It's Time to Make More Room for Program Evaluation in the Education Doctorate Program

Jessica A. Marotta ⓘ
Marymount University
jmarotta@marymount.edu

ABSTRACT

This essay highlights the value of an applied methodology course in program evaluation in the education doctorate program by exploring several benefits that it offers to enhance a doctoral student's ability to solve complex problems of practice. Observations and recommendations are made based on designing and teaching two cohorts of EdD students in a program evaluation course. Improvement science is referenced throughout to highlight how the two may complement each other, not to place a higher value on one or the other. How and where program evaluation and improvement science appear to have possible areas of overlap, along with a brief overview of the major differences, are discussed. The author maintains that both program evaluation and improvement constitute a tremendous capacity to provide the ideas, tools, and approaches to prepare students to be the change agents they hope to aspire to be in their present and future roles as scholarly practitioners.

KEYWORDS

education doctorate, program evaluation, improvement science, Carnegie Project on the Education Doctorate (CPED)

The purpose of this essay is to highlight the value of an applied methodology course in program evaluation in the education doctorate program. It is also to widen the view, or for some to revisit, how program evaluation methodology can enhance doctoral students' research coursework. Improvement science will be used throughout various aspects of this discussion to highlight how the two may complement each other, not to place a higher value on one or the other. Both are considered applied methodology approaches since they are used to solve practical problems through the use of data collection methodologies. As the latest Carnegie Project on the Education Doctorate (CPED) convening in October 2022 showed, improvement science was found in several session titles, while program evaluation had zero title references. Therefore, this essay is written to highlight the benefits program evaluation offers and, in my opinion, why program evaluation should be discussed more widely as an effective applied methodology for doctoral students. I use the term program evaluation loosely to embody the variety of different programs practitioners are involved in observing, leading, evaluating, or participating in at the local, district, state, or national levels. There are a variety of different program evaluations found throughout the vast literature on evaluation such as: formative, summative, outcome, needs assessment, evaluability assessment, and impact evaluation. While not all program evaluation types align with improvement science, there are elements from some, such as formative or implementation evaluations, which tend to have more similarities than other types.

It is important to first provide some background information on the students enrolled in our doctorate program to better appreciate the rationale behind this support for more emphasis on program evaluation. The university's EdD program currently enrolled our sixth cohort with the first cohort graduating Spring 2023, became members of CPED in 2021, and are a fully online program, enrolling students nationally and internationally. The students come from a variety of professional backgrounds and settings. For example, the first cohort (entered Fall 2020) had approximately 25% of students coming from non-educational settings and cohort four (Spring 2022) had approximately 40% of students coming from non-educational settings. While the college does not know if this trend will continue, it certainly recognizes the impact this has on the planning, designing, and teaching of coursework. For example, improvement science is used widely with doctoral students working in educational settings and have access to a sample population for data collection and analysis. Our EdD program does not offer a course in improvement science, and our signature pedagogy is not improvement science; therefore, there is no expectation or arrangement required that our students would conduct research within their local context for their Dissertation in Practice (DiP) or for course assignments. Since improvement science has increasingly been discussed as a possible signature pedagogy for doctoral programs, especially for those working closely with school systems, I will frame program evaluation within this particular discussion.

Due to the number of students working in non-educational settings, our doctorate program needs to consider applied methodologies in addition to improvement science. I will also be reflecting upon my experiences after designing and teaching two semesters of a course titled, Program Evaluation and Decision Making, (n=81) and how those experiences both shaped some unanswered questions, possible future directions for this course, and

its place in our program. Due to the varied nature of the professionals in our program, I needed to provide a wider diversity of non-educational-based evaluation resources and research which is easy to find within the evaluation field. The research on evaluation is intensive and spans multiple professionals and disciplines. It has also been quite rewarding to find resources that meet both the needs of our students and our CPED-influenced program.

