Comparing two types of diagnostic items to evaluate understanding of heat and temperature concepts

Hye-Eun Chu, A. L. Chandrasegaran and David F. Treagust

ABSTRACT The purpose of this research was to investigate an efficient method to assess year 8 (age 13–14) students' conceptual understanding of heat and temperature concepts. Two different types of instruments were used in this study: Type 1, consisting of multiple-choice items with open-ended justifications; and Type 2, consisting of two-tier multiple-choice items. Each of the instruments was administered to two separate cohorts of 173 and 143 year 8 students of similar achievement. The findings indicated that the students were better able to show their understanding in the two-tier multiple-choice items. Hence, based on this investigation, two-tier multiple-choice items may be more suitable for evaluating year 8 students' understanding of science concepts.

Diagnosing students' conceptual understanding in science provides teachers with valuable information when preparing their classroom instruction. This is especially so when the concepts are abstract and students hold many scientifically inappropriate conceptions. Administering a suitable diagnostic instrument before instruction can provide information to teachers about their students' prior knowledge.

An area of concern is students' understanding of thermal concepts in physical science. Students are known to experience confusion between the everyday usage of terms and the relevant scientific terms, and at the same time they display conceptual confusion between heat and temperature, and with regard to heat transfer in equilibrium and non-equilibrium systems (Erickson, 1979; Kesidou and Duit, 1993; Wiser and Amin, 2001).

Several types of diagnostic instruments have been developed by researchers. Traditional multiple-choice items involving a particular topic or concept do not readily reveal students' alternative conceptions because the reason behind a student's selection of a response is not evident. For this reason, Tamir (1971) included an open-ended justification section in multiple-choice items for students to explain their reason for selecting a particular response. Subsequently, science educators have developed several diagnostic instruments consisting of two-tier multiple-choice items (Chu and Treagust, 2014; Treagust, 1995). The first tier consists of four or five content options related to the question. After making their selection in the first tier, students are required to justify their choice by then selecting one of the five or six reasons provided in the second tier. These instruments are convenient to administer and can be readily marked. Information about students' conceptions can then be used for remedial or teaching purposes. Two-tier multiple-choice items ('Type 2') require students to be specific about the reason for their choice of response in the first tier. In 'Type 1' items, their open-ended justifications of their own ideas may not be specific enough to show their understanding (see Boxes 1–4).

However, concerns have been raised about the validity of Type 2 items in identifying students' alternative conceptions because the forced-choice questions tend to limit students to the options that are provided (Griffard and Wandersee, 2001). In their study involving six pre-medical students, five of whom were high-achieving, Griffard and Wandersee (2001) used four two-tier multiple-choice items on photosynthesis that had been developed by Haslam and Treagust (1987). The study showed that none of the students selected their reason from the choices provided; instead, the second tier was regarded as a distinct multiple-choice item and they selected the option that logically followed from their choice in the first tier. Based on the resulting findings, the authors suggested that the items were more a measure of students' test-taking skills than their knowledge about the concept. However, the items were designed for a secondary school curriculum, not university level where more in-depth knowledge would be expected.

In view of the findings by Griffard and Wandersee (2001), we were interested to make a comparison between two-tier multiple-choice items and multiple-choice items with open-ended justifications. Even though several science educators have emphasised the strength of two-tier multiple-choice items in diagnostic assessment (e.g. Treagust, 1988), many teachers and science educators still doubt the ability of two-tier multiple-choice questions to diagnose students' understanding. While using open-ended types of questions may provide teachers with rich information, a lack of time to analyse students' responses may mean that such questions are not very effective in diagnosing students' understanding.

We also argue that the two types of items are asking questions at similar cognitive levels in relation to Bloom's Taxonomy: both items require students to go from analysis to evaluation – reading problem situations, comparing the possible responses in the first-tier options, and evaluating the analysis and first-tier options when completing the second-tier section. The first-tier items were the same in both Types 1 and 2 and were designed, based on previous conceptual studies, to include similar-status alternative conceptions. In the second tier, students should create the response using their own words in Type 1, with evaluation of their analysis of the first-tier options, and students should evaluate the first-tier response via second-tier options when they choose their reasons for first-tier option choices in Type 2. In some Type 2 items, there are more correct responses than in the same items in Type 1, which may be due to scaffolding gained from the second-tier options in Type 2 questions.

