Eurasia Journal of Mathematics, Science & Technology Education, 2014, **10**(5), 509-521

Using Three-Tier Diagnostic Test to Assess Students' Misconceptions of States of Matter

Zubeyde Demet Kirbulut
Harran University, TURKEY

Omer Geban
Middle East Technical University, TURKEY

Received 06 January 2014; accepted 31 July 2014

This study involves the development of a three-tier diagnostic test to measure high school students' understanding of states of matter concepts. The States of Matter Diagnostic Test (SMDT) is a 19-item three-tier diagnostic test consisting of three-tier items for assessing students' understanding of states of matter concepts. The SMDT was administered to 195 10th grade high school students in the pilot study and 102 10th grade high school students in the main study. Cronbach alpha reliability indexes for the SMDT were estimated to be .78 and .83 for the pilot and main study, respectively. Point-biserial coefficients ranged from .20 to .69 with an average of .44 for the pilot study and with an average of .49 for the main study.

Keywords: Diagnostic test, science education, secondary school, states of matter, three-tier test.

INTRODUCTION

In the science education research literature, several studies have been conducted on students' difficulties in learning about various phenomena (see Duit 2007 for a bibliography of literature on students' conceptions in science education). There has been a debate related to the term that is used to describe students' ideas of science concepts that are different from scientifically acceptable understandings. Many researchers have characterized students' ideas that are different from the definitions accepted by experts in various ways, such as misconceptions (Nakhleh, 1992), alternative conceptions (Abimbola, 1988), and children's science (Gilbert et al., 1982). Although there are some differences among these definitions, in this study, the term of misconceptions is used for students' ideas that differ from the definitions accepted by experts. There have been numerous studies indicating that students' misconceptions have considerable influence on students' learning of fundamental science concepts and the subsequent more advanced concepts (Artdej et al., 2010; Ayas et al., 2010; Gabel et al., 1987; Tytler, 2000; Voska & Heikkinen, 2000). Therefore, identification of the students' misconceptions is crucial for the planning of effective instruction and remediation of students' difficulties in understanding science concepts.

Interviews (Bou Jaoude, 1991; Griffiths & Preston, 1992; Osborne, 1980; Osborne & Gilbert, 1980), concept maps (Ingec, 2009; Kaya, 2008; Novak & Gowin, 1984), and multiple-choice tests (Hestenes et al., 1992; Ingram & Nelson, 2006) are the popular tools that are often used to identify students' misconceptions. However, these tools have several limitations. However, there are several limitations of these methodologies. For example, interviews are used for exploring students' ideas because interviews provide detailed understanding of students' conceptions. Apart from bias resulting from

Correspondence to: Zubeyde Demet Kirbulut; Harran University, Osmanbey Campus, Faculty of Education, Elementary Education, 63190, Sanliurfa/TURKEY
E-mail: demetkirbulut@yahoo.com
doi: 10.12973/eurasia.2014.1128a

ISSN: 1305-8223

State of the literature

- It is crucial to determine misconceptions of students in science education. Interviews, concept maps, and multiple-choice tests are the common tools for identifying misconceptions of students. However, these tools have several limitations.
- Two-tier tests were developed to determine the reasoning behind students' answers. Two-tier tests consisted of two parts: conventional multiple-choice question and a set of possible reasons for the given answer.
- Three-tier tests are novel and they provide us to distinguish misconceptions from lack of knowledge via an extra tier that require students to state whether or not they are sure about their answers.

Contribution of this paper to the literature

- Three-tier tests are novel in the literature.
- There are only a few studies in physics on the development and application of three-tier tests.
- There is a gap in the literature regarding the development and application of a three-tier test in chemistry. The current study describes the development of a three-tier diagnostic test to measure students' understanding of states of matter by focusing on the following research question: Is the States of Matter Diagnostic Test (SMDT) a valid and reliable instrument to measure students' conceptual understanding of states of matter?

the personal involvement of the interviewer, interviews are very time-consuming to conduct and to transcribe and analyze the data (Marshall & Rossman, 2006). Also, concept maps are useful tools for identification of misconceptions (Martin et al., 2000; Novak, 1990). Yet, concept maps have disadvantages such as the requirement of training both for teachers and students about how to use concept map and the need for a large amount of time to conduct in the classroom (Kaya, 2008). Multiple-choice tests are the other common tools often used for their better content domain sampling and mechanical scoring. They can also be efficiently administered to large samples of students (Haladyna, 1997). However, multiple-choice tests do not provide reasons for students' holding a particular conception. A student can give a correct answer with a wrong reason or a wrong answer with a correct reason.

Consequently, because of the above limitations of the aforementioned tools, two-tier multiple-choice instruments have been developed by researchers (e.g., Odom & Barrow, 1995; Peterson et al., 1986; Tan et al., 2002; Treagust, 1986; Voska & Heikkinen, 2000). The first part of each item includes a conventional multiple-choice question and the second part of each item contains a set of possible reasons for the given answer in the first part. Two-tier tests are generally superior to conventional multiple-choice tests, since they provide researchers with an understanding of students' reasoning behind their answers (Peterson et al., 1986). Hestenes and Halloun (1995) indicated that the major problem for using conventional multiple-choice tests was to minimize false positives and negatives. Students could provide correct answers with wrong reasoning as "false positives" and wrong answers with correct reasoning as "false negatives". They recommended that minimizing false positives and negatives provides a more valid test. Although a two-tier test eliminates the above-mentioned drawback of a conventional multiple-choice test, it has a limitation: It cannot differentiate misconceptions from lack of knowledge. Three-tier tests enable researchers to address this limitation by adding an extra tier that require students to state whether or not they are sure about their answers to the first two tiers (Caleon & Subramaniam, 2010; Pesman & Eryilmaz, 2010). Three-tier tests are valid tests that can be used efficiently with large samples of students, and help researchers to understand students' reasoning behind their answers without conducting interviews to distinguish misconceptions from lack of knowledge, and to estimate percentages of false positives and negatives (Kutluay, 2005; Pesman & Eryilmaz, 2010).

