

(Re)conceptualizing validity in (outcomes-based) assessment

Lesley Le Grange and Peter Beets

Faculty of Education, Stellenbosch University, Private Bag X1, Matieland, 7602 South Africa
llg@sun.ac.za

Validity is one of the traditional touchstones of assessment practices. However, the definition of validity has evolved over time along with changes in assessment practices. Killen argues that the common view at present that a test is valid when it measures what it is intended to measure provides a necessary, but insufficient basis for considering validity. He proposes that it might be more pertinent to focus on evidence from which valid inferences can be made about learning. We accept Killen's argument but wish to extend his ideas by looking at how validity has been (re)conceptualized in discourses on educational research and to explore how insights gained from this work apply to assessment practices. We examine how triangulation, face validity, catalytic validity, and rhizomatic validity could broaden our thinking about assessment and enhance assessment practices.

Key words: validity, assessment, knowledge traditions, outcomes-based education

Introduction

Changes in assessment theory and practice have become commonplace in many education systems across the globe. Calls have been made for more authentic ways of assessing learning and for assessment to become integral to teaching and learning processes. The international trend towards alternative (to traditional) forms of assessment influenced assessment policies that emerged in South Africa after the country's first democratic elections in 1994.¹ Two waves of change in South African assessment policy may be distinguished: first the introduction of continuous assessment policies which coincided with the introduction of interim school syllabus documents,² and second assessment policies aligned to an outcomes-based education (OBE)³ curriculum.

We agree with Fullan (1991) that the latest policies do not guarantee changes to practices. Changes in South African education policies (including assessment policies) after 1994 found teachers ill-prepared for new demands placed on them. With regard to new assessment policies, Vandeyar and Killen (2003:119) make the important observation on the basis of their study that teachers' responses vary: "a few teachers embraced the changes enthusiastically, many reluctantly accepted the changes, and most resisted". In their article they proceed to provide empirical evidence in support of their claim and conclude that some teachers ignore what they refer to as, "the principles of sound assessment practices". Although we have reservations about referring to "principles of sound assessment" (whatever sound assessment may mean) in the definitive/absolute sense in which they do, we acknowledge that constructs such as reliability, fairness, validity and discrimination are widely recognized as terms closely associated with assessment theory and practice. Based on this acknowledgement, in this article we shall specifically explore ways in which validity in assessment may be (re)conceptualised; an exercise we argue that could have transformative effects for both assessment theory and practice. In our exploration we invoke notions of validity more commonly associated with discourses in social research. In particular, we explore how constructs such as triangulation, face validity, catalytic validity and rhizomatic validity may be applied to assessment theory and practice. Our paper is divided into four main sections. First, we briefly discuss the relationship between validity and assessment. Second, we review how the construct validity has evolved within social research discourses. Third, we invoke particular notions of validity associated with social research and examine what insights they bring to bear on assessment theory and practice. Fourth, we discuss implications that the (re)conceptualization of validity in assessment might have for (outcomes-based) assessment.

Validity and assessment

The conventional definition of validity, "a test measures what it is suppose to measure", is informed by measurement theories from the 1950s. Several validity types have been derived from this notion of

validity and named among others: content validity, predictive validity, concurrent validity and construct validity. Content validity concerns the extent to which a test is representative of the domain that it is supposed to be testing. Predictive validity concerns the extent to which a test score correlates with some criterion measurement made in the future. Concurrent validity concerns the extent to which a test's results correlate with those of a different assessment task performed at the same time. Construct validity refers to the extent to which a test score measures a construct of interest such as intelligence, self-esteem, and so on (for details see Messick 1989; Brady & Kennedy 2001; Killen 2003). The brief descriptions we have provided make it clear that these validity types are rooted in the same idea, or put another way, they are branches of the same tree. Killen (2003) notes that construct validity, predictive validity and concurrent validity have not really appealed to teachers. Applying them has also been constrained by practical difficulties resulting from the way teaching and learning generally occurs in most education systems. On the other hand, as he points out, content validity appeals to teachers essentially because of its simplicity. It is easy for teachers to ascertain whether a test covers relevant content (or outcomes) and to ascertain whether a test has adequate content coverage. However, Killen (2003:3) argues that limiting assessment practices to "uncomplicated and relatively easy-to-apply criteria" exonerates teachers from considerable responsibility, arguing that

[t]he problem is that such an approach ignores the importance of the conditions under which the test is administered, the effect that learners' characteristics will have on their responses, and the responsibility that teachers have to interpret the test results in defensible ways. This simplified approach to validity also overlooks the fact that determining what a test measures requires more than considering just content relevance and representativeness.

