

Learning with and through Evaluation as Social Practice

Joanna Weidler-Lewis, Colby College, jweidlerl@colby.edu
Jimmy Frickey, Trent University, jimmyfrickey@trent.ca

Abstract: Learning as social practice attends to the situatedness of learning and the co-emergence of persons and practices. We argue evaluation should be viewed from this same lens. We present findings from an evaluation project at a high school. Using the project as a case study, we show that evaluators make consequential boundary judgments that either restrict or expand what counts as legitimate knowledge in a given situation as well as whose stakeholder's interests are validated.

Introduction

A salient feature of the Learning Sciences has been—and will continue to be—how we design and organize for learning. Our interventionist approaches to research, such as design-based research (Barab, 2006) and formative interventions (Penuel, 2014), require us to embrace a vision of how we think the world *ought* to be. Our challenge is to make our normative decisions explicit by defining not only *how* we design for learning but more importantly *for whom, with whom* and *for what* of our endeavors (Philip, Bang & Jackson, 2018). In this paper, we analyze how the practice of evaluation and the role of the evaluator contribute to the design and organization of learning. More specifically, we argue that evaluation is itself a social practice enacting boundary judgments that shape practices and it ought to be leveraged to support expansive learning as well as the changes in learning for which we design.

Recognizing evaluation as a social practice is an extension of our view of learning as social practice (Holland & Lave, 2009). We take learning to be situated in activity, mediated by cultural tools, and inherently contextual (Lave & Wenger, 1991; Wertsch, 1998). Learning is an ontological process: a coming to be of valued persons, activities, and objects (Packer & Goicoechea 2000; Weidler-Lewis, Wooten, McDonald, 2020). This process is ongoing and never fully determined; we can analyze how practices and learning outcomes came to be, but we cannot know *a priori* what organizational features will prevail in the future (Latour, 2005). However, as participants in social practices our agency and decision making have the potential to shape and inform practices despite not determining them (Pickering, 2005). We intend to make explicit how theory driven evaluation as a social practice (Schwandt, 2018) plays a role in promoting certain valued practices over others and the consequential decisions an evaluator can make enhancing or denigrating learning related to a research project or a research practice partnership.

Researching evaluation as a particular form of participation in the design of learning contributes to our understanding of how valued and normative practices come to be. Evaluators are often bracketed off from research projects in order to provide an external assessment of the project. Such removal from the ongoing coordination of practice has the potential to contribute to the reproduction of dominant evaluative logics that serve the powerful at the expense of the powerless under the guise of neutrality (e.g., Boyce, 2019). Alternatively, embracing evaluation as being pivotal and consequential in the practices of designing for learning has the potential to create new opportunities for designing towards socially valued futures. In the subsequent sections, we further articulate our view of evaluation as social practice and the importance of judgments that provide the shape and boundaries to our designs. Drawing on data collected from administering and debriefing an evaluation survey in a newly formed and emergent research practice partnership (RPP) between the authors and an experiential learning high school, we highlight key decisions made as evaluators that became consequential in shaping how and what was taken up for future action and inquiry to guide the school's practice, and thereby influencing possibilities for learning and becoming. We conclude by suggesting further lines of inquiry into the social practice of evaluation.

Theory driven evaluation as social practice and boundary judgments

We draw on the scholarship of Thomas Schwandt (2018) to claim that evaluation is a social practice that affords our ability to make evaluative judgments (i.e., claims about merit, worth and/or significance). Theory driven evaluation (TDE) provides information about the performance of a program or practice while also reporting on how and why the program achieved these results in order to guide practitioner actions (Coryn, Noakes, Westine & Schröter, 2011). TDE holds that evaluative evidence should have both scientific credibility (i.e., trustworthiness, validity, and reliability) as well as “practical worth” for stakeholders to improve their practices (Chen, 2013). The inclusion of practical worth for stakeholders in TDE is seemingly aligned with learning practice

theorists who recognize relevance to practice as an important criterion for rigor (e.g., Gutiérrez & Penuel, 2014). However, TDE is routinely conceptualized from cognitive perspectives that promote evaluative thinking, reasoning, and sensemaking as primarily individual accomplishments and ignore the myriad ways that evaluative decisions are made collectively and situated within particular contexts (Schwandt, 2018).

