Participant-Directed Evaluation: Using Teachers' Own Inquiries to Evaluate Professional Development in Technology Integration

Vince Ham
CORE Education

Abstract

Considering the high levels of time and money invested in teacher professional development programmes in information technologies over recent decades, questions arise as to how effective these programmes have been and by whose lights we are to judge. Based on a critical review of the evaluations of several of our own action-research-based professional development programmes in technology integration, this article asks three basic questions about those evaluation processes and, ultimately, about evaluation design itself in such contexts. What specifically should evaluations of professional development in technology integration look at*? What should they look* for*? And who is best located to do the looking? It also considers how our answers to these questions might be adequately represented in a conceptual model for the evaluation of professional development programmes in technology integration. (Keywords: Evaluation, professional development, technology integration, action research)*

In this article I take the evaluation design models presented in two recent meta-analyses of evaluations of the effectiveness of teacher professional development programmes (Timperley et al., 2007; Lawless & Pellegrino, 2008) as a starting point for a discussion of what, conceptually, such evaluations might productively look like in the particular context of professional development in technology integration.

Between them, these meta-analyses present two important challenges to providers and evaluators of professional development (PD) programmes in new technologies (or any other PD, for that matter). One of these challenges is for evaluations of such PD programmes to be duly comprehensive and systematic, especially in terms of tracing the programme's impact along the full length of the chains of influence from PD events, through teachers' changed understandings and practices, to student learning in classrooms. Moreover, as Lawless and Pellegrino (2008) point out, the current international focus on evaluating PD primarily in terms of measurable student outcomes is especially problematic in an area like technology, not least because of the manifold learning outcomes that might be evidenced and the similarly manifold nature of the technologies, pedagogies, and classroom contexts involved. What learning outcomes and which technologies, used in which ways, by teachers of what, to which groups of students, etc.?

The other challenge is to develop conceptual designs for evaluations of technology professional development programmes that adequately represent that comprehensiveness. To start the conversation on the latter, both meta-analyses advance a linear, phased conceptual design for the conduct of PD programme evaluations, based on isolating specific variables at each of the steps involved in the chain of influence, from PD event/activity to eventual student learning. For Lawless & Pellegrino (2008), the phases comprise an investigation of a range of specific variables during the PD events themselves, their effectiveness in shifting the understanding and classroom practices of the teachers, and, finally, the downstream effects on the consequent learning of students (see Figure 1).

As a way of critically reviewing the evaluations of three action-research-based PD programmes that we have been involved in (Ham, Wenmoth, & Davey, 2008), we tried to map the evaluations of those projects onto the generic, linear design above. In doing so, we ended up asking basic questions about our evaluation processes and, ultimately, about evaluation design itself.

The PD Projects and Evaluations

Although the three PD programmes that formed the basis of our reflective review of evaluation in technology PD all had a technology focus, and all included an action research component in some form or another, they differed significantly from each other in their models of delivery, scope and size, participant teacher demographics, and modes of evaluation involved.

The first of these programmes was a large-scale national programme of professional development for teachers on integrating new technologies known as the Information and Communication Technologies Professional Development (ICTPD) programme. In this programme, groups of three to five schools have been clustered together to provide 3-year programmes for their teachers in the use of technology across the curriculum. Over its 10 years of operation, the ICTPD programme has provided ongoing technology professional development for approximately 20,000 teachers in school clusters nationwide.

Each cluster decides the actual model of professional development undertaken, but most have involved some kind of action enquiry approach (Ham, 2005; Ham et al., 2005). In its first few iterations, the evaluation of these nationally offered programmes

