Promising Youth Programs

Personal Responsibility Education Program

OPRE Report #2022-149

November 2022

No Impacts?: How to Enhance What You Learn from Your Evaluation

Written by Jean Knab, Russell Cole and Emily LoBraico, Mathematica

This brief provides suggestions for grantees and evaluators who have completed an impact evaluation but did not find favorable, statistically significant impacts of their program. It provides suggestions for exploratory analyses that could help you understand why you didn't find impacts, or may even have found negative impacts, and how to acknowledge the results of those exploratory analyses in any reporting. The brief also provides suggestions for disseminating your findings and additional supplemental analyses that help you learn as much as you can from the data you've collected.

The primary goal of an impact evaluation is to learn about program effectiveness. Impact evaluations usually aim to credibly show a program achieves impacts on key outcomes. But not all programs demonstrate favorable impacts at the end of an evaluation, and some evaluations yield program impacts that are small, statistically insignificant, or even negative.

There are many reasons why an evaluation might not find favorable impacts for a program (Jacob et al 2019). Perhaps the program was not implemented as intended (for instance with lower-quality or fidelity than planned), or perhaps youth in the comparison group got similar content. Or maybe the evaluation didn't find program impacts because the outcomes were not well matched to the program's theory of change or to the target population. The dosage might have been too low to impact youth behavior, or the content might not have engaged youth. Perhaps the study did not achieve its recruitment targets or had lower-than-expected response rates, which adversely affected the study's power to detect differences as statistically significant (Cole 2020). Though it's disappointing not to find positive impacts, identifying the reasons for the lack of favorable outcomes can help deepen understanding of the program and the needs of youth it served, and possible ways to strengthen and improve the program.

This brief provides suggestions for how you can learn as much as you can from your study and how to share your findings. First, we review ways you can revisit your impact analysis to suggest some potential reasons you did not see favorable results for your program. Then, we discuss ways to disseminate your impact findings thoughtfully and clearly. And finally, we describe a variety of supplemental analyses that help you maximize what you learn from your study, beyond estimating the program's impact. This brief does not provide detailed instructions for how to implement all of these suggestions, but we point the reader to resources and examples.

Mathematica Progress Together

FYSB Family & Youth Services Bureau

OPRE

Revisiting the impact analysis to understand the null effects

Generally, for researchers conducting impact evaluations, when estimating the results that will make up the body of your reporting (that is, the benchmark analyses), you should use the approach you prespecified in your impact analysis plan. However, conducting additional exploratory analyses can provide useful context for understanding the results and for future program or evaluation planning. Based on what you observe in your data, you might want to explore some additional analyses to better understand the influence of the implementation and your measurement decisions to identify ways to revise the program or for future evaluation. In any publication you write, you should clearly state which analyses you originally planned and which analyses you undertook to better understand the results after you knew the results of the study (that is, post-hoc analysis); full transparency in the presentation of both pre-specified and exploratory analyses is best practice against any perceptions of *p*-hacking or data mining.[1] The goal of the recommendations in this brief is not to change your benchmark findings but to supplement them.

When revisiting the impact analysis to understand null, or even negative, effects, begin by examining the program implementation data. If you identify implementation issues, those may explain your lack of impacts. Poor implementation can manifest in a variety of ways that may have consequences on both the program's ability to have impacts as well as your ability to detect impacts.

Examine implementation

It is critically important to carefully assess program implementation data to unpack small, negative, or nonsignificant impacts. Implementation issues reduce the differences in services between the treatment and comparison groups (that is, the effective contrast) being tested and, therefore, reduce opportunities to generate differences in outcomes. First, assess any data you collected on program quality, fidelity, and youth engagement, as well as what the difference in services was between the treatment and comparison groups. If there is only a small difference in content or dosage received across groups, the program quality or fidelity was not strong, or youth report not being engaged, those are possible explanations for why your impact estimates are small or nonsignificant. Also, examine what related content the treatment and comparison groups received outside of the program. If both groups were in a program-rich environment (for instance, you are evaluating an after-school program, and youth receive comprehensive sex education during the school day), you could be running into ceiling effects, even if the related programming was not concurrent.[2] The existing programming might have already affected the behavioral outcomes, leaving little opportunity for further behavior change. In sum, be sure to assess and report implementation and effective contrast findings, as they might help provide context for why the study had smaller-than-expected (and potentially nonsignificant or negative) impacts on outcomes.