Schwandt (2018) identifies three reasons for understanding evaluation as social practice facilitated by an evaluator. First, the identification of facts to measure, and their associated value or worth in a program, generally unfold in an interactive process among the multiple stakeholders who come to agreement or point of view on which to evaluate an intervention, also known as “stakeholder-making.” Second, extensive literature on sensemaking demonstrates that individuals are not lone actors but are rather situated in contexts that impact their choices and action. For example, in education, teachers’ sense-making of policy standards and their subsequent classroom activities are shaped by their organization context (e.g., Allen & Penuel, 2014). Lastly, evaluation should have as its aim practical action such that the evaluator supports “we-judgments” spanning matters of facts and values to answer “what should we do?” Or in the case of learning: *how ought we design for learning?* When evaluation is seen as a social practice and a joint accomplishment of participation by the evaluator and stakeholder community, the decisions made by the evaluator must be analyzed not according to only abstract theory (be it statistical or otherwise), but rather, evaluator decisions should be analyzed within their embedded context and in relation to the practice seeking improvement.

In order to make evaluative decisions for practical action, we must engage in “boundary judgment making.” Any given situation or context cannot be studied in its totality; a boundary judgment is a decision regarding what should be included in the given situation and what should be left out (Schwandt, 2018). Given that evaluation is a social practice including multiple stakeholders, boundaries are not set or given; they are negotiated through collective sensemaking and require normative reasoning. By critiquing multiple boundary judgments, we determine how to proceed. First, judgments are critiqued by invoking morality and asking if a given phenomenon (i.e., policy, program, strategy) is *good* or the *right* thing to be doing. Second, critiques consider alternative arrangements and what *should be* done rather than what is currently happening. Lastly, there is no single or correct answer to what boundaries in an investigation ought to be. Therefore, critique in boundary judgment making is important to: (a) “make sense of a situation,” including its values, motivations, power structures, relevant knowledge, and moral bases as well as “to bear the consequences of what will be done, as well as what we may fail to do;” (b) “unfold multiple perspectives and promote mutual understanding,” by recognizing how different individuals and groups frame situations differently; and (c) “promote reflective practice” through both analysis and change (Schwandt, 2018, p. 132). Because as evaluators we embrace TDE as social practice and recognize the need for reflection in our practice, this research seeks to understand both “what are the consequential boundary judgments for learning we made during our evaluation?” and “who and what practices were served by these boundary judgments?”

Background and methods

As mentioned in the introduction, this work stems from a fledgling RPP involving the authors and an experiential learning high school (ELS). Frickey, the internal coordinator of evaluation & monitoring enlisted the help of Weidler-Lewis to begin research endeavors into both school improvement measures at the local level and more robust inquiry into experiential learning writ large. The timing of our partnership coincided with the 25th anniversary of ELS and was an ideal time to collect data on former students regarding their high school experience. Prior to our work, no systematic follow-up of former students was conducted. Given our mutual goals of serving both ELS and the broader experiential learning community, we decided to survey former students using a combination of psychometric scales used in other educational contexts, items measuring attitudes towards ELS values, practices and other key school indicators, and a single, open-ended prompt for former students to share any information they felt pertinent to our endeavor. In developing the survey, we considered the possibility of creating a single measure of “experiential learning success” to be iterated on by the already validated psychometric scales we employed that could be used in varied experiential learning contexts. We considered our survey to be a pilot of such measure knowing that given limitations, the survey would primarily be used for ELS evaluation. Our purpose in this paper is not to provide an empirical analysis of the survey design, implementation, and/or results. Rather, we endeavor to use the practice of surveying former students and the process of stakeholder-making to underscore how we engaged in TDE social practice and the boundary judgments we made alone and in collaboration with ELS stakeholders.