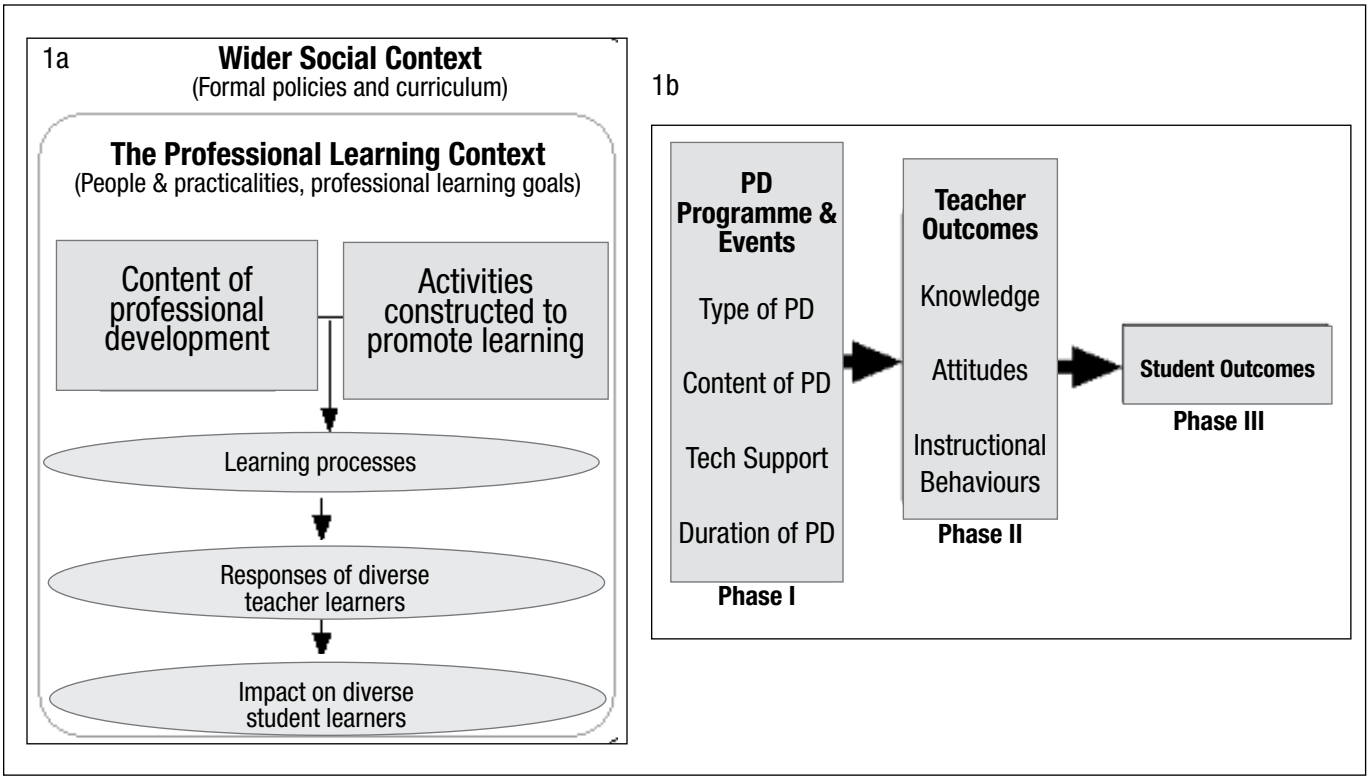

Figure 1. Generic evaluation designs presented in two meta-analyses of PD evaluations. (1a) Timperley et al.'s "Framework for analyzing the effectiveness of professional learning experiences" (2007, p. xxiv). (1b) Lawless & Pellegrino's "Overall evaluation design" for PD in technology integration (2008, p.603).

involved both internal and external monitoring—the former in the form of teachers reporting in milestone documents and conference presentations the results of their own classroom enquiries, and the latter in the forms of pre- and postsurveys of participating teachers and written case studies based on outsider interviews with teachers and hundreds of classroom observations of students engaged in e-learning activities.

The second PD programme was a much more small-scale project that took place more than 2–3 years in one educational institution (the New Zealand Correspondence School). In this programme, a group of 15 teachers in the "e-section" of the school trialed new, online, technology-based distance education methods for teaching their classes of isolated students dispersed around the country, using action research as their form of PD (Ham, Wenmoth, & Davey, 2008). External facilitation of the research projects by an experienced action-research facilitator (myself) provided methodological support, and regular external reviews of key findings doubled as the main evaluation component of the project.

The third programme was a form of collaborative sabbatical known in New Zealand as the E-Learning Fellowships. Under the fellowship scheme, up to 10 innovative teachers with a reputation for effective use of new technologies in their classes conducted year-long research studies of their work with their own students. They also met together for up to 8 weeks per year in facilitated professional learning workshops and worked together as a collective to help each other with their respective enquiries (see http://www.efellows.org.nz). The programme had the joint goals of producing publishable case-study research by teachers identifying the learning outcomes of e-learning activities and providing formative professional learning experiences for the fellows.

For the first three iterations of the e-fellowship programme (2004–2006), an external evaluator was commissioned to identify outcomes for the teachers concerned. But since 2007, the programme has been essentially self-monitoring. The decision to abandon external evaluation was made partly to reduce the cost to the Ministry of Education of funding the scheme, partly because the external evaluations were producing very similar results each year, and partly because better student-impact data was emerging from the teachers' research reports than from the teacher-impact-focused evaluator reports (Ministry Contract Manager, personal communication, December 2006).