To compare the effectiveness of the two types of tests – multiple choice with open-ended responses and two-tier multiple-choice items – we chose the topic of heat energy and temperature because this topic presents many challenges to students and has been an area of recent research in Australia and elsewhere. For example, concepts related to heat and its transfer are first encountered in year 3 (age 8) in the Australian National Science Curriculum (Australian Curriculum Assessment and Reporting Authority, 2012). The research findings will provide information to teachers and science educators for selecting efficient types of diagnostic assessments taking into account their time limitations and students' abilities. This study was guided by the research question aimed at finding out which type of question, i.e. multiple choice with open-ended justifications (Type 1) or two-tier multiple-choice items (Type 2), is more efficient in determining year 8 students' understanding of heat and temperature concepts.

Research methods

The sample

The investigation being described was conducted with a cohort of 143 year 8 students (age 13–14) using 18 two-tier multiple-choice items on heat energy and temperature and comparing their understanding with a cohort of 173 year 8 students using the same multiple-choice items from the first tier but with open-ended justifications for their choice of response in the first tier. Students in both cohorts were from the same school and achievement groups (PSLE T-score range of 200–220) in Singapore. The equivalence in the achieving ability of the two classes of students was based on the similar results that the students had obtained in the Primary School Leaving Examination (PSLE) at the end of their previous year. The students in year 8 were distributed into classes based on the results of this examination. The thermal concepts in this questionnaire are included in the Singapore science/physics curriculum for years 3–12.

After the instruments containing the 18 questions were completed, four volunteer students from each class were interviewed in focus group discussions to obtain a more in-depth insight into their understanding of the associated concepts and their preference for the particular type of items in the instruments.

The two types of tests

Two types of diagnostic instruments to evaluate students' conceptual understanding of thermal concepts have been developed in previous studies (Chu, Treagust, Yeo and Zadnik, 2012; Yeo and Zadnik, 2001). One instrument used multiple-choice items with open-ended justifications that required students to provide a reason for their choice of response to the multiple-choice items (Type 1). The instrument with Type 1 questions was administered to one of the groups ($n=173$); the other instrument, consisting of two-tier multiple-choice items (Type 2), was administered to the second group ($n=143$). As both the groups consisted of equivalent high-achieving students in the school, it could be assumed that their responses were based on the difference in the type of questions (Type 1 or 2) that were administered rather than difference in ability

The 18 items in both the instruments involved four conceptual categories:

1 boiling;
2 melting;
3 heat energy transfer and temperature;
4 thermal equilibrium and conductivity.

An example of each type of item in each of the four conceptual categories is shown in Boxes 1–4.

BOX 1 Two different types of items in the boiling conceptual category

Type 1. Multiple choice item with open-ended explanations

1.A kettle full of water is heated on a stove. When the water starts boiling rapidly, the most likely temperature of the water is about:

A. 88 °C
B. 98 ° C
C. 110 C
D. None of the above answers could be right.
E.

The reason for my answer is:

____

____

____

Type 2. Two-tier multiple-choice item

1.A kettle full of water is heated on a stove. When the water starts boiling rapidly, the most likely temperature of the water is about:

A. 88 °C
B. 98 °C
C. 110 °C
D. D. None of the above answers could be right.

The reason for my answer is:

1. Water always boils at the same temperature.
2. The surrounding pressure affects the boiling point.
3. The temperature of rapidly boiling water will approach its boiling-point.
4. Heat has been transferred to the surroundings.
5. The temperature of rapidly boiling water has exceeded its boiling-point.