Three-tier tests are novel in the research literature (Pesman & Eryilmaz, 2010). There are only a few studies in physics on the development and application of three-tier tests (Caleon & Subramaniam, 2010; Kutluay, 2005; Pesman & Eryilmaz, 2010). No study on the development and application of a three-tier test in chemistry has been reported in the literature. Therefore, this study describes the development and application of a three-tier diagnostic test to measure 10th grade high school students' understanding of states of matter concepts after they were taught that subject.

States of matter is one of the crucial subjects in the 10th grade Turkish chemistry curriculum. It includes fundamental concepts such as solids and liquids, gases, evaporation, condensation, boiling, and vapor pressure which are conceptually related to each other and helpful in explaining several everyday phenomena. Students' understanding of these concepts has attracted considerable research interest over the past 30 years (e.g., Aydeniz & Kotowski, 2012; Bar & Galili, 1994; Bar & Travis, 1991; Canpolat, 2006; de Berg, 1995; Gopal et al., 2004; Johnson, 1998a, b; Novick & Nussbaum, 1978; Osborne & Cosgrove, 1983). Osborne and Cosgrove (1983) examined students' (aged 8 to 17 years old) conceptions about the changes of the states of water by using a clinical interview methodology. They found that conceptions of students of all ages are similar; some nonscientific ideas even were common among older students compared to younger students. For example, both younger and older students have the same idea that water breaks apart into hydrogen and oxygen gases when boiling. In line with this study, Gopal et al. (2004) also conducted interviews with second-year chemical engineering students to investigate their conceptions of evaporation, condensation, and vapor pressure. They found that the following misconceptions were held by students: i) evaporation and condensation require a temperature gradient, ii) evaporation only occurs in a closed system, and iii) the higher the vapor pressure, the faster the evaporation. Bar and Travis (1991) explored children's conceptions of phase changes from liquid to gas. They reported that children's understanding of boiling preceded their understanding of evaporation. Following this line of work, Tytler (2000) also found that younger children did not show greater appreciation related to evaporation and condensation in his study when he compared year one and year six students' conceptions of evaporation and condensation.

Assessment of misconceptions and conceptual understanding is very important for providing effective instruction. By taking into consideration all these issues, this study focuses on the following research question:

Is the States of Matter Diagnostic Test (SMDT) a valid and reliable instrument to measure students' conceptual understanding of states of matter?

METHOD

Sample

In the pilot study, the SMDT was administered to 195 10th grade high school Turkish students aged 15-16 years (49% females and 51% males) after they were taught about the states of matter. In the main study, the SMDT was administered to 102 10th grade high school Turkish students aged 15-16 years (60% females and 40% males) after they were taught the subject. Two types of schools - general high school and Anatolian high school - were included in the pilot study since students in these schools are known to differ in achievement. Every student could be registered to a general high school after graduation from junior high school. However, only the students who are successful in the Secondary Schools Student Selection Examination are able to register in an Anatolian high school. For convenience, three Anatolian high schools and two general high schools were selected for this study.

Procedure

The SMDT was developed using the procedures employed by Kutluay (2005), Pesman and Eryilmaz (2009), and Treagust (1986). The following five stages were pursued for the development of the SMDT: i) defining content boundaries, ii) identification of the reported misconceptions in the literature, iii) conducting interviews to explore whether or not students hold misconceptions different from the reported ones, iv) administering open-ended questions so that students' responses are categorized for writing the distracters of the items, and v) the development and administration of the SMDT for the pilot study.

The content boundaries were defined based on the Chemistry curriculum and textbooks with a list of objectives (see Table 1) that were examined by four chemistry educators and one chemistry teacher. Appropriateness of the content, confirmation of the accuracy, and content validation were established on expert agreement. Students' misconceptions were identified by examining the related literature, conducting interviews, and administering open-ended questions. The interview was semi-structured and consisted of 13 questions and follow-up probes to investigate high school chemistry students' understanding of states of matter, evaporation, condensation, boiling, and vapor pressure. The interview protocol was piloted and revised for face validity. A total of 12 interviews were conducted with each interview lasting up to 50 minutes.

In the light of the findings from the interviews and related literature, 13 multiple-choice items were

Table 1. Objectives of the SMDT

Objectives	Items
To explain the relationship between temperature and volume of an amount of gas at constant pressure. (Charles's Law).	1, 3, 5
To apply the law of conservation of mass in different contexts.	2, 8
To explain the relationship between temperature and pressure of an amount of gas at constant volume. (Gay-Lussac's Law).	4, 6
To explain the relationship between pressure and volume of an amount of gas at constant temperature. (Boyle's Law).	7, 9
To interpret evaporation operationally.	11, 12, 13
To interpret condensation operationally.	14,16
To explain vapor pressure operationally.	15, 18
To explain boiling operationally.	10, 17, 19

constructed with open-ended questions requiring reasons for the selection of a particular response to an item. Most of the questions were the same as the questions in the interview. The questions were examined by the experts (four chemistry educators and one chemistry teacher) to assure that the questions were appropriate and unproblematic, and that the objectives and misconceptions intended to be examined were assessed. The 13 questions were administered to 54 high school students in one hour lasting up 45 minutes. Students' answers to the questions were categorized and the categories with high frequencies were written as the distracters of the second tier of the items to produce 13 two-tier multiple-choice items in the SMDT. The distracters were selected from students' common misconceptions. In the third tier, the students were asked whether they were confident about their answers for the first two tiers with the aim of differentiating misconceptions from lack of knowledge.