The authors' roles in each of these programmes also varied. For the ICTPD evaluation, I have been the chief evaluator, leading the team of outsider researchers that coordinated the teacher surveys, undertaking teacher interviews for the case studies, and conducting the in-class observations. In the correspondence school project, I was more of a critical friend for the project, providing an outsider view of the impact of the programme on the participants, but also extensive mentoring of the teacher-researchers in action-research methods and the like. In the

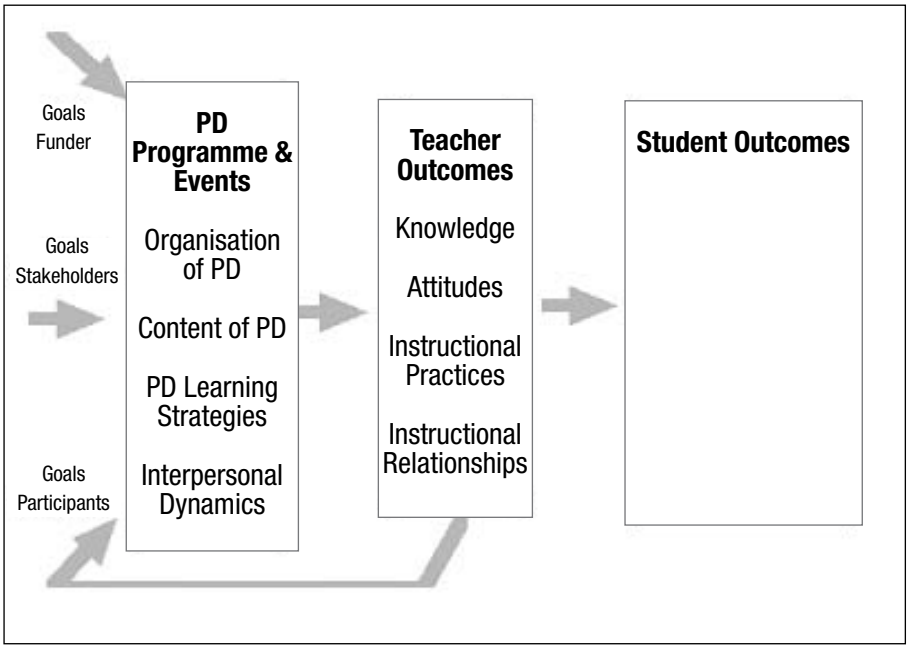

Figure 2. What to look *at* and *for* in post-PD classroom teaching and learning activities using IT.

fellowship programme, I have had no directly evaluative role, but I lead the small team of teacher educators that provide ongoing research mentoring to the fellows as they conduct their various research projects.

Product and Process in the Evolving Theory of Evaluation

The classic definition of evaluation in education, and one possibly undergoing a resurgence given discourses around the commodification of PD, provider accountability, and value for money in PD, is the measurement of outcomes in comparison with goals. As defined in the 1950s by Tyler, Kirkpatrick, and others, evaluation is the process of deciding to what extent predetermined educational objectives are actually being realized (Nevo, 1989). However, evolving theories of programme evaluation in education since then have taken a more comprehensive view of what an evaluation can achieve and therefore, by implication, the range or types of data to be collected, the modes of analysis to be used, and the particular participant/stakeholder/researcher interests that are given credence in the reporting.

In particular, recent generations of writers on evaluation in education have much expanded the traditional objectives-correlated-to-outcomes approach to evaluation, incorporating a greater focus on the process in between the objectives and the outcomes. They have developed models of what might be appropriately called a process evaluation, which traces as much as possible of either or both the 'chain of influence' from a PD event or programme through subsequent teacher actions to student learning and the chain of evidence back from identified student outcomes via a teacher's changed pedagogy to those changes' origins in PD events.

Established evaluation models, such as Stake's (1967) summative and formative evaluation, Scriven's (1967) "goal-free" evaluation, Parlett and Hamilton's (1989) "illuminative evaluation," Kemmis (1989) and Simon's (1987) "emancipator" or "educative" evaluation, MacDonald's political model and the stakeholder model advocated by Weiss (1989), as well as more recent models like Brinkerhoff's (2003) Success Case Methodology and Checkland and Poulter's (2006) Soft Systems Methodologies, all start from the premise that judging the effectiveness of a programme neither rests on nor prioritises the presuppositions, goals, or criteria of any one particular participant or interest group. Rather it derives from, and perhaps can only consist of, an understanding and enunciation of the perspectives, interests, and actions of all of them. Equally important from a methodological perspective is the common assumption in such evaluation models that the more comprehensively the evaluation gathers data on all of the links in such chains, the more likely such internal conflicts of interest are to surface, the more accurately an observer can judge which, and whose, objectives are actually being addressed, and the more valid the evaluation is therefore likely to be as a process of knowing. As Stake and Denny put it, evaluation is an investigation of worth, not just effect.

