Editorial for issue 3

Lejf Moos • John MacBeath

Received: 8 July 2009 / Accepted: 17 July 2009 /
Published online: 29 July 2009
© Springer Science + Business Media, LLC 2009

This issue of the journal contains three articles on assessment instruments and one on policy implementation.

In the first article Lars G. Björk and Joseph Blase investigate one of the links between government and schools within the in the major educational reform *No Child Left Behind* (NCLB) in the USA. One of the aspirations of the NCLB was to make teachers and schools accountable for student learning and thus to make school districts and schools more active in implementing the reform. In this study Björk and Blase explore the case of one school district and ways in which new consortia, a new superintendent and a number of middle managers worked on implementing the reform through distributing leadership and using a Total Quality Management model in the school district and in schools.

An important feature of the school district reform was that middle managers, like deputy superintendents, had to implement new structures and practices that could eventually make their own position redundant. The case study demonstrates that the people who are responsible for implementing policies are very powerful political agents and thus very powerful in forming policies. The middle leaders never implemented all parts of the school district reform. This result may cause politician to reflect on implementation methods when they have decided on a policy and then want it implemented through local, school district layers. Politicians are, of course, the agents who make politics, but may have to reconsider how to get politics implemented because practitioners at all levels are not necessarily guided by the same interests as the politicians are. One practitioner interest may be not to make their own job redundant.

In the second article Francis L. Huang and Tonya R. Moon examine the relations between teacher characteristics and student achievement in low performing schools in the USA. They question traditional knowledge and evidence as to teachers' experiences,

L. Moos (✉)
Aarhus University, Århus, Denmark
e-mail: moos@dpu.dk

J. MacBeath
University of Cambridge, Cambridge, UK

licensing status and educational attainment and student reading achievement. They find that some of the traditional knowledge is not precise because in their study they found that traditional teacher qualifications such as licensing status and educational achievement—in short teacher education—are not statistically significant in producing better student achievement. The total years of teaching experience is also not a significant predictor, but years of experience at a particular grade level is.

It may be surprising that the level of teacher education is not statistically significant for teacher efficacy in the way described here. It also seems surprising that in relation to reading the general level of teacher experience (measured in total years of teaching experience) is not as important as the experience in one grade and one subject. Such results may have political implications if policy makers choose to pay attention to them. However, the authors add a caveat to their conclusions.

Huang and Moon go on to write that the teacher background characteristics used in their study may be too narrow. They have used data from investigations and research available—i.e., teacher certification, possession of advanced degree, attendance at reading conferences and total years of teaching experience—and thus they were not able to include data on factors such as teacher motivation, enthusiasm and presentation skills, attitudes and instructional practices, caring and dedication. From school effectiveness research we know, that factors like these are pivotal to teacher effectiveness and thus partly to student attainment, and therefore the authors are right to pointing out that the results of their research are dependent on the data available or the data that is being produced.

This study reminds us, that using the measures of academic status and certification such as bachelor's and master's degrees are not reliable indicators. Perhaps because these degrees cannot tell us whether the aspiring teacher has the requisite teaching or interpersonal skills, or their level of commitment or interest in the welfare of the children they teach. The study also reminds us that counting the years of teaching experience is also not a good indicator as it does not tell us whether the teacher has learned from prior experiences or the educational contexts in which their skills and attitudes have been shaped. Nor can it tell us about their resilience in the face of challenge and adversity, as the authors express it, '... because they possessed "undampened enthusiasm" ...'

In the third article Peter Denner, Antony Norman and Shu-Yuan Lin discuss fairness and 'consequential validity' of teacher work samples in order to determine whether the results of teacher candidate performance assessment are fair or if they are biased by age, gender or race/ethnicity. They therefore want us to extend the judgement of validity of those assessment tools to encompass the value implications and social consequences of interpretations made on the basis of these tools. In other words, as they write: '... does the assessment system convey the right message about what it means to be a highly qualified teacher?'

As they investigated these issues in two American universities the outcomes of this study may be of more interest to scholars in the USA than elsewhere, where the same kind of assessment tools may not be used. But the intentions and the perspective of the study: to examine the effects of using these assessment tools on diverse segments of students, described by age, gender or race/ethnicity, is very important from a democratic or human right point of view.

And in the final article Thomas M. Haladyna and Audrey Amrein-Beardsley report on the testing an older and a new instrument used to assess teacher educators: two

student surveys of instruction in an American teacher education college. Both surveys are regarded as important sources of information about college teaching in both summative and formative respects. They are intended to serve as the basis for the administrative, New Public Management approach to assessing teacher performance and so carry high-stake implications for promotion and merit pay. At the same time they are intended to serve as instruments for college teachers' teaching practice.

The study demonstrates that neither of the two surveys are able to yield useful subscores for diagnostic purposes, meaning they cannot be used for formative purposes such as assisting teacher educators or colleges of teacher education to improve teaching. The authors conclude *"But unless the instruments with which we measure excellence yield high levels of reliability, are used to make valid inferences about teacher effectiveness, and are used for summative, but more importantly formative purposes, these surveys will do nothing to promote instructional excellence across colleges of higher education."*

For us that the study provides yet further evidence to that the policy maker's dream of using the same kind of instrument for both summative and formative purposes is misguided. As been argued in this journal (see for example Black and Wiliam in the first edition) different kinds of instruments serve differing purposes and we need to be alive to their uses and misuses. While educators are committed to forms of assessment which serve formative purposes they may all too often find themselves using measures which actually distort or undermine that intent. The message to policy makers is to recognise the dysfunctions inherent in single use instruments and the imperative of paying attention to the impact of formative and high-stake assessment.

Lejf Moos & John MacBeath