An additional six questions were written by the researchers using the follow-up questions in the interview guide and the questions in Chemistry textbooks. The content validity of SMDT was established by the experts (four chemistry educators and one chemistry teacher) in terms of the objectives and misconceptions intended to be assessed, and whether the questions are appropriate for the grade level and unproblematic. The 19-item three-tier diagnostic test was administered to 195 10th grade students in the pilot study. In the main study, 102 10th grade students were given the SMDT. The SMDT were completed by the students in one class hour lasting up 45 minutes.

Instrument

The SMDT is a 19-item three-tier diagnostic test consisting of three-tier items for assessing students' understanding of states of matter concepts. The first tier consists of a conventional multiple-choice question with three or four choices. The second tier includes one correct reason and alternative reasons. The alternative reasons are the misconceptions identified from semi-structured interviews and open-ended questions. In addition to alternative reasons, a blank space is provided for the students to write their reasons if their reasoning is different from the given reasons. The third tier requires students to state how confident they are about their answers for the first two tiers.

The SMDT examines the conceptual areas of Charles Law (three items), Boyle Law (two items), Gay-Lussac Law (two items), conservation of matter (two items), evaporation (three items), condensation (two items), boiling (three items), and vapor pressure (two items). Example items of the SMDT are available in the appendix. The original version of the SMDT is in Turkish; however, it was translated into English in order to present the items for this journal submission. The misconceptions that were probed by the SMDT are shown in Table 2.

Data Analysis

The SMDT scores of students were typed into a Microsoft Excel datasheet. Variables were written in the columns and students names were written in the rows of the Excel datasheet. Seven variables were produced: i) one-tier scores, ii) two-tier scores, iii) three-tier scores, iv) confidence tiers, v) misconception one-tier, vi) misconception two-tier, and vii) misconception three-tier.

One-tier scores: This score was created by using students' answers for only the first tiers of items. Correct answers were coded as 1 and others were coded as 0.

Two-tier scores: This variable was based on the first two tiers of items. When a student's answer to both the first and second tiers was correct, it was coded as 1; otherwise, 0.

Three-tier scores: This score was produced by taking all three tiers into account. When a student's answer to all tiers was correct, it was coded as 1; otherwise, 0. This means that it was coded as 1 if the student answers the first two tiers correctly and s/he selects "I am sure" in the third tier.

Confidence tiers: This variable was based on students' answers to only third tiers. When a student was confident about her/his answers for the first two tiers, it was coded as 1; otherwise, 0.

Misconception one-tier: Misconception one-tier was created according to students' answers to the first tiers of items for each misconception in Table 2. When a student's answer to the first tiers was the misconceptions as indicated in Table 2, it was coded as 1; otherwise, 0.

Misconception two-tier: Misconception two-tier was based on students' answers to the first two tiers of items for each misconception in Table 2. When a student's answer to both the first and second tiers was the misconceptions as indicated in Table 2, it was coded as 1; otherwise, 0.

Misconception three-tier: Misconception three-tier was produced by considering students' answers to all tiers of items for each misconception in Table 2. When a student's answer to the first two tiers was the misconceptions and when s/he selects "I am sure" in the third tier as indicated in Table 2, it was coded as 1; otherwise, 0.

The Cronbach alpha reliability was calculated for one-tier scores, two-tier scores, and three-tier scores. Descriptive statistics of the SMDT for three-tier scores were reported (see Table 3). In addition, false negatives and false positives were calculated based on all three

Table 2. The Misconceptions Probed by the SMDT

Misconceptions	Item Choices
When heated, particles expand; when cooled, they shrink.	(1.1.A, 1.2.A, 1.3.A), (5.1.C, 5.2.B, 5.3.A), (2.1.D, 2.2.B, 2.3.A), (3.1.B, 3.2.C, 3.3.A), (6.1.C, 6.2.B, 6.3.A)
In a closed container filled with a gas, when the temperature increases/decreases, the gas pressure always increases/decreases.	(1.1.A, 1.2.B, 1.3.A), (5.1.C, 5.2.A, 5.3.A), (3.1.C, 3.2.A, 3.3.A)
Hot air is lighter than cold air.	(5.1.C, 5.2.C, 5.3.A), (2.1.C, 2.2.C, 2.2.A), (2.1.C, 2.2.D, 2.3.A), (6.1.B, 6.2.C, 6.3.A)
A gas always weighs less than a liquid (or solid).	(2.1.D, 2.2.G, 2.3.A), (8.1.B, 8.2.E, 8.3.A)
Vapor or gas has no weight.	(2.1.A, 2.2.A, 2.3.A), (8.1.A, 8.2.A, 8.3.A)
When the gas pressure increases, the weight of the gas increases.	(2.1.D, 2.2.F, 2.3.A)
In a closed container filled with a gas the volume of a gas always decreases when the temperature decreases.	(4.1.A, 4.2.C, 4.3.A), (6.1.A, 6.2.A, 6.3.A)
When air is compressed, the particles stick together.	(7.1.C, 7.2.A, 7.3.A), (9.1.A, 9.2.A, 9.3.A)
When air is compressed, the particles are all pushed to the end of the syringe.	(7.1.D, 7.2.D, 7.3.A), (9.1.A, 9.2.C, 9.3.A)
When air is compressed, the particles change their shape.	(7.1.A, 7.2.B, 7.3.A), (9.1.A, 9.2.B, 9.3.A)
When water boils/evaporates, it breaks into its components of hydrogen and oxygen molecules.	(8.1.B, 8.2.B, 8.3.A), (11.1.B, 11.2.B, 11.3.A)
In order for evaporation to take place, a liquid has to take in heat from its environment.	(12.1.B, 12.2.D, 12.3.A), (13.1.B, 13.2.D, 13.3.A)
Vaporization starts with boiling.	(12.1.B, 12.2.C, 12.3.A), (13.1.B, 13.2.A, 13.3.A)
Condensation or evaporation requires a temperature gradient.	(12.1.B, 12.2.A, 12.3.A), (13.1.B, 13.2.B, 13.3.A), (16.1.A, 16.2.B, 16.3.A)
Condensation is when air turns into a liquid.	(12.1.C, 12.2.E, 12.3.A), (13.1.C, 13.2.E, 13.3.A), (16.1.B, 16.2.C, 16.3.A)
The cold surface of a closed container and dry air react to form water on the surface via the combination of hydrogen and oxygen.	(14.1.A, 14.2.A, 14.3.A), (16.1.A, 16.2.A, 16.3.A)
When water boils and bubbles come up, the bubbles are oxygen and hydrogen, air, or heat.	(10.1.C, 10.2.A, 10.3.A), (10.1.A, 10.2.B, 10.3.A)
Boiling liquids at atmospheric pressure have different vapor pressures.	(17.1.C, 17.2.A, 17.3.A), (17.1.B, 17.2.B, 17.3.A)
Vapor pressure increases/decreases with height and this causes the water to boil at lower/higher temperatures.	(19.1.A, 19.2.D, 19.3.A), (19.1.C, 19.2.E, 19.3.A)
At constant temperature, the value of the vapor pressure changes with changes in the volume of the vapor in equilibrium with its liquid and in the amount of liquid.	(15.1.C, 15.2.A, 15.3.A), (15.1.C, 15.2.C, 15.3.A), (15.1.B, 15.2.B, 15.3.A), (18.1.A, 18.2.A, 18.3.A), (18.1.C, 18.2.B, 18.3.A)

