

Questioning the validity of inquiry assessment in a high stakes Physical Sciences examination

Umesh Ramnarain

The South African science curriculum advocates an inquiry-based approach to practical work. Inquiry is a complex and multifaceted activity involving both cognitive and physical activity; thus, paper-and-pencil items do not provide the authentic context for this assessment. This study investigates the construct validity of inquiry-related questions in three national Grade 12 Physical Sciences examinations. Clarity about what is being assessed and how well a test samples a construct are critical to validity. The analysis that was guided by Stobart's conceptualisation of construct validity revealed that, to a large extent, inquiry-related questions exhibited threats to validity. The identified threats were categorised as contested validity, unclear validity and construct irrelevance. The findings of this study suggest that greater attention needs to be paid to the formulation of inquiry-related questions in written tests and examinations.

Keywords: science inquiry, assessment, validity, examinations

Introduction

School science curriculum reform in South Africa and other countries throughout the world has focused largely on practical work. Traditionally, practical work, if it did take place, was either in the form of teacher demonstrations or it embodied a cookbook approach. This approach meant learners followed recipes for the execution of procedures handed down by teachers, and gathered and recorded data without a clear sense of purpose (Roth, 1994). 'Inquiry' has become a perennial and central term in the rhetoric of past and present science education reforms (Abd-El-Khalick, BouJaoude, Duschl, Hofstein, Lederman & Mamlok, 2004). According to a report by the Inter-Academy Panel (2012: 19), during inquiry-based learning, learners "use skills employed by scientists such as raising questions, collecting data, reasoning and

Umesh Ramnarain

Department of Science and Technology Education, University of Johannesburg

E-mail: uramnarain@uj.ac.za

Telephone: 011 559 4394

reviewing evidence in the light of what is already known, drawing conclusions and discussing results". This may involve them in first-hand manipulation of objects and materials, and observation of events or it may entail using evidence gained from a range of information sources including books, the Internet, teachers and scientists.

The South African school science curriculum advocates an inquiry-based approach to practical work. The place of scientific inquiry is addressed through Learning Outcome 1 of the National Curriculum Statement (NCS) that is referred to as 'Practical scientific inquiry and problem-solving skills'. This outcome states that 'the learner will be able to act confidently on curiosity about natural phenomena, and to investigate relationships and solve problems in scientific, technological and environmental contexts' (Department of Education, 2003: 13). This imperative is also expressed in the new Curriculum and Assessment and Policy Statement (CAPS) document where Specific Aim 1 states that 'the purpose of Physical Sciences is to make learners aware of their environment and to equip learners with investigating skills relating to physical and chemical phenomena' (Department of Basic Education, 2011: 8). It is evident from these curriculum imperatives that the doing of inquiry does encompass the application of investigative skills such as planning, observing and gathering information, comprehension, synthesising, generalising, hypothesising and communicating results and conclusions (Department of Education, 2003). An inquiry is perceived to be a more encompassing concept than an investigation and includes a range of activities with a focus on describing objects and events, asking questions, constructing explanations, testing those explanations against current knowledge, and communicating their ideas to others (NRC, 1996).

These developments in South Africa mirror the worldwide reform trends in science education. In the United Kingdom, Attainment Target 1 for Science in the National Curriculum has apportioned much priority to inquiry (Department for Education and Employment, 1999). In the United States, the American Association for the Advancement of Science (AAAS, 1993) and the National Research Council (NRC, 2000) endorse inquiry-based science curricula that actively engage learners in investigations.

Inquiry is a multifaceted activity and the widely quoted description given in the National Science Education Standards captures the essence of inquiry:

Inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. Inquiry requires identification of assumptions, use of critical and logical thinking, and consideration of alternative explanations (NRC, 1996: 23).

