The Assessment of Thinking Skills in Chemistry For Secondary School Students in Malaysian Classrooms

Tajulashikin Jumahat[1], Rajendran Nagappan[2], Mohamad Sahari Nordin[1],
Suhailah Hussien[1] & Nooraini Othman[3]
[1]Kuliyyah of Education, International Islamic University Malaysia
[2]Prime Minister's Department, Malaysia
[3]UTM Perdana School, Universiti Teknologi Malaysia

Abstract:

This exploratory study developed a measure of classroom assessment called Test of Thinking Skills in Chemistry (TOTSIC) for secondary school students in Malaysia. The framework used for test specification is based on alternative taxonomy, which derived from Bloom's Cognitive Taxonomy. The alternative taxonomy of thinking had been redefined and categorized into lower-order and higher-order thinking skills. Seventy items of the initial version of TOTSIC were assembled and constructed based on this alternative taxonomy, and have been confirmed by the experts in chemistry subject. Apart from that, item analysis procedure has refined the test into forty items in the final version. The internal consistency of the overall scale as well as LOTS and HOTS items were found to be adequate for multiple-choice format. In the field test, the overall attainment of students from the test was found to be slightly below average. Unfortunately, only a small percentage of the students were having a respectable performance in HOTS items. This study suggests that TOTSIC is a viable preliminary measure for thinking skills in chemistry that can be applied as a formative assessment instrument in classroom settings. However, there is a room for improvement to assess students' level of thinking skills in chemistry primarily for HOTS items. Most importantly, the current study is advocating the immersion of higher-order thinking skills among students as stipulated in the Malaysia Educational Blueprint 2013-2025.

Keywords: Classroom Assessment, Bloom's Taxonomy, Higher-Order Thinking Skills, Science Curriculum, Chemistry

1.0 Introduction

It seems imperative that the attainment of thinking skills in students' learning is a vital aspect in enhancing their thinking capabilities (Maria, 2010). For that reason, it is the main concern among educators to help students to become effective thinkers so as to enable them to think critically, creatively, and to solve problems (Haladyna, 1997). In the Vision 2020, Tun Mahathir Mohamad anticipates the progression of knowledge among Malaysian young generations (Mahathir, 1991). The Vision 2020 calls for a culture and the promotion of thinking skills among students to produce a progressive society by the year 2020. At the moment, this assertion is still relevant due to the emphasis of thinking skills in school-based or classroom assessments (Nooraini & Khairul Azmi, 2014). It requires students to think critically and creatively, which certainly involve higher-order thinking skills, as transpired in the latest installation of the national education policy, namely Malaysia Educational Blueprint 2013-2025 (Ministry of Education, 2013).

2.0 Thinking Skills in Science Education

Scientific thinking among students is a fruitful area that has enticed many educators and psychologists to delve into its significance. In this regard, the needs to improve students' thinking skills and to master the science concepts are some of the vital aims in science education (Lawson, 1995). Indeed, generic scientific skills are applicable in science subjects (Paul & Elder, 2003). Nevertheless, it requires multifarious of thinking activities, such as reasoning and solving problems, whether in the classroom or in experimental environment (Zimmerman, 2007). In the same echo, scientific thinking framework in the science curriculum incorporates critical and creative thinking skills (CCTS) to accomplish conceptualization, problem solving, and decision making with the moderation of reasoning skills (Curriculum Development Centre, 2005). Based on this framework, science teachers are required to infuse CCTS across their science classes and ongoing classroom assessment.

3.0 Scientific Thinking in Chemistry

Learning chemistry is a complex process. It involves various thinking patterns, activities, and abilities, which are combined to comprehend its nature. Moreover, various concepts are highly related to one another. It demands certain prerequisite knowledge in order to move from one level or concept to each other, or even to grasp a simple concept (Gabel, 1998). As a matter of fact, chemistry comprises of higher order concepts, which require students to apply their higher-order thinking capabilities. According to Lawson (1995), the construction of these concepts necessitates the coordination of a large number of separate pieces of information. Thus, students would find that chemistry is one of the toughest subjects at secondary or undergraduate level (Lawson, 1995). In this regard, observation and evaluation of the development of higher-order thinking skill in chemistry need to be wisely addressed through well-planned teaching and learning, as well as in the classroom assessment.

