

LEARNER PERFORMANCE AND FUZZY GENERALISATIONS: KEY ISSUES IN PROFESSIONAL DEVELOPMENT RESEARCH

Jill Adler and Yvonne Reed, Witwatersrand University, South Africa.

In this paper we discuss two methodological issues that emerged through a teacher education research project at the University of the Witwatersrand (Wits). (i) If improved learner performance is an important indicator of professional development impact on the teacher, how can this be established? (ii) Given the complexity of teaching, and related qualitative methods in researching professional development, what kind of knowledge claims can be made? What might be their status? From our empirical experience, we argue that it is feasible to generalise about take-up from, and impact of, a professional development programme through 'fuzzy generalisations' (Bassey, 1999). We also raise concerns about one-off learner tests and inferences about professional development.

In 1996, the University of the Witwatersrand launched a Further Diplomas in Education (FDE) programme in Mathematics, Science and English Language teaching with the following goals: to broaden and deepen teachers' subject knowledge, pedagogic subject knowledge and educational knowledge; to extend teachers' reflective capabilities; to facilitate professional growth; and to enable access to further education. The team responsible for the development of the FDE set out to develop the programme through research¹. Undertaking the simultaneous development of, and research into, a programme was a considerable challenge, exacerbated at the time by a funding climate in South Africa in which teacher education was expected to demonstrate impact: education funders were demanding indicators of impact such as measurements of improved learner performance.

The push to learner performance as a significant component of research in contexts of educational change has recently gained further momentum. In their analysis of education research methodology, Taylor and Vinjevoold (1999) observe that changing forms of educational activity do not necessarily lead to improved learning. From this they assert two implications for research methodology. First, "the final test of the effectiveness of teaching/learning strategies lies in the outcomes of learning" and this is best gauged from learner performance on "carefully designed tasks". Second, observation needs to go beyond descriptions of form (p. 66). These are not new ideas. The issue is how is this testing and observation to be appropriately done, where appropriacy includes ethical research practice and possibilities for capturing the complexity of teaching and learning. In *teacher education* research, the underlying demand for accountability raises a key question. What kinds of data and analysis enable valid claims about learner performance and classroom practice and *how do these lead to inferences about the effects of a teacher education programme?* The double inference from learner performance to INSET effects is particularly problematic. We also need to ask whether INSET programme success can be inferred from qualitative descriptions of classroom practices on the basis of observation and/or videotext? What are the possibilities for generalised claims about substance (as opposed to form) of teaching and learning practices which, we would argue, require in-depth case studies? In short, what methodologies need to be employed to infer the worthwhileness of investment in in-service professional development programmes?

This paper begins with a brief discussion of the research methodology that evolved over the three years of the research project (See Adler and Reed, 2000, for more detail). The

focus then foregrounds mathematics and turns to two broad methodological issues: (i) If learner performance is an important indicator of the impact of professional development, how can this be established? (ii) Given the largely qualitative methodologies employed in professional development research, what kind of knowledge claims can be made and what might be their status?

Research methodology:

The overall aims of the FDE research project were three-fold: (i) to investigate teachers' take-up from the FDE programme in Mathematics, Science and English Language Teaching and to what extent and how this shaped the quality of their classroom practices; (ii) to contribute to knowledge about formal in-service professional development (INSET); and, (iii) to feed back into the FDE programme's curriculum development through research. The research team set out to describe and analyse continuities and changes in classroom practices within and across some participating FDE teachers over time, in relation to conditions in which teachers work and their pupils learn. Methodologically, while the research has 'project evaluation' elements to it, it is more appropriately described as a practice-based (Lampert and Ball, 1998), case study of cases (Bassey, 1999). *The FDE programme is the overall case under study.* It is a case of formal, in-service professional development designed to improve the practice of the practitioners in the programme, but also to contribute to policy and practice in the wider field of teacher professional development. *The teachers we worked with constitute a collection of particular cases.* The selection of 25 teachers across maths, science and English language, and their location in 10 urban/rural, primary/secondary schools is discussed elsewhere (Adler, Lelliott and Slonimsky, 1997). Our unit of study was the 'contextualised teacher', or the 'teacher-in-school'. If take-up from the programme and the quality of classroom practices was to be understood, these needed to be contextualised and personalised, with a description of what happened *over time* with *this* teacher in *these* kinds of circumstances i.e. a set of case studies. However, given the goals of the FDE programme, the study also needed to enable the identification of patterns or trends across teachers and contexts, or a cross-case analysis. The overall research project therefore focussed on a relatively small number of teachers in the programme. There were 10 mathematics teachers in the study, a number small enough to enable us to look in depth at each teacher, and large enough for us to be able to identify patterns and trends across teachers.

