
Teacher & Leadership
Programs



TEACHER
INCENTIVE
FUND

+



TEACHER & SCHOOL
LEADER INCENTIVE
PROGRAM

Dimensions of Dosage: Evaluation Brief for TIF Grantees

White Paper

Jessica Bailey
Karen Shakman
Dana Ansel
Sara Greller
Education Development Center

October 2016

Table of Contents

Introduction	1
Overview of Dosage	1
Dimensions of Dosage.....	2
Data Needed to Study Dosage	3
Additional Considerations.....	4
Conclusion.....	6
Appendix A. Additional Resources.....	7
Appendix B. Key Questions to Guide Discussion of Dosage	8

Introduction

Even when it is not possible to identify a good comparison group for evaluating the effects of Teacher Incentive Fund (TIF) activities, it may be possible to use variations in the “dosage” of the TIF components that educators experienced. Dosage refers to the amount of the intervention that is delivered—for example, the number of hours teachers spend with a teacher leader. While dosage is not a specific research method, understanding and accurately measuring dosage can be an important step toward providing evidence that the activities the TIF grant funded have an impact. In addition, understanding and measuring dosage is important for the replication and scaling-up of effective interventions. While dosage is not a specific research method, understanding and accurately measuring dosage can be an important step toward providing evidence that the activities the TIF grant funded have an impact (Box 1).

As TIF grantees create, implement, and evaluate new interventions with the goal of improving educator effectiveness and student achievement, it is important to understand what dosage is and how it might affect research results. This brief provides TIF grantees with an overview of dosage and the importance of carefully measuring dosage when determining the outcomes of an intervention. It also includes a case study of how the dosage of teacher coaching turned out to be an important consideration for one TIF grantee.

Box 1. Using Measurements of Dosage to Evaluate TIF Programs

In educator evaluation systems, such as those grantees created as part of TIF, dosage is relevant to measuring the impact of a variety of program inputs (e.g., teacher supports or financial incentives) on program outcomes (e.g., teacher ratings or teacher perceptions). In particular, dosage may be a relevant consideration when evaluating **TIF Goal 1** and **TIF Goal 2**.

For TIF Goal 1, “Improving student achievement by increasing teacher and principal effectiveness,” the evaluator should consider the dosage of different types of strategies grantees use to increase teacher and principal effectiveness. For example, grantee programs may include teacher support from teacher leaders or coaches, but these supports may differ in terms of their dosage, such as the frequency of meetings or the length of each session.

For TIF Goal 2, “Reforming teacher and principal compensation systems so that teachers and principals are rewarded for increases in student achievement,” the evaluator might consider any financial incentives as a type of dosage, since the amount of financial incentives or the timing of the dispersal of financial incentives might vary. These variations can be thought of as different doses of the intervention.

Overview of Dosage

The concept of dosage stems from medical research and clinical practice, but is becoming increasingly relevant in the social sciences. Dosage refers to the **quantity or amount of an intervention**. Dosage can be an important element of program implementation, often influencing the intervention’s impact.

Consider the following example:

- An initiative aimed at increasing teacher effectiveness paired a cohort of teachers with coaches who gave them feedback on their instructional practices. In one district, teachers met with their coaches twice

during the academic year, and an evaluation of this program found no increase in teacher effectiveness, as measured by the participating teachers' final teacher evaluation ratings. In another district, teachers met with their coaches every week, and evaluators found a notable increase in the participating teachers' final evaluation ratings.

In both districts, the intervention—using coaches to improve instructional practices—was the same; however, the dosage of the intervention differed. In this situation, dosage refers to the frequency of the meetings with coaches, and the more frequent meetings in the second district may have contributed to the positive impact of the intervention.

This example illustrates the important role dosage can play in understanding the outcomes of an intervention. Before concluding that an intervention does or does not work, evaluators should consider dosage and whether changes in dosage might influence the impact of the intervention.

At the same time, evaluators and TIF grantees would be mistaken to consider dosage in isolation from other factors. While dosage is important, if the content of a professional development intervention is low quality, for example, increasing the dosage will not necessarily improve the outcomes.

