

Documenting and Evaluating Educational Impact Through Program Evaluation

Denise R. Foley
Worcester State University, USA

Abstract

Per federal and state laws, interventions in publicly funded US schools must be evidence-based and decisions to adopt, adapt, continue, or discontinue interventions are data-driven. All educators, regardless of their role or level of direct interactions with students are rated on their "Impact" on student achievement. Properly conducted program evaluations address several key research to practice gaps including over-reliance on high-stakes testing results to measure educator effectiveness. We describe twenty-eight program evaluations conducted by preservice school psychologists in Pre-K through secondary school settings over a five-year period which not only measured intervention effectiveness from the individual student to classroom-wide and district-levels, but also provided models to quantify their effect on PreK-12 student outcomes as required by the state-mandated educator evaluation system. Benefits extended into job search and employment.

1. Introduction

Program evaluation is a broad term describing the systematic collection, analysis, and usage of data to answer a variety of questions about a program. Any combination of qualitative and quantitative methodologies may be incorporated to assess and improve a program's design, implementation, effectiveness or efficiency [1]. Program evaluations are prevalent in educational and school-based programs, as well as, within human service, health-care and charitable organizations and government agencies. The scope of inquiry that program evaluations may address is highlighted when the terms evaluation and program are addressed separately.

An evaluation of a program may be formative or summative in nature. Formative assessments are typically created at the design stage of a program and conducted multiple times such that adjustments may be made during the development and implementation stages of a program. Summative evaluations, by contrast, are typically conducted at the end of a program and focus on outcomes, results and impact relative to previously identified goals [2].

At least in school settings, the term "program" is even broader than "evaluation." Its range encompasses a single child's individual special education services and a social skills intervention with a small group of students, as well as, practices and policies at the classroom, school and/or district-wide levels [3].

As with any research approach, the expectation is that program evaluations reflect the principles of the scientific method such that problems are recognized or questions raised and operationally defined, observations made, reliable and valid data are systematically gathered, and hypotheses formulated and tested, before conclusions or decisions are made [4].

2. Program Evaluation in U.S. Schools (The Reality)

Federal legislation, for example, the *No Child Left Behind Act* of 2002 and the *Every Student Succeeds Act* of 2015 requires that interventions in the publicly funded schools in the United States be supported by research findings and that decisions to adopt, adapt, continue, or discontinue interventions are based on data-driven evidence [5], [6].

Consistent with the federal mandates, the literature bases for school-based professionals including administrators [7], teachers [8], and school psychologists [9], among other professional school-based service providers (e.g., speech-language therapists [10]), strongly endorse the use of evidence-based programs in public schools. Relatedly, the ease with which educators may objectively determine the effectiveness of any program or interventions increases annually (e.g., What Works Clearinghouse; National Center for Intensive Interventions). That said, there is a surprising disconnect between what public school educators understand and embrace about the concepts and components of program evaluation and their regular practices.

Numerous researchers have shown that the majority of decisions around school-based programs are not evidence-based. Rather, decisions to adopt, adapt, continue or discontinue programs from the level of individual child to a district as a whole, are typically "authority-based." In other words, "informed by past practices, experience, and the

advice of authority figures rather than research findings [11, p.314].”

The discrepancy between educator belief and action extends to formative and summative evaluations. Though both are generally considered essential for effectiveness [12], by far the greatest emphasis in schools across the nation has been on narrowly-focused, summative assessments: specifically, on high-stakes tests measuring student progress on standards-based curriculums--not on evaluating interventions [13], [14], [15]. Despite their ubiquity, these evaluation results typically provide too little information, much too late (six to twelve-plus month delay) to determine a student’s academic growth, much less the effectiveness of any intervention.

3. Research to Practice Gaps in Evaluating Effectiveness

One potential driver of this over-reliance and focus can be seen in Massachusetts, that like many states, mandates that educator evaluations be linked to high stakes student outcome data. As shown in Figure 1, ALL Massachusetts educators, regardless of their role or level of direct interaction with students are rated on their “Impact” on student learning and growth (low, moderate or high).