Next, if you have multiple large sites, examine whether there is variation across sites in impacts and look at whether differences in the fidelity or quality of implementation across sites could be related to variation in impacts across sites. This means looking at your implementation data for the full sample but also looking across sites or facilitators to see if some sites or facilitators had lower-quality implementation. Then, look at the impacts across the sites to assess whether the impact findings align with the implementation findings. Seeing that sites with stronger implementation had larger program impacts may help explain the program's overall effectiveness.[3]

It might be useful to examine the take-up rates of the program and consider adjusting for the take-up rate in your impact analysis. Who takes up a program can be defined as youth assigned to the program who ever attended it or you could set a minimum amount of participation cut off (for instance, attended fifty percent of the lessons). If the take-up was low—for instance, if some randomly assigned youth did not attend the program or had very low attendance—it might be possible to estimate the effect of the program among those who took it up, to obtain an adjusted estimate of program effectiveness. (See Luca and Cole 2017 for a description of several approaches to conduct these analyses). This approach enables you to identify the effect of the program for those who chose to participate. This might be particularly useful for a voluntary program in an out-of-school setting when there is less-than-optimal take-up. This approach, however, does not answer the original research question about the impact of the offer of a program under an intent-to-treat framework (that is, including all youth who were assigned to the treatment or comparison group in the analysis). Therefore, this should be a supplemental analysis to your benchmark analysis, which should be an intent to treat analysis.

Finally, based on your observations during implementation, you may have new insights into populations for which the program looked especially promising. For example, perhaps you heard that older youth were receptive to the program content. Many studies pre-specify subgroups for which they expect to find impacts. If the pre-specified subgroups do not reflect the subgroups that were perceived as responding favorably to the program during implementation, you might consider conducting a post-hoc, exploratory subgroup analysis for those additional subgroups. Perhaps the program was a better fit for older youth, and impacts might have been larger where the program was a good fit. It is possible that in the original impact analysis, these effects were washed out by the lack of impacts on the groups for which the program was not a good fit. Exploratory sub-group analysis can be appropriate (Breck and Wakar 2021) if you are transparent about how you determined which subgroups to examine, you examine and present the effects of each subgroup of the whole (for instance, analyze and present the findings for both the older youth and the younger youth), and you are explicit that these analyses are post-hoc and exploratory.

After examining implementation data, you may determine that in fact the implementation was poor in one or several ways. If this is the case, then poor implementation may be responsible for the null or weak or even negative impacts. However, if you find that the implementation was not poor, it may be appropriate to revisit your analytic decisions.

Revisit decisions related to measurement and analytic approaches

If you determine there were few issues with implementation, and they do not explain the lack of statistically significant, favorable findings, revisit your measures. Think about whether they were (1) well aligned with the curriculum (and the effective contrast being tested—that is, the difference in content and dosage between the treatment and comparison group) and (2) sensitive to change, given the characteristics of the population and the length of the follow-up period. If using a scale, perhaps also explore item response theory (Zanon et al 2016) or conduct reliability analysis for the measures, particularly if the measures used were not normed with the population represented in the study. You could also consider post hoc factor analysis to potentially reorganize items into more reliable scales, as appropriate. You might also want to examine additional outcomes that weren't in your original analytic plan but appear to be a better fit given your final sample characteristics and results. For instance, if your sample was younger than you expected during the design

The results of additional analyses that deviate from a pre-specified analysis plan need to be carefully reported as exploratory, post-hoc findings. Often, these additional analyses are necessary to respond to issues that either arose in the implementation of the program and the conduct of the evaluation to help explain surprising findings from the main, benchmark results. While these investigations can help unpack an issue and can be very useful to contextualize null findings from an evaluation, the results of these analyses must be characterized appropriately, and receive less emphasis than the main findings that the study was designed to present, to prevent audiences from speculating that the findings are potentially spurious or due to data mining.

phase and, therefore, sexual initiation rates were lower than you expected, consider looking at precursors to sexual behavior. These might include kissing or sexual touching or being in a risky situation (for example, being unsupervised with a potential romantic partner), if your curriculum was designed to affect those behaviors as well. The benefit of this approach is that you might find outcomes that show larger or significant differences given their proximity to the program's theory of change. These precursors might be predictive of later behaviors—for instance, past the period for which you collected data—and that is a potential hypothesis that you or others could test in future work. The downside of this approach is that some audiences will consider evidence on these additional outcomes (that is, any outcomes other than those you registered as part of the analysis plan) as data mining.