> Considered broadly, evaluation is the discovery of the nature and worth of something. In relation to education, we may evaluate students, teachers, curriculums, administrators, systems, programs and nations. The purposes for evaluation may be many, but always evaluation attempts to describe something and to indicate its perceived merits and shortcomings.... Evaluation is not a search for cause and effect, an inventory of present status, or a prediction of future success. It is something of all of these things but only as they contribute to understanding substance, function and worth. (quoted in Kemmis 1989, pp. 117–118)

Thus, if evaluation is conceived as 'description with value added' and not merely as the measurement of distant consequences against immediate intentions, then any evaluation's broad aim becomes to provide not just a correlation between pre-existing funder-determined goals and consequent teacher or student effects, but also due consideration of the goals and evaluative criteria of all stakeholders, including participants themselves, and a rich description of the chain of events and processes by which such effects are achieved.

What to Look *At* and *For* When Evaluating PD Programmes and Events

At the PD event level (the first stage or phase in the generic evaluation design

model), the question of what to look *at* and what to look *for* involves making a clear distinction between characteristics and criteria as components of the concept of evaluation (see Figure 2). Characteristics, in this sense, means the various features of the PD events and subsequent teacher and student activity that participants or observers tended to make evaluative judgments about. Criteria, on the other hand, were comprised of the often implicit standards those various participants applied in making their judgments with respect to each characteristic. What needed to be looked at, in a descriptive, data-gathering sense, were these characteristics, and what would be looked for, in later data analysis, would be patterns of goal achievement and change in terms of funders, stakeholders, and participants' various criteria or standards of value.

In three studies of participant and stakeholder goals and evaluation criteria, we found that different groups of stakeholders and participants in PD events on technology often value different aspects of inservice activities differently (Ham,1998; Ham et al., 2002, 2005). That is, they do not necessarily share the same goals for the PD programmes and do not necessarily apply the same measures or indicators of value in judging a programme's effectiveness. But we also found that there were a relatively finite and common number of characteristics of those PD events that such judgments were made about. These were the core characteristics of PD events that our evaluations looked at—our version of the possible variables listed in the PD event phase of the generic design. These core groups of characteristics, derived from participant and stakeholder accounts of what they had found effective in the PD, were its:

- Formal organisation
- Content
- Raft of PD strategies employed by the PD facilitators
- Interpersonal dynamics and interactions

Aspects of formal organisation included features such as the location of the professional development activity, the time available and timetabling of the events, its administrative efficiency, and so on. In terms of interpreting criteria from these characteristics, the unifying idea seemed to relate to a general notion of access and availability. In other words, timing, location, and the like could be seen as important less in themselves than as factors affecting the availability of, and participants' access to, people (such as fellow participants, technicians, and collegial experts), information (such as help sheets, timetables, room bookings, and e-mail addresses), or equipment (such as computers, phone sockets, the right version of software, and so on). For almost all participants and stakeholders, the easier such access, the more effective the programme was said to have been.

Content characteristics fell broadly into three subgroups: the acquisition of technical knowledge and skills, the learning or collection of practical classroom teaching strategies and resources using technologies, and a discussion of general pedagogical theory and philosophy as applied to technology-based activities. The criteria most often applied in relation to content was that the PD was wide ranging in its coverage of all three of these areas, and that it did not treat any one of them in isolation, but rather focused on developing conceptual and practical links between and among all three of them. Skills were not learned in isolation from classroom strategies, and neither of them was taught/learned in isolation from the context of what is known about effective pedagogy and powerful learning.

Finally, there was a group of interactional and relationship-focused characteristics that affected participants' effectiveness ratings, most of which had to do with how the particular people involved in the PD related and interacted with each other as a socio-professional group and how needs based the PD was. Criteria applied to these characteristics clustered around high levels of beneficiary control, empowerment and ownership present in the relationship, maximising opportunities for social intercourse through sharing ideas and experiences with other teachers, maximizing the amount of individual attention facilitators give to teachers, issues of compulsion and voluntariness, and establishing positive interpersonal relationships with other participant teachers. The evaluative criteria were often expressed in terms of how such factors affected how they as participants felt about the programme and their sense of ownership of its process and purposes.

Moreover, as Timperley (2007) found in more general contexts, our evaluation studies in technology contexts also found that the organisational characteristics of the PD events were seen, especially by participants, as the least significant in their effects on themselves as teachers, and that the factors related to the content of the programme and, even more, its interpersonal dynamic, as the most significant (Ham, 2006).

Such analyses of the multiple goals operating in our programmes, along with an emerging research consensus around the characteristics of effective professional learning for teachers generally, suggests two revisions of the generic evaluation designs represented above: one to acknowledge the multiplicity of goals and objectives to be considered in evaluations, and another to revise some of the possible variables that could be usefully addressed within the PD events and programmes themselves.

A third "what to look at and what to look for" decision in our evaluations concerned what to investigate in teachers' subsequent classroom teaching and students' consequent learning with IT (the other phases in the generic evaluation design).