BOX 2 Two different types of items in the melting conceptual category

Type 1. Multiple choice item with open-ended explanations

6. Ken takes six ice-cubes from the freezer and puts four of them into a glass of water. He stirs the mixture continuously until the ice-cubes are much smaller but have stopped melting. What is the most likely temperature of the water at this stage?
A. –10 °C
B. 0 °C
C. 5 °C
D. 10 °C

The reason for my answer is:

____

____

____

Type 2. Two-tier multiple-choice item

6. Ken takes six ice-cubes from the freezer and puts four of them into a glass of water. He stirs the mixture continuously until the ice-cubes are much smaller but have stopped melting. What is the most likely temperature of the water at this stage?

A. –10 °C
B. 0 °C
C. 5 °C
D. 10 °C

The reason for my answer is:

1. Water gains heat from surrounding.
2. Water is at a higher temperature than its freezing-point.
3. Water is hotter than ice.
4. Ice and water are at the same temperature.
5. Ice will transfer coldness to water.

BOX 3 Two different types of items in the heat energy transfer and temperature conceptual category

Type 1. Multiple choice item with open-ended explanations

9-10. A can of coke and a plastic bottle of coke were in the refrigerator **overnight**. Sam takes both out and quickly puts a thermometer in the coke in the **can**. The temperature is 7°C

10. A few minutes later, Ned picks up the coke can and then tells everyone that the bench top underneath it feels colder than the rest of the bench. Whose explanation do you think is best?

A. Jon says: "The cold has been transferred from the coke to the bench."
B. Rob says: "There is no energy left in the bench beneath the can."
C. Sue says: "Some heat has been transferred from the bench to the coke."
D. Eli says: "The can causes heat beneath the can to move away through the bench-top."

The reason for my answer is: ____________

____

____

Type 2. Two-tier multiple-choice item

9-10. A can of coke and a plastic bottle of coke were in the refrigerator **overnight**. Sam takes both out and quickly puts a thermometer in the coke in the **can**. The temperature is 7°C

10. A few minutes later, Ned picks up the coke can and then tells everyone that the bench top underneath it feels colder than the rest of the bench. Whose explanation do you think is best?

A. Jon says: "The cold has been transferred from the coke to the bench."
B. Rob says: "There is no energy left in the bench beneath the can."
C. Sue says: "Some heat has been transferred from the bench to the coke."
D. Eli says: "The can causes heat beneath the can to move away through the bench-top."

The reason for my answer is:
1. Convection currents are present.
2. Metal is a good conductor.
3. Heat is lost from the bench when it becomes cold.
4. The bench is hotter than the coke can.
5. Coke can have more coldness than the bench.

BOX 4 Two different types of items in the thermal equilibrium and conductivity conceptual category

Type 1. Multiple choice item with open-ended explanations

11. Vic takes some popsicles from the freezer, where he had placed them the day before. He then discusses the temperature of the wooden stick and the ice part with his friends. Who do you agree with?

A. Deb says: "The wooden sticks are at a higher temperature than the ice."
B. Ron says: "The ice is at a higher temperature than the wooden stick."
C. Ross says: "They are at the same temperature."
D. Ian says: "It depends on the size of the ice."

The reason for my answer is:

____

____

____

Type 2. Two-tier multiple-choice item

11. Popsicles were placed in the freezer the day before. This morning, Vic takes a few popsicles from the freezer. He then discusses the temperature of the wooden stick and the ice part with his friends. Who do you agree with?

A. Deb says: "The wooden sticks are at a higher temperature than the ice."
B. Ron says: "The ice is at a higher temperature than the wooden stick."
C. Ross says: "They are at the same temperature."
D. Ian says: "It depends on the size of the ice."

The reason for my answer is:
1. Wood absorbs heat better than ice.
2. Ice transfers heat better than wood because it has neatly packed particles.
3. Ice and wood were in the freezer for a long time.
4. Ice gives coldness to wood because ice is at a very low temperature.
5. Wood cannot be frozen.
6. The rate of heat inflow and outflow between ice and wood is the same.

