Assessment of students' conceptual understandings in science: The Taiwan National Science Concept Learning Study in an international perspective

David F Treagust and A L Chandrasegaran, Science and Mathematics Education Centre, Curtin University of Technology, Perth, Australia; e-mail: D.F.Treagust@curtin.edu.au

Over the past two decades, several two-tier multiple-choice diagnostic tests in science at a variety grade levels have been reported in the extant literature from several countries. This review summarises these studies, provides an overview of the methodology for their construction, presents two examples of typical items and lists studies published with two-tier tests. The majority of these studies have been conducted with relatively small numbers of students in one school, college or university, though there are exceptions. What distinguishes the research reported in the articles in this Special Issue is that the Taiwanese two-tiered tests for the National Science Concept Learning Study were designed to identify the scientific conceptions of students from large national samples using random sampling procedures similar to those found in studies such as TIMSS and PISA. The article concludes with implications for the use of two-tier tests for pedagogical practice.

Introduction

The nature and extent of students' understanding of scientific concepts and phenomena are key components of any science curriculum. The main emphasis of classroom instruction should, therefore, be to facilitate meaningful learning rather than mere rote-learning of facts that promotes higher levels of understanding of scientific concepts among learners. The effectiveness of classroom instruction to facilitate students' understanding of scientific concepts can be measured by appropriate assessment tools made available for use by classroom teachers. This article addresses the importance of assessment in science learning and presents an overview of the development of two-tier multiple-choice diagnostic instruments that have been reported in the science education research literature since the 1980s and considers the contribution of the Taiwanese National Science Concept Learning Study in this area.

Students' alternative conceptions and the importance of diagnostic assessment

Research data collected over more than three decades has shown that the majority of students come to science classes with pre-instructional knowledge or beliefs about the phenomena and concepts to be taught and that many students develop only a limited understanding of science concepts following instruction (Duit & Treagust, 2003). These students construct what to them are sensible and coherent understandings of phenomena and concepts that do not match universally accepted views held by the scientific community. The resulting misunderstandings or alternative conceptions, if not challenged, become integrated into students' cognitive structures and can interfere with subsequent learning. As a consequence, students may experience difficulty in integrating any new information within their cognitive structures, resulting in an inappropriate understanding of new concepts.

The large body of research literature on students' understanding of science concepts has been categorised, synthesised and summarised by Duit (2006) in a manner that enables researchers to gain a holistic understanding of the field. An important task is to demonstrate how teachers can address students' learning needs by incorporating in their instructional repertoires specially designed assessment procedures. This pedagogical direction is essential if science education research is going to benefit students in their science learning because other research has shown that the majority of teachers do not effectively diagnose students' learning problems, especially at an early stage of the student learning process (see for example, Costa, Marques & Kempa, 2000; Taber, 2001; Widodo & Duit, 2002).

Science teachers' pedagogy can be made more effective by using diagnostic formative assessment methods (Bell & Cowie, 2001). Indeed, current assessment procedures can distort and narrow instruction, thereby misrepresenting the nature of the subject, and maintaining inequities in access to education (Wolf et al., 1991) and are claimed to not provide valid measures of what students know and to provide no opportunity for students and teachers to be involved in discussions about the work being assessed (see for example, Black & Wiliam, 1998). Supporters of alternative approaches to assessment have not specifically elaborated on the value of specially created diagnostic tests but have recommended assessment items that "require an explanation or defence of the answer, given the methods used" (see Wiggins & McTighe, 1998, p. 14) – precisely the information required in the second tier of two-tier test items.

These alternative forms of assessment are different to those generally used by science teachers in that standard tests are largely paper-and-pencil collections of individual items with single correct answers presented without a surrounding context. A change in direction from those assessment procedures generally used by teachers is highly recommended (see Wiggins & McTighe, 1998). The Programme for International Student Assessment (OECD/PISA, 2001), initiated by the Organisation for Economic Cooperation and Development, uses items that are presented in a context and requires students to respond to questions contained in the information presented in the question.

