Rediscovering Educational Purpose in Educational Evaluation

Saville Kushner
University of West England, Bristol – UK

ABSTRACT

This paper argues that evaluation methodology should reflect the character of the field it is proposing to look at. People have the right to be evaluated in their own terms, and evaluation methodologies which are judged for their rigour against abstract canons of validity violate that principle. Two examples are given.

Keywords: Evaluation - Democratic Evaluation - Evaluation Methodology - Curriculum

La metodologia della valutazione dovrebbe riflettere i valori e il carattere del programma che si sta studiando. Le persone hanno il diritto di essere valutate nei loro stessi termini. Le recenti proposte per i 'sistemi metodologici awes' sono una violazione di questo principio. Essi asseriscono criteri di validità per la valutazione che hanno poco a che vedere con la difficoltà che gli individui hanno nel dare senso ai programmi in cui sono coinvolti, e che potrebbero riflettere poco dei giudizi che questi individui stessi danno della qualità del proprio lavoro. Ci sono esempi di valutatori che cercano di disegnare le proprie valutazioni sull'immagine del programma che stanno seguendo e due di questi sono presentati nell'articolo. Uno di questi torna alle origini della Valutazione Democratica (Democratic Evaluation), il cui autore, Barry MacDonald, ha eviitato dai principi per il curriculum di Lawrence Stenhouse, esaminando il suo progetto di curriculum. Una comparazione è fatta tra i principi del curriculum che sostengono al programma e i principi metodologici sviluppati per la valutazione. Ancora il secondo esempio deriva dal lavoro di Barry MacDonald, questa volta utilizzando degli estratti dalla politica di valutazione dell'educazione bilingue negli Stati Uniti d'America. In questa valutazione la scuola al centro dello studio ha rifiutato la valutazione finché la metodologia non fu adattata ai valori e alle pratiche preferenziali della scuola stessa. La scuola sostenne che la metodologia iniziale non rappresentava in maniera efficace i propri sforzi nei confronti dei bambini e delle bambine. La valutazione si focalizzava sulle interazioni all'interno della classe, sebbene queste fossero laddove la scuola soddisfaceva le esigenze di responsabilità formale. La metodologia che loro successivamente hanno proposto era su quei luoghi e quelle interazioni in cui la scuola si sentiva libera di
pensare e disegnare essa stessa, interazioni che erano più appropriate per la cura dei bambini e delle bambine.

L'argomento ha una certa attinenza con il modo in cui la valutazione rispetta i diritti di programma che le persone hanno di essere giudicate contro i criteri e nelle forme di rappresentazione che hanno senso per loro. La valutazione può e, si sostiene, dovrebbe condividere una base di valori con il programma che si sta seguendo – e che ancora rimane indipendente. Questa non è una scelta per chi valuta – è un obbligo. Al cuore del problema, tuttavia, giace una recente tendenza per il dibattito metodologico di essere autoreferenziale – ad esempio cercare la validità all'interno delle proprie logiche interne, scoprire una legittimazione per la valutazione in termini astratti che riguardano che cosa conta come rigore metodologico. E' un pericolo in questi dibattiti che le decisioni riguardanti la metodologia della valutazione siano prese riferendosi alle discussioni dei valori del ricercatore. La metodologia, in altre parole, troppo facilmente ha la tendenza di porre i valori e i significati della ricerca e della comunità di valutazione prima di quelli della comunità di programma che deve essere valutato.

Parole chiave: Valutazione - Valutazione democratica - Metodologia di valutazione - Curriculum
Summary: Coherence and Evaluation Methodology

The argument of this article is that methodological debates around democratic program evaluation have to some extent distracted us from considerations of the basic purposes underpinning what we do. The example used is educational evaluation, one of the primary sources for much of the development of program evaluation (Norris, 1990; House, 1993). But this argument generalizes to other evaluation fields in the public service arena including evaluation of professional action, such as policing and nursing, as well as policy evaluation in professional fields. The challenge of methodological development for evaluation has less to do with the epistemological or ideological integrity of clusters of methods, or with allegiance to ‘schools of thought’, and more to do with evaluation reflecting the ethics of the practice under observation. What makes evaluation coherent with the practice under observation is that it derives its ethical orientation from it, and that it positions itself in relation to social justice with reference to it. For example, if we are considering a participatory approach to evaluation in the field of training for community policing, notwithstanding that we seek our evaluation to reflect general principles of democracy and justice – that we have personal leanings towards participation – our considerations should, nonetheless, focus on how those democratic principles are played out specifically in relation to policing and democratic demands on police service. If a key ethical orientation for community policing is the negotiation of priorities with community, then this gives the rationale to evaluation for adopting a participatory approach. Similarly, if we are evaluating a poverty reduction programme which adopts a rights-based approach, then we would expect the evaluation methodology to be rights-based – however that is interpreted and negotiated.