The complete instruments may be obtained from the first author. The multiple-choice part in the Type 1 items and the first tier of the Type 2 items were modified from the Thermal Conceptual

Evaluation, TCE, developed by Yeo and Zadnik (2001). The second tier of the two-tier multiple-choice items was developed based on known alternative conceptions held by students that had been identified in previous studies on thermal physics (Chu *et al*., 2012; Yeo and Zadnik, 2001).

Analysis

Guidelines were prepared to assess students' open-ended explanations for their responses to each Type 1 item. An example of the guidelines that were used for Item 11 (Box 4) is shown in Box 5.

Only when students selected the correct response and provided a correct justification in the Type 1 items was the response to the combined item considered correct. For the Type 2 items, an item was considered correctly answered only when students selected the correct responses to both the first and second tiers. For both Type 1 and Type 2, if only the response or the justification was correct, the item was considered incorrectly answered. Also, students' inappropriate conceptions in response to the items in the two instruments were compared in order to investigate the effectiveness of the two instruments as diagnostic assessment tools. Students' alternative conceptions displayed in their responses to the Type 2 items were coded based on their response options to both tiers. Similar codes were used to identify students' alternative conceptions in their justifications to Type 1 items. An example of the coding system used for Item 6 is shown in Table 1.

The interviews were recorded and transcribed. These summarised data were used to provide evidence and reasons for the findings in this article.

Table 1 Coding of alternative conceptions in Type 1 responses to Item 6 in Box 2 ($n = 173$)

Coding	Students' scientifically inappropriate conceptions	MC option	%
6AC1	Heat gained by the ice causes its temperature to be above 0 °C.	C	10
		D	6
6AC2	Temperature of the water is above the temperature of ice (0 °C).	C	6
		D	4
6AC3	Cold/hot can be transferred between bodies.	D	4
NA	Students' responses cannot be categorised.	–	17

MC = multiple-choice

Note: There were 11 more such coding categories for Item 6 but they were each for fewer than 3% of students' responses.

BOX 5 Correct response guidelines and examples of students' justifications for Item 11 in Box 4

Guidelines		Students' response
Acceptable scientific conception (ASC)	Students have to show understanding that the popsicle and the wooden stick are in thermal equilibrium with each other since both are placed in the freezer for a long duration and hence are at the same temperature.	No students displayed the acceptable scientific conception.
Incomplete scientific conception (ISC)	Students have to show understanding that the popsicle and the wooden stick are at the same temperature.	Student 15: They are cooled evenly and, therefore, must be at the same temperature. Student 94: Since both are put in the freezer at the same time, they should have the same temperature.

Expected scientifically correct justification

The popsicles have been placed in the freezer for a long duration. The temperature of the ice and stick will be at the same temperature because both have reached thermal equilibrium with each other and there is no more transfer of heat between the two. When in thermal equilibrium, the rate of heat flow between ice and wooden stick and vice versa are the same.

Findings

Comparison of students' correct responses in two types of multiple-choice items

A summary of the findings comparing students' acceptable conceptual understanding about heat and temperature concepts using the two types of items is shown in Tables 2 and 3.

There were significant differences between Types 1 and 2 in the boiling conceptual category. Items 1 to 3 in Type 1 in the boiling category were answered correctly by a very small percentage (fewer than 2%) of students who chose the correct answer of 98 °C for the boiling point of the water and provided reasonable justifications. Most students were not able to explain the relationship between boiling point and atmospheric pressure even at the macroscopic level. In this boiling conceptual category, students' correct responses to items in the Type 1 instrument were in the range 0–2% and about 25% of students' open-ended responses could not be analysed because they repeated the question or provided irrelevant answers. On the other hand, in the Type 2 instrument, students' correct responses to the items were in the range 10–25%, which may be due to scaffolding gained from options in the second tier, and all student responses were able to be analysed.