Dimensions of Dosage

As grantees think about how dosage might be relevant to their TIF program, they should note that there are several dimensions of dosage. The dosage might vary in terms of the *frequency* of the intervention, the length or *duration* of the intervention, or the *intensity* of the delivery. See Box 2 for definitions of some of the key terms related to dosage.

Grantees should also note that they can either hold dosage constant during an intervention or make it one of the variables of an intervention. For example, one program may offer the same financial incentives to all teachers who achieve certain performance goals. In this case, the grantee holds the dosage of the intervention constant. Another program, however, could have two groups of teachers, and the grantee offers each group different financial incentives for the same performance. In this case, the dosage is a variable of the intervention in order to try to determine the threshold of financial incentives that affects performance.

Varying the dosage of an intervention can allow grantees to answer many different research questions. Examples of research questions related to dosage include the following:

- How frequently should teachers meet with teacher leaders?
- What should be the length of each session?
- What amount of incentive pay may relate to improvements in teacher practice?

Answers to these types of questions can help define the threshold where the intervention affects outcomes. These answers can help tailor future interventions to deliver the optimal amount of the intervention. This information may also help preserve resources and enable programs to support more teachers.

Data Needed to Study Dosage

As TIF grantees start to think about the data they will need to study the dosage of an intervention, they will want to clarify the type of dosage to be measured (e.g., frequency, duration, intensity). Using the previous example of instructional coaching meetings with teachers, the grantee must consider whether the dosage of interest is *frequency* (how often the meetings occur), *duration* of the intervention or a session (for how many months the meetings should occur or how long each meeting should be), or *intensity* (the extent to which the meetings have a specific protocol or set of guidelines that must be followed).

The most appropriate dosage to measure in an intervention—and the data required to do so—depends on a grantee’s theory of action (i.e., the assumptions guiding the intervention and the expected changes it will elicit), the grantee’s questions of interest, how the grantee will measure the outcome(s) of interest, and the type of analysis that grantee will conduct to answer the questions.

In the example of an intervention using coaching to improve instructional practice, the outcome of interest might be teacher evaluation ratings. In this case, the grantee might want to study the impact of the intervention by using a rigorous design such as a randomized controlled trial (RCT) or matched comparison-group design to compare the evaluation ratings of teachers who received the coaching with the evaluation ratings of similarly situated teachers who did not receive the coaching. To look at the impact of dosage specifically, the grantee might want to compare the evaluation ratings of teachers who received different amounts (frequency or duration) or types (intensity) of coaching. For more on matched

Box 2.

Definition of Key Terms

Intervention. A program or initiative that is the subject of investigation using research methods.

Dosage. The amount of an intervention that is delivered.

Frequency. How often the intervention is delivered. For example, interventions that require a different number of meetings per month have different frequencies. Even though the total amount of time delivered for an intervention could be the same (e.g., 3 hours), the frequency could differ (e.g., one 3-hour meeting per month versus two 1.5-hour meetings per month).

Duration. Duration refers to the length of the intervention as a whole or of a particular session. For example, if a monthly coaching program is intended to last for 9 months, the duration of the intervention is 9 months. Similarly, if the monthly meeting is 45 minutes, then the duration of the session is 45 minutes.

Intensity. The strength of an intervention or how much of a particular component of an intervention. For example, a coaching program may follow a specific format that requires coaches to provide each participant with a minimum of three pieces of substantive feedback per session.

Threshold. A specific dosage level at which an intervention affects outcomes. For example, a study may find that teachers only need to attend four out of five days of offered professional development to achieve the desired level of change in their practice. Although five days might originally have been the duration of the intervention, the evaluation of the program might find that four days is sufficient for changing outcomes. Therefore, four days of the intervention is the threshold dosage for impact.

Fidelity. Operating a program, initiative, or intervention as intended in order to achieve the desired outcomes.

Adapted from the research brief, Wasik, B. A., Matterna, S. K., Lloyd, C. M., & Boller, K. (2013). Intervention dosage in early childhood care and education: It’s complicated (OPRE Research Brief OPRE 2013-15). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services.

comparison-group design, please see a companion evaluation brief, “Matched Comparison-Group Design Evaluation Brief.”