Nor need this mean that the evaluator simply concedes to the values frame in the field of practice being observed – there are still choices to be made, persuasions to be worked, impartial critique to be rendered. The argument is not that evaluation becomes an advocate for the programme, much less that the evaluation adopts the ‘programme theory’. But the principle of ‘service’ underpinning evaluation demands that we submit our own methodological preferences to the test of coherence with some basic political and ethical orientations of the field in which the programme operates – so as to be what Robert Stake has always called ‘responsive’ (i.e. the programme is the ‘stimulus’ to which the evaluator responds). Evaluation coherence is more a matter of service to practice than a question of theoretical integrity.
'Gold Standards' and the Generation of Knowledge

The call for some to 'return' to methodological gold standards in the field of school curriculum, signified by randomized and controlled studies, is, in its own way, appropriate in that it is coherent. Insofar as these advocates are looking for information to support their curricular aims such approaches are consonant with the way they tend to see school curriculum – this is evidence of straight thinking. There is a congruence between curriculum principles being advocated and methodological principles underpinning its evaluation – they share a common ethic. Those who see curriculum as a site for the transmission of stipulated knowledge, teaching as a compliance activity and knowledge as context-less and subject to pharmaceutical proofs ('what works' treatments), will demand certain things of evaluation such as fixed points of reference for knowledge, productivity measures for pedagogy and criterial standards for learning outcomes. Where they see curriculum as a process for the continued verification of knowledge (rather than its critical scrutiny or alternative-seeking) this principle drives the demand for evaluation that is affirmative of existing knowledge.

Experimental trials for curriculum, to take that example, are important, not for knowledge generation, but for knowledge verification. The generation of new knowledge will have been done as part of framing the hypothesis that drives the experiment. Students and teachers have little role in the development and adaptation of knowledge. There is, too, a separation of knowledge and action – curriculum knowledge is generated in one place and applied in another. Such an approach to curriculum and its evaluation may be moral, humanist, progressive, ethical and honourable – though it is inappropriate to argue that it is democratic in itself, other than in a rudimentary and abstract sense that it might serve the needs of a democratic society. The separation of knowledge and action which removes knowledge generation from classrooms denies teachers, pupils and parents an active role in shaping educational purpose. The best we can hope for in democratic terms is a paternalistic liberalism where curriculum is conceived in our best collective interests, albeit without our participation. Evaluation is implicated in this process, since a central task of any educational evaluation ought to be to support critical enquiry into educational purposes.

Other approaches envision more of what we can call a 'conversational curriculum' and these see curriculum as a site for the generation of knowledge by the student under the guidance of the teacher – in contemporary times we are becoming much interested in the involvement of communities in joining these conversations. For example, where knowledge is seen to be personal, then each learning interaction has uncertain outcomes and that gap between predicted learning outcomes and actual ones is the zone of knowledge generation. I am
Rediscovering Educational Purpose in Educational Evaluation

careful attention here, with curriculum approaches which entertain concerns about democratic justification – such as those curriculum aspirations which see the classroom as a site for the conferment of citizen rights on students, and pedagogy as the husbanding of self-determination and autonomous thinking, each of which gives rise to that zone of uncertainty. These approaches bring preoccupations with educational purposes into the classroom as the substance of pedagogical interactions. This is a quite distinctive ethic to the previous case. The challenge of managing such interactions and uncertainties places certain demands on evaluation which, in order to understand those uncertainties, has to engage in its own form of knowledge generation, enquiry into purposes and engage a participatory ethic.

The problem is that in the evaluation field we have been poorly served by methodological debates over the past 20 years or so in supporting democratic, ‘conversational’ curriculum since debates about educational evaluation have largely taken place in the absence of deliberation over educational (curriculum) principles. Rather, evaluation methodology – particularly what we sometimes too easily call ‘qualitative’ methodologies – has become a self-referencing discourse. Methodological principles are asserted adversarially, say, with reference, to other, less preferred principles. This makes it harder to defend democratic principles in both curriculum and methodology – it certainly makes democratic thinkers look in more disarray than their counterparts who do not so easily lose sight of the conceptual coherence between a transmission curriculum and its ‘gold standard-driven’ evaluation.

Evaluation in Expansive times

The 1960s and 1970s marked a period of fiscal expansion in Britain and the USA. Under the post-war settlement the wealthy had been persuaded to buy in to an opulent society with extended opportunities and a broader base of participation. These were expansive times. The ‘Great Society’ legislation in the USA and the curriculum reform movement in Britain matched with a radical expansion of the higher education sector gave rise to foundational debates about the role of schooling in social reform, the ‘pool of ability’ thesis and the very basis of social stratification.