In one sense, deciding what to look at in classrooms was relatively simple: We needed to look at students as they completed learning activities using computers and at teachers as they taught sessions in which students used digital technologies as the main learning medium. Isolating a manageable set of things to look for within that, however, was more problematic. What was it that the teacher-researchers wanted to know about those teaching and learning activities that could be an index

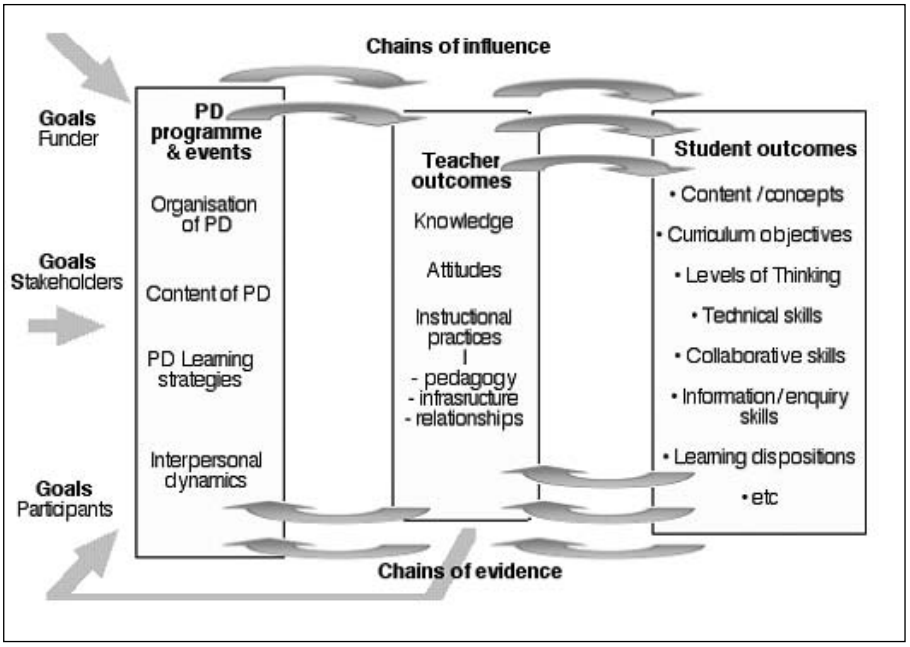

Figure 3. Who is best placed to do the looking?

to their merit or worth as educational interactions? What are the most legitimate kinds of student outcomes that could/should be identified as the likely effects of any changed teaching practice?

Taking a cue again from Stake's definition of evaluation as "the discovery of the nature and worth of something" as description with value added, what the teacher-researchers in our programmes decided to look at and for in their studies of student learning in their classrooms were not comparisons with other (non-technology-focused) practices so much as descriptions of the educative value of the activity per se, as indicated by the particular learning that students demonstrated in their performance of various ICT-mediated activities. The core question that all the action-research studies addressed was: What do our students do when using new technologies of educational worth or value? What, in short, do they learn?

Within that general framework of studying student learning, the teachers defined their own specific research questions (and thus their own evaluation criteria) and created their own idiosyncratic lists of the particular student learning outcomes that they hoped would derive from their particular ICT-based teaching activities. These student outcomes were different for each teacher-researcher, but all, in one way or another, involved gathering evidence of student learning.

To map the various learning outcomes our action researchers researched against the relevant part of the design, we thematically grouped their investigations of learning as follows:

- Studies of student motivation and levels of engagement in learning
- Studies of student achievement of specific curriculum objectives, as stated in formal curriculum documents
- Studies of the taxonomical levels of thinking demonstrated by students in completing the activity
- Studies of the technical skill levels demonstrated by students in the activity
- Studies of the information skills demonstrated by students during the activity

Finally, in mapping student outcome elements, we noted that, like many of the PD programmes evaluated in the meta-analyses, our PD programmes consisted of an ongoing series of PD and classroom events over an extended period of time of between 1 and 3 years. This meant that the programmes themselves could adapt to teachers' responses as they progressed over time, just as the teachers' classroom practices could repeatedly change or adapt in response to students' learning, as that too was evidenced over time and in respect of different ICT-based activities. This cyclical, iterative nature of the interplay between PD events and classroom teaching and between classroom teaching and student learning—between influence and evidence—suggests a cyclical rather than a linear or single-sequence representation in any conceptual designs.

Action research is professional learning done for and by teachers to solve their own situated problems of practice. In action research models of professional development, therefore, the participant teacher is at the centre of both the action and the research. Moreover, both the action and the research are inherently evaluative activities, as the purpose of the research becomes the same as that of a student-outcomes evaluation: to provide evidence that teachers' changed pedagogical practices result in the desired student learning. The participant teacher is responsible for developing his or her own enquiry, plan for data/evidence collection, and, by implication, criteria for evaluating the worth of his or her changed teaching practices for students.