Next, revisit your analytic decisions. When your analysis yields suggestive, but not significant, evidence of program impact across multiple, correlated measures in a common domain, you may want to consider an alternative analytic approach to showcase the promise of your intervention. For instance, if your benchmark results included positive, non-significant results on multiple knowledge measures, you may decide to combine all knowledge measures and use a weighting schema to create a combined, or composite, measure of total knowledge. Estimating impacts on this composite (or using a multivariate analysis of variance, or MANOVA approach) yields a more powerful test of the program (see Cole 2020, Schochet 2008, and Appendix C). The downside of this approach is that a composite measure is often harder to label or describe (for example, a measure of sexual risk that derives from several more tangible, well-defined measures). In addition, some evidence reviews, including the current protocol for U.S. Department of Health and Human Services' Teen Pregnancy Prevention Evidence Review, will not accept a composite behavioral measure as an eligible outcome. Remember to transparently describe these supplemental analyses as post-hoc and exploratory.

Consider additional sensitivity checks (for instance, approaches to missing data or alternate covariate specifications) if the observed data are markedly different from what you had originally expected when writing your analysis plan. Perhaps the original analysis plan stated that your benchmark approach would be a complete case analysis, but you ended up having a substantial amount of missing baseline data. To address this problem, you could consider using imputation (see Kautz and Cole 2017 for guidance). Incorporating sensitivity analyses enables you to re-estimate your impacts and identify a range of results potentially caused by unforeseen characteristics of your data. The disadvantage of conducting analyses beyond those you prespecified in your analysis plan is that some folks might perceive that to be data mining or *p*-hacking. Again, you must be very clear in any products about which analyses were prespecified, which were not prespecified, and why you examined any alternative approaches.

You might wish to supplement your traditional inferential analytic approach with a Bayesian interpretation (Deke and Finucane 2019; Gelman and Weakliem 2009) to provide additional decision support about the promise of a program. This approach is particularly useful for studies with smaller-than-expected sample sizes that might show potentially important effects that are not statistically significant. As Wasserstein and Lazar (2016) note in the American Statistical Association (ASA) statement on *p*-values, scientific conclusions should not be solely based on whether a *p*-value is above or below a specific threshold (for example, $p < 0.05$). Bayesian methods use prior evidence about program effects to determine the probability that a program has a favorable effect on a given outcome and provide insight beyond the information afforded by a *p*-value and a judgment of statistical significance. A Bayesian interpretation of a non-significant impact estimate can still provide useful decision support, with statements such as, “The evaluation finds a 91 percent probability the program reduces the frequency of sex without a condom.”

Given what you have learned from all these explorations, assess whether you think there were issues with the evaluation that limited your ability to obtain a good estimate of the program’s effectiveness, or if it seems the program just did not yield the impacts you expected. That will inform your approach to your report (for instance, adding a section on exploratory findings) and potential next steps, such as planning for another rigorous evaluation, or revisiting program content or facilitator training.

Disseminate your impact findings

Although you might be discouraged by the lack of statistically significant, favorable impacts, disseminating the results of your evaluation is still important. Think about the right combination of products and lessons to share. To achieve a balance, follow these broad principles:

1. **Make sure you unpack the reasons the findings are small, nonsignificant, or negative.** The implementation and effective contrast analyses described earlier help get inside the black box of the impact evaluation to understand why impacts might not be what was expected.

2. **Make sure you discuss the ways your exploratory analyses supplemented your original analysis plan.** Transparency about your analyses will help ensure the reader interprets the supplemental results with caution.

3. **Report the original minimum detectable impacts the study was designed to achieve.** You can report what you powered your study to detect at the design phase (for example, the study was powered to detect impacts of 0.20 standard deviation units, which you thought was justifiable based on a set of assumptions that you should articulate). You can also report what you ultimately observed (for example, you observed impacts of only 0.10 standard deviations).

Reporting your results in a journal article

Although you did not observe favorable impacts of your program, you can (and are encouraged) to publish those results (Wasserstein et al. 2019). Publishing null or negative effects can be helpful for the field to learn what works and what might not work so others can build on that knowledge in their program development. Similarly, if the evaluation was not as well executed as you had hoped, others can learn from your experiences. In addition, even if your study was not sufficiently powered, your results might contribute to a future meta-analysis, which can derive power from multiple studies. For example, the Juras et al. 2019 meta-analysis showed the promise of the Teen Pregnancy Prevention grant program on improving behavioral outcomes, even though few individual studies showed significant impacts on behavior. It also looked across programs to identify particular program components associated with favorable outcomes.