Even without the ingrained scepticism of the wealthy and powerful classes about the value for money of fiscal spend, and a concomitant demand for evaluation-based accountability, there was a need for better information about innovation and change. Donald Campbell (1969) came to write about the ‘experimenting society’ – a society committed to a progressive alliance between science
and government in advancing the welfare state – and this, insofar as it reflected a transatlantic government aspiration, implied an industry of change analysts. Program evaluation as a service to the administrative system was born – but into this epoch of expansive thinking. Remarkably, within only a few years of its 'big-bang' emergence as a professional practice and university discipline, its leaders were already talking about radical revamps of methodological conventions – wanting to “stretch people’s minds as to what should be considered legitimate data” (Stake quoted in Abma & Stake, 2001). Early learning of the shortcomings of evaluation in generating the kind of knowledge being demanded of change programs generated rapid methodological responses. Stake (1975) advocated Responsive Evaluation; MacDonald (1974), Democratic Evaluation; Parlett & Hamilton (1972) Illuminative evaluation; Cronbach (1980) was writing of the evaluator as ‘public scientist’ and urging evaluators to focus on contexts. All of this was happening from the mid-1970s on, much of it driven by the need to broaden the range of data sources available to the evaluator in order to better understand and explain the promise and the failure of innovation. (When one program manager challenged MacDonald over an evaluation account that reported apparently superfluous data on some intimacies of project lives – “We don’t need to know who’s sleeping with whom!” – MacDonald replied, “Oh yes you do.”)

This match of expansive methodological thinking with expansive program thinking was nowhere more marked than in relation to curriculum. In Britain, the curriculum reform movement, led by the Nuffield Foundation and the teacher dominated Schools Council among others, was busy proliferating possible curriculum identities – the Schools Council alone spawned more than 120 national curriculum projects each with its own provenance, theory and pedagogical strategy. What schools and classrooms could be, what teachers could be and, concomitantly, what pupils could be was more open-ended and unpredictable than at any time since. The very titles of some of these projects suggested a new approach to knowledge – Man: a Course of Study, Place, Time and Society, Geography for the Young School Leaver, History 13-16, the Keele Integrated Studies projects and, most celebrated of all, the Humanities Curriculum Project. All of these were concerned with bringing social change into schools – not least changing views on the structure of knowledge – and developing participatory approaches to curriculum interactions. Many envisaged what we might think of as a ‘conversational curriculum’ as opposed to a ‘transmission curriculum’ – ‘conversational’ in the terms Joseph Schwab advocated as deliberation.

A consistent theme across these curriculum projects was the development of autonomous judgement in the young person. Here, for example, is an extract from a conversation with a young student of the History 13-16 curriculum project:
"As I read the booklet on the Red Army, I didn’t realise they were communists and I found myself thinking, ‘Oh, they’re good guys’. But then I thought, ‘but they’re communists – I don’t think much of communists… they are sort of inevitable winners…’ I found myself wanting to know about the losers, what sort of people were they?… was their defeat sort of certain? Why did they lose and not the communists? You could come to your own decision then… and not be brainwashed by the book.” [Quoted in Stenhouse, 1980: 185]

If knowledge is to be personalised in this way curriculum becomes a matter of knowledge generation rather than the verification of stipulated knowledge. Pedagogy becomes a matter of managing those learning processes which lead to the development of judgement in the young person, rather than ensuring that appropriate knowledge is passed to the student. Part of the expansive tenor of the times was a proliferation of possible knowledge-identities among pupils – the very push towards autonomy and the sustained attack on the assumption of a fixed pool of ability that was briefly represented in Britain in the comprehensive school were pointers to young people to aspire beyond the vision of their parents and of a stratified society. Perhaps, had the logic of that period been sustained and explored in greater depth we’d have found that schools can, indeed, compensate for society. As it was, this was a short-lived era.

As curriculum was seen increasingly to be a site for the generation of knowledge, so it was demanded of evaluation that it does the same. Subsequent years saw a parallel proliferation of evaluation identities as utilisation-focused, stakeholder-oriented, empowerment, dialogic, affirmative, participatory and such-like. All of these were the progeny of this fleeting period of expansive thinking. House (1979) talked of this period as one of the transition of evaluation from technical practices through to cultural/political practices – the key point was that evaluation was sponsored to move beyond the technical-craft approach of applied social science to forage in more luxurious intellectual pastures like ethnography, journalism and history. Whatever it took to generate more complex understanding was sought.