Educators earn two separate ratings

Summative Rating	Exemplary	1-YEAR SELF-DIRECTED GROWTH PLAN	2-YEAR SELF-DIRECTED GROWTH PLAN	
	Proficient			
	Needs Improvement	DIRECTED GROWTH PLAN		
	Unsatisfactory	IMPROVEMENT PLAN		
		Low	Moderate	High
Rating of Impact on Student Learning (multiple measures of performance, including MCAS Student Growth Percentile and MEPA where available)				

Figure 1. Massachusetts Educator Evaluation Ratings

In addition, all must demonstrate their level of proficiency (Unsatisfactory to Exemplary) across four standards: Curriculum, planning and assessment; Teaching All Students; Family and community engagement; Professional Culture. Massachusetts divides all school personnel into four categories: superintendents, administrators, teachers and Specialized Instructional Support Personnel (SISP). School psychologists are evaluated using the same rubric for school counselors, nurses, and occupational therapists, among other “SISPs.” The rationale for the common rubric is to “develop a consistent, shared understanding of what proficient

performance looks like in practice [16].” Clearly, the expectations and rationale highlight the persistent, yawning gap in understanding not only what school psychologists are trained to do, but also what they actually do.

4. Bridging the Gaps: Documenting and Evaluating Evidence-Based Interventions

The National Association of School Psychologists (NASP) [17,18] policy documents discuss how school psychologists’ roles and responsibilities support effective educational policies and practices, provide guidelines to link their services to improved student outcomes and advocate for methods to accurately and fairly assess the broad and varied roles they assume across schools and districts. Building on the policy statements, Williams and Monahan [19] provide specific examples of “universal”, “selected” and “indicated” levels of academic and behavioral student outcomes that they might use to measure and document the effectiveness of their programs. Program evaluations raise the profile and expanded the practice of school psychologists locally, more aligned with the NASP Practice Model in which data-based decision making and accountability permeates all aspects of service delivery as depicted in Figure 2 [20].

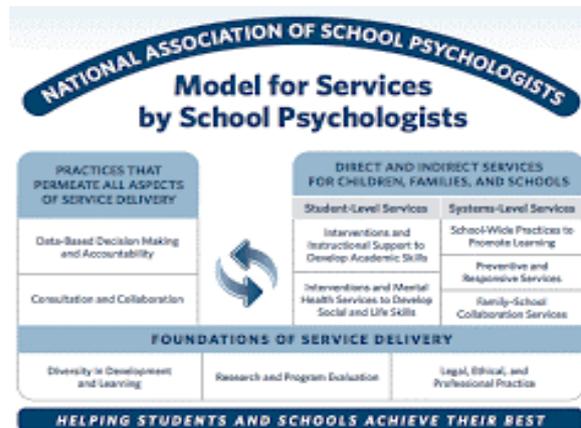


Figure 2. National Association of School Psychologists Practice Model

4.1. Summary of Program Evaluations

Twenty-eight program evaluations conducted in Pre-K through secondary school settings conducted by school psychologists in training over a five-year period which measured program effectiveness from the individual student to classroom-wide and district-levels in twelve K-12 urban and suburban districts surrounding an urban, public university training program in Massachusetts

Though public schools are structured at the macro and micro-levels, for example from the adopted

curriculum to the specific times a school day begins and ends, they are in fact, dynamic and sometimes unpredictable environments. While educators at all levels strive to be preventative and proactive, they are constantly faced with recurring or novel challenges [21]. Mirroring this reality, most programs were somewhat reactive in nature, focusing on a specific group of or individual students with one or more potential or identified difficulties.

Reflecting the differences among the public school practicum settings, each proposed and completed project was unique. All could be variously classified according to several broad characteristics including student grade level, and program focus, type or level. Table 1 Summarizes the characteristics by which the projects were classified.

Table 1. Major Characteristics of Program Evaluations

Characteristic	Definition
Program Focus: Academic	Basic or applied skills in reading, mathematics or written language
Program Focus: Mental/Behavioral Health	Social skills, emotional regulation, on-task behavior, depression, anxiety, attention, etc.
Student Grade Level	Preschool to Grade 12
Program Type	<p><u>Universal:</u> Targeting all students in a specific setting (typically screening at grade or building level)</p> <p><u>Select:</u> Targeting 10-15% of students determined to be at risk for academic and/or mental/behavioral health difficulties</p> <p><u>Intensive:</u> Targeting 1-5% of students with demonstrated academic and/or mental/behavioral health difficulties</p>
Program Level	<ul style="list-style-type: none"> • Individual or small group of students • Classroom(s)- or Grade-level(s) • School- or District-wide

Table 2 summarizes the foci, levels of intervention and the grade levels of the students and schools of the program evaluation completed.