In order to improve assessment practices, Killen (2003) argues that a different understanding of validity needs to be invoked. Drawing on insights from the work of Messick (1989), he notes that the meaning of the term validity has evolved with time and that a more recent understanding of validity concerns the extent to which justifiable inferences can be made on the basis of evidence gathered. The interest therefore is not to validate a test but to validate the inference that can be drawn from the learner's results in the test/assessment task. Viewing validity as inference necessitates a greater responsibility on the part of teachers. As Killen (2003:6) writes:

No longer can a teacher claim to be using a valid assessment task simply because it is clearly linked to the curriculum content, or because someone else has used the test and decided that it is valid, or because it produces results similar to those obtained from other assessment tasks. Instead the teacher must question the validity of the inferences they are making as a result of having used the assessment task.

Validity as inference is a useful idea since it holds the promise of liberating assessment practices from its behaviourist orientations (inform-

ed by measurement theories of the 1950s). Therefore, in this article we argue that a more nuanced understanding of validity as inference may be possible if we invoke notions of validity that have traditionally been more closely associated with discourses on education research. However, we agree with Calfee and Masuda's (1997) suggestion that much of classroom assessment is applied social science research anyway, thus blurring the boundaries between assessment and research.

Validity and knowledge/research traditions

Habermas (1972) argues that different knowledge traditions are linked with particular social interests. He points out that positivistic research employs technical/instrumental reasoning where the ends are pre-defined and are attained by following known rules and predefined means (e.g. the textbook version of the scientific method). Habermas described this kind of knowledge as being informed by technical interest. On the other hand, the interpretive or hermeneutical sciences employ practical modes of reasoning (Habermas, 1972). Thus appropriate decisions are made in the light of the circumstances of the situation and not by pre-defined means and ends. Positivistic research is associated with prediction and control and interpretive research with enlightenment, understanding and communication (Usher, 1996:22). These research traditions, however, do not have an interest in research that changes the world in the direction of freedom, justice and democracy. Habermas (1972) therefore isolates a third type of 'knowledge-constitutive interest', which he links with critical science, that is, an emancipatory interest. This knowledge interest involves the unmasking of ideologies that maintain the *status quo* by denying individuals and groups access to knowledge or awareness about the material conditions that oppress or restrict them (Usher, 1996:22). Importantly, critical science is also concerned with the actions that can be taken to change oppressive conditions. However, critical science has been challenged for positing its emancipatory project as a universal value. Gore (cited in Usher, 1996) uses Foucault's notion of 'regimes of truth' and points out that critical science/theory has its own power-knowledge nexus which may operate in oppressive and repressive ways in particular historical contexts and instances.

Postmodernism reflects a decline of absolutes and questions the belief that research that follows universally defined rules and a universal scientific method will guarantee 'true' results. It questions the dominant (positivist) epistemology's conception of what constitutes scientific knowledge. Postmodern approaches, however, are not anti-science but contend that science should critically self-examine its limitations. Postmodern research approaches also do not uncritically embrace the interpretive research tradition. It sees the interpretive approach as still implicitly operating within the terms and discourse of positivist traditions. In other words, "the emphasis on the 'subjective' instead of the 'objective' is merely a reversal which still works within the framework of the 'objective-subjective' as polar opposites" (Usher, 1996:26). Postmodern approaches seek to subvert this dichotomy and suggest alternatives, which wholly challenge dominant epistemological discourses in all its different forms. According to Hargreaves (1994: 39) adopting a postmodern position involves denying the existence of foundational knowledge on the grounds that no knowable social reality exists beyond the signs of language, image and discourse. This implies that postmodern theorists employ the practice of deconstructing existing representations of social reality, and giving voice to other versions, which are normally marginalised or suppressed. It is in this context that Patti Lather (1992:89-90) extends Habermas's knowledge-constitutive interest and introduces a fourth category, deconstruction (see Table 1).