In order for learners to competently engage in inquiry, they need to have well-developed investigative skills such as 'classifying, communicating, measuring, designing an investigation, drawing and evaluating conclusions, formulating

models, hypothesising, identifying and controlling variables, inferring, observing and comparing, interpreting, predicting, problem-solving and reflective skills' (Department of Basic Education, 2011: 8). In addition to being multifaceted, inquiry is also more complex than popular conceptions would have us believe. According to the AAAS (1993: 9):

It is, for instance, a more subtle and demanding process than the naive idea of 'making a great many careful observations and then organizing them.' It is far more flexible than the rigid sequence of steps commonly depicted in textbooks as 'the scientific method.' It is much more than just 'doing experiments,' and it is not confined to laboratories. More imagination and inventiveness are involved in scientific inquiry than many people realize, yet sooner or later strict logic and empirical evidence must have their day.

Assessing inquiry

The term 'assessment' is nebulous, but for this research I adopt the definition of Nusche *et al.* (2012) cited in Harlen (2013: 24). Here the term is used to refer to "judgements on individual student performance and achievement of learning goals. It covers classroom-based assessment as well as large-scale, external tests and examinations". The issue of assessing for inquiry learning has stimulated much debate amongst scholars in science education. Ketelhut, Clarke, Dede, Nelson and Bowman (2005) pose the question: 'What kinds of assessments will allow valid inferences about whether a student has learned how to engage in inquiry?' They further speculate on whether inquiry can be adequately assessed using a standardised test format. Resnick and Resnick (1992) allude to the complexity of inquiry, and maintain that inquiry involves higher-order thinking skills that are not easily measured through standardised testing. Buckley, Gobert, Horwitz and O' Dwyer (2010) agree that traditional assessments fail to capture the complex understanding of inquiry skills needed to learn from inquiry. This difficulty in assessing due to the complexity of inquiry was highlighted in an international conference held in Helsinki in 2012 that was jointly planned by the Global Network of Science Academies (IAP), ALLEA (All European Academies), the Finnish Academy of Science and Letters, and Finland's Science Education Centre (LUMA). It was recognised there that it is almost impossible to elicit through tests the rich information needed to assess inquiry-based learning goals (Harlen, 2013).

Ketelhut *et al.* express the view that, if teachers are to reinforce an inquiry-based pedagogy in their classrooms, then there needs to be more emphasis on inquiry-based assessment in standardised testing such as examinations. They maintain that, in this way, the dilemma between teaching content knowledge versus scientific process skills can be resolved. This raises the critical issue that, if inquiry is to be tested in examinations, what form should the items assume?

Before engaging with this issue, it is worthwhile to consider more broadly what assessing inquiry science entails. Firstly, inquiry involves the investigation of

phenomena in the natural world, and this requires both physical and mental activity (Hein & Lee, 2000). The physical activity demands the application of process skills such as observation, handling of apparatus and measuring. The mental processing skills include formulating a hypothesis, designing the investigation, drawing valid conclusions and so on. Assessing inquiry would, therefore, require that attention be paid to both aspects. Research into inquiry practice has provided some insight on how inquiry should be assessed. According to Mislevy, Chudowsky, Draney, Fried, Gaffney and Haertel (2003), inquiry skills are developed in scientific contexts and should, therefore, be assessed within these contexts. Context-based assessments are more authentic because they require specific skills to solve real problems (Baxter & Shavelson, 1994; Ruiz-Primo & Shavelson, 1996).

It is clear that, due to the nature of inquiry, traditional forms of assessment, for example, paper-and-pencil items which feature in summative assessments such as tests and examinations, should be scrutinised more closely. Quellmalz and Pelligrino (2009) maintain that this type of assessment has relatively limited possibilities for measuring the complex science knowledge and skills that inquiry instruction was designed to target.

Against this background, the study reported in this article investigated the validity of inquiry tasks in national Grade 12 Physical Sciences examinations. The following question is framed: Can inquiry be adequately assessed in a standardised format such as an examination? In addressing this question, I present a validity analysis of inquiry tasks that have featured in national Grade 12 Physical Sciences examinations. I consider both Physics and Chemistry examination papers for the November examinations of 2010, 2011 and 2012. Before engaging with this, I discuss validity in assessment and present the framework that guided this research.