In addition, students were able to show more correct responses in the Type 1 instrument for Items 6 and 11. These items asked about the temperature of water when ice stops melting, and about the temperatures of the wooden sticks and the ice parts of popsicles that had been placed in the freezer for a day. These items in the Type 2 instrument included options that could only be answered correctly when students had a clear understanding of the thermal concepts. For example, in Item 11 of the Type 2 instrument, few students could correctly answer the second tier because most of them did not know that when two objects were in thermal equilibrium with each other, they were at the same temperature.

For most cases, there were no statistically significant differences between students'

Table 2 Comparison of percentage of correct responses to multiple-choice items with open-ended explanations (Type 1; $n = 173$) and two-tier multiple-choice items (Type 2; $n = 143$) by year 8 students for conceptual categories 1 and 2

Item		Correct MC	Type 1, % correct		Type 2, % correct			χ^2
			MC	MC&Exp.	First tier	Combined tiers		
Conceptual category 1. Boiling								
Q1	The boiling point of water depends on the external pressure.	[B]	24	**0**	52	[B2]	**25**	**47.6***
Q2	The temperature of boiling water remains constant.	[B]	30	**2**	14	[B2]	**10**	**9.8***
Q3	Steam in contact with boiling water is at the same temperature as that of the boiling water.	[B]	29	**2**	39	[B1]	**15**	**15.6***
Q4	The boiling point of water decreases with altitude because of the lower pressure.	[A]	16	**0**	34	[A3]	**25**	**47.6***
Q5	The boiling point of a liquid increases with pressure.	[A]	49	**0**	76	[A3]	**10**	**17.7***
Conceptual category 2. Melting								
Q6	Ice and water in contact with each other are at the temperature of the melting ice.	[B]	26	**16**	22	[B4]	**8**	**4.3***
Q7	Melting ice and ice water are at the same temperature.	[B]	16	**10**	41	[B4]	**7**	**1.1**

Correct MC = correct multiple-choice responses in both Type 1 and Type 2 instruments; MC = multiple choice; MC&Exp. = multiple choice with explanations

*$P < 0.05$

Note: χ^2 was performed to compare MC&Exp. with combined tiers (df = 1).

Table 3 Comparison of percentage of correct responses to multiple-choice items with open-ended explanations (Type 1; $n = 173$) and two-tier multiple-choice items (Type 2; $n = 143$) by year 8 students for conceptual categories 3 and 4

Item		Correct MC	Type 1, % correct		Type 2, % correct			χ^2
			MC	MC&Exp.	1st tier	Combined tiers		
Conceptual category 3. Heat energy transfer and temperature								
Q8	Heat transfer occurs from a hotter body to a colder body.	[C]	68	**3**	70	[C5]	**4**	**0.4**
Q10	Heat transfer occurs from a hotter body to a colder body.	[C]	32	**13**	47	[C4]	**10**	**0.9**
Q13	Heat energy is lost during evaporation resulting in a decrease in temperature.	[D]	24	**0**	22	[D4]	**4**	**7.4***
Q15	Heat transfer occurs from a hotter body to a colder body.	[B]	34	**13**	49	[B4]	**27**	**8.8***
Q16	An insulator reduces heat energy loss by a body.	[A]	19	**7**	17	[A2]	**8**	**0.1**
Conceptual category 4. Thermal equilibrium and conductivity								
Q9	Heat energy is lost by a hotter body until it is in equilibrium with the temperature of its cooler surroundings.	[B]	13	**6**	13	[B5]	**7**	**0.2**
Q11	Heat transfer occurs from a hotter body to a colder body until temperature equilibrium is reached.	[C]	20	**11**	15	[C6]	**5**	**3.8***
Q12	Metals are better heat conductors than plastics.	[B]	9	**4**	8	[B2]	**1**	**2.7**
Q14	Heat energy is lost during evaporation resulting in a decrease in temperature and heat is transferred from a warmer body to a cooler body until equilibrium is reached.	[A]	32	**9**	29	[A4]	**8**	**0.1**
Q17	No net heat transfer occurs between two bodies at the same temperature.	[B]	42	**9**	64	[B4]	**4**	**3.5**
Q18	Metals conduct heat energy better than non-metals.	[B]	8	**8**	28	[B1]	**13**	**1.7**

Correct MC = correct multiple-choice responses in both Type 1 and Type 2 instruments; MC = multiple choice; MC&Exp. = multiple choice with explanations

*$P < 0.05$

Note: χ^2 was performed to compare MC&Exp. with combined tiers (df = 1).

responses to items in the Type 1 and Type 2 instruments (excluding the boiling conceptual category and Items 6 and 11), and students displayed more correct responses in the Type 2 instrument.

