

Diagnosing students' alternative conceptions in science

C.-C. Tsai & C. Chou

Center for Teacher Education and Institute of Education, National Chiao Tung University, Taiwan

Abstract This study described an attempt to develop a networked two-tier test system. A two-tier test is a two-level multiple-choice question that diagnoses students' alternative conceptions in science. Three networked, two-tier test items were presented in this study. Students in Taiwan (555 14 year olds and 599 16 year olds) were asked to answer these items online. An analysis of students' answers suggested that students' alternative conceptions might be retained even after formal instruction about relevant conceptions. Moreover, their responses were related across these three two-tier test items. Further development of the two-tier test system will mainly focus on designing appropriate feedback and guidance that help students overcome their alternative conceptions. In this way, the networked two-tier test system is not only a diagnostic tool, but also an effective instructional tool. This study has illuminated some innovative thoughts for the research and practice of science education.

Keywords: Assessment; Constructivist; Individual; Internet; Quantitative; School; Science; Two-tier test

Introduction

During the last 25 years, many science educators have believed that students' knowledge in domain-specific areas plays a more important role than their general cognitive ability or underlying logical structures on conceptual learning of science (Driver & Easley, 1978). Consequently, science education researchers have widely surveyed students' knowledge in various domains, known as students' 'misconceptions' or 'alternative conceptions' (see Wandersee *et al.*, 1994). This research trend facilitates the practice of constructivism in the field of science education since it is important to know what prior knowledge students bring to a learning environment in order to help them construct new knowledge (Tsai, 1998a; 2000a; 2000b). Researchers have developed methods to explore student alternative conceptions, for instance, interviews (Posner & Gertzog, 1982; Bell, 1995) and concept maps (Novak & Gowin, 1984). However, these methods often require additional training and commitments of time to conduct, analyse and make interpretations (Ruiz-Primo & Shavelson, 1996). So, a two-tier test, a pencil and paper test in a multiple-choice format, has been proposed by science educators to diagnose students' alternative conceptions (Treagust, 1988; Odom & Barrow, 1995).

Accepted 28 August 2001

Correspondence: Dr. Chin-Chung Tsai, Center for Teacher Education, National Chiao Tung University,
1001 Ta Hsueh Rd., Hsinchu 300, Taiwan
Email: cctsaic@cc.nctu.edu.tw

The use of two-tier tests allows teachers to not only understand students' scientifically incorrect ideas, but also to explore students' reasoning behind these ideas. Moreover, it facilitates assessment of alternative conceptions of a larger sample of students in a more efficient and relatively straightforward manner and has been widely used in science education research (Christianson & Fisher, 1999; Voska & Heikkinen, 2000).

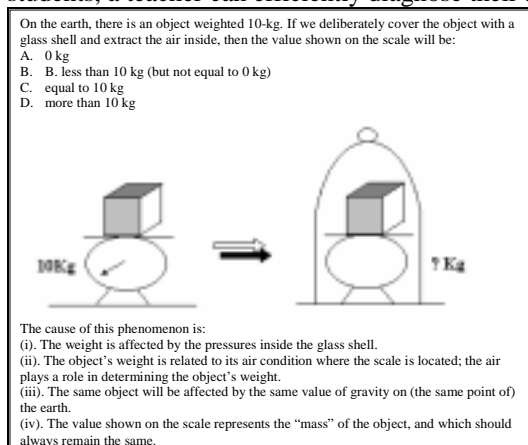
With the rapid development of Internet technology, the two-tier test can be implemented on a web site to further achieve wider accessibility for students and practising teachers. This paper describes an attempt to develop a networked two-tier test system.

Two-tier test

The two-tier test is a two-level question presented in a multiple-choice format. The item presented in Fig. 1 shows a sample that explores students' alternative conceptions about weight. The first tier assesses students' descriptive knowledge about the phenomenon, that is, the weight in a vacuum condition on the earth. The second tier explores students' reasons for their choice made in the first tier. Hence, the second tier investigates students' explanatory knowledge or their 'mental models'. Since the two-tier test is in a multiple-choice format, it is much easier for teachers to score or interpret students' responses. In this way, even with numerous students, a teacher can efficiently diagnose their alternative conceptions.