The data for our analysis comes from email correspondence about the survey between the authors, design documents and notes regarding the survey, the survey itself and its findings, the evaluation report presented to ELS, and the notes and records of the presentation of survey findings to ELS stakeholders. These stakeholders included the head of school, the director of curriculum, the director of student services, the associate head of

student services, the director of professional development, and the former director of professional development who had been an original founder of the school. For purposes of anonymity we do not attribute views to stakeholder role. We first coded our data deductively by analyzing them for boundary judgments according to 1) moral claims (i.e., that which is *right* or *good*), 2) claims about what *should* be occurring at ELS, and 3) claims that interpreted the situation in one of many possible ways that could have been interpreted differently. After coding, we, looked for emergent themes and explicit connections between our codes, the stakeholders, and the different positions the stakeholders held (Saldana, 2009).

We present our initial findings detailing consequential boundary judgments in our evaluation practice. We demonstrate that even though key stakeholders were identified at the outset, whose interests were taken-up—and therefore pursued—emerged in interaction. We show how deference to scientific credibility in TDE can easily usurp its practical worth. This work is still ongoing, as such, we consider it to be exploratory in nature in that we are trying to surface the important tensions in TDE as social practice that will foster further study in this area (Stebbins, 2001).

Findings

From the outset, proponents of social practice theory might be confused as to why we would employ a survey of psychometric measures to evaluate student “success” when such measures assume psychological and cognitive orientations to the world. Indeed, this was one of the first boundary judgments made collectively among the stakeholders insofar as we all agreed that a succinct survey with limited questions would be an ideal starting point to begin an evaluative relationship. Reasons for this decision included time and financial constraints related to gathering data prior to ELS’s anniversary celebration. It is important to note, historically, ELS had a culture of eschewing most forms of quantitative data on students including not assigning students grades and instead favoring students’ lived, experiential data. While such evidence demonstrating student success at ELS was documented in several book publications, little sustained inquiry existed for why some ELS students succeeded when others did not. All parties agreed a survey had the potential to reveal previously undocumented categorical differences (i.e., race, gender, cohort, etc.) with explanatory power for “success” in the ways we defined. The ELS stakeholders saw the potential for quantitative evidence to support their intuitions regarding categorical differences in students’ experience. For example, such a survey might help explain why female students were more likely to graduate despite being more difficult to recruit. The survey might also provide insight into racial difference, which had been a concern of many but not all at the school. As evaluators, we saw the potential for creating a survey measure applicable in other contexts. However, our survey results did not yield any statistically significant difference in any demographic category on any of our items or constructs. As evaluators, we could attribute this lack of evidence to the small student population of the school (266 graduates over its 25 years), the overwhelmingly positive skew of the respondents, and other survey limitations.

While the survey yielded a plethora of descriptive statistics, its lack of any evidence for causal inference was taken up differently by the different stakeholders. For example, the lack of causal evidence seemed to support the historic view that experiential data should be favored over quantitative data, a view contested by current stakeholders. Others remained agnostic of the survey’s usefulness, withholding judgment regarding its potential to compare results over time in relation to program initiatives, so it served its purpose as starting point for future comparisons. Another stakeholder thought the survey data was detecting or could with revision detect differences in student experience. This stakeholder examined the descriptive statistics for evidence to support what they (and many other students and staff) knew to be true and had other forms of evidence regarding the experiences of students of color. These differences no matter how small could and should contribute to the ongoing discussion at the school regarding both how to approach broader, national narratives of race as well as any direct incidents of racism. This stakeholder believed direct anti-racist programming was needed, a position not shared by all, and thus, quantitative data could provide further justification for this targeted programming. As the evaluators, we had the opportunity to contribute to the collective boundary judgment making regarding the survey’s usefulness in supporting conversations on racism in education. However, because no causal inferences (e.g., $p < 0.10$) could be detected from the survey, and because the survey skewed positive, we defaulted to “scientific credibility” as a reason to exclude discussions of students’ racialized experience in ways productive for both this stakeholder and the community. Our unilateral judgment shifted the ways in which we constructed this stakeholder: we focused on the lack of causal inference (one aspect of knowledge present in the situation) rather than what other forms of evidence could or should be brought into our discussion that ultimately shape students’ experience at the school. In answer to the question who was served by this boundary judgment, we conclude that this stakeholder was not, and likely neither were the students for whom they were advocating.