Determining students' alternative conceptions

Analysis of students' responses to the items in both Type 1 and Type 2 instruments can also provide information about their alternative conceptions. In Item 6, for example, students had to decide about the temperature of water when ice cubes mixed with the water had stopped melting. Table 4 provides a summary of the alternative conceptions held by students that were identified using the two types of items. The same alternative conceptions were identified in students' responses to both types of multiple-choice items. Even if the same alternative conceptions were not displayed in the responses to both types of items, it is definitely less time-consuming to analyse the data from the Type 2 item responses.

The alternative conceptions in Table 4 include more alternative conceptions than previous

Table 4 Percentages of students' alternative conceptions in Item 6 from the two types of items

Students' scientifically inappropriate conceptions	Type 1 (*n* = 173), %	Type 2 (*n* = 143), %
Temperature of water is above the temperature of ice (0 °C).	10	27
Heat gained by the water/ice causes its temperature to be above 0 °C.	19	18
Heat gained from the surroundings causes temperature of water/ice to be above 0 °C.	7	13
Cold/hot can be transferred between bodies.	5	38

studies (Chu *et al*., 2012; Yeo and Zadnik, 2001) owing to the second-tier component, which consists of open-ended responses in Type 1 and multiple-choice responses in Type 2. Both types of second-tier components provide justification of the first-tier choices. Other studies did not have this second tier.

Conclusion

The results in Tables 2 and 3 show that, for many of the items, using either Type 1 or Type 2 items provides statistically similar information about students' understanding of thermal concepts. Most students experienced difficulty in understanding:

- melting ice and ice water at the same temperature (item 7);
- heat energy transfer concepts occurring by conduction when there are two objects or systems at different temperatures (Items 8, 10 and 16);
- energy transfer between different materials at the different temperatures until equilibrium is achieved (Items 9, 14 and 17).

Nevertheless, in other instances, the year 8 students displayed better understanding of the associated concepts using two-tier multiple-choice items compared with the multiple-choice items having open-ended justifications, for example with:

- boiling point of water under different pressures (Items 1, 2, 3, 5 and 5);
- decreased temperature due to heat energy loss (Item 13);
- heat energy transfer from a hotter body to a colder body (Item 15);
- the rate of heat transfer and conductivity (Item 18).