Keep in mind your paper should convey the results about your specific program and the lessons learned about evaluation or implementation. Demonstrating the lessons learned might help a paper with null or negative findings get accepted. Although some journals might shy away from publishing a study with null impacts, others are committed to publishing high quality research, regardless of the results, particularly if the study is a replication study. For instance, *Evaluation Review* has a standing call for high quality impact evaluations, which indicates papers with null findings will not be rejected if the study is well executed, sufficiently powered, and has a balanced discussion of the study’s context and findings (Evaluation Review: SAGE Journals, n.d.). Other possible publication avenues include the *Journal of Articles in Support of the Null Hypothesis*, which publishes nonsignificant results twice a year, and the *Public Library of Science*, which publishes nonsignificant findings in its Missing Pieces supplements.

If you aren't successful in getting a journal to publish the paper discussing the impacts, think about other ways to share the results with researchers and program implementers in the field. Consider writing a research brief that you share on your organization's website or an online archive, such as medRxiv or SocArXiv, that accepts unpublished pre-prints or working papers. If the study was well implemented (that is, it was sufficiently powered and program implementation was strong), consider submitting your paper to a relevant evidence review (for example, the What Works Clearinghouse or the Teen Pregnancy Prevention Evidence Review). You could also share key lessons learned in a paper based on the results of your implementation study.

Other products

Be sure to share the findings with community partners as well. Highlight what you learned and what you or the grantee or program developer will do with that information. For instance, if you learned from your implementation study the program wasn't the right fit for the community, you could potentially let community partners know that the implementing organization plans to select another program in the future and will ask for their input and assistance in choosing or adapting one. If you learned your program showed improved outcomes in some domains but not others, you could potentially let them know how the developer will revisit the corresponding content from the domains that did not show impacts. If your evaluation showed evidence of promise but wasn't sufficiently powered, you could potentially use that information to mobilize supporters to help you recruit for a future evaluation.

Consider sharing site-specific results with sites. These might include baseline characteristics and changes in outcomes over time. It is informative for principals and other leaders to see the risk profile and other characteristics of the youth they serve. Remember when sharing these results not to inadvertently identify any youth. If there are small sample sizes at the site, you should not provide site-specific data but could aggregate with another site. If there is sensitive data with small sample sizes (for example, one youth that has been pregnant), consider dropping that data from the presentation.

Consider additional analyses beyond impact estimation

You have collected a lot of valuable data about the program and the youth in it. Regardless of why you didn't find beneficial program effects, you can conduct supplemental analyses to add to what you learn from the main impact study. These supplementary analyses make excellent use of your rich data set and provide valuable insights to your community partners and the field.

Component analysis

Component analysis looks at the relationship between different pieces, or components, of a program with outcomes that are expected to be closely aligned with those factors (Blase and Fixsen 2013; Cole and Choi 2020; Dymnicki et al. 2020). Components might be specific content like a particular lesson or activity, or a feature of implementation, such as fidelity. If you have collected key implementation data on the components you are interested in, you can conduct a components analysis. For example, if you have attendance data for youth across each lesson, you can determine who received a particular lesson or activity (such as a lesson on refusal skills) to see if those youth had greater changes on related outcomes (such as their confidence in being able to say no to risky behavior) than youth who did not receive that lesson or activity.

Other exploratory analyses

It's possible that you have data you could use to answer research questions besides those you set out to explore. For instance, you could use survey data to do the following:

- Test the program's logic model. Most logic models will show proximal outcomes (such as knowledge, attitude, or intended types of outcomes) that are highly related to the content and activities of a program. Logic models also often show more distal outcomes (often sexual behavior outcomes) that are expected to change once the proximal outcomes improve. You can test the logic model by assessing whether the biggest outcome changes are in the proximal outcomes, and if there are smaller changes in the distal outcomes. (See Lee and Cole 2020 for a tip sheet on this topic.)
- Conduct measurement studies. For example, you could document how often youth report inconsistent or unreliable answers within and across surveys. Those results can inform the field about the consistency and reliability with which youth report on sexual behaviors.
- Identify the precursors of behavioral outcomes of interest. For example, you can study what variables at baseline are predictive of risky sexual behavior at follow-up (for example, Goesling and Rangarajan 2008; Marin et al. 2006; Trenholm et al. 2007). These types of analyses can inform future program development by revealing which knowledge, attitudes, or behaviors programs should focus on changing.