There are also two dimensions to the *practice base* of the study. The study is embedded in the practices of the FDE as a mixed-mode delivery, professional development programme. It was carried out by a team of researchers, most of whom are practitioners in the programme. The research was also carried out in classrooms, on and with classroom practitioners. It was our intention from the outset to learn about ours and the teachers' practices by investigating and theorising practice (Lampert and Ball, 1998) in local settings in all their complexity and diversity.

In relation to classroom practice, the key question that framed the overall study was: How are teaching and learning affected by resources (material; time; socio-cultural), by teachers' subject knowledge, by teachers' pedagogic knowledge and by teachers' reflective capabilities? Data was gathered from: school inventories, classroom observation schedules, supplemented by field notes; learner classwork and homework tasks and projects; tests administered to learners; videotapes of some lessons; audiotaped interviews with teachers and school principals; questionnaires and narratives completed by the teachers. The research

team worked in subject teams in the selected schools in the Northern Province and Gauteng for one week in each of the three years (1996-1998) of the project, with the data collected in 1996 serving as a base-line (see Adler, Lelliott and Slonimsky, 1998). The team attempted to develop portraits of each teacher, capturing as fully as possible, the texture of the teacher's practice. Each portrait was similarly structured to facilitate cross-case analysis and the identification of patterns of 'take-up' within and across the three subject areas in the first instance, and then across the whole set of cases. Analysis of the data collected in 1997 and 1998 also began with the development of teacher portraits. By 1998 we had identified key issues that we believed were central to the FDE programme, teacher development and curriculum change in South Africa, and to in-service professional development more widely. We proceeded to analyse the data over the three years according to the following key themes: the nature, availability and use of material and cultural resources as a function of programme take-up and the context of teachers' work; the critical issue of the relationship between teachers' knowledge-bases and development of high order knowledge and skills in their learners; the challenge of language-in-education policy and practice, particularly code-switching as a teaching and learning resource across contexts where English language infrastructure varies; teachers' take-up of the forms and substance of learner centred practice; whether and how in-service professional development plays a role in teachers becoming reflective practitioners.

Discussion of each of these is summarised in Adler, Bapoo, Brodie, Davis, Dikgomo, Lelliott, Nyabanyaba, Reed, Setati, Slonimsky (1999). The research team faced considerable challenges as the research unfolded. Practical constraints continually raised concerns about the potential for mismatch between our epistemological and methodological assumptions, our research intentions and goals, and on the ground realities. Many of these have been discussed elsewhere (Adler et al, 1998; Adler and Reed, 2000). The focus here is on the challenges of an accountability context discussed in the introduction to this paper.

Learner "performance" as indicator of INSET success

What does learner performance on "carefully designed tasks" (Taylor and Vinjevod, 1999, p.66) tell us? About the learner? About the teacher? About the curriculum? About national standards? Learner performance on carefully designed tasks can tell us about any or all of these ... depending. It depends on the nature of the tasks set, when the assessment occurs, where the assessment occurs and how often. These assertions are not new. As Saljö and Wyndam's (1993) study reveals, task "performance" is a function of the task and the learner in a particular setting at a particular time. There are thus significant issues in reading learner competence from single tests or tasks without significant attention to context, let alone moving from learner performance to the teacher's competence.