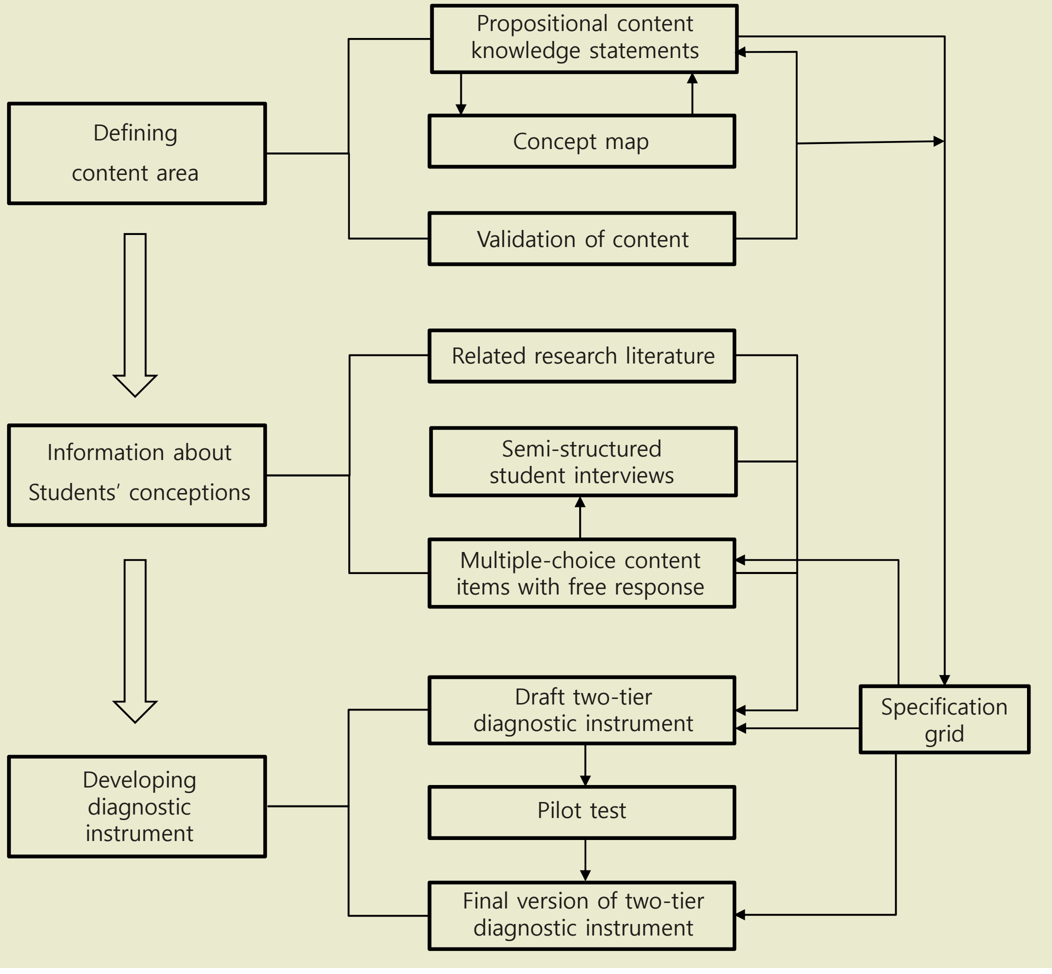
These results from comparison of students' responses to the Type 1 and Type 2 questionnaires indicated that Type 2 items are effective in diagnosing year 8 students' conceptions in terms of easiness of managing and analysing students' responses as well as the similar/larger strength of evaluating students' conceptual understanding. These findings are also supported by students' responses from the focus group discussions with four groups of students (*n* = 16) who had taken the Type 2 questionnaire. Half of these students answered that sometimes they did not know what was expected in the first tier. In such instances, based on the second-tier options they reflected on what they had previously learned and experienced in order to get ideas on how to answer the first tier too. However, about 70% of students in the focus groups complained that the questions and options were very different from examination questions and that the phrases in the options were not familiar to them. The latter point is because the items did not directly provide scientific terms in the options such as thermal equilibrium, heat transfer and conductivity. Students could only answer correctly in the second tier when they knew the meaning of the scientific terms. Rather than this being seen as a disadvantage, Type 2 items do provide thoughtful students with the opportunity to reason through their own thinking.

Chu and Treagust (2014) indicated the importance of following up classroom discussions and designing lessons with activities based on finalised alternative conceptions. There are various ways of doing this; the easiest way for teachers could be by using the diagnostic question contexts and situations to encourage students to apply their prior knowledge and to link it to the newly learnt concepts from teachers.

For analysing students' conceptual learning, teachers should be able to select well-developed diagnostic items from the literature. Treagust and his group (Chu and Treagust, 2014; Damanhuri, Treagust, Won and Chandrasegaran, 2016; Treagust and Chandrasegaran, 2007; Treagust, 1988, 1995) established a process for developing two-tier multiple-choice diagnostic instruments (see Box 6) and this process has implications for the criteria that teachers use to select diagnostic instruments. Over the past 30 years, many diagnostic instruments involving alternative conceptions have become available in the science education literature. Firstly, the options should include alternative conceptions of similar status. Secondly, the items should be written at similar levels of students' understanding in the specific topic and should include students' ways of reasoning. Thirdly, the diagnostic questionnaire should include various situations in everyday contexts so that students can show how they applied their prior scientific knowledge. This situation in everyday contexts can be used in classroom discussions and follow-up lesson development and activities.