Development of two-tier multiple-choice test items

Tamir (1989) found the use of justifications when answering multiple-choice test items to be a sensitive and effective way of assessing meaningful learning among students and addresses, to some extent, the limitations of traditional multiple-choice test items. As a result, he proposed the use of multiple-choice test items that included responses with known student alternative conceptions, and that also required students to justify their choice of option by giving a reason (Tamir, 1971).

The positive outcomes of findings related to students' justifications to test items led to the development of two-tier multiple-choice test items specifically for the purpose of identifying students' alternative conceptions in limited and clearly defined content areas (Treagust, 1988). These paper and pencil tests are convenient to administer and not time consuming to mark. In brief, there are three major aspects to developing these items: (a) the content is defined by the identification of propositional content knowledge statements of the topic to be taught and the development of a concept map that accommodates the propositional statements; (b) information about students' conceptions is obtained from the extant research literature, where available, and where not available by having students provide free response explanations to their answers and conducting unstructured interviews with students who have previously been taught the content/concepts, and (c) the development of the two-tier multiple-choice diagnostic items. The first tier of each multiple-choice item consists of a content question having usually two to four choices. The second tier of each item contains a set of usually four possible reasons for the answer given to the first part.

The reasons consist of the designated correct answer, together with identified students' conceptions and/or misconceptions. The reasons are from the students' responses given to each open response question as well as information gathered from the interviews and the literature. When more than one alternative conception is given, these are included as separate alternative reason responses. Students' answers to each item are considered to be correct when both the correct choice and correct reason are given. Finally, a specification grid is designed to ensure that the diagnostic instrument fairly covers the propositional knowledge statements and the concepts on the concept map underlying the topic. A flow diagram illustrating the various steps in the development of two-tier multiple-choice diagnostic instruments along the lines proposed by Treagust (1988, 1995) is shown in Figure 1. Essentially the authors of the articles in this Special Issue used this methodology and which is not repeated in detail in each of the articles.

Place Figure 1 about here

Two studies conducted in Singapore classrooms serve as examples of item design. Even in an examination-dominated curriculum as exists in Singapore schools, two-tier diagnostic tests can be very useful in helping both teachers and students achieve the expected learning outcomes. In the first study, Tan and Treagust (1999) were interested in 14-16 year olds studying chemical bonding. An earlier test designed by Peterson, Treagust and Garnett (1989) was deemed too difficult for this age group so the first tier response was made relatively easy with a True-False choice but the second tier still probed deeply an understanding behind the first tier response. An example is shown in Figure 2.

The second study by Chandrasegaran, Treagust and Mocerino (2005) sought to identify the main alternative conceptions held by Grade 9 students when describing and explaining chemical reactions using multiple representations. The effectiveness of a planned program of instruction using multiple representations to enhance students' ability to describe and explain some frequently encountered chemical reactions was subsequently evaluated. An item from this study is shown in Figure 3.

Place Figures 2 and 3 about here

As shown in Table 1, several two-tier multiple-choice diagnostic tests have been developed in a variety of content domains and are described in the extant literature. These studies are typically conducted with relatively small numbers of students in one school, college or university. There are exceptions such as the work of Birk and Kurtz (1990) who examined conceptions of molecular structure and bonding (using the instrument of Peterson et al. 1989) with a large number of participants from high school, different levels of college and university chemistry faculty. The studies shown in Table are designed to examine alternative conceptions with the data being able to be used in the teachers' classrooms or claims made that these data can be used in future occasions to inform practice. The relevant bibliographic references to these studies are provided in the Appendix 1.