Because the paradigms mentioned have disparate knowledge-constitutive interests (ontologically, epistemologically and axiologically disparate) their criteria or standards for determining rigour may differ. In Table 2 we have attempted to categorise different validity types within disparate paradigms. However, we acknowledge that any categorical scheme has dangers, in that readers may interpret such schemes narrowly in the sense that items are understood to be rigidly

framed within defined paradigms. However, in many instances the boundaries between different paradigms (and validity types framed within them) are blurred. Also, the same term may be used to describe validity but have different meanings with disparate paradigms. For example, although a conventional understanding of concurrent validity has positivist underpinnings, Lather (1986:67) (re)defines the term for the purposes of emancipatory research. Despite its limitations, we use the categorical framework as a conceptual tool for advancing our arguments, and for facilitating our thinking/learning, that is, we use it for heuristic purposes (see Table 2).

Each of the validity types mentioned could be variously explored in order to provide a more nuanced understanding of validity in assessment, but for the purposes of this article we shall discuss four validity types and examine how these might be brought to bear on assessment theory and practice.

(Re)conceptualising validity in assessment

As mentioned, four validity types used within educational discourses will be invoked to extend and critique Killen's (2003) notion of validity as inference: triangulation, face validity, catalytic validity, and rhizomatic validity.

Triangulation

With reference to social research, Lather (1986:67) argues that triangulation should be expanded beyond the psychometric definition of multiple measures to include multiple data sources, methods, and theoretical schemes for establishing data trustworthiness. She argues that "it is essential that the research design seek counter-patterns as well as convergences if data are to be credible". Triangulation, we argue is a useful construct that teachers can use to give credibility to inferences that they make, especially if an expanded view of triangulation is accepted. Here, we mention two ways in which triangulation could be used to enhance the credibility of inferences teachers make about learners' abilities. Firstly, teachers can use the results of different assessment tasks (e.g. a journal entry, an exhibition/demonstration and a practical investigation) when making inferences. Secondly, assessments by more than one role player (self-assessment, peer-assessment and teacher-assessment) can be used to enhance the credibility of inferences. There may be other ways in which triangulation could be used to enhance the credibility of inferences teachers make, which we invite the reader to explore. Suffice it to say, triangulation opens up greater possibilities of enhancing the credibility of inferences teachers/assessors make and in so doing extends the work of among others, Messick (1989) and Killen (2003). But, Lather (1986:67) gives a different take on validity as inference that we argue requires attention, i.e. that the credibility of data is not only achieved by seeking convergences but also counter-patterns. Killen's (2003) description of validity as inference appears to be founded on the principle of verification whereas Lather's (1986:67) notion brings to bear the importance of falsifying inferences (in line with Popperian thinking). Cronbach (1980, quoted in Lather, 1986:67) neatly captures this notion of validity:

The job of validation is not to support an [inference], but to find out that might be wrong with it. A proposition deserves some degree of trust only when it has survived serious attempts to falsify it.

The notion of validity just described leaves teachers with the challenge of not only making the shift from conventional understandings of validity to embracing the idea of validity as inference, but also to understand that inferences are tentative and that they should be subjected to processes of falsification.

Face validity

Face validity as we use it here does not refer to the conventional understanding of it that a test on first impression appears to measure the intended content or trait (see Hastings & Stewart, 1983:701). Rather we see it as a strategy that is integral to the process of establishing the

Table 1 Paradigms of post-positivist inquiry. Adapted from Lather (1991) and Habermas (1972)

Predict	Understand	Emancipate	Deconstruct
<ul style="list-style-type: none"> • Positivism • Empiricism 	<ul style="list-style-type: none"> • Interpretive • Naturalistic • Constructivist • Phenomenological • Hermeneutic • Symbolic interaction • Micro-ethnography • Macro-ethnography 	<ul style="list-style-type: none"> • Critical • Neo-Marxist • Feminist • Race-specific • Praxis-oriented • Freirean • Participatory 	<ul style="list-style-type: none"> • Post-structural • Post-modern • Post-paradigmatic