Learner *performance* is typically accessed through some form of testing. While it was the original intention in the study to test learners, the complexities of doing this were completely underestimated. We kept coming back to how a particular test or set of tests would be an appropriate or adequate means of assessing learners (and which learners?) over three years, and moreover, in such a way that the impact of the FDE programme on the quality of a particular teacher's practice could be inferred. For example, tests for the same learners at two different times in the year should show learning gains, but we could not see how to legitimately establish any kind of causality between learner gains and their teacher's participation in an in-service programme. More appropriate could be either comparative

testing of programme teachers' classes with other similar classes in their schools, or with similar classes in different schools where teachers were not involved in the programme. But we did not believe we could control for intervening variables, nor that such an endeavour was conceivable. In addition, any of these tests meant the construction of new items in order to assess new forms of knowledge and skills valued in the programme. The reliability and validity of any new test could not be accomplished within the time frame of the project. As Cooper and Dunne (2000) have argued, there are significant validity issues in more complex forms of math assessment, particularly context-embedded items.

We considered the use of existing standardised tests e.g. tests that had been constructed by the Human Science Research Council (HSRC) prior to the first democratic elections in 1994. Aside from inappropriate cultural contexts to contextual mathematics items, there were few that assessed mathematical processes beyond set procedures. Our frustration and difficulties with test development does not mean that it is not important to design effective means for studying learner performance as an indicator of learning gains. As Jansen (1996) and Black and Atkin (1996) argue, there are limitations in evaluations of educational innovations when student learning information is lacking.

Despite difficulties, the FDE team did not abandon assessment of learner performance as an additional indicator of teacher learning. Each teacher observed had a set of his or her classwork books examined to ascertain the kind of written work that was being covered by learners - a coverage that could not be discerned from two or three lesson observations. Learner books are not direct indicators of learner performance. They nevertheless can reflect the kind of mathematics valued by the teachers through inscription and attempts at practice/and mastery. These added to observations of mediated content during lessons. A detailed "Pupils' Written Work Schedule" was constructed, refined in 1997 and 1998 and used to illuminate learner performance in the subject through all their written work accumulated between February and August each year. Classwork books, homework books, test books, exam papers, test papers and scripts of nine learners in a purposefully selected class (three good, three average and three poor) were examined and recorded in the schedule. In addition, some testing was conducted in Grades 7, 8 and/or 9 classes in each of the three subjects. In 1997 the tests used were constructed and conducted by members of the research team. These were exploratory, both in terms of how they were used, and what they revealed. In 1998, we accessed Grade 7 mathematics tests that had been developed as part of a project geared towards more appropriately normed tests than those available to us through the HSRC. We built on Grade 9 tests from 1997 and conducted learner tests in classrooms where teachers were teaching math at either Grade 7 or 9.

We learnt several lessons from this testing. Testing some learners revealed to us how this additional data provides for triangulation of data within case studies. The test performance of learners in different teachers' classrooms by and large confirmed and thus strengthened the accounts of teaching and learning practices analysed and built into the teacher portraits. In instances where there was a mismatch between our independent test assessments and what we observed in learners' written texts, including their in-school testing, we were able to explore these with the teacher and develop insights to enrich the overall portrait of teaching and learning. Testing learners as part of researching teacher development and INSET effects can be illuminating. This is "the good" side of such testing.

The bad side of the testing for us was that, in general, our independent test results confirmed the TIMSS messages with regard to levels of performance in mathematics and science across our schools. Our results were not at odds with test results obtained by most of the teachers themselves in their own testing. The bad is that this situation persists. Our broader data assisted us in seeing that pupil performance was not in any simplistic way, a reflection on the teachers' knowledge-base. For example, one of the secondary level mathematics teachers, a teacher who demonstrated extraordinary take-up from the FDE programme, worked in an over-crowded, impoverished context. Her learners arrived in her Grades 8, 9 and 10 classes considerably under-prepared for the levels at which she was expected to teach and assess them. No wonder then, that on her own tests, let alone the independent tests we administered, performance was extremely poor. This learner performance tells us something about the state of the nation. However, to infer teacher quality and INSET programme quality from such "results" is extremely problematic.