4.0 Assessment of Thinking Skills in Science Classroom

It is undeniable that the infusion of thinking skills in a subject matter should be explicitly imparted as suggested by the curriculum experts (see Rajendran, 2010; Rosnani & Suhailah, 2003). In order to attain these goals, a systematic planning, teaching, and assessment should be prudently done. In addition, assessment could determine if the students' cognitive development has reached the level of higher-order thinking. The integration of high level of thinking questions in classroom assessment is expected to improve teaching and the learning process (Beyer, 1997). It helps educators to determine to what extent their educational activities have so far achieved their goals, as well as to improve the students' thinking abilities from the lower-order to the higher-order thinking skills.

On top of that, classroom teachers need to be competent in developing a quality assessment tool to measure students' thinking skills pertaining to cognitive progresses in students' learning. Accordingly, teachers are required to have the requisite skills in conducting a systematic classroom assessment; which in turn, produce a high quality assessment to obtain the correct information that could contribute to effective classroom instructions (Stiggins, 1992). Quellmalz (1985) pointed out that "Assessments of higher-order skills must be clear, valid, and coordinated if teachers and students are to trust, understand, and use the information they yield." (p. 34). Thus, the need for the assessment of students' thinking skills, particularly in chemistry subject is indeed relevant in science education (Tajulashikin, 2008).

The study demonstrated some scientific approaches to produce a good formative assessment in the chemistry and classroom and to assess the level of students' thinking skills, namely Test of Thinking Skills in Chemistry (TOTSIC).

The purposes of the study were:

a) Employing alternative taxonomy as the conceptual framework to develop the test specification.
b) Developing a measure for the assessment of thinking skills in chemistry.
c) Assessing the overall performance of students' thinking skills and HOTS items.

5.0 Conceptual Framework of Assessment

The choice of an appropriate conceptual framework is deemed important in educational assessment (Rankin, 1987). In fact, the Bloom's Taxonomy (Bloom et al. 1956) has become a dominant conceptual framework for classroom assessment (Kastberg, 2003). However, the categorization of Bloom's Taxonomy is still in puzzle to the practitioners (Marzano et al. 2007). One of the puzzling issues is the vague categorization of 'application'. Some researchers have classified 'application' as LOTS category (Kantasamy & Bhasah, 2006). Meanwhile, Thompson (2008) claimed that 'application' falls into both categories.

In order to overcome this concern, opinion of the curriculum authority is warranted. Therefore, this study had opted Rajendran's (1998) classification of thoughts, which evolved from the Bloom's Taxonomy. Rajendran (1998) redefined the Bloom's framework of cognitive taxonomy in view of Onosko and Newman (1994). In this regard, Bloom's criterion has been divided into two simple levels. The first two hierarchical cognitive levels, namely knowledge and comprehension, were defined as lower-order thinking skills (LOTS), whereas the latter four higher levels of thinking were specified as higher-order thinking skills (HOTS).

As stated earlier, in the national science curriculum, the infusion of critical and creative thinking is an integral part in the Chemistry Curriculum Specifications (Curriculum Development Centre, 2005). Hence, in order to design the test specifications, the researcher had to familiarize with the taxonomy

Commented [DN1]: I would suggest that you move this to the end part of your intro. By doing this, we bring the readers to our study…just before our methods.

Commented [H2]: space

(William, 1991). In doing so, extensive work was done to compare the definition of thinking skills in chemistry by the Curriculum Development Centre (2005), which has a parallel meaning with the taxonomy developed by Rajendran (1998), as well as Onosko and Newmann (1994). After a detailed analysis, Table 1 was established for parallel definitions based on the two classifications of thoughts (LOTS and HOTS) and had been endorsed by an expert in thinking skills. In this manner, the alternative conceptual framework offers a clear guideline to develop test specification for TOTSIC. The researcher also had decided to adopt four scientific thinking abilities from each LOTS (i.e., Recall, Apply Procedural Rules, Relate, as well as Compare and Contrast) and HOTS (i.e., Draw Conclusions, Inference, Analyze, and Predict) categories.