Table 2 Validity types and research paradigms

Paradigm	Positivist/Empiricist	Interpretive	Critical	Postmodern/Poststructuralist
Types of validity	<ul style="list-style-type: none"> • Content validity • Construct validity • Criterion-related validities 	<ul style="list-style-type: none"> • Inter-subjective objectivity • Face validity — (member checking) • Thick description • Triangulation 	<ul style="list-style-type: none"> • Inter-subjective objectivity • Face validity — (member checking) • Thick description • Triangulation • Catalytic validity 	<ul style="list-style-type: none"> • Rhizomatic validity • Transgressive validity • Reflexivity • Verisimilitude • Thick description

credibility of inferences. For Lather (1986) face validity in the context of openly ideological research is closely related to what Guba and Lincoln (1980:110) refer to as "member-checks". Put simply, member-checking involves recycling analyses of data to at least a sub-sample of respondents. With respect to assessment, member-checking could mean going back to learners with tentative inferences and then refining them in the light of learners' responses. Perhaps face validity is more pertinent to higher education where it is conceivable, particularly at the post-graduate level, for lecturers to go back to students with tentative inferences. However, we contend that face validity has relevance at all levels of the education system. To illustrate how face validity might work I refer to an incident that Killen (2003) describes. Killen (2003:4) narrates his involvement with three Australian high school mathematics teachers to help them improve their teaching of algebra. He notes that the three teachers were satisfied that the test that they had set was valid since it measured what they wanted it to measure. The teachers felt justified in concluding that learners who did well in the test had a sound knowledge of basic concepts in algebra and those who obtained low test scores had a poor knowledge of basic algebra. Killen later interviewed some of the learners who performed poorly in the test and found out that these learners did not have the prior arithmetic knowledge to perform the algebra tasks. He concluded that their inability to perform the algebra problems had very little to do with the extent to which the test was representative of the domain of algebra concerned, and that their lack of arithmetic knowledge would have also produced poor results in a disparate domain that required the same prior arithmetic knowledge. So, not only was the teachers' understanding of validity parochial but also their inferences did not take into account of the prior knowledge learners needed to perform the algebra tasks. Killen's interview with the learners might be viewed as a form of member-checking or least it enables us imagine what this form of validity could be in practice.

Catalytic validity

Catalytic validity concerns the degree to which the research process has transformative or empowering outcomes. Lather (1986:67) writes:

[Catalytic validity] is premised on a recognition of the reality-altering impact of the research process itself, but also on the need to consciously channel this impact so that respondents gain self-understanding and, ideally, self-determination through research participation.

In the South African context, it is necessary that social practices such

as assessment are transformative in the sense that they heighten learners' self-understandings — how dominant social, economic and political discourses influence the construction of learners' identities. In South Africa there is a tension between goals of social relevance (issues such as equity and redress) and what is referred to as academic merit/standards. Guba (1981) importantly reminds us that, "relevance without rigor is no better than rigor without relevance". The dominant, "value-neutral" content validity with its emphasis on whether a test is representative of a particular knowledge domain does not generally provide space for addressing issues of social relevance. However, catalytic validity holds the promise of addressing both the concerns of social relevance and academic standards. When inferences are therefore made about learners' abilities they should be validated in terms of all three knowledge-constitutive interests that Habermas (1972) identifies: the technical, the practical, and the emancipatory. Inferences should not only be based on learners' abilities/achievements of "value-neutral" content knowledge but also on the extent to which assessment practices have been empowering for learners by being more democratic and in the South African context, perhaps also incorporating the extend to which they foster what is captured in the Bill of Rights of the South African Constitution.

Rhizomatic validity

We use the rhizome as a metaphor for the reinscription of validity in assessment. Deleuze and Guattari (1987) suggest the tree as a modernist model of knowledge and the rhizome as a model for postmodern knowledge. Rhizomes are elongated underground stems with aerial roots, as well as leaves and flower stalks found at their growing tips. Lather (1994:45) argues that to act rhizomatically,

is to act via relay, circuit, multiple-openings, as crabgrass in the lawn of academic preconceptions ... There is no trunk, no emergence from a single root, but rather arbitrary branchings off and temporary frontiers that can only be mapped, not blueprinted ... Rhizomatics are about the move from hierarchies to networks and the complexity of problematics where any concept, when pulled, is recognised as connected to a mass of tangled ideas, uprooted, as it were, from the epistemological field.

Rhizomatic validity troubles the single rootedness of validities underpinned by positivist assumptions. As mentioned content validity, predictive validity, concurrent validity and construct validity are branches of the same trunk rooted in the same notion of validity concerned with the idea of a test being measured against some fixed or known entity

external to the test itself. Concerning validity as inference, postmodern inferences are "without firm foundations or fixity", but rather in constant movement (Gough & Price, 2004). Rhizomatic validity dissolves inferences "by making them as temporary, partial [and] invested" (Lather, 1994:46). Teachers/assessors might therefore self-reflexively engage the inferences they seek to draw. By this we mean that teachers/assessors acknowledge they have an autobiography marked by the significations of gender, sexuality, ethnicity, class, and so on (Usher, 1996:38), that impact on the work they perform, in this instance drawing (temporary) inferences. Furthermore, teachers might view inferences as networks of seemingly unrelated ideas or performances. These networks are assemblages of un(planned) manifestations of learning, ideas that spring up at different places and (un)expected times. Inferences become a "rhizomatic journey among intersections, nodes and regionalizations through a multicentred complexity" (Lather 1994: 46).