The "normed" tests we used were themselves problematic. We found that some of the math items were ambiguous. This raised serious questions about reading off learner competence from their performance on tests that included such items. In addition, and as is the case with reform anywhere, for some of the learners, the form of the test items was unfamiliar. A simple prompt by a researcher on one occasion enabled a correct response by the learner. Crude analysis of test performance then will misrepresent learner knowledge. Testing is not simply a matter of "carefully designed tasks" but crucially a function of the testing context, including learners' familiarity with the tasks. Test validity is a serious research endeavour. Our concern as a result of our experience is not that testing should not be done, but that research and development be undertaken to develop instruments appropriate to various processes of research. Wilson and Berne note similar in professional development research in the USA where the capacity of researchers to tie measures of teacher learning to measures of student learning is challenged by "the lack of robust and standardised measures of student learning" (1999, p. 197).

Testing is not the only indicator of learning gains. Our experience suggests that close analysis of learner written material is more illuminating than one off tests in the contexts of teacher education research. Close examination of learners' classwork and test books, alongside the tests and examinations that teachers set and their marksheets were illuminating of depth and breadth of coverage over time by the teacher and the kinds of knowledge forms that were inscribed by learners hence indicating what is valued as knowledge within the school setting. We did not need independent tests, over and above such analysis, to reveal to us key challenges for teachers and hence the FDE programme. Learner written texts revealed learners' limited exposure to knowledge, and how teachers, for a range of reasons, were not covering required areas of learning, nor enabling learners to engage with knowledge at anything beyond the most superficial levels of recall and repetition. We noted difficulties with selection, sequencing and grading of tasks across most of the teachers, and this was evidenced across learner written texts. This observation tells us not only about teaching practices, but also about assumptions in the FDE programme. Like the government of the day, the programme had not grasped the full extent of the breakdown of a culture of teaching and learning across schools in struggles fought against apartheid education in the 1970s and 1980s.

Teacher education research: the status of resulting knowledge claims Bassey (1999) defines a case study as a “Study of singularity conducted in natural settings” (p.22). Earlier, we discussed how and why it was necessary to work in depth with a few teachers: qualitative multifaceted observation was required if we were to do any justice to the complexity of teaching as a social practice. How then do we, on the basis of diverse case studies, even with cross case analysis, make claims about teacher up-take from the FDE programme that extend beyond the specific research teachers, and hence to overall programme effects? Bassey (1995) distinguishes two kinds of empirical study in educational research: the search for generalisations (requiring investigation of large populations through carefully selected samples), and the study of singularities (case studies). The implication here is that case studies cannot lead to generalisations, and thus that they are limited in their use in educational policy and planning. Bassey argues that it is possible to develop what he describes as fuzzy generalisations from carefully conducted case studies. He uses the term “fuzzy generalisation” for a statement that makes no absolute claim to knowledge but hedges its claim with uncertainties. It arises when an empirical finding from a case study such as *In this case it has been found that* is turned into a qualified general statement such as *In some cases it may be found that* or *If we do x rather than y then teachers may learn more*. Bassey suggests that if educational researchers disseminate their findings in the form of “fuzzy generalisations” they are inviting teachers and education policy makers to enter into a discourse with these generalisations. Entry into such discourse is likely to be facilitated by access to an ‘audit trail’ - the evidence in support of the fuzzy generalisations which the case study has produced.