Table 2. Program Evaluations by Focus, Level of Intervention & Grade Level

	Percentage
Focus	
Academic	18
Mental/Behavioral Health	50
More than 1 Area	25
Other	7
Level	
Universal	33
Targeted	43
Intensive	24
Grade Level	
Early Childhood/Elementary (PreK-6)	48
Middle & High School (7-12)	39
Multiple PreK-12 grades	3
District-Wide	10

In terms of traditional Response to Intervention (RTI) tiers, 33% of the interventions for academic and behavioral student outcomes were universal, 43% were targeted/select and 24% were intensive/indicated. In terms of grade levels, 48% of the interventions were with early childhood and elementary students (PreK-6th); 39% were in middle and secondary schools (grade 7-12); 10% were at the district level; and 3% were across multiple grades. Reflecting the mental health crisis in and complex needs of school aged youth, **half** of all projects focused on programs addressing mental and behavioral health concerns and 25% of programs focused on more than one area of need. Twenty-five percent focused exclusively on academic interventions.

Consistent with a significant proportion of published studies conducted in school settings, no program evaluations proposed or executed applied a true experimental design (i.e., included a control group). Rather, most projects employed pre- or quasi-experimental designs. The data and collection processes varied from archival data (e.g., attendance records), to self-reports, survey responses and systematic observations. Analyses varied from simple descriptions and percentage of non-overlapping data points for single subjects to t-tests and ANOVAS.

4.2. Representative Sample of Program Evaluations

What follows are descriptions of three program evaluations which highlight both the breadth and depth of the projects completed. These programs were implemented at the three intervention levels (universal, intensive, and select) with three different foci. The three evaluands were in the behavioral (positive behavioral supports), social emotional (social skills), and academic (oral reading fluency) realms respectively. One was conducted over the course of one full academic year (reading), while the other two were completed over the course of one half academic year or roughly five to sixth months.

4.2.1. Universal Behavioral Health Program. This program evaluation was conducted on a school-wide Positive Behavioral Interventions and Supports (PBIS) program that had been existence for about eight years in a suburban elementary school. The evaluation sought to determine staff “buy-in” to the principles and philosophy of the PBIS model; and the consistency with which the program was implemented across the school. Two types of data were collected to evaluate the PBIS program. The first were an analysis of “GOTCHA” tickets per grade level. GOTCHA tickets were designed for distribution by teachers, related service providers and administrators to students observed engaging in positive behaviors corresponding reflecting the values of the school. These were respectful, responsible, intelligent, courageous and kind behaviors. Consistent with data over previous academic years, Table 3 summarizes how the distribution of “GOTCHA” ticket averages varied widely across grades. On average, kindergartens received the most tokens of positive reinforcement and the average across fourth grade classrooms was lowest.

Table 3. Mean PBIS Tickets Distributed by Grade

Grade	Number of Classrooms	Mean “GOTCHAs” per grade
K	5	244
1	4	79
2	4	87
3	5	191
4	5	40
5	4	83
6	4	64

To measure staff buy-in (a critical element of implementation fidelity and potential program success) was an anonymous survey with nine,

five-point Likert-scale and two open-ended questions. The survey was completed eight percent of school staff members. Key finding included the following:

- 95+% /strongly agreed with basic PBIS premises & practices, e.g., behavior is best taught directly, using positive reinforcement, need for consistency among staff & settings
- 90% believed common data-collection procedures are necessary
- Though 87% regarded PBIS as having a positive impact on student behavior, only 75% reported consistent use of the common strategies (e.g., “GOTCHA” tickets; teaching expectations, rules).
- In contrast to responses to their perceptions of PBIS principles, answers to statements regarding PBIS as it relates to RTI were less clear. Only 70% reported understanding the tiered interventions in place at the school and their roles at each tier of the framework.