Validity in assessment

In order to develop any assessment, the most important issue to resolve is determining what is going to be assessed (Hein & Lee, 2000). Clarity about what is being assessed and how well a test samples a construct are critical to validity (Stobart, 2001). Kane (2006: 17) defines validity as 'the extent to which the evidence supports or refutes the proposed interpretation and uses'. Validity is, therefore, not a property of tests, or even of test outcomes, but a property of the inferences made on the basis of these outcomes (Wilian, 1998).

This assertion is also expressed in the definition of validity by Messick (1989: 13):

Validity is an integrative evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment.

Sometimes the demands of a test may measure something other than what the test claims to be measuring, which threatens the validity of the claims made. For example, a science test item may contain text that has linguistic features which are synonymous with scientific writing. These may be complex sentences, density of information, subordinate clauses, unfamiliar words and ambiguous phrases. A study conducted by Ramnarain (2012) revealed that the complexity of science text places a great demand upon the linguistic proficiency of learners. 'Variable' is a word that is seldom used in the everyday language of students, and this study reported cases where learners misinterpreted the word to mean some action that needs to be taken to ensure the validity of the experimental results. In science, the term 'variable', in fact, refers to a factor or condition that is subject to change, especially one that is allowed to change in a scientific experiment to test a hypothesis. Dempster and Reddy (2007: 920) point out that readability measures do not prove to be reliable predictors of learners' choosing the correct answer, and the problems of readability of items 'overlie a lack of knowledge, skills, and reasoning ability in science'. Consequently, the validity of assessment of students' knowledge and skills embodied in the science curriculum becomes questionable.

Stobart (2001) presents a validity framework of items in national curriculum assessments in the UK. A key concept in establishing the validity of assessment items is construct validity. According to Stobart (2001: 167), 'clarity about what is being assessed and how well a test samples this construct are critical to validity'. Construct validity refers to how adequately an assessment reflects the full range of outcomes of learning in a particular subject domain (Harlen, 2013). The accuracy of the results as a measure of the construct will depend on how the assessment is administered as well as what it contains. A threat to construct validity is that a paper-and-pencil item may measure something other than what it claims to measure. The following questions may be used in an inquiry into the construct validity of an assessment: What is being assessed? Does the assessment do what it claims to do? Stobart lists the following as potential threats to construct validity: contested construct; unclear construct; and construct irrelevance. A threat to construct validity may be evident when a question relates to a construct, but the inferences made from learner results on this construct are contested. This threat is referred to as a contested construct. An unclear construct is perceived as a threat to construct validity when the construct featured in the question is not well defined and is open to multiple interpretations. An assessment is denounced for construct irrelevance when it claims to measure certain skills or knowledge but, instead, measures something else.

In this article, I invoke construct validity in my analysis of inquiry items in national Grade 12 Physical Sciences examinations.

Method

My analysis focused on the Physics and Chemistry papers of the Grade 12 Physical Sciences examinations of 2010, 2011 and 2012. I firstly identified all inquiry-related questions in these six papers and compiled these into a set. I was guided in this process by an assessment standard for Learning Outcome 1 of the NCS that addresses inquiry in the Physical Sciences curriculum. This standard specifies the following stages in inquiry that learners should be assessed on: identify and question phenomena; design/planning of an investigation; drawing graphs; arriving at results; and drawing a conclusion (Department of Education, 2003). It is elaborated in the NCS that in 'identify and question phenomena the learner can be examined on formulating an investigation question, listing all possible variables, and formulating a testable hypothesis. In examining for the 'design/planning of an investigation' learners can be asked to identify variables (dependent, independent and controlled), plan the sequence of steps, and suggest an appropriate method of recoding results. With regard to 'drawing graphs', learners can be examined on drawing accurate as well as sketch graphs from the given information. In 'arriving at results' learners can be questioned on identifying patterns in the data and interpreting results. Finally, learners can be examined on 'drawing a conclusion' from the information given graphically. All questions that addressed these stages was classified as inquiry related and identified for further analysis. I sought validity in this process of identifying the inquiry questions by asking two researchers in science education to independently do the same. There was 94% agreement amongst the three of us on this classification. We resolved the small discrepancy through discussion.