tiers. For "false positives", if a student who was confident about the responses given to the first two tiers gave a correct response to the first tier with an incorrect reasoning in the second tier, it was coded as 1; otherwise 0. For "false negatives", if a student who was confident about the responses to the first two tiers gave an incorrect response to the first tier with a correct reasoning in the second tier, it was coded as 1; otherwise 0. Furthermore, the correlation between two-tier scores and confidence tiers was investigated for the validity of the SMDT.

RESULTS AND DISCUSSION

In this part, first, the results of the pilot study are given and then, the results of the main study are reported. Cronbach's alpha reliability coefficients of the SMDT were estimated to be .61, .70, and .78, respectively for one-tier scores, two-tier scores, and three-tier scores in the pilot study. Table 3 summarizes the descriptive statistics of the SMDT for the three-tier scores in the pilot study.

Table 3. Descriptive Statistics of the SMDT for Three-Tier Scores in the Pilot Study

Number of students	195	
Number of items	19	
Maximum possible score	19	
Mean	4.57	
Standard deviation	3.58	
Minimum	.00	
Maximum	15.00	
Skewness	.90	
Kurtosis	.19	
Reliability	.78 (for three-tier scores)	
Point-biserial coefficients	Mean	Range (items)
	.44	below .20 (1)
		.20-.29 (1)
		.30-.39 (4)
		.40-.49 (7)
		.50-.59 (3)
		.60-.69 (3)
Difficulty Level	Mean	Range (items)
	.23	.00-.09 (4)
		.10-.19 (3)
		.20-.29 (3)
		.30-.39 (8)
		.40-.49 (1)

Table 4. The Percentages of False Negatives, False Positives, and Lack of Knowledge in the Pilot Study

Items	False Negatives	False Positives	Lack of Knowledge
Item 1	1.0	41.0	20.5
Item 2	2.6	1.5	52.3
Item 3	2.1	13.3	36.4
Item 4	2.6	14.9	36.9
Item 5	1.5	54.9	23.1
Item 6	4.6	2.6	42.6
Item 7	5.6	2.1	37.4
Item 8	1.0	2.1	43.1
Item 9	1.5	15.4	44.1
Item 10	3.1	8.2	43.6
Item 11	7.7	3.1	29.7
Item 12	0.5	8.7	29.7
Item 13	2.1	9.2	41.5
Item 14	24.1	5.1	32.8
Item 15	4.6	3.6	43.1
Item 16	5.1	2.6	47.7
Item 17	1.5	1.5	50.3
Item 18	2.1	4.1	45.1
Item 19	3.1	9.7	38.5
Average	4.0	10.7	38.8

Table 3 shows that the Point-biserial coefficients except for two items (items 6 and 16) based on ITEMAN are good with an average of .44 (Ebel, as cited in Crocker & Algina, 1986). This shows that items are functioning quite satisfactorily. Ebel (as cited in Crocker & Algina, 1986) proposed that if the item-scale correlation value was greater than .40, the item was functioning quite satisfactorily. If it was between the values of .30 and .40, the item was functioning somehow good. If it was between the values of .20 and .30, the item needed revision. If it was below .19, the item should have been deleted or completely revised. Items 6 and 16 were revised after the pilot study. The Turkish grammar in these sentences was changed. It was also seen that the difficulty levels of items except one item were below .40 with an average of .23. The mean score was found to be 4.57 and the possible maximum score was 19. The skewness of the three-tier scores was found to be .90. The difficulty level and positive skewness explains the low mean value of 4.57.