Since most educational doctorates offer part-time programs within an approximate three-year completion, it is worthy to recognize the limitations of our students possibly either designing, implementing, or finishing a program evaluation for coursework. Program evaluations can vary greatly in scope and length, and it is reasonable to acknowledge the difficulty of completing one within the three-year program, let alone for coursework during one semester. While students could focus on a process or formative evaluation and not necessarily see the evaluation to its completion, it still might prove difficult if they do not have the influence or capacity to use their professional context, as with improvement science DiPs. However, the topic of this essay will focus more on the program evaluation coursework students take as a required course as opposed to examining program evaluation DiPs. Therefore, granting students the choice to study a previous, ongoing, or a planned evaluation allows for maximizing their ability to examine the complexities of an evaluation they have or have had direct experience with throughout their years as a practitioner.

Program evaluation and its importance as an applied methodology course have several benefits. If one of the goals of an effective, practice-based DiP is its utility to society or to the local context, then I believe that program evaluation skills and knowledge should be included in this dialogue. However, I will first discuss what program evaluation coursework can offer doctoral students, followed by how and where program evaluation and improvement science appear to have possible areas of overlap. I am not claiming that I have explored all possible overlaps and differences between program evaluation and improvement science, but rather I am recommending that CPED-influenced institutions further integrate program evaluation approaches within their education doctorate conversations. In fact, upon a brief search of program evaluation and improvement science to see if this topic of overlap has been widely and explicitly investigated, there appears to be limited work in this area. For example, in several resources on program evaluation and improvement science, there does not seem to be much mention of each other in the literature (Newcomer et al., 2015; Peurach et al., 2022). In one resource, however, there is a distinct call for how improvement science can benefit the evaluation field and how to increase the "cross-talk and perhaps even cross-fertilization of ideas, techniques, and tools between evaluation and improvement science" (Christie et al., 2017a, p. 7). Ultimately, the goal of this essay is to open this possible crosstalk of what skills, abilities, and knowledge applied evaluation approaches can enhance the current improvement science discussions without the exclusion of one over the other.

WHAT PROGRAM EVALUATION COURSEWORK OFFERS DOCTORAL STUDENTS

Leaders engage with program evaluation in varied ways that require a complex understanding of the intricacies of effective evaluation as an observer, participant, leader, evaluator, communicator, or recipient of evaluation findings and impact. I would argue that most practitioners engage with program evaluation at some level in their daily lives, but many do not see themselves as evaluators or are aware of the methodologies of program evaluation. Given that schools, as well as non-educational organizations and settings, are filled with multiple and various programs, it appears that having a knowledge and skill set in program evaluation would be a significant asset to educational leaders. When reading about the skills and knowledge EdD scholarly practitioners need in their real-world contexts, program evaluation research methodology quickly emerged and captured my attention. For example, research is often explained as seeking new knowledge, while evaluation research tends to provide information for decision-making and determining the merit or worth of a program (Giancola, 2021). Solving complex problems of practice is not just figuring out how to get things done but "about what to get done, why to get it done, who decides, and whose interests are served" (Cochran-Smith, 2009, p. 21). Since CPED has been working diligently to provide a distinct identity of the EdD from the PhD, it appears that a bigger focus on decision-making within real-world settings aligns more with evaluation research than traditional research, just like improvement science is highly regarded as a distinctive approach to analyze problems of practice and use multiple frames to develop meaningful solutions. Studies have examined how EdD students were developing their professional identities as educational leaders, researchers, and action researchers and how students appear to not see themselves as such, in relation to being a learner or leader (Buss & Avery, 2017; Zambo et al., 2015). This makes me wonder if there is an opportunity to help students see themselves as evaluators and if that would make it a more easily relatable professional identity since the word evaluation is commonly used in educational, corporate, and non-profit settings.