Place Table 1 about here

Taiwanese research using two-tier diagnostic tests

As described by Chiu, Gou and Treagust (this issue), a series of workshops held over the length of the NSCL study with numerous participants ensured a systematic and thorough development of the research goals. Treagust was involved early in one of these workshops in September 2000, providing an overview of the approach to developing two-tier multiple choice diagnostic test items. What distinguishes the research reported in the articles in this Special Issue is that the Taiwanese two-tiered tests were designed not for diagnostic formative assessment but to identify the scientific conceptions of students from a large national sample with random sampling procedures found in studies using large data collecting procedures such as TIMSS and PISA (see Tam in this issue). These data from Taiwanese students were collected to gauge performance at elementary/primary, junior high school and senior high school, as well as compare male and female students (for example, see Wang et al.), rural and urban students (for example, see Wang et al.) and aboriginal and non-aboriginal students (see Kao) in a wide range of chemistry, physics and biology topics. In addition, these studies examined correlations of outcomes on the diagnostic tests such as biology self-efficacy (for example, see Wang et al.). In order to place the findings from the Taiwanese students in an international context, the articles also refer to findings of related concepts from other studies (not necessarily with two-tier tests) in a range of countries.

The result is a large number of additional tests (now translated into English) that can be used by other researchers either for small classroom studies or with larger groups. Cross comparisons across nations may also be possible and be a useful avenue for future research collaboration. This research initiative is a significant contribution by Taiwanese researchers to the development of two-tier diagnostic tests. As previously stated, these two-tier multiple-choice tests are readily administered and can be easily scored to ascertain students' understanding, and thus are particularly useful for classroom teachers enabling them to use the findings of research to inform their teaching – a principal desired outcome of formative assessment. However, the contribution by Taiwanese researchers has enhanced this research agenda by illustrating how these tests can be used in a program of national assessment that includes not only science concept learning but also allows correlation with factors that impact upon or influence science concept learning.

Implications of the use of two-tier tests for pedagogical practice and research

The use of two-tier diagnostic tests discussed in this article and described in this Special Issue can help to address concerns about current assessment practices by overtly assessing the outcomes of thinking about scientific concepts within a specified context rather than assessing knowledge of information. By using these diagnostic instruments at the beginning or on completion of a specified topic, science teachers can achieve better understanding about the nature of students' understanding and the existence of any alternative conceptions or misconceptions in a particular topic being studied. Once students' alternative conceptions are identified, science instruction can be modified by incorporating conceptual change teaching approaches (see Duit & Confrey, 1996) that can help challenge students' thinking by developing and/or utilising alternative teaching approaches that specifically address students' non-scientifically acceptable conceptions. In the diagnostic instrument on qualitative analysis, for example, students had difficulty understanding the concepts of oxidation and reduction (Tan et al., 2002), as at least three models of redox reactions are commonly encountered in a chemistry course (de Jong & Treagust, 2002). Teachers need, therefore, to pay more attention to clarifying these models during instruction so that students will be in a position to confidently use them to decide if a chemical reaction is a redox reaction.

Indeed, students' conceptions identified by two-tier diagnostic instruments are known to many experienced teachers but are only appreciated by less experienced teachers after instruction has been completed. In the latter instance, there is no possibility to consider students' conceptions of the phenomena and to incorporate these in the teaching process. However, research evidence also suggests that even experienced teachers frequently do not appreciate the problems encountered by students in learning complex science concepts (Widodo & Duit, 2002). There are two reasons for this. First, regular approaches to instruction do not probe sufficiently for reasoning of answers; second, the usual assessment procedures do not demand such detailed explanations of concepts. The importance of the national Taiwanese data using these diagnostic tests is that they can be generalised across the nation. Consequently, there is much less likelihood for teachers to believe that alternative conceptions identified in responses from students in one teacher's class or one school will be rejected and not considered as being relevant to their classes.