In order to check the validity of the SMDT, the relationship between the two-tier scores and the confidence tier scores was investigated in the pilot study. In addition, the probabilities of false negatives and positives were calculated. The correlation between two-tier scores and confidence tier scores was examined as a quantitative approach to provide evidence for the validity of the SMDT (Cataloglu, 2002; Pesman & Eryilmaz, 2010). Cataloglu (2002) and Pesman and Eryilmaz (2010) reported that there should be at least a moderate positive correlation between two-tier scores and confidence tier scores since students with high scores are expected to be more confident than students with low scores. Pearson-product moment correlation coefficient between two-tier scores and confidence tier scores of the SMDT was calculated. It was found that there was a moderate positive correlation between two-tier scores and confidence tier scores ($r= .34$, $n= 195$, $p< .01$). The moderate positive correlation provides validity evidence in that more confident students have higher scores in the SMDT.

Hestenes and Halloun (1995) reported that minimizing the probabilities of false negatives and positives was important for validity of the test. They suggested that the probability of false negatives needs to be less than ten percentages. In addition, they added that minimizing the probability of false positives is more difficult because of the chance factor. Table 4 demonstrates the percentages of false negatives and positives in the pilot study. When the items were checked for false negatives, it was found that all the items, except for item 14, were below 10 with the average of 4.0. Item 14 is related to the condensation in an open system. When item 14 was examined, it was seen that most of the students chose one of the wrong alternatives –hot air condenses and water droplets form on the outer surface of the bottle- for the first-tier although they gave the correct answer for the second-tier. The correct answer for the first-tier is "water vapor in the air condenses and water droplets form on the outer surface of the bottle". This could have resulted from students' carelessness in that they could not differentiate between those two alternatives. When the percentage of false positives was checked, it was seen that items 1 and 5 had the highest percentages. These items were checked and no problem was found related to these items. However, this result could be attributed to the students' misunderstanding of the constant pressure in a system. The students selected correct answers for the first-tier of items 1 and 5; however, they did not give a correct reason for their answer to the first-tier since they may think that "in a closed container filled with a gas, when temperature increases/decreases, the gas pressure always increases/decreases".

In terms of lack of knowledge values, it was found that all values were high with the average of 38.8. This

Table 5. The Percentages of Misconceptions for One-tier, Two-tier, and Three-tier Scores in the Pilot Study

Misconceptions	Percentages of Misconceptions		
	Misconception one-tier	Misconception two-tier	Misconception three-tier
Misconception1	50	9	6
Misconception 2	83	31	24
Misconception 3	41	12	8
Misconception 4	32	15	8
Misconception 5	13	9	4
Misconception 6	23	11	6
Misconception 7	59	27	19
Misconception 8	40	8	3
Misconception 9	47	12	7
Misconception 10	42	10	4
Misconception 11	25	3	2
Misconception 12	27	7	3
Misconception 13	27	7	4
Misconception 14	39	16	10
Misconception 15	11	4	1
Misconception 16	66	26	19
Misconception 17	26	21	11
Misconception 18	40	29	17
Misconception 19	47	5	2
Misconception 20	30	14	7

Table 6. Descriptive Statistics of the SMDT for Three-Tier Scores in the Main Study

Number of students	102	
Number of items	19	
Maximum possible score	19	
Mean	7.96	
Standard deviation	4.30	
Minimum	0.00	
Maximum	18.00	
Skewness	0.02	
Kurtosis	-0.66	
Reliability	.83 (for three-tier scores)	
Point-biserial coefficients	Mean	Range (items)
	.49	.20-.29 (3)
		.30-.39 (2)
		.40-.49 (4)
		.50-.59 (6)
		.60-.69 (4)
Difficulty Level	Mean	Range (items)
	.42	.00-.09 (1)
		.10-.19 (2)
		.20-.29 (1)
		.30-.39 (5)
		.40-.49 (3)
		.50-.59 (4)
		.60-.69 (2)
		.70-.79 (1)

Table 7. The Percentages of False Negatives, False Positives, and Lack of Knowledge in the Main Study

Items	False Negatives	False Positives	Lack of Knowledge
Item 1	0.0	10.2	12.7
Item 2	1.0	2.0	35.3
Item 3	0.0	6.9	18.6
Item 4	4.9	8.8	33.3
Item 5	0.0	10.1	13.7
Item 6	2.0	4.9	24.5
Item 7	1.0	2.0	17.7
Item 8	0.0	3.9	23.5
Item 9	2.0	12.7	31.4
Item 10	2.0	3.9	31.4
Item 11	5.9	2.0	15.7
Item 12	1.0	7.8	14.7
Item 13	1.0	3.9	28.4
Item 14	6.0	4.9	27.5
Item 15	1.0	1.0	43.1
Item 16	8.8	2.0	21.6
Item 17	1.0	2.0	40.2
Item 18	5.9	2.9	31.4
Item 19	10.2	12.7	14.7
Average	3.8	8.9	25.2

result supported the advantage of using three-tier tests rather than conventional multiple-choice tests. Three-tier tests provide more accurate results for students' misconceptions by differentiating misconceptions from lack of knowledge. In other words, through the conventional or two-tier tests, misconceptions are overestimated since false responses due to lack of knowledge are evaluated as misconceptions.

Table 5 presents the percentages of misconceptions for one-tier, two-tier, and three-tier scores in the pilot

 © 2014 iSER, *Eurasia J. Math. Sci. & Tech. Ed.*, **10**(5), 509-521

study. This table shows that three-tier tests predict students' misconceptions more accurately compared to two-tier and conventional multiple-choice tests since three-tier tests include two-tier and confidence tier scores. The percentages of misconceptions decrease from one-tier to the three-tier scores. Two-tier tests are superior to conventional multiple-choice tests in that two-tier tests enable us to calculate false-negative scores (Hestenes & Halloun, 1995) and thus, two-tier tests could predict students' misconception scores more precisely. By the same token, three-tier tests are more advantageous than two-tier tests since they give us a chance to differentiate students' misconceptions from lack of knowledge. This result of this study corroborated the studies conducted by Kutluay (2005) and Pesman and Eryilmaz (2010).