I just finished teaching program evaluation to our first two cohorts of EdD students in Summer 2022 (n=81) which has inspired my thinking on the value and importance of teaching program evaluation as an applied methodology course and even placing it earlier in their course sequencing. What I found quite surprising was the number of students who admitted that the programs they are currently or have been involved with were never formally or informally evaluated. For these reasons, this course helped reshape my thinking on the importance of program evaluation in what I found to be a gap in their knowledge and skill set that could have significant implications for their abilities to solve complex problems of practice. Relatedly, I found this course was impactful and relevant since it allowed students to bring in their rich experiences as practitioners and have an assignment that asked them to apply their professional experiences from working with program evaluations. Students are asked throughout the semester to create program evaluations from start to finish, beginning with an explanation of a context-specific problem, a program theory, a logic model, an evaluation matrix (with evaluation questions, SMART (Specific, Measurable, Achievable, Relevant, Time-Bound) indicators and targets, data collection methodology and data analysis), a stakeholder analysis, a monitoring and evaluation plan, a communications plan, and a dissemination plan. Every student easily found a program to reference, either an existing one or was able to create a new one within their contexts, which supports the argument that programs are prevalent in a diversity of settings. Moreover, I have been energized by their abilities to align or complement their possible DiP concepts to the course assignments. Several students have even altered their DiP trajectories to a program evaluation DiP, or they have been able to enhance their DiP by incorporating

aspects of program evaluation. Doctorate programs are called upon to develop the tools our EdD students need, and I would focus on the word need in that statement and support the idea that program evaluation knowledge and skills are certainly needed for students to be able to determine effects validly and reliably or influences of initiatives and programs to avoid wasting resources, efforts, and political capital (Hochbein & Perry, 2013).

An aspect to be reemphasized is that our students do not all have access to or the influence to implement a new program or to revise an existing program. However, a definite future direction is to track how many of our students' DiPs are within their contexts, have implemented a change within their spheres of influence at their workplaces, had the time to collect data from implementations, or have implemented program evaluations. Without these data yet, it is difficult to determine if and how many students chose aspects of program evaluation, but it certainly is a program goal to collect these data for each cohort. I am hopeful that this type of analysis of their DiPs will shed additional light on whether my observations are confirmed for the continued support of program evaluation as an applied methodology course.

Another aspect to be reemphasized in that I am not calling for program evaluation to be a signature pedagogy for our doctorate program, but rather I am calling attention to the worthiness of a program evaluation applied methodology course into the required coursework. I fully support the relevancy that program evaluation knowledge and skills offer to our practitioners who will most likely all interact with a program evaluation and could benefit significantly from being able to ask the right kinds of questions related to an evaluation - whether it is looking at data from a previous one, analyzing a current one, or planning a future one. While an EdD student may not have the opportunity to engage in an improvement science Plan-Do-Study-Act (PDSA) cycle or in a completed program evaluation, I believe that there are enough programs for our students to examine at any given point in their careers that would provide them experiences to draw from to understand how both might work. However, I argue that more students have had experiences with program evaluations than with the PDSA cycles, which I believe makes program evaluation coursework more inclusive of what our students can relate to and more relevant to the transfer of these skills to future program evaluations. These PDSA cycles, though, offer students an alternative to the lengthier, complex, time-consuming program evaluations. Evaluation efforts can be short, localized efforts to iteratively collect data on a smaller scale initiative. Improvement science and program evaluation both offer students the skills, knowledge, and professional competencies to understand how to undertake future evaluations. It is valuable to expose students to both of these methodologies and the rich resources available within each approach. One accessible point for students learning evaluation approaches could relate to the three essential questions that guide improvement science processes: 1) What are we trying to accomplish?, 2) How will we know that a change is an improvement?, and 3) What change can we make that will result in an improvement? (Langley et al., 2009). While the exact wording of these questions may not be used in their program evaluations, these are certainly questions that program designers, evaluators, and stakeholders should collaboratively discuss. These questions would be meaningful guideposts for the logic model, program theory, and evaluation matrix development.

In addition to the required semester-long program evaluation assignment, the course covers additional topics such as the history of the evaluation field, the variety of evaluation ideologies, designs, and approaches to provide a groundwork for how and why they may approach evaluation in the ways they do and why and how the field of evaluation has grown over time. While most improvement science research has recently come from the healthcare and educational fields, evaluation research spans a greater variety of professions. While the task of evaluation could easily be argued as being around much longer, the evaluation profession emerged primarily in the 1970s, with the early evaluators coming from a wide range of disciplines, such as psychology, sociology, educational psychology, education, and statistics. As they brought their methods of evaluation into the field and the profession grew, evaluation approaches and methodologies continued to be debated, articles were published, and as a result, there is an influential amount of research available for students to explore today.