Bassey’s argument for “fuzzy generalisations”, and even weaker claims in the form of “fuzzy propositions” arises out of his extensive educational research experience, where he has seen numerous studies of quality not impacting on teachers and policy makers precisely because findings are deemed too specific. Our findings in the FDE research project, and the status of the related claims we believe we can and should make about the FDE programme as a whole, INSET practice in South Africa, and INSET practice more widely, resonate with Bassey’s notion of “fuzzy generalisations”. Indeed, “fuzzy generalisations” appear to be constitutive and reflective of other teacher education research.

In their review of “highly regarded” published research on “teacher acquisition of professional knowledge in” the USA, Wilson and Berne (1999, p. 194) identify a number of common themes, one of which is particularly pertinent here. They identify a concern with the labour intensity entailed in the qualitative nature of the research (hence expensive in human and related financial terms from our point of view), and the substantial commitment it demands in terms of examining teacher talk, and classroom practices. They point out that “[E]ach research project *struggles with ways to document teacher knowledge*” (1999, p. 195, emphasis added). Because of the complexity of classroom practice and the qualitative, case study nature of much of the research, documenting and hence evidencing teacher professional development is difficult. Claims made (“... programs ... were likely to ...”) are tentative, “fuzzy generalisations” in Bassey’s terms. To return to the FDE research, we will draw on our analysis of teachers’ take-up of language practices (Setati, Adler, Reed and Bapoo, forthcoming) to illustrate how we documented, evidenced, and then drew out recommendations at the level of the FDE programme, and fuzzy generalisations in relation to INSET policy in South Africa, and INSET research and development more widely.

We documented extent and frequency of teachers' and learners' code-switching practices over the three years of study. In the final year we also examined the production and reception of expressive and discourse-specific language. We used structured classroom observation schedules, unstructured videotape of lessons, structured observation of learners' written texts and teacher interviews. We found increased use of code-switching by teachers and learners in most classrooms, in particular increased drawing on learners' main language(s) as a resource. We learned from the teachers that their code-switching practices are intentional but dilemma-filled, particularly in the face of the dominance of English in the South African context. We also found widespread 'take-up' by most teachers of forms such as group work, and hence increased possibilities of learning from talk (i.e. of learners' using language as a social thinking tool). However, our observations in 1998 raised the question as to whether most of the teachers were complementing this shift to *learning from talk* with strategies for *learning to talk* i.e. learning to talk and write formal mathematical discourse. We also found that while the above were general patterns across all the teachers, they concealed important attenuated differences across teachers in different contexts, levels and subjects. For example, because their primary goals differ, there was more code-switching by math teachers than English language teachers. There was less code-switching and more focus on using and modelling English in primary than secondary mathematics classes as primary teachers carry out their dual functions of teaching the subject, and developing learners' proficiency in English. This dual role and emphasis on English was complicated further in rural schools i.e. schools with limited English language infrastructure.

These 'findings' from our case study of cases, led us to the following recommendation for the FDE programme, and fuzzy generalisations for INSET policy and practice. The intent here is to inform ongoing curriculum review in the FDE programme, and to invite teachers and policy makers to "enter into a discourse" (Bassey, 1999, p.52) with these generalisations.

* At the level of the FDE programme, we need to pay more explicit attention to possible journeys from exploratory and informal talk in the main language towards formal mathematical talk and writing in English

*At the level of educational policy in South Africa, findings from our research *suggest* that some of the dominant 'messages' in current curriculum documents *may* need to be reviewed.

For example, one of these messages in Curriculum 2005 is that group work is 'good' as it encourages exploratory talk and co-operative learning. The issue of how teachers and learners are to navigate the journey from informal spoken language (in the learners' main and/or additional languages) to formal, written mathematics in English is not addressed.

* At the level of INSET: What we have shown from our study of FDE teachers in multilingual contexts is that firstly, take up is attenuated across contexts. This *suggests* the need for more serious engagement in teacher education with the possibilities of, and constraints on, what are typically presented as panaceas for 'good practice'. The different English language infrastructures, levels and subjects in and with which teachers work *appear to be* significant for shaping INSET possibilities and constraints. We need further research and development programmes that dis-aggregate schools and classrooms along these three different axes (Setati, Adler, Reed and Bapoo, forthcoming, emphases added).