Some implications for outcomes-based assessment

A narrow interpretation of outcomes-based education sees its epistemology as instrumentalist, which reinforces behaviourist assessment practices. When outcomes-based education is narrowly defined, its interpreters easily invoke validities such as content validity — the interest being, does the test measure the outcome? Killen (2003:9) uses the term "outcome-related evidence of validity" to describe content validity's variant within an OBE framework. Outcome-related evidence of validity has two components, outcome relevance and outcome coverage. Outcome relevance concerns the extent to which assessment items require learners to demonstrate one or more pertinent outcomes. Outcome coverage refers to the extent to which the assessment task or test is a representative sample of a set of outcomes. Also, when outcomes-based education is narrowly defined, outcomes are described in absolute terms so that an outcome is either achievable or not. As Killen (2003:10) writes:

One of the sources of invalidity in inferences about learner achievement in OBE is the belief that it is possible to make clear-cut decisions about whether or not a learner has achieved a particular outcome. In outcomes-based education it is often said that the principal reason for assessment is so that we will know whether or not learners have achieved the outcomes we wanted them to achieve. This view of assessment is unfortunate because it can mislead teachers into thinking that it is possible to make clear distinctions between those learners who have achieved certain outcomes and those who have not.

However, he argues that the problem might be overcome if we "think of assessment as the process of determining how well learners are able to demonstrate what they have learned, rather than trying to determine in some categorical sense which learners have or have not learned" (Killen 2003:10). Triangulation can strengthen the credibility of inferences drawn, about what learners have learned, by providing them with different opportunities to demonstrate what they have learned, and by assessing what they have learned through the use of different assessment methods as well as the use of different assessors. Face validity can enhance the credibility of inferences about what has been learned, by way of teachers sharing initial impressions of learners' work with them and receiving feedback from learners about such impressions. Face validity in the sense that we use it here serves as a type of formative assessment. When using triangulation and face validity, outcomes are not to be viewed as parochial descriptions of what can be demonstrated or not. Rather the outcome is what the teacher has in mind when designing curricula and assessing learners, shifting the emphasis to "determining how well learners are able to demonstrate what they have learned" (Killen, 2003:10). Catalytic validity invites us to rethink what we mean by "what has been learned" or what is meant by "know" when we say that an outcome by definition is, what learners know and can do. Catalytic validity in this instance concerns the extent to which assessment practices are catalysts for change, emancipation, and empowerment. In other words, the inferences drawn

here are concerned with whether assessment practices have developed a heightened consciousness of how particular assessment practices are oppressive and others liberating. For example, they may (not) understand how assessment tasks or tests are gender, race or culturally biased. Clearly, the view of knowledge referred to here is distinctly different to that of propositional knowledge (knowledge expressed as statements, facts or theories).

Rhizomatic validity troubles the notion that outcomes or inferences drawn about learning can be defined in any absolute sense. Viewing outcomes as rhizomes enable us to understand them as being in constant movement, that is, without fixity. They are always tentatively understood as moments that emerge during pedagogical episodes when teachers observe learners' performances. Inferences drawn about what is learned becomes an art of assembling momentary or emerging performances in a classroom. The inference gives meaning to the outcome and in a sense, tentatively "defines" the outcome. Even though the teacher may start his/her curriculum planning with nationally defined outcomes, representations of these outcomes in policy documents are paramount. Instead outcomes gain significance through learner performances that are often unplanned and that might occur at unexpected times.