In the course, I also focus on the American Evaluation Association's (AEA) mission, vision, and guiding principles. It is interesting to look at both CPED and the AEA, because while the function of each organization is vastly different, there are similarities worthy of noting. Both appear to emphasize the generation of new knowledge, preparation of people to learn and apply professional skills to make an impact (AEA calls out evaluation specifically, while CPED references appropriate and specific practices), and a desire for them to become stewards of their particular professions.

AEA Mission: The American Evaluation Association's mission is to improve evaluation practices and methods, increase evaluation use, promote evaluation as a profession, and support the contribution of evaluation to the generation of theory and knowledge about effective human action. (https://www.eval.org)

CPED Definition of Professional Doctorate: "The professional doctorate in education prepares educators for the application of appropriate and specific practices, the generation of new knowledge, and for the stewardship of the profession." (https://www.cpedinitiative.org)

Without a detailed analysis here, one can find additional similarities in the terminology and wording used within the AEA and CPED mission, vision, and guiding principles, such as systematic inquiry; equity and ethics; program improvement; and application of knowledge to improve the lives of others. If doctoral students chose to conduct evaluation research, the AEA can provide a variety of support through their publications and website. Through conducting their evaluation research, the principles of our CPED-influenced program would also be integrated. The AEA is a well-established organization established in the 1980s with many resources available, such as a guiding principles training packet, self-assessment on the AEA competencies, detailed explanation of their cultural competencies, publications, and their efforts with policy advocacy. Another benefit of showing students the AEA website is the diversity of job opportunities listed within both educational and non-educational settings. There are no certifications, license exams to take, or required coursework listed for most of these positions. This exposure to the evaluation job market could possibly open up their curiosities to investigate further or validate for many that they already have the competencies of being evaluators through their previous experiences. The AEA quarterly sourcebook, *New Directions for Evaluation*, provides over three decades' worth of a wide range of articles on evaluation that could possibly benefit those involved in improvement science since the literature on improvement science is

not extremely robust (Hinnant-Crawford, 2020). While very few of the articles directly mention improvement science, reading the challenges, concepts, and applications of previous studies helps bring to light how evaluations can be a mechanism to facilitate learning at the individual, team, and organizational levels. "Transformative learning can be facilitated when employees seek to understand something, address critical organizational issues, and improve their work through a participatory, dialogic, reflective, and inquiry-oriented approach to evaluation and the use of findings" (Preskill & Torres, 2000, p. 29).

AREAS OF OVERLAP AND DIFFERENCES BETWEEN PROGRAM EVALUATION AND IMPROVEMENT SCIENCE

T The next section is what I hope to be the beginning of a continued conversation on how and where program evaluation and improvement science appear to have possible areas of overlap, along with a brief overview of the major differences. In reviewing the AEA mission and CPED's definition of the professional doctorate, one initial common link would be the value that a participant places on the initiative or program. To reflect upon this concept of value, there is a need to engage in conversations about who the initiative or program aims to help, who is not included, how power influences decision-making, and how the skill of cultural competency will impact the processes that take place before, during, and after an evaluation. These are complex conceptual questions that must be addressed before any new initiative begins in either improvement science or program evaluation.

Next, it appears that for program evaluation and improvement science, the starting point is doing a heavy conceptual lift in learning about the problem to be examined. Both appear to use similar possible sources of data, such as previous evaluations or previous data collection, existing documentation, conversations and interviews with stakeholders, process mapping, relevant research studies, and a review of the literature. If that is the case, then the skills needed to examine these kinds of data sources, such as how to conduct a needs-assessment for example, would be cross-cutting and a benefit for both. After that, defining the problem and what should be done to address the problem appears to be the next goal of each. The tools may differ in helping to see the system and understand the complexities of where the problem is located. A driver diagram or a root cause analysis may be used in improvement science, while a logic model is commonly used in program evaluation. However, both can use program theory, theory of change, theory of action, theory of knowledge, or theory of improvement to explain why an identified strategy might be effective (Giancola, 2021; Hinnant-Crawford, 2020; Langley et al., 2009) Stakeholders and the value of recognizing researcher positionality both play a crucial role in developing and defining the problem (Herr & Anderson, 2005). Both also value stakeholder engagement and buy-in, such as developing shared goals, implementing the initiative, and analyzing the data, and therefore, both include ways to connect them to these processes. Next, both call for defining and developing measures that help with defining what success looks like under certain conditions. Both program evaluation and improvement science support using SMART indicators and targets that provide guidelines for what data will be collected and how it will be collected, analyzed, communicated, and disseminated so others can learn from the data cycle. There is a shared goal of systematically applying and using data to help bring