Program evaluation resulted in the following recommendations for next academic year:

1. Annual review of the PBIS model for all staff
2. RTI training as relates to behavior
3. Conduct parent training
4. Create common language around behaviors
5. Increase consistency of expectations/use of rewards across grades/settings
6. Provide data collection training and support regarding behavior for teachers

4.2.3 Intensive Social Skills Program. This program evaluation sought to objectively measure generalization of specific social skills outlined in students’ IEPs *outside* of the small group setting where the intervention was implemented, specifically in classrooms. The students were 6th grade boys (N=5) with Social Communication IEP goals. Four were dually-diagnosed with Autism Spectrum Disorder (ASD) and Attention Deficit Hyperactivity Disorder (ADHD). One was eligible for special education services under the category of neurological impairment. Students participated in weekly group sessions following the *Social Thinking* curriculum conducted by a supervising school psychologist and practicum student who alternated leading and recording each students’ use of three-to-five “expected” social communication skills during group. Beginning at week 15, weekly generalization probes began in classrooms.

Analysis of multiple baseline single-subject data comparing structured (group) and unstructured (classroom) settings over two months, found Conversation skills (initiating, maintaining, listening) to be most easily taught, maintained

(100%) and generalizable (100%), This was followed by labelling appropriate Problem Solutions (80%, 40%) and accurately Reading Social Cues (80%, 33%). No gains or generalizations were found for making inferences and predictions or demonstrating flexible thinking.

The collected program evaluation data led to changes in type and level of in-class supports for three students; discussion of current consultation model; discussion of supports needed for classroom teachers.

4.2.3. Targeted Academic Program. This final program evaluation was unique in the literature in that *Read Naturally* was used as a sole intervention (16 to 30 minutes per day over the course of one academic year) and was found to be an effective intervention for closing the achievement gap of students (N=32) in grades two through six with deficits (but not disabilities) within the area of reading fluency. The students attended a small, rural elementary school.

As shown in Figure 3, all but two students either closed the achievement gap based on benchmark and progress monitoring scores from the Dynamic Indicators of Basic Early Literacy- Sixth Edition or made “ambitious” growth as defined by leaders in the field of reading remediation [21].

The program evaluation results had direct impact on tiers of support decisions for students with and added to the professional literature on effective interventions.

5. Conclusions: Evaluating the Program Evaluations

The program evaluations were successful on several levels. First, many of the interventions were not only shown to be effective for PreK through secondary students, but were also efficient and promoted real-time positive changes in programs. Second, these program evaluations helped move the preparation of preservice school psychologists towards the ideal model of school psychology