On the earth, there is an object weighted 10-kg. If we deliberately cover the object with a glass shell and extract the air inside, then the value shown on the scale will be:

A. 0 kg
 B. less than 10 kg (but not equal to 0 kg)
 C. equal to 10 kg
 D. more than 10 kg



The cause of this phenomenon is:

(i). The weight is affected by the pressures inside the glass shell.
 (ii). The object's weight is related to its air condition where the scale is located; the air plays a role in determining the object's weight.
 (iii). The same object will be affected by the same value of gravity on (the same point of) the earth.
 (iv). The value shown on the scale represents the "mass" of the object, and which should always remain the same.

Fig. 1. A two-tier test about students' concepts of weight (item 1)

more than 25 secondary school students. Students' interview responses were analysed and categorised, and these analyses and categorisation helped the researchers decide the choices of the test items.

The networked two-tier test

In the networked two-tier test system, only one item per screen was presented. For every presented item, there are two steps. For instance, in an item exploring students' alternative conceptions of sound which is adapted to a network presentation, the first step presents only the first tier of the item (Fig. 2). After the student makes a choice and confirms it within the first tier, then the system shows the second tier, with the first tier remaining on the screen (Fig. 3).



Fig. 2. The first tier of item two in the networked system.

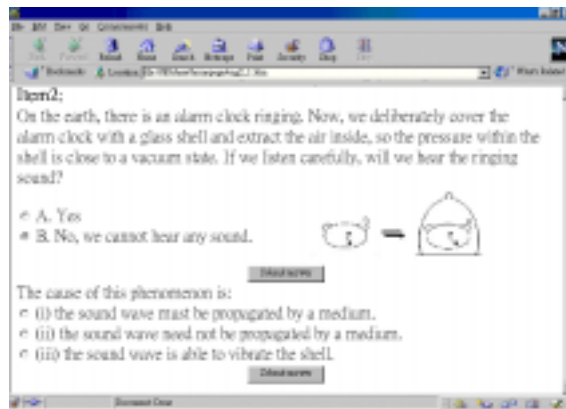


Fig. 3. The second tier of item two with the choice of the first tier still visible.

The item is presented in two steps to prevent influence of the first response by information contained in the second step in the sequence. This helps to ensure that the student's first response to an item represents their initial preconception and is not influenced by other possibly confounding information. Nevertheless, when making the choice of the second tier, the first tier is kept on the screen. This may help students to select a reason (i.e. the choice in the second tier) that is consistent with their choice made in the first tier. If, in some rare cases, students really want to change their initial choice after viewing the second tier, they could exit the system and then re-enter and change their answer. Although there is some inconsistent choice combinations across two tiers made by students later in this study, the initial separation of

the first tier's presentation from the second should ensure that some students' inconsistent ideas are really 'inconsistent', and that these inconsistencies do not necessarily come from cross-contamination within the system.

Three networked two-tier test items and their results

This study placed three two-tier test items on the web to be accessed by a group of Taiwan's students to survey their alternative conceptions. The first item (Fig. 1) explored students' concepts about weight, and the second item (Fig. 2 and Fig. 3), assessed students' ideas about sound. The final item is presented in Fig. 4, investigating students' alternative conceptions about heat and light. The items were presented in Chinese, and the English version of these items reported in this paper were an equivalent version after translation.

Research data for these two-tier test items were gathered from 15 8th-grade classes (14-year-old students) and 13 10th grade classes (16-year-old students). These students came from 11 schools, across different demographic areas of Taiwan, and this sample included 555 8th graders and 599 10th graders. Each student keyed in his/her ID and assigned password in school computer classrooms to complete three two-tier test items. The networked two-tier test items were administered to

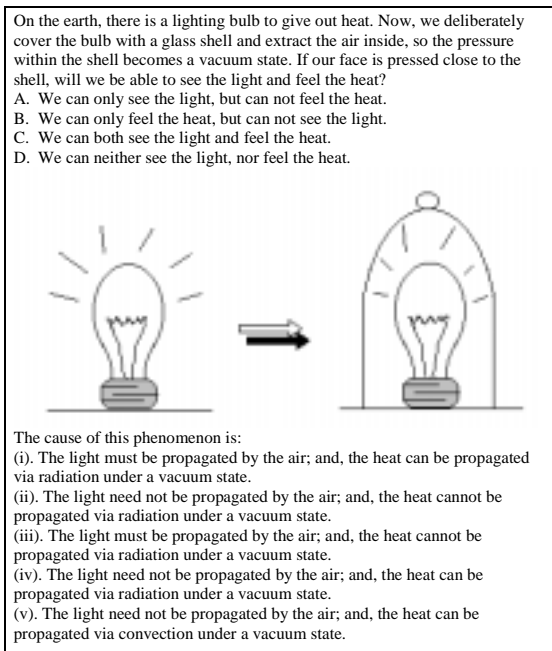


Fig. 4. A two-tier test about students' concepts of heat and light (item 3)

