

In the text below, I summarize the results of this study in more detail. Please refer to the attached poster for more information on the background, research questions, and methodology. I have also included the synthesis of the literature in the poster.

Results

Following the literature search, five studies were selected as recent examples of interventions illustrating the use of a FOI framework. While additional studies were found in the literature search, most echoed similar interventions and approaches to measuring fidelity.

Commonalities

All of the five studies represent interventions where teachers are considered program providers, meaning they are responsible for implementation. This is unsurprising given the nature of K-12 educational interventions, which typically involve the implementation of a new curriculum or additional professional development learning. Teachers are evidentially the primary providers of education and therefore the target of interventions that seek to improve student learning.

Existing FOI frameworks all stress the need to specifically list out the essential components of a program. The emphasis on structure makes these interventions suitable for evaluating fidelity, and the use of dosage-related indices are thus quite ubiquitous across FOI studies. Perhaps most surprising among the evaluations reviewed is the pervasiveness of measures relating to quality of delivery. Measuring quality, even outside the realms of fidelity, is a difficult concept to articulate. In the context of a curriculum intervention, what constitutes high quality of implementation beyond simple measures of dosage? Each of the four studies measuring a curriculum intervention operationalized quality by indicating essential components of best pedagogical practice, in line with the recommendation of Century and colleagues (2010). The common method for evaluating fidelity to instructional-pedagogical critical components is the use of observation ratings.

Limitations

Most notably missing from the reviewed studies were indices of participant responsiveness. Contrary to instructional-pedagogical critical components, engagement measures what the students experience through the intervention, such as increased teamwork and discussion. In a sense, engagement represents what the participant – the individual on the receiving end of the intervention – is supposed to experience.

Another limiting feature of the FOI evaluations presented is the data collection demand. Consider the measurement of fidelity to quality of delivery. As mentioned previously, the method by which most evaluators in the studies examined measured this construct was through observation ratings. The extent to which one or two class sessions represent the entirety of a teacher's fidelity to a curriculum is certainly questionable. The financial resources and time necessary to conduct an increased number of observations, however, is generally difficult to attain (Lee, Penfield, & Maerten-Rivera, 2009). Additionally, there is concern as to whether a teacher might increase fidelity for that particular class knowing they are being observed. Perhaps this explains why quality of delivery was generally unrelated to the measured outcomes across the studies examined. An additional consideration with regards to data collection is self-reported

data, which can certainly be biased. This is especially problematic when accountability is tied to fidelity.

A final consideration is the methodological link between fidelity and intended outcomes. Among the many purposes for FOI evaluations is to determine which components of a program are most essential and attributable to potential impacts. Ironically, the extent to which most analyses of FOI are capable of indicating such a causal inference are very limited. Compliance to a program or curriculum is often endogenous to unobserved factors.

Rethinking fidelity

Given the aforementioned tension between fidelity and teacher independence, it is unclear how the FOI frameworks accommodate programs and interventions that seek to alter teacher practices while respecting context and teacher expertise. How should evaluators handle diversity in implementation when evaluating educational interventions, specifically when implementation is not intended to follow a structured pattern?

In the ensuing case study, I offer an alternative perspective in the form of *integrity of evaluation* (see Byrk, 2016), whereby participant responsiveness can be leveraged as a key indicator of implementation for interventions that do not follow a structured form of practice. The case study refers to an evaluation of the Proficiency Based Education (PBE) curriculum reforms that have taken place in Maine throughout the last two decades (Shakman et al., 2018). The evaluators were tasked with understanding the implementation and impact of the reform. The PBE reform provided flexibility to schools and teachers with regards to implementation, only providing a holistic ideal of how classroom instruction should look like. As such, the existing FOI frameworks were deemed inappropriate for such an evaluation.

Instead of studying the program providers – as is typical in the aforementioned FOI studies above – the evaluators surveyed students concerning their exposure to aspects of classroom instruction and experience that may have differed under the new reform. Latent Profile Analysis (LPA) on the survey results was used to classify distinct subgroups of students distinguished by their exposure to elements of PBE.

The participant-focused approach mitigated the prominent limitations of the aforementioned FOI evaluations. Firstly, data collection simply required survey administration to *participants* (the students) as opposed to self-reporting by the program providers (the teachers) or observation ratings. Secondly, by surveying students, the quantitative measure of implementation reflects the *experience* of the reform felt by the students, which is more indicative of the goals outlined by the state's legislature. Finally, by structuring the PBE policy in a way that granted flexibility to schools and teachers, additional stress was not placed on the teachers to implement the policy in a way that did not speak to their respective contexts. As such, the measure of implementation reflects a truer reality of the program in practice as opposed to one that is mandated and perhaps ill-suited for certain contexts.