The significance of the data reported in this Special Issue is to provide a national overview of Taiwanese students' understanding in a range of science subjects and underlying concepts. In this way, although using a different assessment system, the data can be used in a similar way to the data from TIMSS and PISA to identify directions for future developments in science education. This issue is certainly not trivial as in several countries there is much concern about the low participation rate of students taking science, especially in higher levels of secondary school education (OECD Global Science Forum, 2006). In particular, Australian enrolments are on the decline in the more conceptually demanding calculus-based options that lead to acceptance in university science and engineering courses (Dobson, 2003; Hassan & Treagust, 2003) and there are enrolment declines in countries such as Great Britain and France (Nature, 2002). Indeed, the success, and even the continuation, of science programs at university are dependent on foundational improvements in science education in primary and secondary schools.

A major challenge for the science education community is how to arrest this decline in enrolments in science subjects. One way of encouraging more students to study science is by presenting science in such a way that is understandable and interesting and it is our contention that formative assessment can play a key role. For example, by using multiple-choice diagnostic items, students can begin to question and understand the underlying science concepts based on their responses to the questions. Through formative assessment incorporated into teaching, students are encouraged to think about the concepts and consider alternative explanations rather than memorise basic facts for a test or examination which are then forgotten.

Knowing what can work and how to implement it are two different issues. Although there is evidence from several studies that formative assessment using tests like the ones in this Special Issue can increase students' comprehension of concepts, there is an extensive body of knowledge that teachers' knowledge and awareness of the findings of science education research in general is still very limited (Gilbert et al., 2004). Those teachers who find the time to be involved in research studies are likely to be early advocates. Others do not have the time or means of accessing science education research journals and even when these research outcomes are available there is a considerable time lapse before teachers become aware of the research findings.

There is, therefore, a need for initiatives that will facilitate speedy communication of science education research findings to the classroom. For example, the Evidence-based Practice in Science Education (EPSE) project coordinated by Millar (2003) in the United Kingdom has resulted in the development of several formative diagnostic instruments consisting of various types of items including two-tier multiple-choice items and others requiring students to provide reasons for their answers to particular items. These instruments are available on the internet. In the context of this article, by making information on the findings about the alternative conceptions related to particular concepts available on CD-ROMs or the internet, for example, teachers may have a better understanding of students' learning difficulties in those concepts. The ready availability of this information has the potential to be of invaluable assistance to teachers in the planning and implementation of classroom instruction.

In addition, at a research level, the data from the Taiwanese National Science Concept Learning Study can be the beginning of a concerted examination of the curriculum and teaching approaches using the model of educational reconstruction (Kattmann, Duit, Gropengießer, & Komorek, 1995) which comprises an analysis of content structure, empirical investigations of student's understanding of science content and the construction of instructional modules. Taken to a classroom level, the inclusion of relevant multiple-choice diagnostic instruments for administration by the teacher can provide feedback on the type of follow-up remediation that needs to be undertaken. The success of any such initiative will, however, depend on strong links between academic researchers and classroom practitioners.

References

Bell, B., & Cowie, B. (2001). *Formative assessment and science education.* Dordrecht, The Netherlands: Kluwer.

Birk, J. P. & Kurtz, M. J. (1999). Effect of experience on retention and elimination of misconceptions about molecular structure and bonding. *Journal of Chemical Education*, 76, 124-128.

Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5(1), 7-74.

Costa, N., Marques, L. & Kempa, R. (2000). Science teachers' awareness of findings from educational research. *Chemistry Education: Research and Practice in Europe,*1(1), 31-36.

De Jong, O., & Treagust, D. F. (2002). The teaching ad learning of electrochemistry. . In Gilbert, J. G., De Jong, O., Justi, R. Treagust, D. F. & van Driel, J. H. (Eds.). *Chemical education: Towards research based practice* (pp. 317-338). Dordrecht, The Netherlands: Kluwer