about improvement. As a result, both call for methodological rigor and knowledge generation from these data collection cycles. Consequently, both benefit from having an established monitoring and evaluation system which are well-documented and easily found in the evaluation field. Having a system, either for a 90-day data cycle or longer, focused on monitoring and tracking crucial information on whether the theory of change, logic model, and evaluative questions guiding the intervention are accurate, appropriate, and adequate, provides extremely useful information. Most plans have components for establishing clear roles and responsibilities, accountability checkpoints, and processes for sustaining capacity which I believe can easily be applied to improvement science efforts.

The ability to communicate and disseminate effective ongoing program efforts and program results needs to be a crucial skill set within the educational doctorate. The importance of communication is also found in the literature, as authors in one study conclude that "the area for the greatest development is actually about action and use [in which students could have] hands-on opportunities to practice communicating findings and advocacy for conclusions" (Firestone et al., 2021, pp. 97-98). Others emphasize how "effective communication requires some depth of understanding of research, policy, and practice to help persuade others in decision-making processes" (Firestone & Leland, 2021, p. 14). This idea of persuading others to use results for social betterment can be found in a review of utilization evaluation research (Henry & Mark, 2003; Patton, 1997; Weiss, 1988). Furthermore, Weiss (1988) discusses the functions that evaluation serves, the structures that hinder the use of evaluation findings, and ways to get around these barriers and influence policy that can also be examined in the improvement science dialogues.

In an attempt to cover a brief overview of these two, Table 1 highlights a comparison of several elements discussed throughout this article.

Table 1. Comparison of elements of Program Evaluation and Improvement Science

Element	Program Evaluation	Improvement Science
Types	Varies such as: formative, needs assessment, summative, process, impact, implementation, outcome, evaluability assessment	Formalized PDSA cycles to test for improvement within a local context
Purpose	Can vary such as: improvement in average outcomes, decision-making, degree of change, social accountability, use of findings, influence on social, economic, or political policy	Improvement through systematic study within a localized context: testing a change to help reduce variability in outcomes
Theoretical foundations	Similar such as: visualizations (Logic models), theory of action, theory of change, theory of improvement	Similar such as: visualizations (driver diagrams), theory of action, theory of change, theory of improvement
Testing and Data Progression	Can vary: small to large scale testing and data collection of early, intermediate, and long-term outcomes	Smaller scale testing and data collection of early and/or intermediate outcomes within the PDSA cycle
Predominant fields using approach	Various fields: psychology, sociology, education, government, nonprofits, corporate	Education, healthcare

Element	Program Evaluation	Improvement Science
Size	Spans from large-scale to small-scale contexts	Small scale in localized contexts
Methodological Procedures	Similar data collection methods (qualitative, quantitative, mixed methods)	Similar data collection methods (qualitative, quantitative, mixed methods)
Process	Can vary but mainly a one-time occurrence	Multiple PDSA cycles
Time allotment	Can vary greatly: months to decades	Iterative PDSA cycles
Stakeholder participation	Can vary depending on the process selected, participatory and realist evaluation highly encourage stakeholder involvement	Necessary for improvement, need their local knowledge of how the systems work
Reporting	Can vary in format, length, audience dissemination, degree of fostering participatory analysis	Format is tailored to foster collaborative discussion to those involved in the PDSA cycles
Fear of failure	High: if dealing with large-scale and long-term programs	Low: failure is expected in order to learn from the PDSA cycles