References

- Abry, T., Hulleman, C. S., & Rimm-Kaufman, S. E. (2015). Using indices of fidelity to intervention core components to identify program active ingredients. *American Journal of Evaluation*, 36(3), 320-338.
- Benner, G. J., Nelson, J. R., Stage, S. A., & Ralston, N. C. (2011). The influence of fidelity of implementation on the reading outcomes of middle school students experiencing reading difficulties. *Remedial and Special Education*, 32(1), 79-88.
- Berman, P., & McLaughlin, M. W. (1976, March). Implementation of educational innovation. In *The educational forum* (Vol. 40, No. 3, pp. 345-370). Taylor & Francis Group.
- Buxton, C. A., Alleksaht-Snyder, M., Kayumova, S., Aghasaleh, R., Choi, Y. J., & Cohen, A. (2015). Teacher agency and professional learning: Rethinking fidelity of implementation as multiplicities of enactment. *Journal of Research in Science Teaching*, 52(4), 489-502.
- Bryk, A. (2016). Fidelity of implementation: Is it the right concept. *Carnegie Commons Blog*, 17.
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation science*, 2(1), 40.
- Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. *American Journal of Evaluation*, 31(2), 199-218.
- Crawford, L., Carpenter, D. M., Wilson, M. T., Schmeister, M., & McDonald, M. (2012). Testing the relation between fidelity of implementation and student outcomes in math. *Assessment for Effective Intervention*, 37(4), 224-235.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: are implementation effects out of control?. *Clinical psychology review*, 18(1), 23-45.
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American journal of community psychology*, 41(3-4), 327-350.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: implications for drug abuse prevention in school settings. *Health education research*, 18(2), 237-256.
- Fetters, M. D., & Molina-Azorin, J. F. (2020). Utilizing a mixed methods approach for conducting interventional evaluations. *Journal of Mixed Methods Research* 14(2), 131-144.

- Gates, E. F. (2016). Making sense of the emerging conversation in evaluation about systems thinking and complexity science. *Evaluation and Program Planning*, 59, 62-73.
- Ginsburg, A., & Smith, M. S. (2016). Do randomized controlled trials meet the “Gold Standard”. *American Enterprise Institute*. Retrieved March, 18, 2016.
- Hanita, M., Bailey, J., Khanani, N., & Bocala, C. (2020). *Connecticut Teacher Education and Mentoring Program (TEAM): Fidelity of Implementation and Teacher Retention*. (REL 2020–015). Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northeast & Islands.
- Keller-Margulis, M. A. (2012). Fidelity of implementation framework: A critical need for response to intervention models. *Psychology in the Schools*, 49(4), 342-352.
- Kim, J. S. (2019). Making every study count: Learning from replication failure to improve intervention research. *Educational Researcher*, 48(9), 599-607.
- Lee, O., Penfield, R., & Maerten-Rivera, J. (2009). Effects of fidelity of implementation on science achievement gains among English language learners. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 46(7), 836-859.
- Lynch, S., & O'Donnell, C. (2005, April). The evolving definition, measurement, and conceptualization of fidelity of implementation in scale-up of highly rated science curriculum units in diverse middle schools. In *Annual meeting of the American educational research association*.
- Meyer, A., Miller, S., & Herman, M. (1993). Balancing the priorities of evaluation with the priorities of the setting: A focus on positive youth development programs in school settings. *Journal of Primary Prevention*, 14, 95-113.
- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American journal of evaluation*, 24(3), 315-340.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2010). A Procedure for Assessing Fidelity of Implementation in Experiments Testing Educational Interventions. *Society for Research on Educational Effectiveness*.
- Nelson, M. C., Cordray, D. S., Hulleman, C. S., Darrow, C. L., & Sommer, E. C. (2012). A procedure for assessing intervention fidelity in experiments testing educational and behavioral interventions. *The Journal of Behavioral Health Services & Research*, 39(4), 374-396.

- O'Donnell, C. L. (2008). Defining, conceptualizing, and measuring fidelity of implementation and its relationship to outcomes in K–12 curriculum intervention research. *Review of educational research*, 78(1), 33-84.
- Shakman, K., Foster, B., Khanani, N., Marcus, J., & Cox, J. (2018). " In Theory It's a Good Idea": Understanding Implementation of Proficiency-Based Education in Maine. *Education Development Center, Inc.*
- Tyack, D. B., & Cuban, L. (1995). *Tinkering toward utopia*. Harvard University Press.
- Weiss, C. H., Murphy-Graham, E., Petrosino, A., & Gandhi, A. G. (2008). The fairy godmother—and her warts: Making the dream of evidence-based policy come true. *American journal of evaluation*, 29(1), 29-47.
- Zvoch, K. (2012). How does fidelity of implementation matter? Using multilevel models to detect relationships between participant outcomes and the delivery and receipt of treatment. *American Journal of Evaluation*, 33(4), 547-565.