Duit, R. (2006). Bibliography: Students' alternative frameworks and science education (IPN Reports-in-Brief). University of Kiel. http://www.ipn.uni-kiel.de/aktuell/stcse/stcse.html Retrieved 12.09.06
Duit, R. & Confrey, J. (1996). Reorganising te curriculum and teaching to improve learning in science and mathematics. In Treagust, D.F., Duit, R. & Fraser, B.J. (1996). *Teaching and learning of science and mathematics* (pp. 79-93). New York: Teachers College Press.
Duit, R. & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education*, 25(6), 671-688.
Dobson, I. R. (2003). *Science at the crossroads? A study of trends in university science from Dawkins to now – 1989-2002.* (Melbourne, Vic: Monash University). (A study commissioned by the Australian Council of Deans of Science).
Garnett, P.J., & Treagust, D. F. (1992). Conceptual difficulties experienced by senior high school students of electrochemistry: Electrochemical (galvanic) and electrolytic cells. *Journal of research in Science Teaching,* 29, 1079-1099.
Gilbert, J. K., Justi, R., van Driel, J. H., de Jong, O., & Treagust, D. F. (2004). Securing a future for chemical education. *Chemistry Education: Research and Practice, 5*, 1, 5-14.
Hassan, G. & Treagust, D. F. (2003). What is the future of science education in Australia? *Australian Science Teachers Journal,* 49(3), 6-15.
Kattmann, U., Duit,R., H Gropengießer,H., & M Komorek, M. (1995). *Educational reconstruction – bringing together issues of scientific clarification and students' conceptions* Paper presented at the annual meeting of the National Association for Research on Science Teaching, St Louis, Mo, April.
Millar, R. (2003, Jan). *EPSE teaching resources* [internet]. http://www.york.ac.uk/depts/educ/projs/DiagQuestionsIntro_Dec02.htm. Retrieved 5.9.06, from the World Wide Web.
Nature (2002). Selling science to the young. *Nature,* 417, 1.
OECD Global Science Forum (2006). *Evolution of Student Interest in Science and Technology Studies Policy Report* [internet]. http://www.oecd.org/dataoecd/16/30/36645828.pdf. Retrieved 12.9.06, from the World Wide Web.
Peterson, R. F., Treagust, D. F. & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching,* 26, 301-314.
Taber, K. S. (2001). Constructing chemical concepts in the classroom: Using research to inform practice. *Chemistry Education: Research and Practice in Europe,* 2(1), 43-51.
Tamir, P. (1971). An alternative approach to the construction of multiple-choice test items. *Journal of Biological Education,* 5, 305-307.
Tamir, P. (1989). Some issues related to the use of justifications to multiple-choice answers. *Journal of Biological Education,* 23(4), 285-292.
Tan, D. K-C. & Treagust, D. F. (1999). Evaluating students' understanding of chemical bonding. *School Science Review,* 81, 75-83.
Treagust, D. F. (1988). The development and use of diagnostic instruments to evaluate students' misconceptions in science. *International Journal of Science Education,* 10, 159-169.

Treagust, D. F. (1995). Diagnostic assessment of students' science concepts. In S. Glynn and R. Duit (eds.) *Learning science in the schools: Research reforming practice* (New Jersey: Lawrence Erlbaum Associates), 327-346.

Widodo & Duit, R. (2002). Conceptual change views and the reality of classroom practice. In S. Lehti & K. Merenluoto (Eds.), Proceedings of the Third European Symposium on Conceptual Change: A Process Approach to Conceptual Change (pp. 289-297). Turku, Finland: University of Turku, Faculty of Education.

Wiggins, G. & McTighe, J. (1998). *Understanding by design.* (Alexandria, VA: Association for Supervision and Curriculum Development).

Wolf, D., Bixby, J., Glen III, J. & Gardner, H. (1991). To use their minds well: Investigating new forms of student assessment. In G. Grant (ed.) *Review of Research in Education* (Washington, DC: American Educational Research Association), 31-74.