One key skill I believe both improvement science and evaluation can benefit from is the growing field of data visualization in the communication and dissemination of user-friendly and accessible data. Both methods collect, analyze, and allow the data to tell a story, so it would be meaningful for students to be aware of, practice creating, and practice disseminating a variety of effective data visualizations. I understand that time is always an issue and making room for this skill may prove difficult, even when I know that students still struggle with creating simple charts in Word. While I imagine this skill will stretch both the professors and students, it is nonetheless a useful skill to help prepare students make data accessible and meaningful to stakeholders, so that they can apply minimal effort in deciphering the data. Additionally, if the data are to be analyzed by the stakeholders, such as hosting data parties, the data must be clear and presented in an accessible way for the best use of time when analyzing data, rather than struggling with unorganized, poorly labeled data. As data are shared with stakeholders, there are cross-cutting skills such as how to manage and facilitate a group to get them engaged, motivated, and skilled in data analysis. Moreover, I would imagine that being skilled in data visualization is a cross-cutting theme that could be used in courses beyond evaluation and improvement science.

Furthermore, it appears that dealing with unexpected and unintended consequences are additional areas that overlap. In both program evaluation and improvement science, the following can occur: unanticipated program or initiative changes, strained relationships with the stakeholders, negative findings, null findings, insufficient data, and barriers to obtain data. Students working with program evaluations or improvement science would benefit from understanding these possibilities and the various misconceptions about obtaining and making meaning from null results.

I wonder what the major differences would look like between a mini-PDSA cycle and a mini-evaluation. Since the traditional time frame for a PDSA cycle is 90 days, this would then need to be shortened and scaled down to fit within a semester. Whether or not a mini-PDSA or mini-evaluation could actually occur within a semester

still brings up the challenge of a student having the influence, ability, and capacity to see one through to completion in time and conducted effectively. Consequently, would it be worthwhile having a student plan for one, rather than be expected to actually implement it? I contend that it is extremely worthwhile based upon the last two semesters of having students create an evaluation plan, but would they get the same benefit from planning for a 90-day PDSA cycle? Most likely, the answer would depend on the student's professional context and which would provide them with more valuable information to prepare them for the day when these skills will used. Perhaps then covering both the challenges and promises of improvement science and program evaluation might be a balance suited for doctoral students? Common challenges that rise to the top for practitioners involved in either PDSA cycles or a program evaluation quickly discover the challenge of finding the time and resources needed to conduct either one effectively or efficiently without a full commitment from all those involved in each process connected to this initiative or program. Developing and sustaining a learning organization with a growth mindset is another common challenge worth discussing. Evaluation certainly has its fair share of being seen as a box-checking, compliance-driven, time-consuming, irrelevant, political, punitive, and useless activity that improvement science would like to evade (Eagleston, 2020). Lastly, creating the knowledge on how to scale programs to new settings can certainly cause challenges and opportunities for both the improvement science and evaluation field.

Not all evaluation types will have areas of alignment with improvement science, but I find that participatory, utilization-focused, empowerment, and realist evaluations have the most common elements with improvement science and, therefore, opens the door for further possible crossover learning. In particular, realist evaluations have a similar approach to improvement science in that it calls for asking how and why a program or initiative works, for whom, and under what conditions, in hopes of understanding how to adapt and scale up in new contexts (Byrk et al., 2015; Pawson & Tilley, 1997; Westhorp, 2014). Pawson and Tilley (1997) provide a framework for understanding how context, mechanisms, and outcomes explain why and how the program works, for whom, and under what conditions. Realist evaluations and improvement science both give emphasis to the complexities and wide variations of contexts, the practical implications for society, and a call for interdisciplinary or networked communities to help carry out the evaluations (Byrk et al., 2015; Pawson, 2006; Pawson & Manzano-Santaella, 2012). While improvement science calls for iterative PDSA cycles, Pawson and Tilly (2001) similarly call for speaking of evaluations in the plural, as to acknowledge this same iterative cycle and to ask big questions of small interventions. Their belief also acknowledges the learn fast, fail fast, and improve quickly principle embedded in the improvement science methodology.