graders. The scientifically correct answer, that is (C)(iii), was chosen by only 8% of eighth graders and 12% of 10th graders. Some logically inconsistent answer combinations were still chosen by a few students, for instance, the answers of (A)(iv) (B)(iv), and (D)(iv) (not shown in the Table). These answers may come from students' mindless guessing. A χ^2 test comparing 8th graders' answer selection and that of 10th graders indicated that there were no significant differences between these two grades of students ($\chi^2 = 17.05$, d.f. = 15, $p \gg 0.05$). That is, even after formal instruction about weight, many 10th graders may have the same alternative conceptions as those expressed by 8th graders.

Table 1. The frequent responses on the two-tier test items

Items	Answer combination	8th grade (n, %)		10th grade (n, %)	
Item1	(A) (ii)	173	(31.2%)	145	(24.2%)
	(B) (ii)	80	(14.4%)	72	(12.0%)
	(C) (iii)*	45	(8.1%)	70	(11.7%)
	(C) (iv)	60	(10.8%)	80	(13.4%)
Item2	(A) (ii)	149	(26.8%)	103	(17.2%)
	(A) (iii)	89	(16.0%)	78	(13.0%)
	(B) (i)*	213	(38.4%)	347	(57.9%)
Item3	(A) (ii)	92	(16.6%)	78	(13.0%)
	(B) (i)	78	(14.1%)	74	(12.4%)
	(C) (iv)*	88	(15.9%)	115	(19.2%)
	(D) (iii)	47	(8.5%)		(< 8%)

Percentages smaller than 8% have been omitted

* scientifically correct answer

Students' answers to the second item (Fig. 2 and Fig. 3) revealed that 38% of 8th graders as well as 58% of 10th graders selected a scientifically correct answer, that is (B)(i). Students seemed to have more accurate ideas about the concepts of sound.

these students at the beginning of the fall semester. According to Taiwan's national curriculum, the beginning of eighth grade is the stage prior to receiving formal instruction about these scientific concepts, while the tenth graders had been taught about these concepts during the study of 8th and 9th grade science.

Table 1 presents the students' frequent answers on the two-tier test items. The most favourable answer combination of item 1 selected by students was (A)(ii), indicating that many students had an 'air implying weight' alternative conception. This finding was found for both 8th graders and 10th

However, 17% of 10th graders, after receiving formal instruction, still expressed a view that sound could be propagated without a medium. Nevertheless, the χ^2 test suggested that the answer selection of 10th graders were statistically different from that of 8th graders ($\chi^2 = 46.42$, d.f. = 5, $p < 0.001$). More 10th graders selected correct answers than 8th graders. Formal instruction seemed to have some impacts on students' ideas of sound in terms of the results of the second two-tier test item.

Students' responses to the third two-tier test item (Fig. 4) were presented below. Although the scientifically correct answer (C)(iv) was expected to be one of the most frequent choices selected by students, the proportion of these students was not very high (16% and 19% for 8th graders and 10th graders, respectively). An almost equal proportion of students selected the answer of (A)(ii) and they believed that the light can be propagated in a vacuum condition but the heat could not. Still a notable proportion of students held an almost opposite alternative conception that the heat can be propagated in a vacuum condition but the light could not. Hence, they chose (B)(i) as their answer (14% of 8th graders and 12% of 10th graders). A χ^2 test also showed that the responses of the 10th graders were not significantly different from those of the 8th graders ($\chi^2 = 14.93$, d.f. = 19, $p \gg 0.05$). In summary, the results derived from these two-tier test items suggest that in many cases, students' alternative conceptions may be retained even after formal instruction about relevant conceptions. This finding is consistent with a wide base of findings in science education research literature that alternative conceptions are tenacious and resistant to extinction by conventional teaching strategies (Wandersee *et al.* 1994).