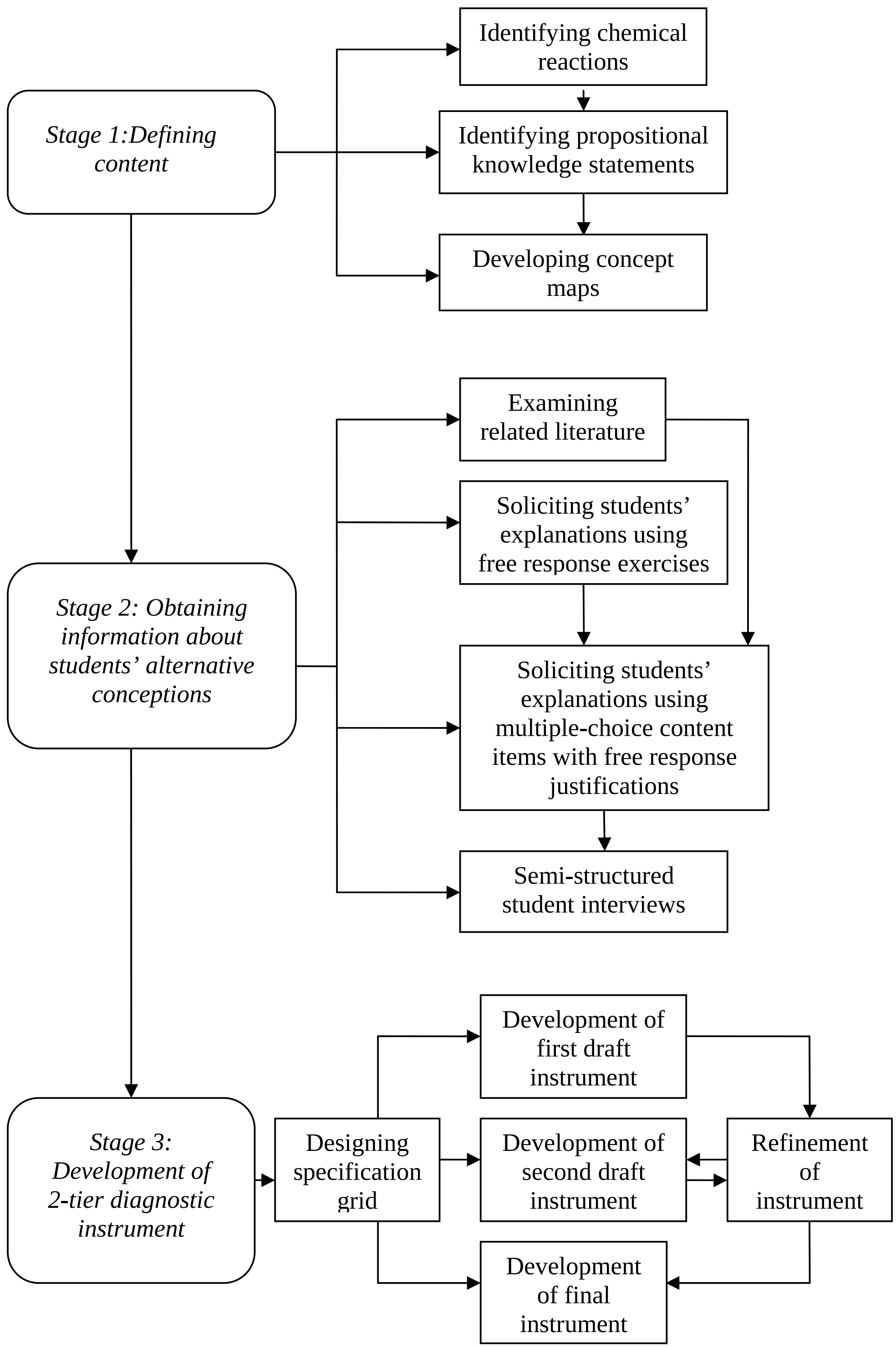

Figure 1. Steps in the development of two-tier multiple-choice diagnostic instruments based on the methodology proposed by Treagust (1988, 1995).

The compound formed between magnesium and oxygen can be used as a heat-resistant material to line the walls of furnaces.