We live in different times. The new technocratic progressivism of Blair-type reforming administrations, in particular, lead to powerful alliances between science and government (as in Campbell’s ‘experimenting society’) within an officially-sponsored mistrust of intellectualism and independent professional action. Reforming zeal persuades government into firm confidence in policy and solution along with an intolerance of delay. Control replaces understanding as a basis for curriculum development, and curriculum itself reflects that same certainty over goals and purposes. Where, earlier, there was a tolerance – if not always a taste – for complexity and an expansive view of legitimate knowledge, now there are more proscribed views. Knowledge that is useful is tightly drawn to the parameters of the problem as defined.
Saville Kushner

In this contemporary context the narrow remit of the information held to be useful by government administration speaks of more proscribed evaluation re-
mits. Broad ranges of data sources are not as valued – but more to the point, the
discovery of new knowledge is often only barely valued, and educational pur-
pose are generally expected to be reaffirmed by evaluation rather than brought
into question. Where governments are confident of asserting the nature of the so-
cial problem and its solution what they seek is verification rather than genera-
tion of knowledge. The information requirements of technocratic progressivism
are highly focused (e.g. on implementation issues) and in place of expansive
views we have diminished aspirations for knowledge. These, then, are times when
evaluation is asked to drawn in its own aspirations, return to what we sometimes
talk of as 'methodological gold standards'. Where there is official certainty over
social goals all evaluation theorising can do is to interrupt the process of admin-
istration of reform and undermine the authority of government.

Curriculum and its Evaluation in parallel

What is happening to curriculum as its knowledge base expands and contracts has
profound implications for evaluation, as evaluators seek to respond to the kinds
of information and information processes asked of us. Let me take the emergence
of Democratic Evaluation as an example. The rubrics of Democratic Evaluation
emerged in response to – to some extent in parallel with – the development of
curriculum principles. Put simple, the demands of an expansive, conversational
curriculum stimulated the development of expansive (in data terms) 'conversa-
tional' evaluation. First, then, the curriculum project which was conceived by
Lawrence Stenhouse and evaluated by Barry MacDonald.

The curriculum project: Lawrence Stenhouse was directing the Humanities Cur-
riculum Project, sponsored by the Schools Council between 1968 and 1972. This
project realised Stenhouse's ambitions to make of classrooms 'laboratories' for
the generation of knowledge among young people and sites for the intellectual
development of teachers and teaching. The project, at its core, created resources
and conditions for young people of school-leaving age to deliberate over 'con-
troversial issues' which they would encounter on their release from school – Re-
lations Between the Sexes, Race, War, Poverty, etc. The project team developed
resource packs – evidence-bases, we might call them today. Central to the phi-
losophy of the project was that the teacher was cast in the role of 'neutral chair',
managing exchanges, the review of materials and subsequent conversations. The
teacher could express no personal view to avoid the inevitable effect of giving an
authority cue to the students and interrupting their search for an autonomous, independent judgement. If a young person expressed a racist view, say, the pedagogical task was to help elaborate that view and expose it to contestation among other views, equally well elaborated, and against the evidence base. In radical constructivist terms, the task was to test out that theory of the world for its 'functional fit' (Von Lasarsfeld, 1991) in the lives of the young people. Most challenging of all, if a student left these discussions with confirmed racist views that had been informed with evidence and properly contested through rational deliberative processes, then the teacher had to rest content – i.e. the teacher had no material interest in learning outcomes¹.

In Stenhouse’s project, then, the teacher was the guardian of process. Stenhouse, in fact, used this project as a site within which to elaborate his argument that we had to shift our attention from learning outcomes to ‘principles of procedure’ (Stenhouse, 1975) – i.e. be less concerned with destinations than with the quality of the journey. The teacher’s task, in a Dewey-an sense the most intellectual of all, was to create the conditions for the rational and detailed exchange of views so as to develop a sophistication of judgement in the young person. The key idea was that refined, independent judgement was a handier tool for living and citizenship than any defined body of substantive knowledge. Democracy in the classroom as accomplished through the discovery of individual intellectual freedom and judgement – but, importantly, the means of agency (and the underpinning of curriculum) was the collective and the reaching for intersubjective – in Rawlsian terms, overlapping – consensus (i.e. overlapping in the sense that consensus on controversial issues need not violate individual differences and variations).

The evaluation: Barry MacDonald had already made the decision that he could not give sufficient coverage in his evaluation to the 36 schools involved in piloting this project. The pedagogical, moral, ethical, psychological, political dimensions of the project experience (the neutral chair role, in particular, put great strain on the teachers) implied more intensive study of teachers, classrooms and contexts than evaluation resources allowed. MacDonald decided to adapt his methodology (using, for example, attitude testing) and to conduct 'condensed' studies of a small sample of schools from which he hoped to generalise in terms of learning about the project as a whole. He had made the intellectual journey from what Stake (2003) characterises as 'scalar/criterial' evaluation to a 'case-based/experiential' approach (see Simons, 1987, for a more detailed account of this journey). He had, that is to say, discovered case study.