In any evaluation design for action-research-based PD programmes, the teacher is at the very centre of both the chain of influence (as an actor) and the chain of evidence (as a researcher) (see Figure 3). The teacher directly experiences the PD as its immediate beneficiary, directly determines any consequent changes in pedagogical practice, and directly observes student outcomes in relation to that practice. As a matter of data collection, the teacher is in a position to provide first-order data about the PD events, first-order data about their own pedagogical practices, and second-order (or third-order) data about children's learning. An external researcher observing PD and classroom events or interviewing participants post hoc about them has at best second- or third-order access to any of these. Teachers are uniquely positioned to gather

and interpret evidence related to their own experiences in and after PD events and to gather and reflect on evidence of student outcomes.

As a matter of value, moreover, the teachers in such PD models set their own goals and questions, and thus by implication their own criteria for evaluation. By definition, they evaluate the achievement of their own goals through the research process itself. Through rigorous research activity focused on their own particular puzzles of practice, participant researchers provide empirical evidence of the impact of the PD on those classroom practices, which is in turn based on empirical evidence on the impact over time of those changed practices on their students' learning.

That is not to say that only participant researchers should be involved in the evaluation of PD programmes. In two of our three PD projects (Ham, Wenmoth, & Davey, 2008), the evaluation of student outcomes was entirely internal in this way—that is, it was done by the teacher participants themselves through their research projects. But in one case the evaluation was both internal (consisting of the volunteer teachers' action research studies) and external (consisting of a series of teacher surveys and observational case studies conducted in participants' classrooms by outside researchers), and we arguably got the best of our three programme-level evaluations from the one that involved both participant and external observer evaluators. But it is to say that as a matter of evaluation design, action-research models of PD have both teacher and student outcome elements inherently built into the PD design itself.

Nor is it to say that participant evaluation methods are unproblematic in terms of validity. Rather, participant evaluation through action research and external evaluation through external observation and interview both suffer from converse, but equal, advantages and limitations (Hammersley, 1993; Ham & Kane, 2006). On the one hand, teacher-action-researchers are closer to the action and have a richer understanding of their own and their students' learning, but they face a number of practical data gathering and analysis difficulties as participants. They may not see the woods for the trees. On the other hand, external researchers, even if they adopt an ethnographic approach, are much further from the action. They are in a good position to see all three or four phases of activity, but only at a more superficial level of understanding than that of the critically reflective participant. External evaluators may not see the trees for the woods (see Figure 4).

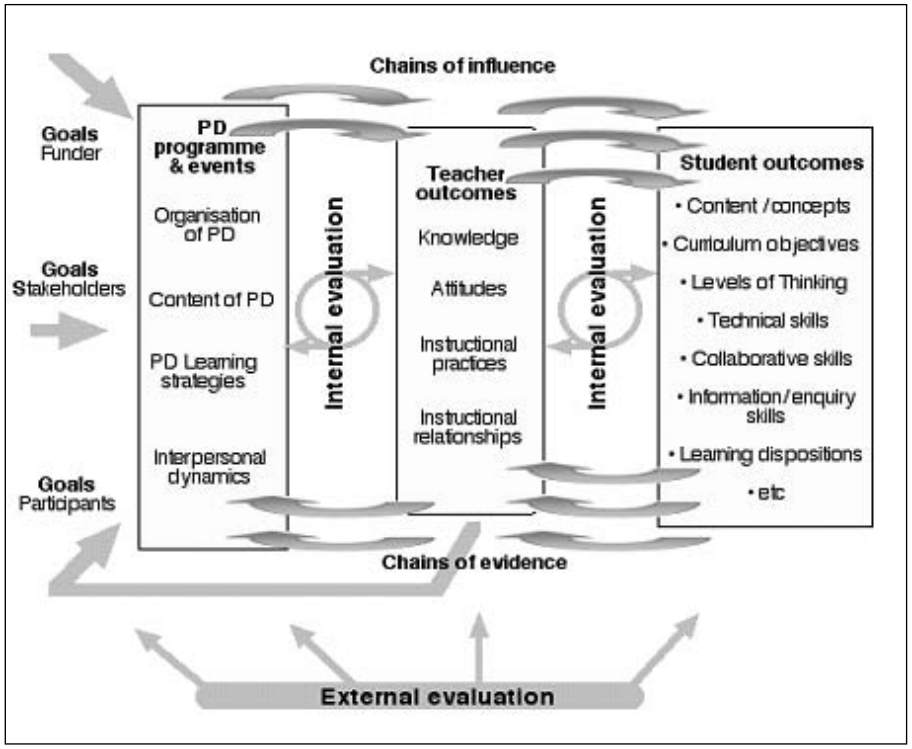

Figure 4. Internal and external evaluation involvement in PD programmes.