You can also use your implementation data to answer implementation-related research questions or identify lessons learned. For example, you could:

- Assess the association between training and professional development on program delivery (such as quality or fidelity) if you have sufficient data to do so. You would need to have a large number of facilitators, data on the training received (type and dosage), and measures of program implementation to use as outcome data.
- Think about the lessons you learned about implementing the program with your population that would be useful for others in the field. For example, if you experienced key challenges with participant attendance or in-session engagement, you might have tried new strategies to engage youth and have some information to share about what worked best during your implementation. This type of information, when shared with individuals who have a similar population and face similar challenges, could support future implementations to have stronger retention, and potentially stronger program impacts as a result.

Conclusion

You learned a lot from your evaluation. The goal of this brief is to help you think about how to maximize your learning and how to share what you've learned with the field and your community partners. Be proud of the work you've accomplished and that you are sharing it with others to support continuous learning and improvement.

References

Blase, Karen, and Dean Fixsen. "Core Intervention Components: Identifying and Operationalizing What Makes Programs Work." ASPE Research Brief. Washington, DC: Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, 2013. Available at https://aspe.hhs.gov/reports/core-intervention-components-identifying-operationalizing-what-makes-programs-work-0

Breck, A., and B. Wakar. "Methods, challenges, and best practices for conducting subgroup analysis." OPRE Report #2021-17. Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2021. Available at https://www.acf.hhs.gov/opre/report/methods-challenges-and-best-practices-conducting-subgroup-analysis

Cole, Russell. "Small Samples Due to Lower-than-Planned Enrollment in Impact Evaluations: What to Do?" Evaluation Technical Assistance Brief No. 5. Washington, DC: Children's Bureau, U.S. Department of Health and Human Services, 2020. Available at https://www.mathematica.org/publications/small-samples-due-to-lower-than-planned-enrollment-in-impact-evaluations-what-to-do

Cole, Russell, and Jane Choi. "Understanding How Components of an Intervention Can Influence Outcomes." OPA Brief. Washington, DC: Office of Population Affairs, U.S. Department of Health and Human Services, 2020. Available at https://opa.hhs.gov/sites/default/files/2020-07/corecomponentsbrief.pdf

Deke, John, and Mariel Finucane. "Moving Beyond Statistical Significance: The BASIE (BAyeSian Interpretation of Estimates) Framework for Interpreting Findings from Impact Evaluations." OPRE Report 2019-35. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services, 2019. Available at https://www.acf.hhs.gov/opre/report/moving-beyond-statistical-significance-basie-bayesian-interpretation-estimates

Dymnicki, Allison, Lisa Trivits, Cheri Hoffman, and David Osher. "Advancing the Use of Core Components of Effective Programs: Suggestions for Researchers Publishing Evaluation Results." ASPE Research Brief. Washington, DC: Office of Human Services Policy, Office of the Assistant Secretary for Planning and Evaluation, U.S. Department of Health and Human Services, 2020. Available at https://aspe.hhs.gov/sites/default/files/migrated_legacy_files//195646/ASPE-Brief-Core-Components.pdf

Evaluation Review: SAGE Journals. "ER Standing Call for Papers." N.d. Available at https://journals.sagepub.com/pb-assets/cmscontent/ERX/ERX_2017StandingCFP.pdf. Accessed November 9, 2021.

Gelman, Andrew, and David Weakliem. "Of Beauty, Sex and Power: Too Little Attention Has Been Paid to the Statistical Challenges in Estimating Small Effects." American Scientist, vol. 97, no. 4, 2009, pp. 310–316. Available at https://www.americanscientist.org/article/of-beauty-sex-and-power

Goesling, Brian, and Anu Rangarajan. "Early Predictors of Girls' Adolescent Sexual Activity: Longitudinal Findings from the Girls Shape the Future Study: Final Report." Princeton, NJ: Mathematica Policy Research, 2008. Available at https://www.mathematica.org/publications/early-predictors-of-girls-adolescent-sexual-activity-longitudinal-findings-from-the-girls-shape-the-future-study

Head, Megan L., Luke Holman, Rob Lanfear, Andrew T. Kahn, and Michael D. Jennions. "The Extent and Consequences of p-Hacking in Science." *PLoS Biology*, vol. 13, no. 3, 2015. Available at https://doi.org/10.1371/journal.pbio.1002106

Jacob, Robin T., Doolittle, Fred, Kemple, James, & Somers, Marie-Andrée. (2019). A Framework for Learning From Null Results. Educational Researcher, 48(9), 580–589. https://doi.org/10.3102/0013189X19891955