Subsequently, I analysed the items for construct validity. I was guided in this analysis by the threats to construct validity identified by Stobart. I did this by reading the items and then checked to see whether they exhibited any of the threats to construct validity. Then, the items were classified in terms of whether they conformed to construct validity. The set of inquiry items was then analysed by the same two researchers in science education who were involved in the compilation of this set. Inter-rater reliability in the classification of the questions was established due to a 90% agreement in the classification. Again, differences in the classification of the questions were resolved through discussion.

Findings

Table 1 below informs on the extent to which the inquiry items in the Physical Sciences examinations violated construct validity. The Physics and Chemistry papers for each examination are considered jointly.

Table 1: Classification of inquiry questions according to construct validity

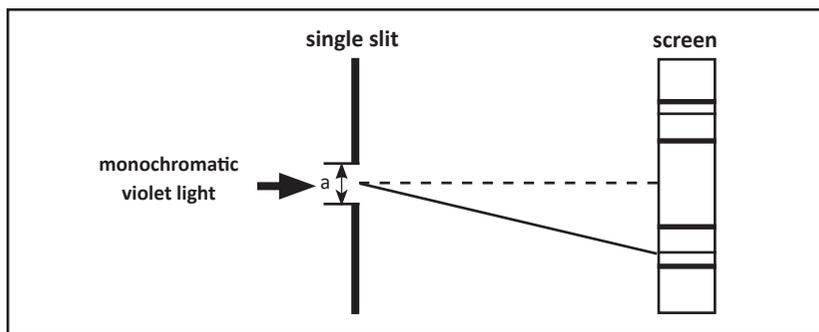
	Number of inquiry questions	Number of inquiry questions violating construct validity			Percentage of inquiry questions violating construct validity (%)
		Contested construct	Unclear construct	Construct irrelevance	
November 2010 examination	8	4	1	0	63
November 2011 examination	9	3	2	1	67
November 2012 examination	11	4	2	2	73
Total	28	11	5	3	68

It is evident from this table that there are a substantial number of the inquiry items that exhibited the threats to construct validity identified by Stobart. In particular, there was a relatively high prevalence of questions that were classified as ‘contested construct’. These were questions where the inferences that could be made on the target construct were open to challenge.

The following cases are presented to illustrate the threats to construct validity reported above.

Contested construct

In question 7.1 of the 2011 Physics examination learners were required to ‘Write down the investigative question’ for an investigation on the “change in broadness of the central bright band in a diffraction pattern when light passes through single slits of different widths’. This question relates to the stage ‘Identify and question phenomena’ in an assessment standard of Learning Outcome 1 of the NCS (Department of Education, 2003). It is specified here that learners can be examined on formulating an investigation question. The following diagram of the apparatus for the experiment was provided:



The marking memorandum provided by the Department of Basic Education gives the following criteria in the allocation of marks:

<i>Criteria for investigative question</i>	<i>Mark</i>
The dependent and independent variables are stated	1
Asks a question about the relationship between dependent and independent variables	1

The target construct that is being tested in this question is the learner’s ability to ‘formulate an investigation question’. The formulation of an investigation question is part of the problem-finding or problem-posing phase, and the question drives the subsequent investigative process. This is an ‘initial phase of problem-solving involving the construction of an internal, mental representation of the problem using existing schemata perceived as relevant by the problem solver’ (Appleton, 1995: 383). If this is the primary purpose of asking learners to formulate investigation questions, the task of asking learners to write down a question does not meet this objective. The investigation is based on the topic of diffraction and refers to the relationship between the width of the central band in this diffraction pattern and the width of the slit through which light passes. It is clear that this topic would already have been dealt with in class and the envisaged outcome would be that learners would have acquired an understanding of diffraction and the relationship between central band width and slit width. The target construct of asking learners to formulate an investigation question in driving an investigation is contested and, hence, this question poses a threat to the inferences that can be made from inquiry performance.