Table 1 Parallel Definitions of Thinking Skills in Science and LOTS and HOTS

Thinking Skills in Science	Description (Curriculum Development Centre, 2005)	(Rajendran, 1998)
Lower Order Thinking Skills (LOTS)		
Attributing	Identifying criteria such as characteristics, features, qualities and elements of a concept or an object.	Recall of recognize information
Comparing and contrasting	Finding similarities and differences based on criteria such as characteristics, features, qualities and elements of a concept or event.	Compare, contrast
Grouping and Classifying	Separating and grouping objects or phenomena into categories based on certain	Summarize

	criteria such as common characteristics or features.	
Sequencing	Arranging objects and information in order based on the quality or quantity of common characteristics or features such as size, time, shape or number.	Applying procedural rules
Prioritizing	Arranging objects and information in order based on their importance or priority.	Applying procedural rules
Detecting bias	Identifying views or opinions that have the tendency to support or oppose something in an unfair or misleading way.	Compare, contrast
Relating	Making connections in a certain situation to determine a structure or pattern of relationship.	Relate
Visualizing	Recalling or forming mental images about a particular idea, concept, situation or vision.	Recall of recognize information
Making analogies	Understanding a certain abstract or complex concept by relating it to a simpler or concrete concept with similar characteristics.	Relate
	Higher Order Thinking Skills (HOTS)	
Generating ideas	Producing or giving ideas in a discussion.	Produce original communications
Making inferences	Using past experiences or previously collected data to draw conclusions and make explanations of events.	Inferences
Making hypotheses	Making a general statement on the relationship between manipulated variables and responding variables in order	Draw conclusions

	to explain a certain thing or happening. This statement is thought to be true and can be tested to determine its validity.	
Making generalizations	Making a general conclusion about a group based on observations made on, or some information from, samples of the group.	Generalizations
Synthesizing	Combining separate elements or parts to form a general picture in various forms such as writing, drawing or artifact.	Produce original communications
Predicting	Stating the outcome of a future event based on prior knowledge gained through experiences or collected data.	Make predictions
Inventing	Producing something new or adapting something already in existence to overcome problems in a systematic manner.	Create
Analyzing	Examining information in detail by breaking it down into smaller parts to find implicit meaning and relationships.	Analyze
Evaluating	Making judgments on the quality or value of something based on valid reasons or evidence.	Express opinions and make choices and decisions
Making conclusions	Making a statement about the outcome of an investigation that is based on a hypothesis.	Draw conclusions

(Source: Tajulashikin, 2008)

5.1 Test Specification and Item Development for TOTSIC

Five initial topics in the Chemistry Curriculum Specifications for Form 4, namely Introduction to Chemistry, The Structure of the Atom, Chemical Formulae and Equations, Periodic Table of Elements, and Chemical Bonds (Curriculum Development Centre, 2005), were taken into consideration based on the scheme of work. Accordingly, the test specification was formed based on the topics as well as categorization of LOTS and HOTS for scientific thinking as exhibited in Table 1.

The TOTSIC scale employed multiple-choice format. Thus, the TOTSIC scale was subjected to dichotomous scoring. In fact, each item was constructed to measure an attribute based on the content and the level of thought (Grondlund, 1982). This format indeed proffers a parsimonious way of assessment in education (Mohamad Sahari, 2008). Despite of some criticisms, if the items are carefully constructed, a multiple-choice item could indeed measure higher-order thinking (Popham, 2007).