An additional research question can be examined if variations in teaching and learning strategies may have effects on students' conceptual understanding of these topics as assessed by the two-tier test format. Students who experienced different teaching approaches, e.g. traditional vs. so-called constructivist, may have different conceptions about the content as examined by the test items following instruction. More recently, the effects of learning strategies on the development of students' conceptions also have recently gained attention among science educators (Tsai, 1998b; Tsai *et al.*, 2001). Hence, further research may be profitably directed toward gathering information about students' learning strategies in relation to their development of scientific conceptions as measured by networked two-tier items.

Moreover, given the fact that teaching in Taiwan's science classrooms is mainly traditional, this study was not able to assess students' responses in relation to different teaching strategies. A promising way of exploring this issue is to conduct carefully designed controlled experiments, comparing students' responses after traditional instruction with those after other, perhaps more progressive, instructional approaches. Over the past two decades, science educators have developed some teaching strategies to overcome students' alternative conceptions (e.g. Driver & Oldham, 1986; Tsai, 2000a) and these may be profitably examined for effectiveness in conceptual change learning using the networked two-tier test system. Another potential way of exploring this issue is to repeat this study by comparing the results of this study with the results from a similar set of students in another country, which is very different from Taiwan, for example, the UK or USA. The comparison may give clues about how education in different contexts, or using different instructional regimes, may affect students' knowledge development in science.

Responses across items

This study further examined another important research question; that is, students' responses across three test items. In other words, the relationships among students' understandings of the concepts of weight, sound, and heat and light were investigated. In order to explore these relationships, this study used a simple framework that categorised students' responses on each networked two-tier test item, using the following three levels of accuracy:

- Incorrect: students' responses are incorrect in both tiers, allocated for one point.
- Partially correct: students' responses are correct in one and only one tier of the item, allocated for two points.
- Correct: students' responses are correct in both tiers, allocated for three points.

For example, students with a response of (C)(iv) on item 1 were classified as 'partially correct', while those with a response of (A)(ii) on the same item were categorised as 'incorrect.' These data were used in correlation analyses of response accuracy across the three test items for 8th graders and 10th graders.

Table 2. The correlation of levels of correctness across three test items

	Item 1		Item 2 ^b		Item 3 ^a	
	8th grade	10th grade	8th grade ^b	10th grade ^c	8th grade ^b	10th grade ^c
Item 1	1	1	0.23***	0.26***	0.21***	0.29***
Item 2			1	1	0.43***	0.50***
Item 3					1	1

*** $p < 0.001$

Notes:

- Item 1 exploring students' alternative conceptions about weight.
Item 2 exploring students' alternative conceptions about sound.
Item 3 exploring students' alternative conceptions about light and heat.
- The t -value comparing the correlation coefficients between 0.43 and 0.23 is 4.19 ($p < 0.001$), while t -value comparing the correlation coefficients between 0.43 and 0.21 is 4.65 ($p < 0.001$).
- The t -value comparing the correlation coefficients between 0.50 and 0.26 is 5.73 ($p < 0.001$), and the t -value comparing the correlation coefficients between 0.50 and 0.29 is 4.96 ($p < 0.001$).

The results in Table 2 indicate that students' understandings about weight (item 1), sound (item 2), and heat and light (item 3) were statistically related ($p < 0.001$). Students having better understandings on the concept of weight tended to have better understandings in the other two domains. For both eighth graders and tenth graders, the correlation between item 1 and item 2, and that between item 1 and item 3, though significant, were not very high. However, students' levels of correctness on item 2 (about sound) and item 3 (about heat and light) were highly related ($r = 0.43$ for eighth graders, and $r = 0.50$ for tenth graders). A series of t -tests comparing the correlation coefficients also showed that the correlation coefficient between item 2 and item 3 was significantly higher than the others ($p < 0.001$). In the light of studies on physics knowledge frameworks, these findings were plausible because both the concepts of sound and those of heat and light are related through a common theoretical mechanism of 'wave' propagation, while the concepts of weight are derived from the foundations of classical mechanics. However, a careful look about 8th-grade students' responses across item 2 and item 3 revealed another interesting finding. Among the 213 eighth graders who selected (B)(i) as their response (i.e. the correct answer) on item 2, 33 (15.5%) of them selected (D)(iii) as their answer in item 3. This percentage was much higher than that calculated from the whole 8th-grade sample in this study (8.5% listed in Table 1). This implied that the 8th graders

who knew that sound required an elastic medium to be transferred sometimes got confused and thought light and heat must be the same. Nevertheless, the data of the tenth graders did not show a similar finding. Researchers are encouraged to look into the complexities of students' mental models more carefully by looking back at the responses of students on the items and how they may be related. Since the networked two-tier test system has the capacity to record students' responses within a digital database, educators can effectively conduct such correlation analyses even when additional items are included.