A similar threat to validity is reflected in question 7.2 where learners were required to ‘Write down TWO variables that are kept constant’ during this investigation. The identification of variables is a process skill that is indicated in Learning Outcome 1 of the NCS and also specified in an assessment standard for this outcome (Department of Education, 2003). The listing of variables is key in the construction of a hypothesis and investigation questions in fair testing investigations. Owing to a prior conception of the diffraction phenomenon, the validity of the inferences based on the learner performance through the question is threatened.

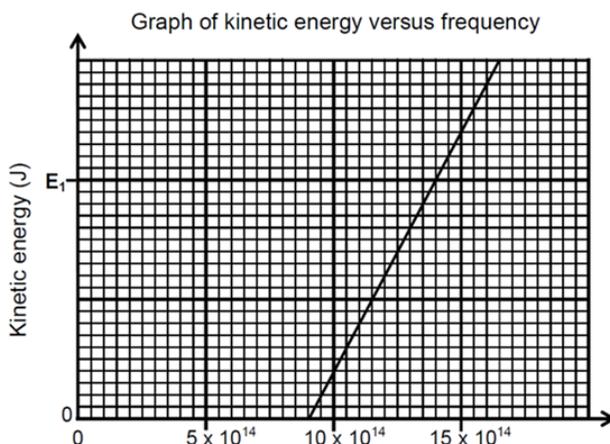
Another case where the purpose of the assessment is questionable is question 4 of the November 2012 Chemistry examination. The question refers to a practical investigation on the boiling points of alkanes and size of molecules. A table of results is presented.

ALKANE	MOLECULAR FORMULA	BOILING POINT ($^{\circ}\text{C}$)
Methane	CH_4	-164
Ethane	C_2H_6	-89
Propane	C_3H_8	-42
Butane	C_4H_{10}	-0.5
Pentane	C_5H_{12}	36
Hexane	C_6H_{14}	69

In question 4.2.3, learners are asked to write down a conclusion that can be drawn from the results. The examination guidelines state that learners should be able to 'Draw conclusions from information' (Department of Education, 2010: 3). This question is supposedly assessing the ability of learners to analyse the data and draw a conclusion based on this analysis. This is not an authentic task, as learners should already have a conception of the relationship and be able to correctly answer it by merely recalling their knowledge. As a result, the validity of inferences that can be based on learners' performance becomes questionable.

Unclear construct

In question 11 of the 2012 Physics examination, an investigation is described on the relationship between the light of different frequencies shone onto a metal cathode and the kinetic energy of the emitted electrons. A graph of results obtained is presented.

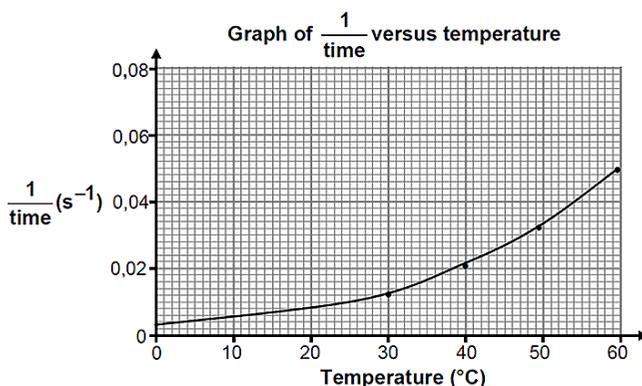


In question 11.1, learners are asked to write down the dependent, independent and controlled variables. As indicated already, the identification of a variable is a skill that is related to the broader skill of formulating a hypothesis or an investigation question in the planning stage of an inquiry. In drawing a graph on the relationship between dependent and independent variables it is the accepted practice in science to label the x-axis the independent variable and the y-axis the dependent variable. In this question learners can quite conceivably correctly identify these variables from the labelling of the two axes. If the intended construct for which learner performance is being measured is 'to identify variables in an investigation' this would require learners to engage conceptually with the inquiry problem. Instead, learners in this case can correctly answer this question without conceptual understanding by simply referring to the physical quantities indicated on the axes. It is, therefore, unclear what construct is being tested through this question. In terms of the Stobart framework, the unclear construct that is evident here is a threat to construct validity, and the validity of inferences that could be made on learners' performance on the target construct is being threatened. A more valid assessment would be made if learners are assessed on this skill when presented with a problem that is located in a meaningful context, and then asked to identify variables and formulate a question and a hypothesis on this. An example of this could be a teacher who demonstrates the photoelectric effect using a photocell, and then learners are asked to plan investigations related to this effect. This adds authenticity to the inquiry, and will give learners some insight into the nature of science at the same time.