As a matter of fact, the development of multiple-choice question is a tedious and time consuming exercise (Haladyna, 1999). In order to mitigate this impediment, items were gathered from text and reference books, as well as self-constructed. For certain items, necessary modifications were required to tap the higher-order thinking of students; as students need to be given unfamiliar questions to accomplish a valid assessment of HOTS (King et al. 1998). Finally, the initial item-pool for TOTSIC consisted of 70 items and these items were allocated in the test specification based on LOTS and

Commented [H3]: delete

HOTS classification. Sample of items are presented in APPENDIX 1.

6.0 Content Validity

Content validation is an initial part of the validation process in developing a test (AERA, APA, & NCME, 1985). In this study, it involved a logical evaluation review conducted by experts of chemistry content's knowledge and curriculum. This is to determine whether the items and test are aligned to the test specification and theoretical underpinning of thinking skills in chemistry. It is also used to ascertain the sufficiency of coverage of the content areas, the appropriateness, and the representativeness of a test (Reynolds et al. 2006). This procedure enables in revising any faulty item to make necessary amendments.

The content of TOTSIC was validated by two chemistry lecturers from a higher institution. Both experts were asked to verify if the items in the TOTSIC are compatible with the chemistry content and the level of thinking skills as defined by Rajendran (1998). These items were also reviewed by them; whereby the first expert commented on the content, including the stem and response options for each item. Meanwhile, the second expert looked into the technical aspects of the item structure. As a result, both experts approved the content and the face validity of TOTSIC. Therefore, the refined version of TOTSIC is ready for pilot testing.

7.0 Methods

This study employed cross-sectional survey method. The data were analyzed by using two softwares, namely ITEMAN version 3.5 and SPSS version 11.5. It involved two phases of studies; Study 1 was meant for the pilot study and Study 2 was intended for the final test of TOTSIC.

7.1 Sample for Study 1

The main purpose of the first study was to select credible items from the students' responses. This study involved 46 students from two classes of Form 4 science, which comprised of equally distributed 23 males (50%) and 23 females (50%). The test was administered simultaneously in these two classes. The students were given 1 hour and 15 minutes to give them ample time to answer all 70 test items in the initial version of TOTSIC.

Commented [H4]: space

7.2 Item Analysis for Study 1

The data from Study 1 were analyzed using classical analysis software, namely ITEMAN version 3.5 for dichotomous scoring items (i.e. multiple chioice question). ITEMAN is a classical test theory (CTT) tool that provides comprehensive statistical item analysis and test information concerning descriptive statistics, item discrimination, test reliability, etc.

7.3 Sample for Study 2

The Study 2 data was collected from another school in the same district which shared almost similar demographic features. It involved 57 number of students from two classes in Form 4 Science. The students comprised of 18 males (31.6%) and 39 females (68.4%). The administration of TOTSIC took for about one hour to complete. The data contained the current attainment of students' thinking skills in chemistry.

8.0 Results And Discussion

Based on ITEMAN output in Study 1, the discrimination power of items was obtained from the point-biserial correlation (*r-pbis*) index that reflected the item-to-total score correlation (Quek et al. 1998). According to Reynold, Livingstone, and Wilson (2006), this method has become a dominant approach in obtaining item discrimination. The cut off point for selecting a good item is in the range of 0.30 and above which indicates a good discriminating power (Sumner, 1987).

8.1 Item Discrimination and Reliability of TOTSIC

In the first round, 37 items with *r-pbis* of 0.30 and above were retained; 19 items were from LOTS and 18 items from HOTS. Meanwhile, items that fell below this threshold value were considered as weak items or possessed low discriminating power. These items were discarded from the TOTSIC item pool as depicted in APPENDIX 2. However, three items were retained with *r-pbis* value of 0.25 and above (Hogan, 2003; Mohamad Sahari, 2008). Consequently, the final version of TOTSIC comprised of 19 LOTS items and 21 HOTS items. Nevertheless, the arrangement of 40 items in the final version of TOTSIC remained unchange. Hence, the final version of TOTSIC was ready for the final test.