Then came MacDonald’s response to the political and ethical challenge these condensed studies presented. Psychometric studies carry their own authority which
cannot be negotiated with the respondent. MacDonald’s condensed studies, however, carried only the authority conferred on them by his respondents — i.e. whereas psychometric data and analysis are relatively untouched by respondent appeals, case study data is easily invalidated by the respondent’s claim that it is untrue or inaccurate — or even unfair. This required the development of an evaluation ethic, something which became the subject of a series of international seminars (the Cambridge Evaluation Conferences — see Elliott & Kushner, 2007) and publications which became seminal to the development of case-based and democratic approaches to program evaluation (see Simons, 1987 and Norris, 1990 for summaries and Hamilton et al., 1977 for an edited collection). Evaluation accounts had to be negotiated with the respondents — in fact, a democratic ethic imposed an obligation to distribute ownership of the evaluation.

As MacDonald developed his rubric for Democratic Evaluation — published in 1974 — his thinking was closely allied to that of Stenhouse’s curriculum thinking. I will categorise them for ease of comparison.

<table>
<thead>
<tr>
<th>CURRICULUM</th>
<th>EVALUATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stenhouse was inventing the teacher as ‘neutral chair’ in classroom discussion</td>
<td>MacDonald was proposing the evaluator as ‘neutral broker’, protecting the democratic rights of those being evaluated</td>
</tr>
<tr>
<td>Stenhouse aimed to develop autonomous, independent judgement in the student</td>
<td>MacDonald argued that evaluators fed the judgement of decision makers but did not subvert them with their own expressed judgements</td>
</tr>
<tr>
<td>For Stenhouse, substantive curriculum knowledge belonged to the student: pedagogical knowledge was the preserve of the teacher</td>
<td>For MacDonald, program knowledge belonged to program people: the evaluator moved in the preserve of understanding change processes</td>
</tr>
<tr>
<td>Stenhouse was devising pluralistic and inter-subjective approaches to classroom deliberation</td>
<td>MacDonald was promoting values-pluralism and an inclusive (stakeholder-type) approach to evaluation.</td>
</tr>
<tr>
<td>At the heart of classroom learning in Stenhouse’s project was mutual understanding for overlapping consensus</td>
<td>MacDonald argued for information exchange between stakeholders for program transparency and mutual accountability</td>
</tr>
<tr>
<td>Stenhouse moved attention away from learning objectives to process factors and from learning outcomes to principles of procedure</td>
<td>MacDonald urged evaluators to treat program goals as unstable and not to make recommendations</td>
</tr>
</tbody>
</table>
I am not arguing that one maps onto the other precisely, nor that one was derived from the other. More to the point was that both Stenhouse and MacDonald – as close friends and colleagues – shared a sensitivity to an emergent educational ethic (both could source their ideas in contemporary writing and events) and a prevailing morality. Both were responding to the expansive context they found themselves in, taking advantage of the opportunity to promote progressive, liberal democratic ideals. Simultaneous invention was as much a part of the story as mutual ‘borrowing’. But, then, others were making similar discoveries at the time, and had done so earlier (e.g. notably Schwab, 1969, in curriculum; but Parlett & Hamilton, 1972, in evaluation among others – Philip Jackson and Louis Smith were busy rediscovering the sociological realities of schooling). What I am also implying is that we might analyse contemporary movements in evaluation in similar ways – i.e. mapping across from the substance of programs to methodological developments in their evaluation.

Nor am I arguing that evaluation, in responding to a common educational ethic, should adopt the value system of the educators being evaluated. This places evaluation in an advocacy stance which may be justifiable under limited circumstances, but, as a general rule, can only undermine the credibility of independent evaluation and deny privileges to those whose values lie outside of those espoused by the evaluation. While evaluation must be broadly in sympathy with the justification of the evaluated (i.e. all sometimes competing stakeholders) and their values, this need not amount to endorsement. An important ingredient of independence is impartiality.

Self-referencing Methodologies

I am not arguing, either, that this ability to map across from programs to evaluation methodologies is a story of consistency. In fact, it is as likely that methodological deliberations in recent years have become detached from educational purposes. The continuing debate about qualitative as against quantitative approaches, overlaid with postmodernist critique has fuelled methodological debates which are essentially self-referencing in that foundational statements may refer to theory and methodology more than to practice. The justification of one methodological assertion is often made with reference, not to the practices being evaluated, or to an educational ethic, but to other methodological statements. This can provoke that sense of existential ‘nausea’ some might have occasionally noticed where one TV soap opera refers in its script to another.