Therefore, our final version of the generic evaluation design includes both internal and external loci of evaluation.

Conclusion

In this article, I have argued that in methodological terms evaluation can be conceived as "description with value added"—as the systematic investigation of both the procedural and the consequential worth of some educational practice or system, seen from a variety of perspectives and value sets. By placing participant teachers at their centre, models of PD based on action research have inherent potential to closely link both teacher effects and student outcomes directly back to aspects of the PD experience to provide a rich evidence base about those effects and outcomes from the participants' (as opposed to the PD providers' or even a researcher's) perspective, and to allow the PD to be formative, responsive, and iterative in its progressions over time. They are arguably more likely to be comprehensively valid, and to provide rich multiperspective data on the full chain of influence in what are usually highly situated contexts. But they are also arguably less likely to be universally reliable, or to provide convenient, reproducible recipes of effective PD delivery based on reliable instruments, large sample sizes, and standardized scoring procedures that are often the milieu proposed for more linear, cause-effect, or correlative evaluation designs.

As a matter of validity and reliability in evaluation, therefore, the potential upsides of action-research PD models in technology are:

- Ability and opportunity to trace the impact of the PD right through to student outcomes that is inherent in the PD process

- High validity in establishing an empirical evidence base of both PD impact on individual teachers and teacher impact on individual students
- Clear articulation of the evaluative criteria that participants apply and the application of their particular measures of success
- Rich descriptive evidence of the constituent elements (variables) of all three key steps in, and all iterations of, the PD, as a developmental process evolving over time

The potential downsides, however, are that:

- The research may not be focused on all stakeholder perspectives. Who assesses the goal achievement of the funders when these may not be the same as those of the participants?
- To provide genuine research rigour, it is often necessary to confine action researchers' questions to narrower rather than broader student outcome phenomena.
- It is logistically difficult to be both actor and researcher at the same time, as a practical matter of data collection. Data gathering, analysis, and synthesis of results can often become extra work unless adequate time or resourcing is allowed for it in the design.

When considered from the perspective of beneficiary-oriented models of PD such as action research, generic, linear designs for evaluating PD in technology begin to look incomplete. They represent the chain of influence implied by the PD process but lack adequate representations both of its iterative nature and the complex chain of evidence that would make an evaluation design truly comprehensive. Teacher professional development is too manifold and complex an intervention in teachers' professional lives to be conducive to evaluation through simplistic, goals-outcomes correlations. Any final assessment of its effectiveness or worth should involve a richer evidence base than is often the case for all of what teachers learn and do when they take part in it, what they understand and do as a matter of changed classroom practice as a result, and what students in those teachers' classrooms do and learn as a consequence of that.

A discussion of what gets evaluated in professional development programmes is not only a timely reminder of the need to judge technology PD by its effects on student learning. It is also a timely reminder of the need for comprehensiveness in the design of the evaluations of such programmes. It is a reminder of the need to collect student outcome data, but not only student outcome data. We also need rich process data. In this respect, it is also a reminder of the key questions that we need to ask ourselves in designing future evaluations in the area:

- Is the evaluation gathering data about the whole of the process and about all of the key elements and activities in that process? Are we covering everything that needs looking at?
- Is the evaluation applying indicators of worth or just or measures of effect, and do those indicators or measures acknowledge goals and intended outcomes for all key participants? Are we being inclusive or exclusive in what we are looking for?
- Are those conducting the evaluation, or providing data for it, in a good position to know? Are the right people doing the looking and the judging?

Though somewhat glibly outlined above, these are not unchallenging questions for the evaluation community. They challenge our frequent acceptance of some quite fundamental assumptions about evaluation design, particularly the design of evaluations of externally funded, policy-inspired, large-scale PD programmes. In terms of what to look for, do we have to accept that only that which can be (statistically) measured is worth evaluating? In terms of what to look at, do we have to accept that only the goals of the funder and only the actions of the PD provider are legitimate subjects of evaluation? And, with respect to both, do we have to accept that only external researcher experts are capable of doing the evaluating?

Author Note

Vince Ham is director research at CORE Education, a not-for-profit educational research and development organisation based in Christchurch, New Zealand. He has particular research interests in educational research methods, especially action research; the professional learning of teachers and teacher educators; and the use of new technologies for teaching and learning. Correspondence regarding this article should be addressed to Vince Ham, Director Research, CORE Education, Level 7, NatColl House, 151 Kilmore Street, PO Box 13678, Christchurch, New Zealand. E-mail: vince@core-ed.net

References

Brinkerhoff, R. (2003). *The success case method: Find out quickly what's working and what's not.* San Francisco: Berrett-Koehler.