Construct irrelevance

Question 6.2 of the 2010 November Chemistry examination paper refers to the investigation of the rate of reaction between hydrochloric acid and sodium thiosulphate, and temperature.

Hydrochloric acid and a sodium thiosulphate solution are used to investigate the relationship between rate of reaction and temperature. Learners are presented with the experimental procedure for this investigation and a graph of results is presented.



Sub-question 6.2.1 required learners to 'write a possible hypothesis for the investigation'. This is again a requirement in assessing inquiry where it is stated that learners should be able to 'Formulate a testable hypothesis' (Department of Education, 2010: 3). A scientific hypothesis is a proposed explanation of a phenomenon which still has to be tested. Most of the time a hypothesis is written like this: If ____ [I do this] ____, then ____ [this] ____ will happen. In view of the results being presented in the graph above, the target construct of hypothesising is irrelevant and misplaced here. The graph of results already reveals the relationship between temperature and rate of reaction. It is, therefore, nonsensical to ask learners to now formulate a hypothesis on this relationship.

Discussion

This study revealed that inquiry questions in national Physical Sciences examinations lack construct validity. The analysis that was guided by Stobart's conceptualisation of construct validity revealed that, to a large extent, the inquiry questions exhibited threats to validity. The inferences that could be made on learners' performance in inquiry were found to be misleading due to the inquiry construct's being contested. In certain cases it was not clear what construct was being targeted in the question and, as a result, no valid inferences could be made on learner performance. There were also questions where the supposed construct that was being targeted was not really being addressed, and displayed construct irrelevance.

The findings of this study suggest that greater attention needs to be paid to the formulation of inquiry-related questions in written tests and examinations. As with any pedagogical approach, it is important to align learning outcomes, and teaching and learning activities with assessment. A major reason why inquiry items have been incorporated in written examinations is to incentivise the teaching of inquiry at school. It is common practice that teachers will 'teach to the test' (Phelps, 2011), which means that teachers place heavy emphasis on preparing learners for a standardised test. 'Teaching to the test' does have a negative connotation in education, but when the assessment tasks reflect accurately the constructs inherent to the learning outcomes then 'teaching to the test' is appropriate (Hein & Lee, 2000). Hein and Lee also affirm that, if assessment criteria for inquiry are shared with learners, this may encourage learners to practise what will be assessed, leading to an improvement in achievement.

Inquiry is a complex and multifaceted activity involving both cognitive and physical activity, and paper-and-pencil items do not provide the authentic context for this assessment. As pointed out already, it is preferable to assess inquiry skills in the same context in which these skills are developed (Mislevy *et al.*, 2003; Quellmalz & Pelligrino 2009; Ruiz-Primo & Shavelson, 1996). However, the implementation of standardised testing of inquiry in a school context is faced with challenges such as the lack of physical resources, the large classes and the related organisational difficulties.

Given this scenario, paper-and-pencil items, although not the ideal, may still prove to be the most feasible way to do standardised testing of inquiry. It is, therefore, recommended that further research be undertaken on the development of such written tasks that adhere to validity requirements. This recommendation is affirmed by delegates of the IAP conference in Helsinki referred to earlier who suggest that more studies are needed in order to improve the validity of assessment tools so that unrealistic assumptions of accuracy in assessment can be deterred (Harlen, 2013).