The following types of data might be important to collect when studying dosage:

- Characteristics of teachers and coaches (e.g., grade level, experience level)
- School characteristics (e.g., level and grade span)
- Frequency of meetings (e.g., once a week or once a month)
- Duration of meetings (e.g., 30 minutes or 1 hour)
- Intensity of conversation (e.g., did staff have in-depth discussions guided by a protocol or more open-ended conversations without specific guidelines?)
- Measures of instructional practice (e.g., teacher observation score)
- Teachers’ or other stakeholders’ perceptions regarding the intervention. These could be collected via survey or interviews (e.g., responses may include information about frequency, duration, and intensity, as well information on how supported teachers felt and the types of support they received)
- Time logs (e.g., participants could track the frequency and duration of the intervention)
- Other outcome data, such as teacher evaluation ratings or student achievement

Additional Considerations

Here are some additional considerations to keep in mind when thinking about measuring dosage of an intervention and its impact:

- **Intended, Offered, and Received Dosage.** It is important to distinguish between the *dosage intended* in the intervention, the *dosage offered*, and the *dosage* program participants *received*. The intended dosage may differ from the actual delivered dosage. For example, an instructional coaching program may be designed to offer teachers support during one meeting per week for 6 months, but a coach may only be able to schedule meetings every other week for 6 months because of staffing shortages or scheduling conflicts. In such a case, the offered dosage would be half the intended dosage. In addition, the dosage received by participants may vary from either the intended or offered dosage. Using the same example, although the meetings may be scheduled every other week, one of the teachers in the program may have repeated absences; as a result, she may only attend half the scheduled meetings. In this case, the received dosage is one-quarter of the intended dosage. It is important to collect data on all three—intended, offered, and received dosage—if possible.
- **Fidelity.** Dosage and fidelity of implementation link closely. Fidelity refers to whether an intervention is implemented as it was intended, and dosage can be an important element of fidelity. In the previous example of intended dosage versus offered dosage, the intended coaching intervention was scheduled weekly, but that did not occur. Therefore, there was not complete fidelity of implementation of the intervention. If the dosage is intended to be a constant (e.g., all participants receive the same amount of the intervention), mechanisms must be in place to measure what actually occurs. If there is not fidelity of

implementation and an intervention does not achieve the desired outcomes, evaluators have difficulty disentangling whether the intervention was ineffective or the lack of fidelity caused the intervention to be ineffective.

- **Optimal Dosage.** Grantees should consider how they will know when they have reached the optimal dosage. Reviewing the existing literature on the intervention and dosage is a useful first step. For example, research shows that single-day teacher training workshops do not provide sufficient dosage to improve teachers' classroom practices or affect teacher learning over the long term (Joyce & Showers, 2002). When designing, developing, and evaluating interventions, grantees should have a theory of action regarding the optimal dosage and identify indicators that will determine when this threshold has been reached.

Grantee Spotlight: New York City Department of Education

TIF Grantee New York City Department of Education (NYC DOE) uses teacher leadership as a means to increase student achievement. The NYC DOE program increases teacher leadership capacity through professional learning opportunities for teachers. Teacher leadership roles include peer coaching and master teacher roles. Citywide, NYC DOE has almost 700 teacher leaders; the TIF grant funds 370.

While NYC DOE did not explicitly set out to study dosage per se, one of its TIF evaluation questions was about the relationship between the frequency of collaboration among teacher leaders and teachers and (1) the effectiveness of the collaboration and (2) use of the Danielson Framework. (The Danielson Framework for Teaching is a normative description of teaching practices that NYC DOE has incorporated into its teacher evaluation system.)

To answer this question, NYC DOE examined dosage, in this case, how frequently teachers worked with their teacher leaders, using data collected in the district's teacher survey. NYC DOE surveyed over 4,000 teachers in the 2014–15 school year. The survey measured dosage using a 6-point Likert scale (a scale commonly used in surveys that requires the respondent to indicate the extent to which he/she agrees or disagrees with a statement) asking how often teachers worked with their teacher leaders. The survey also probed how effective the teachers reported that work to be.