I True II False

Reason

A. The lattice of magnesium oxide resembles that of silicon.

B. The covalent bonds between magnesium and oxygen atoms are strong.

C. The intermolecular forces between the magnesium oxide molecules are weak.

D. There are strong ionic forces between magnesium and oxide ions in the lattice.

Figure 2. Item 4 from the chemical bonding diagnostic instrument (Tan & Treagust, 1999)

Colourless aqueous solutions of lead(II) nitrate and potassium iodide, KI, are mixed together. A yellow powdery precipitate of lead(II) iodide, PbI_2, is immediately produced.
The chemical equation for the reaction is
$Pb(NO_3)_2(aq) \quad + \quad 2KI(aq) \quad \rightarrow \quad PbI_2(s) \quad + \quad 2KNO_3(aq)$.

Which of the following is the ***ionic equation*** for the reaction that has occurred?

A $\quad Pb^{2+}(aq) \quad + \quad 2I^{-}(aq) \quad \rightarrow \quad PbI_2(s)$

B $\quad K^{+}(aq) \quad + \quad NO_3^{-}(aq) \quad \rightarrow \quad KNO_3(aq)$

C $\quad Pb^{2+}(NO_3^{-})_2(aq) \ + \ 2K^{+}I^{-}(aq) \ \rightarrow \ Pb^{2+}(I^{-})_2(s) \ + \ 2K^{+}NO_3^{-}(aq)$.

The ***reason*** *for my answer is:*

1 Potassium nitrate is produced in the chemical reaction.

2 The K^{+} and NO_3^{-} ions remain unchanged in aqueous solution as 'spectator ions'.

3 All ions involved in the reaction must be represented in the ionic equation.

Figure 3. Item 12 from the multiple representations diagnostic instrument

(Chandrasegaran, Treagust & Mocerino, 2005)

Table 1 Summary of the development of diagnostic instruments since the 1980s

Topic/concept	Authors
Photosynthesis and respiration	Haslam and Treagust (1987)
Photosynthesis	Griffard & Wandersee (2001)
Diffusion and osmosis	Odom & Barrow (1995)
Breathing and respiration	Mann &. Treagust (1998)
Internal transport in plants and human circulatory system	Wang (2004)
Flowering plant growth and development	Lin (2004)
Covalent bonding and structure	Peterson, Treagust &. Garnett (1989)
Chemical bonding	Tan & Treagust (1999)
Qualitative analysis	Tan, Treagust, Goh & Chia (2002)
Chemical equilibrium	Tyson, Treagust & Bucat (1999)
Multiple representation in chemical reactions	Chandrasegaran, Treagust &. Mocerino (2005)
Ionisation energies of elements	Tan, Taber, Goh & Chia (2005)
Acids and bases	Chiu (2001, 2002)
States of matter	Chiu, Chiu & Ho (2002)
Light and its properties	Fetherstonhaugh & Treagust (1992)
Formation of images by a plane mirror	Chen, Lin & Lin (2002)
Forces	Halloun &. Hestenes (1985) Hestenes, Wells & Schwackhamer (1992)
Electromagnetism	Paulus & Treagust (1991)
Electrical circuits	Millar & Hames (2001)
Force, heat , light and electricity	Franklin (1992)

Appendix 1

Published and available two-tier tests in science content areas

Chandrasegaran, A. L., Treagust, D. F. & Mocerino, M. (2005, July). *Diagnostic assessment of secondary students' use of three levels of representation to explain simple chemical reactions.* Paper presented at the 36th annual conference of the Australasian Science Education Research Association (ASERA), Hamilton, New Zealand.

Chen, C-C., Lin, H-S. & Lin. M-L. (2002). Developing a two-tier diagnostic instrument to assess high school students' understanding – the formation of images by a plane mirror. *Proceedings of the National Science Council, ROC (D)*, 12, 3, 106-121.