Christie et al. (2017b) assert that there are four areas, for what they refer to as a cross-fertilization discussion, between the two with more of a focus of what evaluation researchers can learn from the theory and application of improvement science including learning from error, examination of variation, appreciating context, and focusing on systems change. While each area can be explored at greater depth than provided here, I will discuss briefly how these four areas highlight possible future discussions. The authors assert that the evaluation field can learn from how improvement science carves out a trial-and-error environment that has a welcoming learning space for error instead of trying to rule out or adjust for error like

many evaluations do (Christie et al., 2017b). I would add though that realist evaluations may be a bridge in this case since formative evaluations can offer real-time feedback to adjust if an initiative is not working as intended and if the involvement of the evaluator has this agreed-upon-role to make adjustments before starting the evaluation. Most evaluators understand that variation among and between contexts will occur, that context matters, the problems that they attempt to solve exist within complex systems, and that systems thinking is vital to the conceptualization of the problem definition process (Christie et al., 2017b).

One key difference outlined by Christie et al. (2017b) is that while improvement science has one purpose of seeking continuous improvement through systemic inquiry, the evaluation field has a variety of models which have different goals. While the goals may differ, evaluation methodological procedures may be similar, and in fact, be quite like those used in improvement science. Since improvement science focuses on rapid, iterative PDSA cycles at the local level, Christie et al. (2017b) claim that social accountability does not play as much of a role as it does in the evaluation field. Accountability does not appear to have a major role in improvement science since the emphasis is on looking at what needs to improve, how to bring about that improvement, determining if improvement occurred, and planning the next steps. Failure is expected and accepted. On the other hand, evaluators tend to experience internal and external pressure, perceived or actual, to demonstrate or prove that an implementation of an initiative worked, which can cause feelings of high-stakes. This is an interesting topic warranting a deeper discussion, since I imagine that at some point, the issue of accountability might arise throughout these iterative, rapid PDSA cycles. If the time, resources, and effort are afforded for these PDSA cycles, I could imagine that someone needs to account for each steps in the cycle being implemented as intended and that documentation of each step was thorough, detailed, and timely.

A second key difference between program evaluation and improvement science is noted in the time frame in which both are conducted. Improvement science PDSA cycles offer quick feedback on whether an initiative was an improvement, under what conditions, and for whom. This then allows for continued PDSA to help minimize variation in hopes of being able to scale the initiative reliably in new contexts. I wonder though if this difference can run into similar challenges as evaluations tend to show. For example, understanding the change process illustrates the complexity of attempting to measure change, let alone in a 90-day cycle. Change takes time, and there may not be enough time to see the improvement actualize or if change did occur and if it had lasting results. The implementation of the program or intervention may not have been consistently implemented or implemented as planned. Data collection may have been faulty, or the data analyzed inaccurately. Human error can greatly impact the program or intervention from the very beginning to the very end. With careful analysis of why the results occurred as they did, even more, detailed and careful planning in advance for next time might result in a more successful outcome. Possible action plans, such as piloting additional revisions, creating different target goals, or altering the data collection method, can be designed to help ensure that a program does not get discontinued. Even with participatory or realist evaluations, people need time to process what the new initiative will look like, and how it will impact them. Improvement science tends to support these possibilities by the very nature of why and how rapid, iterative cycles can collect this data to be used in the next cycle. However, Hall and

Ford (1987) offer a Concerns-Based Adoption Model (CBAM) to better recognize where individuals can be in a change process, as detailed in the Stages of Concern framework. The framework helps evaluators better acknowledge the training and support that may be needed. Whether conducting evaluations or PDSA cycles, change is a highly personalized experience, people learn at different rates, and in different ways. Reflection on what the data shows takes time. A healthy and heavy dose of patience is required across the board.

IMPLICATIONS

Due to the various professions and contexts of our students, improvement science not being the university doctorate program's signature pedagogy, and the prevalence of organizational, local, or state programs that so many students engage with as part of their day-to-day responsibilities, I call upon CPED-influenced institutions to engage in conversations about how doctoral students can benefit from learning both about improvement science and program evaluation philosophies, strategies, and tools. I believe this is a powerful opportunity to expose students to the richness, complexities, similarities, differences, and benefits of each as they attempt to solve problems of practice. Scholarship coming from either improvement science or program evaluation both emanate from "curiosity, a love of learning, an appreciation for complexity, a tolerance for ambiguity and a relentless need to make sense of one's experience" (Piantanida et al., 2019, p. 17). In conclusion, I maintain that both program evaluation and improvement science have a tremendous capacity to provide the ideas, tools, and approaches to prepare students to be the change agents they hope to aspire to be in their present and future roles.