What the second half of this paper argues is that we need to return to deriving educational evaluation methodology from restated or reaffirmed education-
principles and purposes. We need to bring curriculum and its evaluation into closer realignment. That realignment needs to be part of a rediscovery of those earlier, expansive versions of education possibility which encourage the proliferation of possible identities. This implies a reemphasis on evaluation as knowledge generation rather than as knowledge verification. Educational evaluation needs to become, once again, an instrument for the questioning and rediscovery of educational purposes, and its methodologies designed for this. I will provide one example, taken from my own work, of such a shift. This is the case of an evaluation of bilingual education policy in the USA conducted some years ago (MacDonald & Kushner, 1982).

The Case

There is space here for only the briefest of introductions to the context. No matter – the observations below are merely illustrative of an argument. The team had been commissioned by the Ford Foundation to conduct an evaluation of bilingual schooling policy in the USA on the grounds that (a) it was culturally/politically neutral in this area of heightened civil rights tensions (contestable); (b) its members are the early pioneers of case study, an approach thought capable of revealing the lived experiences of bilingual teaching, learning and advocacy where previous (e.g. sociolinguistic) technologies had failed; and (c) the team are advocates of Democratic Evaluation and a negotiated base for enquiry. A single school was chosen as the principal site for direct observation.

The school was given negotiating rights over the generation and use of the data, enshrined in an ethics protocol that laid out the mutual rights and obligations of evaluation participants and evaluators. The school case study would only be published on an agreed basis, the school (and its individuals) having all opportunities to amend and add to the case study account. The team’s task was to represent the school in its own terms and at its best – the challenge was to reveal the promise of bilingual education, and to demonstrate the issues and dilemmas that afflicted even the best practitioners.

The data: What follows, then, are two examples drawn from the final report: the first, an extract from the rejected draft, the second, an extract from the second round of observations. The first observation begins with the voice of the evaluator.
"Literacy is premised on a knowledge of the cultural components that the terms, sentences, paragraphs and texts refer to. Literacy in English requires that the Spanish monolingual and dominants from Hispanic homes learn about the referents of the Anglo culture. English literacy — understanding the cultural significance of what is read — is vital for academic achievement in an English-dominant state. In this lesson the beginnings of literacy, learning to read, are being taught at the beginning of the new school year.

Wendy asks the class to recite the sounds of the English vowels so that she may hear them. As they recite, Wendy points with a stick to the letters on display. Each of these letters has a picture of an object in which the vowel sound is made — long 'a' as in 'cake', long 'o' as in 'rose', short 'o' as in 'top'.

Teacher: Can you repeat that? Long 'o' says 'o' — try it. Long 'o' says 'o', just like its name. Okay? Do you know the name of this letter in English? That's 'a'. Long 'a' says a. What is it? Long 'a' says a. Okay! This is 'e', the name of this letter in English. Long 'e' says 'e'. Say it.

Pupil: Long 'e' says 'e'.

Teacher: Good. Okay. This is 'i'.....

The sound/letter correspondence section concludes. Reading whole words and sentences commences. These words incorporate the letters and sounds of the parts of this and preceding lessons.

Teacher: Alright, now let's look at our first picture, here. Take a good look at the picture. We have three sentences, remember what are sentences? Let's look up and remember what a sentence is. It's a group of words that tell a little story. ...I want everyone trying, please. Sound out your words if you don't know them. Do you, Gabriel? Read it.

Pupil: See Jim ride.

Teacher: Very good! Let's look at the picture. Do you think that sentence goes with the picture? See Jim ride? Who's in the picture? Jim. What is he doing?

Class: Riding!

Teacher: Do you think that sentence tells about the picture? Yes, I think it does. But don't circle it yet. Let's read all of them in case there's one that's better."
Here is an example of how these observations were analysed in the account – once again, the voice of the evaluator.

"By pointing to pictures of objects and getting the children to recite the appropriate vowel sounds the teacher helps the children to memorise the sound that matches the word that fits the picture. Although the formulation of the rule (e.g. in the form long 'o' says 'o', long 'u' says 'u') might be memorised, the application of the rule is dependent upon the student's recognition of a word to which the rule applied, and the learner's ability to reproduce the sounds that are appropriate in that language. ... The acquisition of the connection between sound and letter-symbols is manifest in the appropriate utterance of the sound. This approach to the teaching of reading through initiation into sound/letter relations is known as the Phonic method. A main alternative method in use is usually called 'Look and Say'. The sound/letter relationships in Look and Say are acquired inductively."