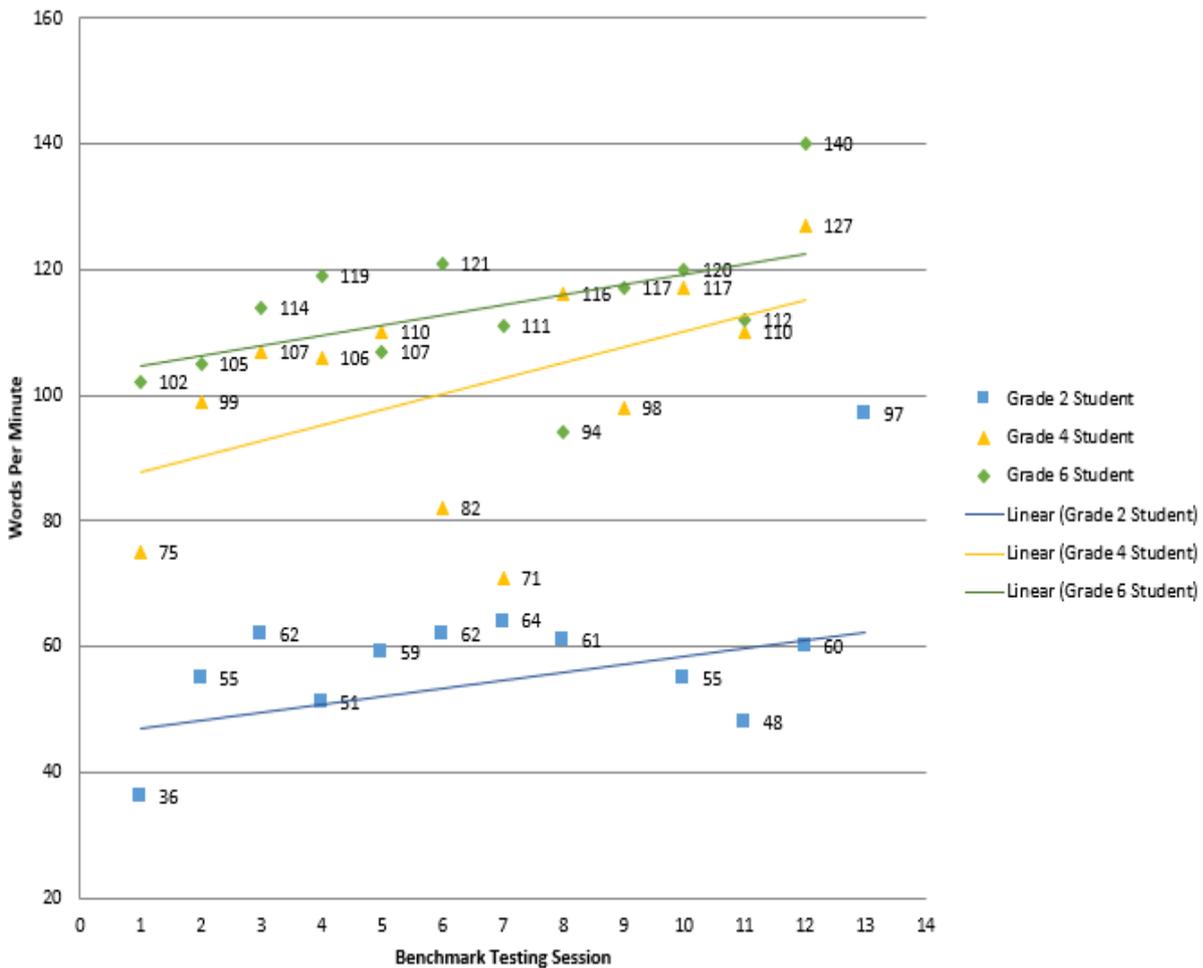


Figure 3. Students Reaching Benchmarks for Oral Reading Fluency

practice model as articulated by the National Association of School Psychologists and demonstrate the positive, powerful (but often invisible) work school psychologists do. Third, through the program evaluations, both preservice and in-service school psychologists determine ways to quantify their “impact” on PreK-12 student “outcomes” as required by the state-mandated) educator evaluation system.

Though most students might not have realized or appreciated it until they were actively seeking employment, given this author’s experience in public schools, this author knew that engaging in research projects had the strong potential to make graduates “stand out” in an otherwise homogenous field of newly minted school psychologists. Perhaps the words of the preservice educator who completed the program evaluation on the targeted academic program (discussed above), upon securing their first job as a school psychologist, best highlight the powerful yet not-quite tangible forth benefit of the program evaluation projects. They wrote:

Hello Dr. Foley, So I had my official interview at [redacted] elementary school... And 2 hours later they offered me the position!! I'm going to take it. The school is really moving in a positive direction with regards to [Response to Intervention] and progress monitoring. I'll get to do lots of consultation and other pieces than just assessment. I will meet with the superintendent next week to sign the contract!

Thank you so much for the letter of reference and all of the experience you included me in. The [Program Evaluation] opportunities with the Read Naturally research and DIBELS really gave me an edge and allowed me to demonstrate that I know how to “Walk the Walk”. See you at graduation!