Concluding reflections

It is widely accepted that validity is one of the important characteristics of "good" assessment practices. But, validity does not have a single meaning and is a term whose meaning has evolved over time. A conventional understanding of validity concerns, whether a test or an assessment task measures what it is supposed to measure. Killen (2003) argues that this conventional understanding of validity, informed by measurement theories of the 1950s, has come under scrutiny, and that it might be more appropriate to think of validity as the extent to which justifiable inferences can be made on the basis of evidence that has been gathered. The shift in the understanding of validity at a theoretical level has not necessarily translated into practice and as a consequence the conventional view of validity still dominates assessment practices. However, Killen's (2003) invocation of validity as inference, opens up an opportunity to further reconceptualise validity in assessment. We respond to this challenge by invoking notions of validity more commonly associated with social research so as to extend Killen's idea of validity as inference. We specifically explore how triangulation, face validity, catalytic validity and rhizomatic validity can be brought to bear on assessment theory and practice. Although our argument is chiefly theoretical we contend that our exploration invites us to imagine alternative assessment practices to conventional ones so that outcomes-based education (assessment) can be liberated from its behaviourist roots. Outcomes should not be viewed as fixed entities that are to be achieved through instrumentalist means, rather they should be the product of critical engagement between teacher and learner and the knowledge they co-produce. Education should not serve a purpose external to itself (such as achieving narrowly defined and pre-determined outcomes). In a context where nationally defined outcomes exist (such as in South Africa), these should be used merely as entry points to pedagogical episodes and that its multiple meanings should emerge from pedagogical interaction. We invite you, the reader to imagine alternative assessment practices so as to liberate outcomes-based education from what often circumscribes it.

Notes

1. We acknowledge that changes in assessment theory and practice are embedded in global and local geo-sociopolitical forces.
2. Interim school syllabus documents represented mainly cosmetic revisions of apartheid school syllabi. The only major revisions to the apartheid syllabi in the middle 1990s was the excising of, in particular, racial, cultural and gender biases in subjects such as History.
3. South Africa's first post-apartheid Minister of Education launched a model of outcomes-based education entitled Curriculum 2005 in March 1997, ostensibly marking a distinctive departure from apartheid educa-

tion. Although Curriculum 2005 has undergone revision, its more recent variant the Revised National Curriculum Statement (RNCS) for GET as well as the National Curriculum Statement for FET remain underpinned by an OBE "philosophy".

References

- Brady L & Kennedy K 2001. *Celebrating student achievement: Assessment and reporting*. Sydney: Prentice.
- Calfee R & Masuda W 1997. Classroom assessment as inquiry. In: Phye G (ed.). *Handbook of classroom assessment*. San Diego, CA: Academic Press.
- Deleuze G & Guattari F 1987. *A thousand plateaus: Capitalism and schizophrenia*. (Trans. B Massumi). Minneapolis: University of Minnesota Press.
- Fullan M 1991. *The new meaning of educational change*. Cassell: London.
- Gough N & Price N 2004. Reworking the world: poststructuralism, deconstruction and the 'real' in environmental education research. Unpublished paper.
- Guba E 1981. Criteria for assessing the trustworthiness of naturalistic inquiries. *Educational Communications and Technologies*, 29:75-81.
- Guba E & Lincoln Y. 1981. *Effective evaluation*. San Francisco: Jossey-Bass.
- Habermas J 1972. *Knowledge and human interests*. London: Heinemann.
- Hargreaves A 1994. *Changing teachers changing times: Teacher's work and culture in the postmodern age*. London: Cassell.
- Hastings J & Stewart J 1983. An analysis of research studies in which "homemade" achievement instruments were utilized. *Journal of Research in Science Teaching*, 20:697-703.
- Killen R 2003. Validity in outcomes-based assessment. *Perspectives in Education*, 21:1-14.
- Lather P 1986. Issues of validity in openly ideological research: Between a rock and a soft place. *Interchange*, 17:63-84.
- Lather P 1991. *Feminist research in education: Within/against*. Geelong: Deakin University Press.
- Lather P 1992. Critical frames in educational research: Feminist and post-structural perspectives. *Theory into Practice*, 31:87-99.
- Lather P 1994. Fertile obsession: Validity after poststructuralism. In: Gitlin A (ed.). *Power and method: Political activism and educational research*. London: Routledge.
- Messick S 1989. Validity. In: Linn R (ed.). *Educational measurement*. New York: Macmillan.
- Usher R 1996. A critique of the neglected epistemological assumptions of educational research. In: Scott D & Usher R (eds). *Understanding educational research*. London: Routledge.
- Vandeyar S & Killen R 2003. Has curriculum reform in South Africa really changed assessment practices, and what promise does the revised national curriculum statement hold? *Perspectives in Education*, 21:119-143.