According to the evaluation report the external evaluator provided:

“Of the teachers who met with their teacher leaders more than once a month, a majority agreed that it helped them to see the strengths and weaknesses of their practice more clearly (55%) and to improve their own instructional practice (59%). Teachers who met with teacher leaders once a month or less, in contrast, were half as likely to report such impact.”

Thus, NYC DOE found preliminary evidence of a threshold effect for the meetings between teachers and their teacher leaders—meeting more than once a month. In addition, although survey respondents' perception of the Danielson Framework was mixed, the evaluation showed that teachers who worked more frequently with teacher leaders were significantly more likely to report it was a helpful tool for improving practice.

NYC DOE has used the findings about the frequency of collaboration to inform its conversations with school principals. In these conversations, NYC DOE stresses that working with a smaller group of teachers more frequently may be more important than spreading the work across more teachers with less frequent contact. As a result of this finding, the district also decided to survey teacher leaders to learn which teachers they worked with and how frequently. This additional survey will enable the district to triangulate the results across the two groups. They also hope to examine the long-term relationship between the contact of teacher leaders with teachers and student achievement.

Conclusion

Understanding the dosage of an intervention is necessary for effective program implementation. Dosage can refer to the frequency, duration, or intensity of an intervention, such as the number and length of professional development meetings or the dollar amount of financial incentives. As TIF grantees plan to implement new programs or initiatives to increase teacher effectiveness, considering the dosage of the intervention is an important aspect of program evaluation.

The following key points will support TIF grantees considering measuring dosage:

- Existing research or a theory of action may offer grantees guidance regarding how much of an intervention is needed to achieve the desired outcome(s).
- Dosage may be held constant across all intervention participants or may differ within an intervention. Varying the dosage may allow a grantee to investigate how much of an intervention might be necessary to achieve the desired outcome(s).
- Monitoring mechanisms must be in place to accurately measure dosage. During the implementation of an intervention, the intended dosage may vary from the actual dosage, and different participants may receive different dosages.
- The data required to measure dosage depends on the program evaluation questions and the type of data analysis that TIF grantees will conduct.
- The dosage of an intervention might affect the outcomes measured.

Appendix A. Additional Resources

Eskolta School Research and Design for the NYCDOE Office of Teacher Recruitment and Quality. *A study of teacher leadership in New York City. Finding #3*. Retrieved from <http://schools.nyc.gov/NR/rdonlyres/C9D2AB65-E24E-47D5-9ACA-348E44AA24E7/196441/TRQ1Satisfaction.pdf>

Downer, J., & Yazejian, N. *Measuring the quality and quantity of implementation in early childhood interventions* (OPRE Research Brief OPRE 2013-12). Washington, DC: Office of Planning, Research, and Evaluation, Administration for Children and Families, US Department of Health and Human Services (2013). Retrieved from <http://www.researchconnections.org/childcare/resources/25564/pdf>

Joyce, B. R., & Showers, B. (2002). *Student achievement through staff development*. Alexandria, VA: Association for Supervision and Curriculum Development.

Wasik, B. A., Mattera, S. K., Lloyd, C. M., & Boller, K. (2013). *Intervention dosage in early childhood care and education: It's complicated* (OPRE Research Brief OPRE 2013-15). Washington, DC: Office of Planning, Research and Evaluation, Administration for Children and Families, U.S. Department of Health and Human Services. Retrieved from http://www.acf.hhs.gov/sites/default/files/opre/dosage_brief_final_001_0.pdf

Appendix B. Key Questions to Guide Discussion of Dosage

- Is there an aspect of your TIF program for which you might study dosage (e.g., professional development, teacher coaching, financial incentives)?
- What is the outcome of interest (e.g., teachers' perceptions, teacher evaluation ratings, student achievement)?
- Is it possible that the dosage might be affecting the outcomes of the intervention?
- Is there existing research or a theory of action that offers guidance regarding how much of the intervention is needed (the threshold) in order to achieve the desired result?
- Is the dosage intended to be held constant across all participants or is it intended to vary? How will you monitor the dosage?
- What is your intended level of dosage (frequency, durations, and intensity)? Was the actual level of dosage the same as the intended level, and did all participants receive the intended dosage?