The voice of the evaluator here is authoritative and the narrative is constructed as a reinforcement of that authority. The observation is rendered and reported in such a way as to exemplify the dominant theory of language acquisition. There is an implicit hierarchy of knowledge – theorist at the top, theory interpreter following, and practitioner whose actions are to be interpreted at the bottom – and here lies a democratic dilemma. In relation to the theme of this paper, we notice that there is a congruence between the construction of the curriculum and that of the evaluation methodology. Both reinforce that external structure of authority, the knowledge hierarchy. The curriculum presupposes the primacy of formal (rather than situated) theory; the evaluation is verifying of the theory. The key element of this analytic strategy is the separation of knowledge and action – i.e. the tools for understanding the case are derived in contexts (perhaps university offices, other field sites) apart from the action. We learn nothing new about bilingual pedagogy, nor, incidentally, of the pedagogy.

In interview following this observation, the teacher was asked about the cultural component of bilingual education – what recognition, we asked, does the school give to the student's cultural background? Her response was that we skewed the answer by looking too selectively – guided in our sampling by theory and not by curiosity.

"But where do you see [culture]? How, when I am teaching English reading, am I going to have 90% of that lesson in 'culture'? Come to culture classes, okay, and you'd see it. Come to the Music class, see it...they're using a Spanish instrument – they're not using an American instrument..."
So the analysis was limited by the sampling – and the sampling was itself informed by the theoretical constructs guiding the evaluators. The observation and its subsequent analysis would, perhaps, have scored highly on construct validity – i.e. the theoretical assumption underpinning the enquiry found proper expression in the enquiry technology. But such measures of internal validity prove inadequate to protect the enquiry in a world of real consequences and possible ethical damage. We were looking, at best, in the right way but at the wrong things; at worst, in the wrong way at the wrong things. The teachers felt that they had little possibility of appeal against the methodology other than to dismiss the entire study – which they did.

But, perhaps central to our purposes here, we may note that this approach to the analysis of action does not provide robust opportunities for knowledge generation. The process of analysis is one, essentially, of verification. The practices of the Hernandez School are, in a sense, another test of the language acquisition theory and insofar as the theory comfortably explains the action it is verified. No new knowledge base is being built, other than descriptive data on this particular context. The knowledge which might be generated would be a product of enquiry into the context – what we learn afresh about the particularities of bilingual pedagogy in this place at this time. However, though the application of language acquisition theory may be unique in this context, the dominance of the theory in its application reduces that uniqueness in relation to universal understanding. The impact of context is diminished, and opportunities to explore the impact of uniqueness narrowed.

The team was ejected from the school and had to renegotiate access – allowed back in on the grounds that the school guided the sampling and the methodology. Here are extracts from an observation of another teacher’s class conducted under these new conditions and in search for a situated, practical theory of bilingual pedagogy.
Grade 5 Social Studies Lesson

Sheila explains some of the curriculum problems involved in Elementary School Social Studies in a bilingual programme:

"In an elementary school such as ours where you're dealing with two languages, and a kid has to learn Spanish reading and oral English, and progress towards English reading, that takes time, so they're going to lose something in the curriculum... several of the students miss out on Social Studies... I spent the first three months of the 5th Grade getting across the idea of city, state, capital, that Puerto Rico was a commonwealth, not a state, what an island is, okay? Why? Because as these kids come up, if they've been programmed into Title 1 classes, English and Spanish, they don't have everything a regular English speaker would have, had they been in class all the time."

The students who work at their own desks have paired up to share a printed sheet upon which is a map of the United States. Alongside the map are some questions about the geography of the United States and some questions about the early explorers and developers of the country. Sheila says, "we are going to learn about the fine state of Virginia," and translates this into Spanish.

[…]

Sheila described to us another curriculum problem:

"At one point this year, I was almost going to do Social Studies three days in English and two days in Spanish... but you've got to remember that my kids, even my Hispanic kids, except for two, are fairly good in English at this point, and I watered the concept down into their simple English understanding..."

Sheila goes on to the next question which is about the explorers, She reads from the text and mentions the London Virginia Company. "What's a company?" she asks. Mike replies, "Like the Boston Herald America." "How do you know it's a company?" Sheila asks. Mike replies, "I do." The word 'company' is translated by and for some of the Hispanic pupils.

[…]

If there is actual watering down of the curriculum Sheila believes it relates not only to language problems but also to the availability of time. The conversation develops into specific problems of bilingual pedagogy:

"Okay, but the problem with teaching Social Studies here is the language...?"
No, no. The problem with teaching Social Studies is that we’re dealing with two languages and it’s taking time...

[...]