Conclusions and further development of networked two-tier test system

It is clear that the network technology, being easily administered and time efficient, can help educators collect a large number of students' responses on the items and then explore students' alternative conceptions. Many past studies on students' alternative conceptions mainly used interview or concept mapping techniques (Wandersee *et al.*, 1994). Due to their time consuming quality, the number of students that can be assessed and the range of alternative conceptions analysed are usually limited. The use of a networked two-tier test system can help researchers and educators investigate students' conceptions for a large sample (even national sample) and it also provides a deeper analysis for students' conceptual frameworks than traditional multiple-choice tests. In addition, the networked two-tier system can be used prior to instruction to explore students' prior knowledge, or can be used after instruction to assess students' learning outcomes. The system provides a quick way to gain a relatively detailed picture about students' prior knowledge or learning outcomes for teachers and researchers.

Clearly, the next step for this study is to develop more two-tier test items to be evaluated with a range of students. Researchers are encouraged particularly to develop more items within a conceptual domain to fully explore students' alternative domain-specific conceptions. For example, a series of two-tier tests about weight or broadly about gravity (such as the first one presented previously in this paper) can be developed to more thoroughly explicate students' mental models about gravity. Test items with multiple tiers (not only two tiers) can also be considered.

It is well recognised that computers can be effectively used as a medium to test individuals as well as large and small groups without a substantial increase in time or expense compared to paper and pencil formats. Test content and feedback can be customised to meet the individual needs of students by providing different difficulty levels, emphases, suggestive feedback and remedial materials. Further development of the system should focus on designing and providing appropriate feedback for students' incorrect answer combinations. Coe (1998) summarised studies of giving feedback and concluded some conditions that produced improved performance. For example,

- Feedback should be given to individuals on their individual performance.
- Feedback should be given as soon as possible after performance.
- Feedback should be specific and focused on the task.
- Feedback should aim to correct errors or inadequacies or feedback should have a diagnostic function.

Azevedo & Bernard (1995) also proposed that effective feedback in computer-assisted instruction involves the computer's ability to evaluate both the correctness of the learner's answer and the underlying causes of error. Clearly, the networked

two-tier test system can provide feedback that fulfills these conditions. For example, if one student has an alternative conception that 'air implies weight' (as elicited in the first two-tier test in this study), the system can promptly provide a feedback that asks the student to think about the weight condition on the moon. Most students are well informed that the moon is in a vacuum and the weight of an object is about one sixth when compared to that on the earth. Hence, an object still has its weight in a vacuum condition, such as on the moon. As another example, if one student has an alternative conception that heat and light should be propagated by a medium and they cannot be delivered in a vacuum condition, the system can be designed to provide a series of instant feedback that leads the student to reflect on how we can see the light as well as feel the heat from the sun. By addressing students' underlying errors or inadequate explanations, such discrepant feedback, directly linked to the student's incorrect answer combination can challenge their existing conceptions and thus facilitate the process of conceptual change. Chi (1996) has developed a fourfold categorisation of feedback: corrective feedback, reinforcing feedback, didactic feedback, and suggestive feedback. It is expected that the networked two-tier test system, based on students' incorrect (or even correct) responses, can promptly provide either one or more types of feedback of these kinds for learners. It should be noted that the feedback should not be limited in a simple reply to the error, rather, it could be a series of interactive questioning, demonstrations, and open-ended explorations, which make the system more 'constructivist.'

Moreover, the networked two-tier test system can be designed to present the test materials and relevant feedback in a multimedia format. With feedback that addresses students' alternative conceptions, whether corrective, reinforcing, didactic or suggestive, the system will become much more of an interactive, multimedia type. When designed as an interactive, multimedia learning environment, the networked two-tier test system becomes not only a diagnostic tool, but also an effective instructional tool, helping students overcome their alternative conceptions. Finally, the networked system can record students' learning paths when navigating the system. Future studies of networked two-tier tests can effectively explore the relationship between students' navigation modes and paths and their test responses on the two-tier test.