In the presentation of this paper, I will present additional information on the overall research methodology employed in the FDE professional development project, as well as some of the "audit" trial that lead to the fuzzy generalisations about changing language

practices and their effects. Hopefully, through this reflection on issues emerging from a specific research project, this paper will contribute to the development of theory, practice and research, in their inter-relationship, in mathematics teacher education more broadly.

References

Adler, J., and Reed, Y., (2000) Researching teachers' take-up from a formal, in-service professional development programme. *Journal of Education*. 25. 192-226.

Adler, J., Lelliott, T. and Slonimsky, L. with Bapoo, A., Brodie, K., Reed, Y., Setati, M., Mphunyane, M., Nyabanyaba, T., Van Voore, M. and Davis, H. (1997) *A baseline study: teaching and learning practices of primary and secondary mathematics, science and English language teachers enrolled in the Wits Further Diploma in Education*. Report. Johannesburg: University of the Witwatersrand.

Adler, J. Lelliott, T and Reed, Y., with Bapoo, A., Brodie, K., Dikgomo, P., Nyabanyaba, T., Setati, M., Slonimsky, L., Davis, H. and De Wet, H. (1998) *Mixed-mode FDEs and their effects: an interim report on the teaching teaching and learning practices of primary and secondary mathematics, science and English language teachers enrolled in the Wits Further Diploma in Education*. Report. Johannesburg: University of the Witwatersrand.

Adler, J., Bapoo, A., Brodie, K., Davis, H., Dikgomo, P., Lelliott, T., Nyabanyaba, T., Reed, Y., Setati, M., and Slonimsky, L. (1999) *Mixed-mode further diplomas and the effects: Summary report on major findings of a three year research project*. Johannesburg: University of the Witwatersrand.

Bassey, M. (1999) Case study research in educational settings. Buckingham: Open University Press.

Black, P. and Atkin, M. (1996) (Eds.) *Changing the Subject: innovations in science, mathematics and technology education*. London: Routledge.

Cohen, D. K. and Ball, D. L. (1990) Relations between policy and practice: A commentary. *Education Evaluation and Policy Analysis*, 12, 331-338.

Cooper, B. and Dunne, M. (2000) *Assessing children's mathematical knowledge: Social class, sex and problem-solving*. Buckingham: Open University Press.

Jansen, J. (1996) Does teacher development work? True confessions of a hardened evaluator. In JET, *Quality and Validity in INSET Evaluations*. Johannesburg: Joint Education Trust.

Lampert, M. and Ball, D. (1998) *Teaching, multimedia and mathematics: investigations of real practice*. New York: Teachers' College Press.

Saljö, R. and Wyndhamn, J. (1993) Solving everyday problems in the formal setting: An empirical study of the school as context for thought. In Chaiklin, S. and Lave, J. (Eds.) *Understanding practice: Perspectives on activity and context*. Cambridge: Cambridge University Press. Pp.327-342.

Setati, M., Adler, J., Reed, Y. and Bapoo, A. (forthcoming) Incomplete journeys: Code-switching and other language practices in mathematics, science and English language classrooms in South Africa. *Language and Education*.

Taylor, N. and Vinjevoold, P. (1999) *Getting Learning Right*. Johannesburg: JET/DoE.

Wilson, S. and J. Berne (1999) Teacher learning and the acquisition of professional knowledge: An examination of research on contemporary professional development. In Iran-Nejad, A. and P. David Pearson. (Eds.) *Review of Research in Education # 24*, 1999. Washington: American Education Research Association. Pp. 173 - 209

NOTES

1. The team that worked on the research over its three years, and to whom debt is owed for all research outcomes, including this paper, was: J Adler, A Bapoo, K Brodie, H Davis, P. Dikgomo, T Lelliott, T Nyabanyaba, Y Reed, M Setati, and L Slonimsky.