References

- Abd-El-Khalick F, BouJaoude S, Duschl RA, Hofstein A, Lederman NG, & Mamlok R 2004. Inquiry in science education: International perspectives. *Science Education*, 88(3): 397-419.
- American Association for the Advancement of Science 1993. *Benchmarks for science literacy*. New York: Oxford University Press.
- Appleton K 1995. Student teachers' confidence to teach science: Is more science knowledge necessary to improve self-confidence? *International Journal of Science Education*, 19: 357-369.
- Baxter G & Shavelson R 1994. Science performance assessments: Benchmarks and surrogates. *International Journal of Education Research*, 21(3): 279-298.
- Buckley BC, Gobert JD, Horwitz P & O' Dwyer L 2010. Looking inside the black box: Assessing model-based learning and inquiry in BioLogica. *International Journal of Learning Technologies*, 5(2): 166-190.
- Dempster ER & Reddy V 2007. Item readability science achievement in TIMSS 2003. *Science Education*, 91: 906-925.
- Department for Education and Employment 1999. *The national curriculum for England: Key stages 1-4*. London: Qualifications and Curriculum Authority.
- Department of Basic Education 2011. *Curriculum and assessment policy statement: Grades 10-12 Physical Sciences*. Pretoria: Government Printer.
- Department of Education 2003. *National curriculum statement grades 10-12: Physical Sciences*. Pretoria: Government Printer.
- Department of Education 2010. *Examination Guidelines: Physical Sciences Grade 12*. Pretoria: Government Printer.
- Harlen W 2013. *Assessment & inquiry-based science education: Issues in policy and practice*. Global Network of Science Academies (IAP) Science Education Programme (SEP): Trieste, Italy.
- Hein GE & Lee S 2000. Assessment of science inquiry. In National Science Foundation (ed.), *Foundations inquiry: Thoughts, views, and strategies for the k- 5 classroom* (2: 99-108). Arlington, VA: National Science Foundation.
- Inter-Academy Panel 2012. Taking inquiry-based science education into secondary education. Report of a global conference. Retrieved from [2 February 2013] at <http://www.sazu.si/files/file-147.pdf>

- Kane MT 2006. Validation. In RL Brenman (ed.), *Educational measurement*. Westport, CT: American Council on Education/Praeger.
- Ketelhut DJ, Clarke J, Dede C, Nelson B & Bowman C 2005. *Inquiry teaching for depth and coverage via multi-user virtual environments*. Paper presented at the National Association for Research in Science Teaching, Dallas.
- Messick S 1989. Validity. In R Linn (ed.), *Educational Measurement* (3rd ed.). Washington: American Council on Education and Macmillan.
- Mislevy R, Chudowsky N, Draney K, Fried R, Gaffney T & Haertel G 2003. *Design patterns for assessing science inquiry*. Menlo Park, CA: SRI International.
- National Research Council 1996. *National science education standards*. Washington: National Academy Press.
- National Research Council 2000. *Inquiry and the national science education standards: A guide for teaching and learning*. Washington: National Academy Press.
- Phelps RP 2011. Teach to the test. *The Wilson Quarterly*, 35(4): 38-42.
- Quellmalz ES & Pelligrino JW 2009. Technology and testing. *Science*, 32: 75-79.
- Ramnarain, U. (2012). The readability of a high stakes physics examination. *Acta Academica*, 44(2): 110-129.
- Resnick LB & Resnick DP 1992. Assessing the thinking curriculum: New tools for education reform. In B Gifford & M O' Connor (eds), *Changing assessment: Alternative view of aptitude, achievement and instruction*. London: Kluwer Academic Publishers.
- Roth W-M 1994. Experimenting in a constructivist high school laboratory. *Journal of Research in Science Teaching*, 31(2): 197-223.
- Ruiz-Primo MA & Shavelson RJ 1996. Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, 33(6): 569-600.
- Stobart G 2001. The validity of national curriculum assessment. *British Journal of Educational Studies*, 49(1), 26-39.
- Wiliam D 1998. *The validity of teachers' assessments*. Paper presented at the 22nd annual conference of the International Group for the Psychology of Mathematics Education, Stellenbosch, South Africa, July 1998.