Further studies need to be conducted with students at higher grade levels in order to identify any differences in the suitability or otherwise of two-tier multiple-choice items compared with multiple-choice items with open-ended justifications. Even if the latter items are found to be more appropriate for students at higher levels, analysis of the open-ended justifications is certainly more time-consuming for the science teacher. Also, analysing Type 2 items provides a more convenient means of identifying students' alternative conceptions. An obvious solution to the problem is to ensure that the items in a two-tier multiple-choice diagnostic

BOX 6 Stages in the development of two-tier multiple-choice diagnostic instruments



instrument are pitched at the appropriate level of understanding for the students. It would therefore be inappropriate to use a two-tier diagnostic instrument that was developed to evaluate year 8 students' understanding of a concept with, say, year 12 students who would have more advanced understanding of the particular concept and would hence be restricted to the options provided and would be forced to make a decision that might not cover their own explanations.

Acknowledgements

This work was supported by the Ministry of Education in Singapore (Grant Scheme: National Institute of Education (NIE), Office of Education Research, OER 36/12 ZHE).

References

Australian Curriculum Assessment and Reporting Authority (2012) *National Science Curriculum*. Canberra. Available at: www.australiancurriculum.edu.au/download?view=f10.

Chu, H.-E., Treagust, D.F., Yeo, S. and Zadnik, M. (2012) Evaluation of students' understanding of thermal concepts in everyday contexts. *International Journal of Science Education*, **34**(10), 1509–1534.

Chu, H.-E. and Treagust, D.F. (2014) Secondary students' stable and unstable optics conceptions using contextualized questions. *Journal of Science Education and Technology*, **23**(2), 238–251.

Damanhuri, M.I.M., Treagust, D.F., Won, M. and Chandrasegaran, A.L. (2016) Development of an achievement test to evaluate acid-base concepts in a Malaysian context. *International Journal of Environmental and Science Education*, **11**(1), 9–27.

Erickson, G.L. (1979) Children's conceptions of heat and temperature. *Science Education*, **63**(2), 221–230.

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

Haslam, F. and 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**(3), 203–211.

Kesidou, S. and Duit, R. (1993) Students' conceptions of the second law of thermodynamics—an interpretive study. *Journal of Research in Science Teaching*, **30**, 85–106.

Tamir, P. (1971) An alternative approach to the construction of multiple choice test items. *Journal of Biological Education*, **5**(6), 305–307.

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.

Treagust, D.F. (1995) Diagnostic assessment of students' science concepts. In *Learning Science in the Schools: Research Reforming Practice*, ed. Glynn, S. and Duit, R. pp.327–346. Mahwah, NJ: Lawrence Erlbaum Associates.

Treagust, D.F. and Chandrasegaran, A.L. (2007) The Taiwan National Science Concept Learning Study in an international perspective. *International Journal of Science Education*, **29**(4), 391–403.

Wiser, M. and Amin, T. (2001) Is heat hot? Inducing conceptual change by integrating everyday and scientific perspectives on thermal phenomena. *Learning and Instruction*, **11**4–5), 331–355.

Yeo, S. and Zadnik, M. (2001) Introductory thermal concept evaluation: assessing students' understanding. *The Physics Teacher*, **39**(8), 496–504.

Hye-Eun Chu is a science education lecturer in the Department of Educational Studies at Macquarie University, Sydney, Australia. Email: hye-eun.chu@mq.edu.au

A.L. Chandrasegaran, now retired, was a research associate in science education in the School of Education, Curtin University, Perth, Australia. Email: alchandran41@gmail.com

David F. Treagust is a professor of science education in the School of Education, Curtin University, Perth, Australia. Email: d.treagust@curtin.edu.au