After the items 6 and 16 were revised by changing their grammar and wording in the pilot study, the SMDT was administered to 102 10th grade students. Cronbach's alpha reliability coefficients for the main study were estimated to be .62, .73, and .83, respectively for one-tier, two-tier, and three-tier scores. Table 6 shows the descriptive statistics of the SMDT for three-tier scores in the main study.

Table 6 shows that Point-biserial coefficients based on ITEMAN are good with an average of .49 (Ebel, as cited in Crocker & Algina, 1986). This shows that items are functioning quite satisfactorily. It was also seen that the difficulty levels of items was medium with an average of .42. The mean was found to be 7.96 and the possible maximum score was 19. The mean explains the difficulty level of items. The skewness of the three-tier scores was found to be 0.02. Since the skewness value is close to 0, the distribution of the scores is nearly symmetrical. The kurtosis value is negative and this means that the distribution of the scores is rather flat.

Table 7 shows the percentages of false negatives, false positives, and lack of knowledge for the SMDT for three-tier scores in the main study. When the items were checked for false negatives, it was found that the average value of false negatives was 3.8. When the percentages of false positives were checked, it was seen that the average value of false positives for all items was 8.9. The average values of false negatives and positives were below 10 and this showed the validity of the test (Hestenes & Halloun, 1995). In terms of lack of knowledge values, it was found that the average of lack of knowledge scores was 25.2. Pearson-product moment correlation coefficient between two-tier scores and confidence tiers of the SMDT was also calculated. It was found that there was a high positive correlation between two-tier scores and confidence tiers ($r= .57$, $n= 102$, $p< .01$). The high positive correlation provides validity evidence in that more confident students have higher scores in the SMDT.

There are various reasons to use three-tier diagnostic instruments. Apart from their objectivity in scoring, broad content domain sampling, mechanical scoring, and generalizability, they have advantages in terms of enabling researchers to examine the validity of the instrument and estimate misconception scores. The correlation between two-tier scores and confidence tiers and the percentages of false negatives and false positives provide evidence for validity of the test. Three-tier tests estimate misconception scores more accurately compared to one-tier and two-tier tests since they differentiate misconceptions from lack of knowledge (Caleon & Subramaniam, 2010; Kutluay, 2005; Pesman & Eryilmaz, 2010).

It could be concluded that the SMDT provides a valid and reliable three-tier diagnostic instrument for evaluating students' misconceptions and conceptual understanding of states of matter concepts as Caleon and Subramaniam (2010); Kutluay (2005), and Pesman and Eryilmaz (2010) indicated. Furthermore, this study demonstrated that the three-tier test seems to be the most reliable one among all types of instruments since the reliability coefficients for the SMDT in the pilot study were estimated to be .61, .70, and .78, respectively for one-tier, two-tier, and three-tier scores and Cronbach's alpha reliability coefficients for the main study were estimated to be .62, .73, and .83, respectively for one-tier, two-tier, and three-tier scores.

Consequently, three-tier tests are superior to the tools such as interviews, multiple-choice tests, and two-tier tests due to their broad content domain sampling, mechanical scoring, validity evidence, and differentiation of misconceptions from lack of knowledge (Kutluay, 2005; Pesman & Eryilmaz, 2010). Further studies could use the SMDT as a tool for assessing students' misconceptions of States of Matter subject. In line with Caleon and Subramaniam's (2010) study, the SMDT could be used as pre- and post-test to assess students' understanding of the subject. Researchers may prefer the SMDT, evidently a valid and reliable diagnostic instrument, to evaluate the effectiveness of an instruction designed for helping students remediate their misconceptions they hold about the states of matter, and acquire a better understanding. In addition, with similar purposes, teachers also would like to use the SMDT. Three-tier tests provide opportunity for teachers to gain deeper insight about understandings of their students. By using the percentages of lack of knowledge, teachers can evaluate their instruction. The large percentage of lack of knowledge may mean that the instruction did not facilitate students' understanding of the related concepts. The science education research literature lacks three-tier tests. In this study, in order to measure 10th grade high school students' understanding of states of matter concepts, the SMDT was developed by the researcher. There is need for more studies to develop three-tier tests in other subject areas.

REFERENCES

Abimbola, I. O. (1988). The problem of terminology in the study of student conceptions in science. *Science Education, 72*(2), 175-184. doi: 10.1002/sce.3730720206

Artdej, R., Ratanaroutai, T., & Coll, R. K. (2010). Thai grade 11 students' alternative conceptions for acid-base chemistry. *Research in Science & Technological Education, 28*(2), 167-183. doi: 10.1080/02635141003748382

Ayas, A., Ozmen, H., & Calik, M. (2010). Students' conceptions of the particulate nature of matter at secondary and tertiary level. *International Journal of Science and Mathematics Education, 8*(1), 165-184. doi: 10.1007/s10763-009-9167-x

Aydeniz, M., & Kotowski, E. L. (2012). What do middle and high school students know about the particulate nature of matter after instruction? Implications for practice. *School Science and Mathematics, 112*(2), 59-65. doi: DOI: 10.1111/j.1949-8594.2011.00120.x

Bar, V., & Galili, I. (1994). Stages of children's views about evaporation. *International Journal of Science Education, 16*(2), 157-174. doi: 10.1080/0950069940160205

Bar, V., & Travis, A. S. (1991). Children's views concerning phase changes. *Journal of Research in Science Teaching, 28*(4), 363-382. doi: 10.1002/tea.3660280409

BouJaoude, S. B. (1991). A study of the nature of students' understandings about the concept of burning. *Journal of Research in Science Teaching, 28*(8), 689-704. doi: 10.1002/tea.3660280806

Caleon, I., & Subramaniam, R. (2010). Development and application of a three-tier diagnostic test to assess secondary students' understanding of waves. *International Journal of Science Education, 32*(7), 939-961. doi: 10.1080/09500690902890130

Canpolat, N. (2006). Turkish undergraduates' misconceptions of evaporation, evaporation rate, and vapour pressure. *International Journal of Science Education, 28*(15), 1757-1770. doi: 10.1080/09500690600779957

Cataloglu, E. (2002). *Development and validation of an achievement test in introductory quantum mechanics: The quantum mechanics visualization instrument (QMVI).* Unpublished doctoral dissertation, Pennsylvania State University, Pennsylvania, U.S.A.

Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory.* Florida: Holt, Rinehart, and Winston, Inc.

De Berg, K. C. (1995). Student understanding of the volume, mass, and pressure of air within a sealed syringe in different states of compression. *Journal of Research in Science Teaching, 32*(8), 871-884. doi: 10.1002/tea.3660320809

Duit, R. (2007). *Bibliography. Students' and Teachers' Conceptions and Science Education.* Retrieved March 28, 2008, from http://www.ipn.uni-kiel.de/aktuell/stcse/

Gabel, D. L., Samuel, K. V., & Hunn, D. (1987). Understanding the particulate nature of matter. *Journal of Chemical Education, 64*(8), 695-697. doi: 10.1021/ed064p695

Gilbert, J. K., Osborne, R. J., & Fensham, P. J. (1982). Children's science and its consequences for teaching. *Science Education, 66*(4), 623-633. doi: 10.1002/sce.3730660412

Gopal, H., Kleinsmidt, J., & Case, J. (2004). An investigation of tertiary students' understanding of evaporation, condensation and vapor pressure. *International Journal of Science Education, 26*(13), 1597-1620. doi: 10.1080/09500690410001673829

Griffiths, A. K., & Preston, K. R. (1992). Grade-12 students' misconceptions relating to fundamental characteristics of atoms and molecules. *Journal of Research in Science Teaching, 29*(6), 611-628. doi: 10.1002/tea.3660290609

Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking.* Needham Heights, MA: Allyn & Bacon.

Hestenes, D., & Halloun, I. (1995). Interpreting the force concept inventory: A response to March 1995 critique by Huffman and Heller. *The Physics Teacher, 33*(8), 502-506. doi: 10.1119/1.2344278

Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *The Physics Teacher, 30*(3), 141-151. doi: 10.1119/1.2343497

Ingec, S. K. (2009). Analysing concept maps as an assessment tool in teaching physics and comparison with the achievement tests. *International Journal of Science Education, 31*(14), 1897-1915. doi: 10.1080/09500690802275820

Ingram, E. L., & Nelson, C. E. (2006). Using discussions of multiple choice questions to help students identify misconceptions and reconstruct their understanding. *American Biology Teacher, 68*(5), 275-279. doi: 10.1894/0038-4909

Johnson, P. (1998a). Children's understanding of changes of state involving the gas state, Part 1: Boiling water and the particle theory. *International Journal of Science Education, 20*(5), 567-583. doi: 10.1080/0950069980200505

Johnson, P. (1998b). Children's understanding of changes of state involving the gas state, Part 2: Evaporation and condensation below boiling point. *International Journal of Science Education, 20*(6), 695-709. doi: 10.1080/0950069980200607

Kaya, O. N. (2008). A student-centred approach: Assessing the changes in prospective science teachers' conceptual understanding by concept mapping in a general chemistry laboratory. *Research in Science Education, 38*(1), 91-110. doi: 10.1007/s11165-007-9048-7

Kutluay, Y. (2005). *Diagnosis of eleventh grade students' misconceptions about geometric optic by a three-tier test.* Unpublished master's thesis, Middle East Technical University, Ankara, Turkey.

Marshall, C., & Rossman, G. B. (2006). *Designing qualitative research.* Thousand Oaks, CA: Sage.

Martin, B. L., Mintzes, J. J., & Clavijo, I. E. (2000). Restructuring knowledge in biology: Cognitive processes and metacognitive reflections. *International Journal of Science Education, 22*(3), 303-323. doi: 10.1080/095006900289895

Novak, J. D. (1990). Concept mapping: A useful tool for science education. *Journal of Research in Science Teaching, 27*(10), 937-949. doi: 10.1002/tea.3660271003

Novak, J. D., & Gowin, D. B. (1984). *Learning how to learn.* New York: Cambridge University Press.

Nakhleh, M. B. (1992). Why some students don't learn chemistry. *Journal of Chemical Education, 69*(3), 191-196. doi: 10.1021/ed069p191

Novick, S., & Nussbaum, J. (1978). Junior high school pupils' understanding of the particulate nature of matter: An interview study. *Science Education, 62*(3), 273-281. doi: 10.1002/sce.3730620303

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*(1), 45-61. doi: 10.1002/tea.3660320106

Osborne, R. J. (1980). Some aspects of the students' view of the world. *Research in Science Education, 10*(1), 11-18. doi: 10.1007/BF02356304

Osborne, R. J., & Cosgrove, M. M. (1983). Children's conceptions of the changes of state of water. *Journal of Research in Science Teaching, 20*(9), 825-838. doi: 10.1002/tea.3660200905

Osborne, R. J., & Gilbert, J. K. (1980). A technique for exploring students' views of the world. *Physics Education, 15*(6), 376-379. doi: 10.1088/0031-9120/15/6/312

Pesman, H. (2005). *Development of a three-tier test to assess ninth grade students' misconceptions about simple electric circuits.* Unpublished master's thesis, Middle East Technical University, Ankara, Turkey.