In conclusion, although the items in this paper may as yet be limited in scope, this paper has illuminated some innovative thoughts, maybe not for web-based testing, but certainly for the research and practice of science education. The current system can overcome an unsolved issue for contemporary science educators, that is, to investigate students' alternative conceptions with a much larger sample of students. The future system, which incorporates adaptive feedback for students, may help students overcome alternative conceptions, an important instructional goal shared by many modern science educators.

Acknowledgements

Funding of this research work is supported by National Science Council, Taiwan, ROC, under grants NSC 89-2511-S-009-005 and NSC 90-2511-S-009-005. The authors also express their gratitude to Professor O. Roger Anderson at Teachers College, Columbia University and three anonymous referees' helpful comments for the further development of this paper.

References

- Azevedo, R. & Bernard, R.M. (1995) A meta-analysis of the effects of feedback in computer-based instruction. *Journal of Educational Computing Research*, **11**, 2, 111–127.
- Bell, B. (1995) Interviewing: A technique for assessing science knowledge. In *Learning Science in the Schools: Research Reforming Practice* (eds. S.M. Glynn & R. Duit), pp. 347–364. Lawrence Erlbaum Associates, Mahwah, NJ.
- Christianson, R.G. & Fisher, K.M. (1999) Comparison of student learning about diffusion and osmosis in constructivist and traditional classrooms. *International Journal of Science Education*, **21**, 6, 687–698.
- Chi, M.T.H. (1996) Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology*, **10**, S33–S49.
- Coe, R. (1998) Can feedback improve teaching? A review of the social science literature with a view to identifying the conditions under which giving feedback to teachers will result in improved performance. *Research Papers in Education*, **13**, 1, 43–66.
- Driver, R. & Easley, J. (1978) Pupils and paradigms: a review of literature relate to concept development in adolescent science students. *Studies in Science Education*, **5**, 61–84.
- Driver, R. & Oldham, V. (1986) A constructivist approach to curriculum development in science. *Studies in Science Education*, **13**, 105–122.
- Driver, R., Guesne, E. & Tiberghien, A. (1985) *Children's ideas in science*. Open University Press, Philadelphia, PA.
- Novak, J.D. & Gowin, D.B. (1984) *Learning How to Learn*. Cambridge University Press. Cambridge.
- Odom, A.L. & Barrow, L.H. (1995) The development and application of a two-tiered diagnostic test measuring college biology students' understanding of diffusion and osmosis following a course of instruction. *Journal of Research in Science Teaching*, **32**, 1, 45–61.
- Posner, G.J. & Gertzog, W.A. (1982) The clinical interview and the measurement of conceptual change. *Science Education*, **66**, 2, 195–209.
- Ruiz-Primo, M.A. & Shavelson, R.J. (1996) Problems and issues in the use of concept maps in science assessment. *Journal of Research in Science Teaching*, **33**, 6, 569–600.
- Treagust, D.F. (1988) Development and use of diagnostic tests to evaluate students' misconceptions in science. *International Journal of Science Education*, **10**, 2, 159–169.
- Tsai, C.-C. (1998a) Science learning and constructivism. *Curriculum and Teaching*, **13**, 1, 31–52.
- Tsai, C. (1998b) An analysis of scientific epistemological beliefs and learning orientations of Taiwanese eighth graders. *Science Education*, **82**, 4, 473–489.
- Tsai, C.-C. (2000a) Enhancing science instruction: The use of 'conflict maps'. *International Journal of Science Education*, **22**, 3, 285–302.
- Tsai, C.-C. (2000b) Relationships between student scientific epistemological beliefs and perceptions of constructivist learning environments. *Educational Research*, **42**, 2, 193–205.
- Tsai, C.-C., Lin, S.S.J. & Yuan, S.-M. (2001) Students' use of web-based concept map testing and strategies for learning. *Journal of Computer Assisted Learning*, **17**, 1, 72–84.
- Voska, K.W. & Heikkinen, H.W. (2000) Identification and analysis of student conceptions used to solve chemical equilibrium problems. *Journal of Research in Science Teaching*, **37**, 2, 160–176.
- Wandersee, J.H., Mintzes, J.J. & Novak, J.D. (1994) Research on alternative conceptions in science. In *Handbook of Research on Science Teaching and Learning* (ed. D.L. Gabel), pp. 177–210. Macmillan, New York.