Chiu, M. H. (2001). Exploring mental models and causes of high school students' misconceptions in acids-bases, particle theory and chemical equilibrium. *Project Report in National Science Council.*

Chiu, M. H. (2002). Exploring mental models and causes of high school students' misconceptions in acids-bases, particle theory and chemical equilibrium. *Project Report in National Science Council.*

Chiu, M. L., Chiu, M. H. & Ho, C. Y. (2002). Using cognitive-based dynamic representations to diagnose students' conceptions of the characteristics of matter. *Proceedings of the National Science Council Part D: Mathematics, Science and Technology Education*, 12(3), 91-99.

Fetherstonaugh, T. & Treagust, D. F. (1992). Students' understanding of light and its properties: Teaching to engender conceptual change. *Science Education,* 76(6), 653-672.

Franklin, B. J. (1992). The development, validation and application of a two-tier diagnostic instrument to detect misconceptions in the areas of force, heat, light and electricity. Unpublished PH.D. thesis, the Louisiana State University and Agricultural and Mechanical College.

Griffard, P.B. & Wandersee, J.H. (2001). The two-tier instrument on photosynthesis: what does it diagnose? *International Journal of Science Education*, 23, 10, 1039-1052.

Halloun, H. & Hestenes, D. (1985). The initial knowledge state of college physics students. *American Journal of Physics*,53, 1043-1055.

Haslam, F. & Treagust, D. F. (1987). Diagnosing secondary students' misconceptions of photosynthesis and respiration in plants using a two-tier multiple choice instrument. *Journal of Biological Education,* 21, 203-211.

Hestenes, D., Wells, M. & Schwackhamer, G. (1992). Force concept inventory. *The Physics Teacher*, 30, 141-158.

Lin, S-W. (2004). Development and application of a two-tier diagnostic test for high school students' understanding of flowering plant growth and development. *International Journal of Science and Mathematics Education*, 2, 175-199.

Mann, M. & Treagust, D. F. (1998). A pencil and paper instrument to diagnose students' conceptions of breathing, gas exchange and respiration. *Australian Science Teachers Journal*, 44(2), 55-59.

Millar, R. & V. Hames. (2001, August). *Using diagnostic assessment to improve students' learning in science: some preliminary findings from work to develop and test diagnostic tools.* Paper presented at the Third conference of the European Science education Research Association (ESERA), Thessaloniki, Greece.

Odom, A. L. & Barrow, L. H. (1995). Development and application of a two-tier diagnostic test measuring college biology students' understanding of diffusion and osmosis after a course of instruction. *Journal of Research in Science Teaching,* 32, 45-61.

Paulus, G. M. & Treagust, D. F. (1991). Conceptual difficulties in electricity and magnetism. *Journal of Science and Mathematics Education in South East Asia,* 14(2), 47-53.

Peterson, R. F., Treagust, D. F. & Garnett, P. (1989). Development and application of a diagnostic instrument to evaluate grade 11 and 12 students' concepts of covalent bonding and structure following a course of instruction. *Journal of Research in Science Teaching,* 26, 301-314.

Tan, D. K-C., Taber, K. S., Goh, N. K. & Chia, L-S. (2005). The ionisation energy diagnostic instrument: a two-tier multiple-choice instrument to determine high school students' understanding of ionisation energy. *Chemical Education Research and Practice,* 6(4), 180-197.

Tan, D. K-C. & Treagust, D. F. (1999). Evaluating students' understanding of chemical bonding. *School Science Review,* 81, 75-83.

Tan, D. K-C., Treagust, D. F., Goh, N-K. & Chia, L-S. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic qualitative analysis. *Journal of Research in Science Teaching,* 39(4), 283-301.

Tyson, L., Treagust, D. F. & Bucat (1999). The complexity of teaching and learning chemical equilibrium. *Journal of Chemical Education,* 35, 1031-1055.

Wang, J-R. (2004). Development and validation of a two-tier instrument to examine understanding of internal transport in plants and the human circulatory system. *International Journal of Science and Mathematics Education,* 2, 131-157.