Pesman, H., & Eryilmaz, A. (2010). Development of a three-tier test to assess misconceptions about simple electric circuits. *The Journal of Educational Research, 103*(3), 208-222. doi: 10.1080/00220670903383002

Peterson, R., Treagust, D. F., & Garnett, P. (1986). Identification of secondary students' misconceptions of covalent bonding and structure concepts using a diagnostic instrument. *Research in Science Education, 16*(1), 40-48. doi: 10.1007/BF02356816

Tan, K. C. D., Goh, N. K., Chia, L. S., & Treagust, D. F. (2002). Development and application of a two-tier multiple choice diagnostic instrument to assess high school students' understanding of inorganic chemistry qualitative analysis. *Journal of Research in Science Teaching, 39*(4), 283-301. doi: 10.1002/tea.10023

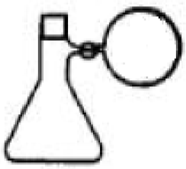
Treagust, D. F. (1986). Evaluating students' misconceptions by means of diagnostic multiple choice items. *Research in Science Education, 16*(1), 199-207. doi: 10.1007/BF02356835

Tytler, R. (2000). A comparison of year 1 and year 6 students' conceptions of evaporation and condensation: Dimensions of conceptual progression. *International Journal of Science Education, 22*(5), 447-467. doi: 10.1080/095006900289723

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. doi: 10.1002/(SICI)1098-2736(200002)37:2<160::AID-TEA5>3.0.CO;2-M

APPENDIX

The States of Matter Diagnostic Test



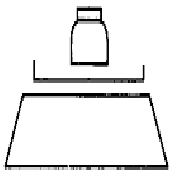
The system shown on the left figure represents a closed flask connected with an elastic balloon and filled with air at 40 ^{0}C. If the temperature of the system was decreased from 40 ^{0}C to 5 ^{0}C, how would the volume of the balloon change? (Atmospheric pressure is 1 atm for both temperatures.)

1.1
- **A.** It would decrease.
- **B.** It would not change.
- **C.** It would increase.

1.2 Which one of the followings is the reason of your answer for the previous question?
- **A.** If the temperature was decreased, air particles in the system would shrink.
- **B.** If the temperature was decreased, the pressure in the system would decrease.
- **C.** If the temperature was decreased, the pressure in the system would increase.
- **D.** If the temperature was decreased, the distance between air particles would decrease.
- **E.** Atmospheric pressure is constant.
- **F.** ……………………………………………………………………..

1.3 Are you sure about your answers for the previous two questions?
- **A.** I am sure.
- **B.** I am not sure.

2.1.

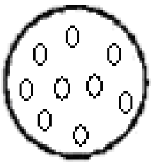


Buket weighs a closed jar filled with cold air as shown on the left figure. Then, she exposes the jar to the sun and weighs it again. Buket also knows the weight of the jar after it is vacuumed and subtracts the weight of the container from the weight results of cold and hot air in the jar. What could be said about the weight of cold and hot air in the jar?

- **A.** The weight of cold and hot air is the same and equal to zero.
- **B.** The weight of cold and hot air is the same and greater than zero.
- **C.** Cold air is heavier than hot air.
- **D.** Hot air is heavier than cold air.

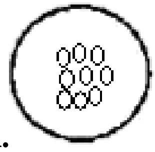
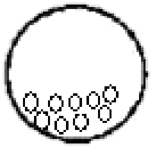
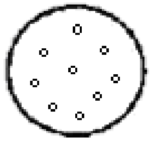
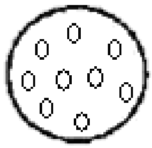
2.2 Which one of the followings is the reason of your answer for the previous question?
- **A.** Gases are weightless.
- **B.** If you heat a gas, the particles of the gas expand.
- **C.** When air is warmed up, it rises up.
- **D.** When air is cooled down, it becomes dense and its particles get closer to each other.
- **E.** There is no substance which goes out or goes into the jar.
- **F.** When a gas is warmed up, the movement and pressure of the gas particles increase and the number of collisions on the walls of the jar increases.
- **G.** Hot air is more moisturized than cold air.
- **H.** ....

2.3 Are you sure about your answers for the previous two questions?
- **A.** I am sure.
- **B.** I am not sure.



The figure shown on the left represents a closed steel tank containing hydrogen gas at 20 ^{0}C and 3 atm. The balls in the drawings represent the distribution of hydrogen molecules. Which one of the followings represents the distribution of hydrogen molecules if the temperature is lowered to -5 ^{0}C? (Note: At -5 ^{0}C, hydrogen is still a gas.)

6.1

A.  **B.**  **C.**  **D.** 

6.2 Which one of the followings is the reason of your answer for the previous question?
- **A.** If the temperature was decreased, the distance between hydrogen molecules and the volume of the gas would decrease.
- **B.** If the temperature was decreased, hydrogen molecules would shrink.
- **C.** If the temperature was decreased, hydrogen molecules would become heavy and sink to the bottom of the tank.
- **D.** If the temperature was decreased, the movement of hydrogen molecules would slow down and the pressure of the gas would decrease.
- **E.** ....

6.3 Are you sure about your answers for the previous two questions?
- **A.** I am sure.
- **B.** I am not sure.

10.1 When an amount of water is boiling, you see bubbles coming from the boiling water. What do you think that the bubbles are made of?
- **A.** Air
- **B.** Oxygen gas
- **C.** Oxygen and hydrogen gas
- **D.** Water vapor

10.2 Which one of the followings is the reason of your answer for the previous question?
- **A.** When water evaporates, it breaks into oxygen and hydrogen gas.
- **B.** There is air in water and bubbles are made of air.
- **C.** When water evaporates, the distance between water molecules increases.
- **D.** ……………………………………………………………………………..

10.3 Are you sure about your answers for the previous two questions?
- **A.** I am sure.
- **B.** I am not sure.