6. References

- [1] Mills, G.E. & Gay, A.R. (2016). Educational Research: Competencies for Analysis and Applications, 11th Ed. New York: Pearson.
- [2] Harlen, W. (2007). Teachers' summative practices and assessment for learning – tensions and synergies. *The Curriculum Journal*, 16(2), 207-223. DOI:10.1080/09585170500136093.
- [3] Armistead, R. L. & Smallwood, D. L. (2014).The national association of school psychologists model for comprehensive and integrated school psychological services. In P.L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp.9-24). Bethesda, MD: National Association of School Psychologists.
- [4] Pluymer, K. (2014) Problem-Solving Foundations for School Psychological Services. In P.L. Harrison & A. Thomas (Eds.), *Best practices in school psychology: Data-based and collaborative decision making* (pp.25-39). Bethesda, MD: National Association of School Psychologists.
- [5] No Child Left Behind: A Desktop Reference. (2002). Washington, D.C.: U.S. Department of Education, Office of Elementary and Secondary Education.
- [6] Every Student Succeeds Act, 20 U.S.C. § 6301 (2015). Congress.gov/114/plaws/publ95/PLAW-114publ95.pdf.
- [7] Honig, M.I. & Coburn, C. (2008). Evidence-Based Decision Making in School District Central Offices: Toward a Policy and Research Agenda. *Educational Policy*, 22(4), 578-608.DOI: 10.1177/0895904807307067.
- [8]Krelow, A.G. & Blatz, S.L. (2011). The ABCs of Evidence-Based Practice for Teachers. *Teaching Exceptional Children*, 43(5), 8–19.DOI:10.1016/j.teln.2010.01.004.
- [9]Hicks, T.B., Shahidullah, J.D., Carlson, J.S.. & Palejwala, M.H. (2014). Nationally certified school psychologists’ use and reported barriers to using evidence-based interventions in schools: The influence of graduate program training and education. *School Psychology Quarterly*, 29(4), 469-487.
- [10] American Speech-Language-Hearing-Association (2022). Evidence based practice. Retrieved from: <https://www.asha.org/research/ebp/evidence-based-practice-process/>. (Access Date: 2 February, 2021).
- [11] Powers, J.D., Bowen, J.K. & Bowen, G.L. (2010). Evidence-based programs in school settings: Barriers and recent advances. *Journal of evidence-based social work*, 7, 313-331.
- [12] Garrison, C., & Ehrlinghaus, M. (2007). Formative and summative assessments in the classroom. Retrieved from <http://www.amle.org/Publications/WebExclusive/Assessment/tabid/1120/Default.aspx>[13]Au, W. (2007). High Stakes Testing and Curricular Control. *Educational Researcher*, 36(5), 258-267. DOI: 10.3102/0013189X07306523.
- [14] Center on Education Policy. (2012). State high school exit exams: A policy in transition. Washington, DC: Center on Education Policy, Graduate School of Education and Human Development, George Washington University.
- [15] Polsel, J, Rice, S. & Dulfer, N. (2014). The impact of high-stakes testing on curriculum and pedagogy. *Journal of Educational Policy*, 29(5), 640-657.
- [16] Massachusetts Department of Elementary and Secondary Education. (2019). The educator evaluation Massachusetts model system: Evaluating teachers and specialized instructional support personnel. Malden, MA: Author.
- [17]National Association of School Psychologists. (2008). Ready to learn, empowered to teach: Excellence in education for the 21st century. Bethesda, MD: Author. Retrieved from: <http://www.nasponline.org/advocacy/2008educationpolicydocument.pdf>. (Access Date: 12 December, 2021).

[18] National Association of School Psychologists. (2012). A Framework for the personnel evaluation of school psychologists utilizing the NASP Practice model. *Communiqué* 41(3), 26-27. Bethesda, MD: Author.

[19] Williams, B.B. & Monahan, L.W. (2014). Best practices in school psychologists' self-evaluation and documenting effectiveness. In P.L. Harris & A. Thomas (Eds.), *Best practices in school psychology VI* (pp. 581-598). Bethesda, MD: National Association of School Psychologists.

[20] Skalski, A. K., Minke, K., Rossen, E., Cowan, K. C., Kelly, J., Armistead, R., & Smith, A. (2015). *NASP Practice Model Implementation Guide*. Bethesda, MD: National Association of School Psychologists.

[21] Fuchs, L.S., Fuchs, D., Hamlett, C.L., Walz, L., & Germann, G. (1993). Formative evaluation of academic progress: How much growth can we expect? *School Psychology Review*, 22, 27-48.

[22] Darling-Hammond, L. (2006). Constructing 21st-Century Teacher Education. *Journal of Teacher Education*, 57(3), 1-15 DOI: 10.1177/0022487105285962.