Juras, Randall, Emily Tanner-Smith, Meredith Kelsey, Mark Lipsey, and Jean Layzer. "Adolescent Pregnancy Prevention: Meta-Analysis of Federally Funded Program Evaluations." *American Journal of Public Health*, vol. 109, no. 4, 2019, pp. e1–e8. https://ajph.aphapublications.org/doi/10.2105/AJPH.2018.304925

Kautz, Tim, and Russell Cole. "Selecting Benchmark and Sensitivity Analyses." Evaluation Technical Assistance Brief. Washington, DC: Office of Adolescent Health, U.S. Department of Health and Human Services, 2017. Available at https://opa.hhs.gov/sites/default/files/2020-07/selecting-benchmark-and-sensitivity-analyses.pdf

Lee, Joanne, and Russell Cole. "Pre-Post Outcome Study How To Guide." Washington, DC: Office of Population Affairs, Office of the Assistant Secretary for Health, U.S. Department of Health and Human Services, 2020. Available at https://rhntc.org/resources/pre-post-outcome-study-how-guide-tpp-programs

Luca, Dara Lee, and Russell Cole. "Estimating Program Effects on Program Participants." Technical Assistance Brief. Washington, DC: Office of Adolescent Health, U.S. Department of Health and Human Services, 2017. Available at https://opa.hhs.gov/sites/default/files/2020-07/estimating-program-effects-on-program-participants-brief.pdf

Marin, Barbara VanOss, Douglas B. Kirby, Esther S. Hudes, Karin K. Coyle, and Cynthia A. Gómez. "Boyfriends, Girlfriends, and Teenagers' Risk of Sexual Involvement." *Perspectives on Sexual and Reproductive Health*, vol. 38, no. 2, 2006, pp. 76–83. https://doi.org/10.1363/3807606

Schochet, Peter. "Guidelines for Multiple Testing in Impact Evaluations of Educational Interventions." Washington, DC: Institute of Education Sciences, U.S. Department of Education, 2008. Available at https://www.mathematica.org/publications/guidelines-for-multiple-testing-in-impact-evaluations-of-educational-interventions

Trenholm, Christopher, Barbara Devaney, Ken Fortson, Lisa Quay, Justin Wheeler, and Melissa Clark. "Impacts of Four Title V, Section 510 Abstinence Education Programs: Final Report." Princeton, NJ: Mathematica Policy Research, 2007. Available at https://www.mathematica.org/publications/impacts-of-four-title-v-section-510-abstinence-education-programs

Wasserstein, Ronald L. and Nicole A. Lazar. "The ASA's Statement on p-Values: Context, Process, and Purpose." *The American Statistician*, vol. 70, issue 2. 2016, pp. 129–133. https://doi.org/10.1080/00031305.2016.1154108

Wasserstein, Ronald L., Allen L. Schirm, and Nicole A. Lazar. "Moving to a World Beyond "p< 0.05"." The American Statistician, vol. 73, sup1. 2019, pp. 1–19. https://doi.org/10.1080/00031305.2019.1583913

Zanon, C., Hutz, C. S., Yoo, H. (Henry), & Hambleton, R. K. (2016). An application of item response theory to psychological test development. *Psicologia: Reflexão e Crítica, 29*(1), 18. https://doi.org/10.1186/s41155-016-0040-x

Endnotes

[1] p-hacking or data mining refers to an unethical practice of selecting or cherry-picking only the favorable results from your analysis, or conducting a large number of analyses until a statistically significant result emerges (Head et al. 2015).

[2] Ceiling effects happen when outcome scores are high across all participants. In such a situation, where all individuals have high outcome scores, by definition, there cannot be any difference in the treatment and comparison group averages, because they are all at the ceiling.

[3] If you had differences in control services across sites, for instance you randomized within schools and each school offered different control services, the differences in the effective contrast may also help explain differences in impacts across sites.

Recommended citation: Knab, Jean, Russell Cole and Emily LoBraico. (2022). "No Impacts?: How to Enhance What you Learn from your Evaluation." OPRE Report Number 2022-149. Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

Submitted to:
Selma Caal, Project Officer
Kathleen McCoy, Project Monitor
Office of Planning, Research, and Evaluation
Administration for Children and Families
U.S. Department of Health and Human Service

Contract number:
HHSP233201500035I/HHSP23337008T

Submitted by:
Jean Knab, Project Director

Mathematica
600 Alexander Park, Suite 100
Princeton, NJ 08540 P.O. Box 2393
Telephone: (609) 799-3535

Mathematica reference number:
50238.01.L42.101.000