Accordingly, the Cronbach's alpha reliability was obtained from the final 40 items of final TOTSIC as displayed by ITEMAN output. With regards to TOTSIC subscales, the LOTS reliability was 0.72 and the HOTS was 0.80. Whereas the internal consistency of the overall test was 0.86. This indicated that TOTSIC has adequate reliability and it is homogenous in measuring the construct of thinking skills in chemistry (Cohen et al. 2010).

8.2 Students Performance on HOTS Items

The study also observed the current status of students' HOTS which measured by TOTSIC. This will evaluate the students' performances on thinking skills in chemistry. The mean score of the final TOTSIC in Study 2 was 20.75 (*N*=57) with standard deviation of 6.61. This total average score was approximately 50 percent of the total test score (40). Based on this average score, the students were considered to have low performance in terms of thinking skills in chemistry. According to Haladyna (1997), a low average score might indicate problems in the classroom instruction such as difficulty in learning chemistry content, ineffective teaching, poor student participation, or items in the test are difficult. In addition, Gabel (1998) revealed that the difficulties faced by the students were likely due to non-scientific conception in chemistry that they acquired while learning. Hence, these possibilities need to be determined and addressed by classroom teachers respectively.

With regards to HOTS item performance, it was found that only 4% (n=2) of the students managed to answer 80% and above correctly for HOTS items (80% normally indicating A grade in Malaysian secondary schools). This indicated that only a few students had higher achievement score for HOTS items in the TOTSIC. Meanwhile, more than half of the students obtained below the average of 50% for HOTS items. This result suggested that the majority of students failed to master higher-order thinking skills or assimilate with the content in chemistry satisfactorily. In this regards, chemistry teachers should overcome this setback and at the same time they should undertake proactive measures to increase students' ability in HOTS items. This would enable students to obtain better performance in the chemistry subject. However, this requires concerted effort and it would take some time to achieve this objective primarily in science subjects (Lawson, 1995). In addition, Stiggins and Chappuis (2005) advocated that practicing student-involved classroom assessment could possibly narrow the achievement gaps; probably would lead to better overall students' performance.

9.0 Implications

The study corroborated the utility of alternative taxonomy instead of the ordinary Bloom's. Moreover, parallel analysis of the alternative taxonomy of thinking level (i.e. LOTS and HOTS), critical and creative thinking in the science curriculum have proffered a number of advantages in constructing test specification. Adaptation of items from available sources has also expedited the generation of item pool in the test. In spite of this, one should not ignore the importance of guidelines in developing multiple-choice questions (Haladyna, 2004).

Evaluation of items from content experts could enhance the soundness of items in measuring thinking skills pertaining to the content of the items. Besides, the item analysis could also improve the quality of the items and increase the validity and the reliability of the test. The integrity of the test could also be upgraded by gathering the evidence of construct validity for the purpose of classroom formative assessment in the future. Even though it is time consuming for classroom teachers to conduct formative assessments, this effort is indeed worthwhile in order to carry out a quality classroom assessment and enhance students' performance.

Apart from that, the results of the study suggesting that the students' ability in HOTS were in the range of unsatisfactory level. Therefore, in order to enhance the level of students' thinking skills generally and HOTS specifically, active learning should be encouraged in teaching and learning so as to align with the implementation of thinking skills in school curriculum. Moreover, according to Williams (2007), thinking skills needed to be taught, and explicit instruction in thinking skills must be a priority goal to all teachers. This assertion is to reprimand the notion that the teacher merely anticipates his/her students that could readily read and think critically. Using a single format of assessment, however, is insufficient to evaluate the students' level of thinking skills. Indeed, much has to be done by classroom teachers to inculcate and infuse HOTS among students with the right strategies.