Discussion and conclusion

While our findings may be brief, we believe they offer the powerful conclusion that just as social practice views of learning guide learning designers to create expansive, rather than reductive, opportunities for learning, so too can evaluators and their evaluations support this cause. In order to do so, evaluators must recognize the ways in which their decisions and judgments hamper or enable reflective inquiry into the very practices they are purporting to serve. The stakeholders came together to support inquiry into understanding varied student experience and implicit in this coming together was a hope to have further answers to and further evidence for, “what should we do?” One stakeholder had evidence, including individual experience as a student, that they were bringing to the process of evaluation; we failed to legitimize it by limiting the discussion of ‘evidence’ only to the causal claims missing from the survey. We bounded the survey discussion to only include causal inference reliably shown in the survey data despite knowing the school legitimized many other forms of data and the survey could have been used to further those alternative conversations. Moving forward, our role in facilitating stakeholder-making is to continually ensure that we are accurately representing the mutual agreement among stakeholders and to whatever extent possible include multiple forms of evidence in addition to multiple stakeholder perspectives. Part of our evaluation practice, then, must include processes to reengage and reassess if and how our shared agreement has been sustained as well as what dominant logics persist at whose expense. Because stakeholder-making is an ongoing practice, a productive line of future inquiry is to trace this process over time and continue to iterate on how best ensure a relationship of reciprocity that values and builds on each other’s expertise.

References

- Allen, C. D., & Penuel, W. R. (2015). Studying teachers’ sensemaking to investigate teachers’ responses to professional development focused on new standards. *Journal of Teacher Education*, 66(2), 136-149.
- Barab, S. (2006). *Design-Based Research: A Methodological Toolkit for the Learning Scientist*. Cambridge University Press.
- Boyce, A. (2019). A Re-imagining of evaluation as social justice: A discussion of the Education Justice Project. *Critical Education*, 10(1).
- Coryn, C. L., Noakes, L. A., Westine, C. D., & Schröter, D. C. (2011). A systematic review of theory-driven evaluation practice from 1990 to 2009. *American Journal of Evaluation*, 32(2), 199-226.
- Gutiérrez, K. D., & Penuel, W. R. (2014). Relevance to practice as a criterion for rigor. *Educational Researcher*, 43(1), 19-23.
- Holland, D., & Lave, J. (2009). Social practice theory and the historical production of persons. *Actio: An International Journal of Human Activity Theory*, 2(1), 1-15.
- Latour, B. (2005). *Reassembling the social: An introduction to actor-network theory*. Oxford, England: Oxford University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press.
- Packer, M. J., & Goicoechea, J. (2000). Sociocultural and constructivist theories of learning: Ontology, not just epistemology. *Educational Psychologist*, 35(4), 227–241.
- Philip, T. M., Bang, M., & Jackson, K. (2018). Articulating the “how,” the “for what,” the “for whom,” and the “with whom” in concert: A call to broaden the benchmarks of our scholarship. *Cognition & Instruction* 36 (2): 83–88.
- Pickering, A. (1995). *The Mangle of Practice: Time, Agency, and Science*. University of Chicago Press.
- Penuel, W. R. (2014). Emerging forms of formative intervention research in education. *Mind, Culture, and Activity*, 21(2), 97-117.
- Saldaña, J. (2013). *The Coding Manual for Qualitative Researchers* (2nd Ed.). In SAGE Publications Inc.
- Schwandt, T. A. (2018). Evaluative thinking as a collaborative social practice: The case of boundary judgment making. *New Directions for Evaluation*, 2018(158), 125-137.
- Stebbins, R. A. (2001). *Exploratory research in the social sciences* (Vol. 48). Sage.
- Weidler-Lewis, J., Wooten, M., & McDonald, S. P. (2020). The ontological construction of technology and behavior through practice. *Human Behavior and Emerging Technologies*, 2(4), 377-386.
- Wertsch, J. V. (1998). New York, NY: Oxford University Press.