Checkland, P., & Poulter, J. (2006). *Learning for action: A short definitive account of soft systems methodology and its use for practitioners, teachers, and students.* New Jersey: Wiley.

Ham, V. (1998). T*owards criteria for evaluating teacher professional development in information technologies: A pilot study of four New Zealand models of in-service education.* Unpublished MSc Thesis, Department of Educational Studies. Oxford University, UK.

Ham, V. (2005). Clustering schools for teacher development in ICT: The "grand eExperiment." In Lai K. (Ed.), *E-learning communities: Teaching and learning with the Web.* Dunedin: Otago University Press.

Ham, V., & Kane, R. (2004). Finding a way through the swamp: A case for self-study as research. In Loughran et al. (2004), *International handbook on self-study in teaching and teacher education.* Amsterdam: Kluwer.

Ham, V., Gilmore, A., Kachelhoffer, A., Morrow, D., & Moeau, P. (2002). *What makes for effective professional development in ICT? An evaluation of the 23 school clusters ICTPD programme, 1999–2001.* Wellington: Ministry of Education.

Ham, V., Moeau, P., Williamson-Leadley, S., Toubat, H., & Winter, M. (2005). *ICTPD through three lenses: An evaluation of the ICTPD School Clusters programme 2001–2003. Supplement: Action researcher reports.* Wellington, New Zealand: Ministry of Education.

Ham, V., Wenmoth, D., & Davey, R. (2008). Teachers doing IT for themselves: Action research as professional development. In Borthwick and Pierson (Eds.), *Transforming classroom practice: Professional development strategies in educational technology.* Eugene, Oregon: ISTE.

Hamilton, D., Jenkins, D., King, C., MacDonald, B., & Parlett, M. (Eds.) (1977). *Beyond the numbers game.* London: Macmillan.

Hammersley, M. (1993). On the teacher as researcher. In M. Hammersley (Ed), *Educational research: Current issues, volume 1* (pp. 211–231). Buckingham: Open University Press.

Kemmis, S. (1989). Seven principles for programme evaluation in curriculum development and innovation. In House (Ed.), *New directions in educational evaluation.* London & Philadelphia: Falmer.

Lawless, K. A, & Pellegrino, J. W. (2007). *Professional development in integrating technology into teaching and learning: Knowns, unknowns, and ways to pursue better questions* and answers. *Review of Educational Research, 77*(4), 575–614.

Nevo, D. (1989). The conceptualisation of educational evaluation: An analytical review of the literature. In House (Ed.), *New directions in educational evaluation.* London & Philadelphia: Falmer.

Scriven, M. (1967). The methodology of evaluation. In Tyler, R., Gagné, R., and Scriven, M. (Eds.), *Perspectives of curriculum evaluation.* Chicago: Rand McNally.

Simons, H. (1987). *Getting to know schools in a democracy. The politics and process of evaluation.* London: Falmer.

Stake, R. (1967). The countenance of educational evaluation. *Teachers College Record, 68,* 523–40.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. (2007). *Teacher professional learning and development: Best evidence synthesis iteration (BES).* Wellington: Ministry of Education.

Weiss, C. (1989). The stakeholder approach to evaluation: Origins and promise. In House (Ed.), *New directions in educational evaluation.* London & Philadelphia: Falmer.

Appendix

Characteristics (What to Look At) and Criteria (What to Look For)

A characteristic of something (such as a PD programme or event) is one of its discernible features. A criterion is a standard against which an individual judges a characteristic or group of characteristics of something to be good or bad, acceptable or unacceptable, successful or unsuccessful , effective or ineffective, etc. For example, when someone sets out to buy a house and says to the land agent, "I want a north San Diego property with a large garden, at least three bedrooms, and made of brick," the characteristics to be evaluated may be conceived as location, garden size, number of bedrooms, and type of building material.

By contrast, the criteria, or standards, being applied are north San Diego, large, three, and brick. A consensus among home buyers is much more likely to exist with regard to characteristics than with regard to criteria: They will tend to look at the same or similar aspects of a property, even though they may as individuals make very different judgments about it. That, presumably, is why land agents tend to describe in advertisements the same set of features for every property they put on the market.

In terms of research design, data collection, and analysis, therefore, evaluated characteristics are those observable features of the professional development programmes or classroom activities about which evaluators, or participants, make judgments, whereas their criteria are those specific, often idiosyncratic, standards by which particular participants and stakeholders finally judge a programme or its components to have been successful or unsuccessful from their perspective. The characteristics of PD programmes or classroom teaching and learning are what a programme evaluation looks at. Criteria are the measures or indicators of value that it looks for in judging its effectiveness.