Commented [H5]: skills

10.0 Conclusion

Classroom assessment is an ongoing endeavour. This study offered scientific methods for teachers to conduct a formative classroom assessment on thinking skills for the chemistry subject as part and parcel of the teaching and the learning processes. The classification of thoughts, namely LOTS and HOTS, have made the assessment framework more practical and approachable for classroom assessment. The adequacy of preliminary psychometric properties are required to obtain an accurate information on students' performance. Nevertheless, at this point of time at least, the study provided an initial information on students' level of thinking and knowledge attainment in a particular curriculum within the classroom context. This study also has proffered some ways to attain this goal; primarily in the milieu of school-based assessment in which teachers should be competent to conduct their own classroom assessment. Hence, this study is hoped to advocate the enhancement of higher-order thinking skills among students as stipulated in the National Education Blueprint 2013-2025 in Malaysian educational system.

Commented [H6]: Blueprint's

REFERENCES

AERA, APA, & NCME (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.

Beyer, B.K. (1997). *Improving student thinking: A comprehensive approach*. Boston: Allyn & Bacon.

Bloom, B. S., Engelhart, M. D., Furst, E. J., Hill, W. H., Krathwohl, D. R., & David M. (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York, NY: Longman.

Cohen, R.J. & Swerdlik, R.J. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th Ed). New York: McGraw-Hill.

Curriculum Development Centre (2005). *Curriculum Specifications Form Chemistry 4*. Kuala Lumpur: Ministry of Education Malaysia.

Gabel, D. (1998). The complexity of chemistry and implications for teaching. In B.J. Fraser & K.G. Tobin (Eds), *International handbook of science education* (pp. 233-246) London: Kluwer Academic Publishers.

Grondlund, N. E. (1982). *Constructing achievement tests*. New Jersey: Prentice-Hall.

Haladyna, T. M. (2004). *Developing and validating multiple choice test items* (3rd ed.). New Jersey: LEA.

Haladyna, T. M. (1999). *A complete guide to student grading*. Boston: Allyn & Bacon.

Haladyna, T. M. (1997). *Writing Test Items to Evaluate Higher Order Thinking*. Boston: Allyn & Bacon.

Kantasamy, S.R. & Bhasah Abu Bakar (2006). Pengujian item kemahiran berfikir aras tinggi (KBAT) dalam pengujian bilik darjah: Satu kajian kes. In Malaysia Examination Syndicate (Eds). *Humanising Assessment: Generating Excellence* (pp.1-10). Putrajaya: Ministry of Education Malaysia.

Kastberg, S. (2003). Using Bloom's Taxonomy as a framework for classroom assessment. *Mathematics Teacher*, *96*(6), 402-405.

Lawson, A. E. (1995). *Science teaching and the development of thinking*. Belmont, CA: Wadsworth.

Mahathir Mohamad (1991). *Malaysia: The Way Forward (Vision 2020)*. Kuala Lumpur: Malaysian Printing Department.

Maria Salih (2010). Developing thinking skills in Malaysian science students via an analogical task. *Journal of Science and Mathematics Education in Southeast Asia*, *33*(1), 110–128.

Marzano, R.J. & Kendall, J.S. (2007).*The New Taxonomy of Educational Objectives* (2nd ed). Thousand Oaks, CA: Corwin Press.

Ministry of Education Malaysia (2013). *Malaysia Education Blueprint 2013-2025.* Retrieved on Dis 23, 2015 from www.moe.gov.my/userfiles/file/PPP/Preliminary-Blueprint-Eng.pdf

Mohamad Sahari Nordin (2008). *Pengujian dan penaksiran di bilik darjah*. Kuala Lumpur: International Islamic University Malaysia Press.

Nooraini Othman & Khairul Azmi Mohamad (2014). Thinking skill education and transformational progress in Malaysia. *International Education Studies*, *7*(4), 27–32. http://doi.org/10.5539/ies.v7n4p27

Onosko, J.J. & Newmann, F. M. (1994). Creating more thoughtful learning environments. In J.N. Mangieri& C.C. Block (eds). *Creating powerful thinking in teacher and students: Diverse perspectives*. Fort Worth: Hart Brace College Publishers.

Quellmalz, E. S. (1985). Needed: Better methods for testing higher-order thinking skills. *Educational Researcher*, *17*(5), 5-14.

