether the statistically significant impacts of



---



# STATISTICAL VS. SOCIAL SIGNIFICANCE

---

Reading Partners has any social value. My point here is that while I believe that indeed Reading Partners is effective...I am not (at least yet) persuaded that it is useful.

Looked at from this perspective I find the report's conclusion bit too sanguine: "Overall, the evaluation finds that the Reading Partners model is effective. The program produced measurable impacts on reading skills among students who had a fairly broad range of reading abilities when the study began and among students from a wide range of grades (grades 2 through 5)" (ES-8). I do not mean to suggest that there is any kind of shortcoming in the research conducted by MDRC. Indeed, as far as I could tell the research was very robust and of the highest quality. The matter here is one of interpretation and the assigning of meaning. The phenomenological problem that I have been discussing (treating statistical significance as ipso facto equivalent to social significance) seems to be standard practice in the program evaluation domain. Caveat emptor.

## **RESPONSE FROM MARIO MORINO (FACILITATOR), JULY 13, 2015**

David, I much appreciated the insights you shared and highlighted these points in particular:

- "... the difference between statistical significance and social significance...and how very frequently (including in this instance) the former is treated as if it were inherently the latter."
- ...asking if "having statistically significant impacts ... has any social value."
- the more difficult question is questioning that even if a program is effective, "is it useful?"
- "The phenomenological problem that I have been discussing (treating statistical significance as ipso facto equivalent to social significance) seems to be standard practice in the program evaluation domain."

While a strong proponent of being evidence-based in understanding if a program "works," you raise the additional question is that even if it works, we need to determine if what it does is, indeed, useful.

You've raised a very important point of view and would welcome hearing how others think about this.

Thoughts?

# STATISTICAL VS. SOCIAL SIGNIFICANCE

## RESPONSE FROM KRIS MOORE, CHILD TRENDS, JULY 13, 2015

While I agree with these cautions, I want to add another consideration, which is that the ability to read, even a little better, makes it more likely that a child will be able to do science and math problems, and perhaps they are more likely to read on their own. And, hopefully, if school is less frustrating, a child might even behave a little better. A longitudinal follow-up with data on exploratory (as well as confirmatory) outcomes could answer these kinds of questions.

## RESPONSE FROM DEAN FIXSEN, UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, JULY 13, 2015

Ways to describe and assess social validity have been around for a while (references below). They help assure social significance of interventions and outcomes in a variety of fields. Present day views of social sciences have wed themselves to randomized group designs as the only way of knowing and in that mode statistical significance is the only test of what matters. Science is broader than that, and any science aimed at improving human service outcomes definitely goes far beyond that narrow view.

Wolf, M. M. (1978). Social Validity: The case for subjective measurement or how applied analysis is finding its heart. *Journal of Applied Behavior Analysis*, 11, 203-214.

Schwartz, I. S., & Baer, D. M. (1991). Social Validity Assessments: Is Current Practice State-of-the-Art. *Journal of Applied Behavior Analysis*, 24(2), 189-204.

Wolf, M. M. (1991). Why did social validity become a classic? *Current Contents*, 23, 8.

Sudsawad, P. (2005). Concepts in clinical scholarship - A conceptual framework to increase usability of outcome research for evidence-based practice. *American Journal of Occupational Therapy*, 59(3), 351-355.

Hurley, J. J. (2012). Social Validity Assessment in Social Competence Interventions for Preschool Children: A Review. *Topics in Early Childhood Special Education*, First published online April 6, 2012.

## RESPONSE FROM MARI KURAISHI, CO-FOUNDER AND PRESIDENT, GLOBAL GIVING, JULY 13, 2015

Kristin, I agree with you. And David is entirely right in separating out social significance. In fact I might even riff off of his framework to say that this also fits the classic economics divide of private good vs. social good, in that the individual student is receiving an undeniable private good in making some gains, even if insignificant in the broader scheme of things. As a parent with a kid with an IEP, I know that ANY gains is sought after and appreciated, precisely for the reasons Kristin lays out. But it is frustrating not to know how to get the gains faster and/or cheaper.