REFERENCES

Buss, R. R., & Avery, A. (2017). Research becomes you: Cultivating EdD students' identities as educational leaders and researchers and a "learning by doing" meta-study. *Journal of Research on Leadership Education*, *12*(3), 273–301.

Byrk, A., Gomez, L., Grunow, A., & LaMahieu, P. (2015). *Learning to improve: How America's schools can get better at getting better*. Harvard Education Press.

Christie, C. A., Inkelas, M., & Lemire, S. (2017a). Editors' notes. *New Directions for Evaluation*, *153*, 7–10.

Christie, C. A., Inkelas, M., & Lemire, S. (2017b). Understanding the similarities and distinctions between improvement science and evaluation. In C. A. Christie, M. Inkelas, & S. Lemire (Eds.), *New Directions for Evaluation*, *153*, 11–22.

Cochran-Smith, M., & Lytle, S. L. (2009). *Inquiry as stance: Practitioner research for the next generation*. Teachers College Press.

Firestone, W. A., & Leland, A. S. (2021). Strategies for promoting evidence use through the education doctorate. *Impacting Education*, *6*(4), 8–15.

Firestone, W. A., Perry, J. A., Leland, A. S., & McKeon, R. T. (2021). Teaching research and data use in the education doctorate. *Journal of Research on Leadership Education*, *16*(1), 81–102.

Giancola, S. P. (2021). *Program evaluation: Embedding evaluation into program design and development* (1st ed.). SAGE Publications.

Hall, G. E., & Ford, S. M. (1987). *Change in schools: Facilitating the process*. State University of New York Press.

Henry, G. T., & Mark, M. M. (2003). Beyond use: Understanding evaluation's influence on attitudes and actions. *American Journal of Evaluation*, *24*(3), 293–314.

Herr, K., & Anderson, G. L. (2005). *The action research dissertation: A guide for students and faculty*. SAGE Publications.

Hinnant-Crawford, B. N. (2020). *Improvement science in education: A primer*. Myers Education Press.

Langley, G., Moen, R., Nolan, T., Norman, C., & Provost, L. (2009). *The improvement guide: A practical approach to enhancing organizational performance* (2nd ed.). Jossey-Bass.

Newcomer, K. E., Hatry, H. P., & Wholey, J. S. (2015). *Handbook of practical program evaluation*. Jossey-Bass.

Patton, M. Q. (1997). *Utilization-focused evaluation* (3rd ed.). SAGE Publications.

Pawson, R. (2006). *Evidence-based policy: A realist perspective*. SAGE Publications.

Pawon, R., & Manzano-Santaella, A. (2012). A realist diagnostic workshop. *Evaluation, 18*(2), 176–191.

Pawson, R., & Tilley, N. (1997). *Realistic evaluation*. SAGE Publications.

Pawson, R., & Tilley, N. (2001). Realistic evaluation bloodlines. *American Journal of Evaluation*, *2*(3), 317–24.

Peurach, D. J., Lin Russell, J., Cohen-Vogel, L., & Penuel, W. R. (2022). *The foundational handbook on improvement research in education.* Rowman & Littlefield.

Piantanida, M., McMahon, P. L., & Llewellyn, M. (2019). *On being a scholar-practitioner: Practical wisdom in action*. Wisdom of Practice Series, Learning Moments Press.

Preskill, H., & Torres, R. T. (2000). The learning dimension of evaluation use. *New Directions for Evaluation*, *88*, 25–37.

Weiss, C. H. (1988). Evaluation for decisions: Is anybody there? Does anybody care? *Evaluation Practice*, *9*(1), 5–19.

Westhorp, G. (2014). Realist impact evaluation: An introduction. Methods lab. Overseas Development Institute.

Zambo, D., Buss, R. R., & Zambo, R. (2015). Uncovering the identities of students and graduates in a CPED-influenced EdD program. *Studies in Higher Education*, *40*(2), 233–252