Quek, Bong Cheang & Hasny Zakaria (1998). Analisis item dan pembinaan ujian: Satu perbandingan antara pendekatan rujukan norma dan rujukan kriteria. *Jurnal Pendidikan*, *2*(2), 112–120.

Paul, R. & Elder, L. (2003). A miniature guide for students and faculty to scientific thinking. Retrieved on October 23, 2015 from www.criticalthinking.org.

Popham, W. J. (2007). *Classroom assessment: What teachers need to know* (5th ed). Boston: Allyn & Bacon.

Rajendran Nagappan (2010). *Teaching and acquiring higher-order thinking skills: Theory and practice*. Tanjong Malim: Sultan Idris University of Education Malaysia.

Rajendran Nagappan (1998). Teaching high order thinking skills in classrooms: The need for transformation of teaching practice. *Unpublished PhD Thesis*, Michigan: Michigan State University.

Commented [H7]: ...language classrooms

Rankin, S. (1987). Assessing higher-order thinking skills: issues and practices. In J.A. Arter, & J.R. Salmon, (ed). *Assessing higher order thinking skills: Issues and practices conference proceeding*. (Eric Reproduction Service: ED 290760).

Reynolds, C. R., Livingstone, R. B. & Wilson, V. (2006). *Measurement and assessment in education*. Boston: Allyn& Bacon.

Rosnani Hashim & Suhailah Hussien (2003). *The teaching of thinking in Malaysia*. Kuala Lumpur: International Islamic University Malaysia.

Stiggins, R. J. (1992). High quality classroom assessment: What does it really mean? *EducationalMeasurement: Issues and Practice, 11*(2), 35–39.

Stiggins, R. J. & Chappuis, J. (2005). Using Student-Involved Classroom Assessment to Close Achievement Gaps. *Theory Into Practice, 44* (1), 11-18.

Sumner, R. (1987). *The role of testing in school*. Berkshire: NFER-Nelson.

Tajulashikin Jumahat (2008). The development and validation of a test for the assessment of thinking skills in chemistry for students in a Malaysian secondary school. *Unpublished Master Thesis*, Tanjung Malim: Sultan Idris University of Education Malaysia.

Thompson, T. (2008). Mathematics teachers' interpretation of higher-order thinking in Bloom's Taxonomy. *International Electronic Journal of Mathematics Education*, *3*(2), 96–109.

William, J.M. (1991). *Writing quality teacher-made test: A handbook for teachers*. (Eric Reproduction Service: ED 349 726).

Williams, R. B. (2007). *Higher order thinking skills: Challenging all students to achieve*. California: Corwin Press.

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27*(2), 172–223.

APPENDIX 1

Sample of Items for LOTS and HOTS

Thinking Skills Category		Sample of Item
LOTS	Recall	Due to the stability of the inert gases I they exist as monoatomic gases II they cannot accept, share or lose electrons III they can only form bonds with one another IV their valence electron shell are completely filled with electrons A I, II and III only C II, III and IV only **B** I, II and IV only D I, II, III dan IV only
	Apply Procedural Rules	Which relation is true about molar mass of two carbon atom to its relative atomic mass? [Relative atomic mass: C=12] A 1 B 2 C 12 **D** 24
	Relate	2.8.1 is an electron arrangement of sodium. Which is true about its chemical property? A The electropositivity is decreasing down the group. **B** The electropositivity is increasing down the group. C The melting and boiling point are decreasing down the group. D The melting and boiling point are increasing down the group.
	Compare and Contrast	Which pair of elements will combine to form a compound that is not dissolve in water?

<table>
<tr><td></td><td></td><td>A Calcium and oxygen
B Carbon and chlorine
C Calcium and chlorine
D Sodium and chlorine</td></tr>
<tr><td rowspan="2">HOTS</td><td>Draw Conclusions</td><td>The rate of a chemical reaction involving a solid reactant increases by decreasing the size

The above statement is hypothesis of an experiment where 100 cm^3 of dilute hydrochloric acid is added to 10 g calcium carbonate lumps.
Which one of the following changes has to be made when repeating the experiment to test the hypothesis above?