# STATISTICAL VS. SOCIAL SIGNIFICANCE

## RESPONSE FROM GORDON BERLIN, PRESIDENT, MRDC, JULY 13, 2015

I am just catching up on this thread and would make several points:

1. Yes, **statistical significance and policy relevance are not the same thing. One has to reach a judgment about any project's results by separately considering both forms of significance.**
2. **The modest 2 month impact in Reading Partners is an average, and, like any average, there is a lot of dispersion around that mean.** Some children gained more, some children gained nothing; in fact, some might have lost ground. The analysis and the numbers do not really tell us anything about individual children. Given how unusual it is to find positive findings in a rigorous, careful evaluation, one should not put too much weight on the point estimate. The program's likely true effect is a range bounded by the confidence interval.
3. **In considering policy relevance, cost matters. And because Reading Partners relies on volunteers, the cost is low.** If policymakers can get a two month boost on average for all served children for very low cost, in the resource constrained world we inhabit, policy relevance would rise. **Benefits outweigh the modest costs.**
4. As the authors explain, the addition of RP to the school's repertoire for children who were behind, had the effect of **freeing up the "remedial" resources that would have been spent on these children, allowing enterprising administrators and teachers to concentrate those freed up resources on the other children who were behind,** the children in the control group. In analyzing the results of an experiment one always wants to understand the control group context, keeping in mind that the program's impact is driven by the treatment difference between those in the program group and those in the control group. **In this case, both groups of children got extra attention and resources above and beyond what they would have normally gotten. That RP still made a difference despite the fact that control group children also received school provided special resources, adds to the author's and MDRC's confidence that RP is policy relevant.**

Hope this helps.

## RESPONSE FROM DAVID HUNTER, JULY 14, 2015

Gordon's points are both illuminating and helpful. However, his point #3 is not as obviously true as it might seem. Yes, as he says, "In considering policy relevance, cost matters." Of course. But he then goes on to observe that "If policymakers can get a two month boost on average for all served children for very low cost, in the resource constrained world we inhabit, policy relevance would rise. Benefits outweigh the modest costs."

# STATISTICAL VS. SOCIAL SIGNIFICANCE

The problem is, that Gordon is looking at the issue of cost only within the context of a given program and school, or perhaps even a school district. But what if we expand the unit of analysis to our society? Is it really good for our society and its people to have relatively inexpensive programs that produce weak results at best? Results that are unlikely in to exert a meaningful influence over the life prospects of the people who most need a helping hand? Wouldn't we want our social policy analysts to take this wider view? **And might it not well be true that paying for inexpensive but weak programs will, in the end, cost our society more (in lost opportunities for people who needed more robust help than the weak programs provide - and the loss in contributions they might have made had they received such help)?**

Consider the example of home visitation programs for single mothers. If there were someone in your family who qualified for such a program, would you want her to receive services from the Nurse-Family Partnership or Healthy Families America? In the former there are proven and significant (life-changing) impacts both for the mother and sixteen years later for their children. In the latter, there also are proven impacts - but the effects are very modest in comparison and don't seem to be strong enough to change life trajectories. The Nurse-Family Partnership is more expensive. Is that a good reason for policy analysts and makers to favor Healthy Families America when looking to spend public moneys?

Now one might well say that MDRC had no necessary reason to comment on such matters. They were hired to evaluate a program for its impacts and they did so very well. Although I wish evaluators would in fact consider social validity issues when reporting the statistical validity of their findings, some might argue this goes beyond the scope of their work. Fair enough. But then for certain it is **not** beyond the scope of work of the Social Innovation Fund. **In fact, understanding social validity and societal impact are what the fund is supposed to be all about. And there was no discussion of this issue either by the SIF or in the evaluation of it that was produced by The Social Innovation Research Center. Them, I simply cannot let off the hook.**