But what’s the difference between bilingual teaching and teaching in English and Spanish alternately the same lesson? Or is there none — is that what bilingual teaching is?

I don’t know. See, to me it’s very hard to describe bilingual teaching, because every year my bilingual teaching changes.

Does it? Can you give me an example?

It has to change. This year I have nine speakers of English out of a class of 25....

[...]

The ‘homeroom’ principle, the homeroom bilingual principle — is that the same in all bilingual programmes?

No, it varies.

Peculiar to the school?

Peculiar to this school, yes, and to some other programmes. Some other Principals say that the homerooms must be mixed, in mother classes they are totally segregated.

Right! Hang on. So another principle of your bilingual teaching is that both language groups have to hear the other language being spoken in the class, right?

Yea — oh, yeah.

[...]

I think, you know, there are some people — you look at them and you say, because of the situation we have here, because I’m a native English speaker, and I’ve got the 5th Grade, I’d do English first. But suppose I had these same kids, exact grouping in grade 2, the majority would be Spanish and I would do Spanish first and English second.

So even though it is bilingual teaching, within the pedagogic exercise there is still a dominant language? And that’s another principle.”
In the first account the purpose of field-based observations is to generate data which would subsequently be analysed with a theoretical lens – but analysed elsewhere. The theory was derived ‘elsewhere’ and the subsequent analysis happens elsewhere, too.

This second account is quite different. The analysis is started in situ – without the aid of theory – though, no doubt, with the ‘theoretical sensitivity’ of the evaluator as a resource. In fact, in place of theory we have theorising, the construction of theory from systematic reflection on experience. As the evaluator asks Sheila to reflect on her experience, the attempt is to discover theoretical ‘building blocks’ – here, enunciated as ‘principles’ underpinning her pedagogical action. As more principles emerge from these reflections so the elements of a theory begin to emerge.

At this point we may note that the narrative is constructed differently. In the first example we saw data from three discrete sources – theory, observation, respondent reflections. These were separated out – partly, as we saw, on the principle of the separation of knowledge/action, theory/practice. Here, however, the different kinds of data are interpolated in a way that violates the chronology of the event (the teacher was interviewed after the observation) but which ‘plots’ the emergence of an analysis. This narrative structure is itself analytic and draws him reader into the analytical task. In the first set of observations the reader has to follow the lead and the thinking of the evaluator; in this one, the reader is invited to join the analysis as it emerges.

Another aspect of this second approach is the integration of observation and interview into observation-based interviewing. The observation alone is not sufficient to feed the evaluation analysis and requires explanation and then further reflection from Sheila. The analysis is collaborative, with checking and re-checking happening at the same time. In fact, the analytical interview plays a number of functions at once – to generate data, to subject that data to a deliberative process, to theorise about those deliberations and then to verify the results.

To summarise, the former observation shows extant sociolinguistic theory being tested and verified: the second shows a process in which pedagogical theory is being generated. In the former, there is a separation of knowledge and action (knowledge is generated elsewhere, in other projects and imported into this action context); whereas in the second we see the conflation of knowledge and action in a process of reflection in and on practice. The resulting theory of bilingual pedagogy elaborated by the evaluation combined both in dynamic tension and extended conventional theory as it applied to this particular context.
Deliberation and Purpose

The second observation shows how methodological interactions give rise to collaborative theorizing between evaluator and practitioner in such a way as to expose educational purposes to scrutiny and reflection. In the course of the interactions that gave rise to this second example we see the evaluation methodology being reinvented as well as the situated theory of bilingual education. Here is an example of methodology being constructed historically, a reflection on what eventually worked rather than a prescription for action. As the methodology did emerge it did so in response to the need to develop independent educational understanding on behalf of the teachers. The task was to generate theory of bilingual pedagogy that was situated and existentially located in the meaning and the challenge facing teachers – i.e. practical (pragmatic) theory developed in the context of action. The evaluation methodology qualifies as being educational in two respects, first, in that both practitioners and evaluators learn from the experience and carry that learning forward into other contexts – second, in that the methodology is informed by the educational principles under observation.

Central to the theme of this article, however, is this comparative illustration of how different curriculum aspirations and purposes are reflected in correspondingly different evaluation approaches.

Note

1 There is an extended argument to be stated elsewhere grounding this difficult conclusion in pragmatic philosophical approach to education and moral development. Rorty (1998) for example, argues that moral progress is, in a pure Darwinian sense, a natural – i.e. un-sought – effect of social evolution. The kind of rationality underpinning our historical social development – what Rorty defines as ‘tolerance of difference’ – is likely to guarantee that liberal morality is more likely to find ‘functional fit’ in a social life if subjected to rational deliberation.

References


Contatti: Saville.Kushner@uwe.ac.uk