A Increase the mass of calcium carbonate to 20 g.
B Increase the concentration of the same acid.
C Use 10 g calcium carbonate powder.
D Use 200 cm^3 of the same acid.</td></tr>
<tr><td>Inference</td><td><table><tr><th>Mass of salt added to pure water (g)</th><th>Boiling point of water (^0C)</th></tr><tr><td>0</td><td>100.0</td></tr><tr><td>3</td><td>100.5</td></tr><tr><td>6</td><td>101.0</td></tr></table>The table shows the result of an experiment. Which of the following inferences is correct.

A The boiling point of water is affected is not a constant.
B The boiling point of water is affected by the amount of salt in it.
C There is a relation between the temperature of water and the amount of</td></tr>
</table>

<table>
<tr><td></td><td></td><td>salt in it.
D The higher the boiling point of water, the larger the amount of salt dissolved in it.</td></tr>
<tr><td></td><td>Analyze</td><td>The fixed variables in this experiment are

I size of marble chip
II mass of marble chips used
III temperature of hydrochloric acid
IV concentration of hydrochloric acid

A I, II and II   C II, III and IV
B I, II and IV   D I, II, III and IV</td></tr>
<tr><td></td><td>Predict</td><td>The table below shows the melting points and boiling point of four substances, W, X, Y and Z

<table>
<tr><td>Substance</td><td>Melting points (^0C)</td><td>Boiling Points^0C</td></tr>
<tr><td>S</td><td>-191</td><td>-117</td></tr>
<tr><td>T</td><td>-53</td><td>72</td></tr>
<tr><td>W</td><td>115</td><td>443</td></tr>
<tr><td>X</td><td>80</td><td>125</td></tr>
</table>

Which of the substance is a liquid at room temperature?
A S   B T   C W   D X</td></tr>
</table>

APPENDIX 2

Item Discrimination Based on ITEMAN Output

Item No.	**Item Discrimination (*r-pbis*)**	**Remarks**	**Item No.**	**Item Discrimination (*r-pbis*)**	**Remarks**	**Item No.**	**Item Discrimination (*r-pbis*)**	**Remarks**
1	.21	Rejected	25	.10	Rejected	49	.22	Rejected
2	-.09	Rejected	26	.10	Rejected	50	.21	Rejected
3	.39	Accepted	27	.50	Accepted	51	.13	Rejected
4	-.15	Rejected	28	.38	Accepted	52	.13	Rejected
5	.04	Rejected	29	.65	Accepted	53	.21	Rejected
6	.06	Rejected	30	.37	Accepted	54	.13	Rejected
7	.30	Accepted	31	.30	Accepted	55	.39	Accepted
8	.30	Accepted	32	.42	Accepted	56	.41	Accepted
9	.00	Rejected	33	.33	Accepted	57	.32	Accepted
10	.39	Accepted	34	-.18	Rejected	58	.22	Rejected
11	.51	Accepted	35	.49	Accepted	59	.07	Rejected

12	.45	Accepted	36	.48	Accepted	60	.35	Accepted
13	.27	Rejected	37	.29	Rejected	61	.26*	Accepted
14	.33	Accepted	38	.44	Accepted	62	.00	Accepted
15	.41	Accepted	39	.41	Accepted	63	.19	Accepted
16	.57	Accepted	40	.28	Rejected	64	.17	Accepted
17	.18	Rejected	41	.38	Accepted	65	.23	Rejected
18	.30	Accepted	42	.37	Accepted	66	-.07	Rejected
19	.13	Rejected	43	.43	Accepted	67	.36	Accepted
20	.32	Accepted	44	.28*	Accepted	68	.10	Rejected
21	.39	Accepted	45	.45	Accepted	69	.04	Rejected
22	.36	Accepted	46	.17	Rejected	70	.25*	Accepted
23	.43	Accepted	47	.49	Accepted			
24	.43	Accepted	48	.37	Accepted			

Note: Items with * were accepted in the final TOTSIC with consensus