



## Use of ICT tool for Item Analysis of a Science Performance Test

**Ng Khar Thoe, Fong Soon Fook & Soon Seng Thah**

*Specialist, R&D Division, SEAMEO RECSAM, Penang, MALAYSIA*

*School of Educational Studies, Universiti Sains Malaysia, 11800 Penang, MALAYSIA*

*Educational Planning and Research Division, Ministry of Education, Putrajaya, MALAYSIA*

### Abstract

The availability of technological tools for research activities made it possible for the researcher to analyze large data effectively. This paper reviews the two stages of the development of test for multiple choice questions (MCQ) and item analysis. The technical features of ITEMAN software will also be explored for item analysis activities of an instrument namely 'Higher Order Thinking Test in Science' (HOTTIS) with illustration on construction and analysis of items. The article also reports on the pilot studies that were conducted including the procedures to develop and refine test questions as well as establish reliability and validity of the instrument. The implementation of Item Analysis using ITEMAN includes preparation of a template using SPSS software to serve as data file for the items to be analyzed and the responses by students. It will be saved as fixed ASCII file and will be opened as notepad for editing of the data file transferred from SPSS. The researcher then click on the 'ITEMAN3' application software and follow the instructions to key in files and other prompted responses. Then the output will appear that include Difficulty Index, Discrimination Index and Alpha value. A table will be prepared to summarize the findings of item analysis.

---

**Ng, K.T., Fong, S.F. & Soon, S.T. (2009). Use of ICT tool for Item Analysis of a Science Performance Test. *Malaysian Journal of Educational Technology*, 9(1), pp. 5-15.**

---

### Introduction

The advent of digital era has paved the way for the development of numerous innovative Information and Communication Technology (ICT) software applications for the storage, analysis, and display of various useful information and research data. For example, with the availability of presentation software such as graphic tools, it is possible to prepare illustrative posters that could be stored and retrievable whenever required. The invention of ICT-based research tools such as Excel, SPSS and ITEMAN software also expedite/ease the data management processes and reduce the stress of researcher to undergo the tedious tasks of having to manually analyze data collected from large samples.

This paper reports part of a bigger scale of study related to Problem-based Learning (PBL). The main aim of this article is to explore the technical features of ITEMAN software for item analysis activities of an instrument namely 'Higher Order Thinking Test in Science' (HOTTIS) with the following the objectives:

- To illustrate the construction and analysis of items in HOTTIS that were devised to evaluate higher cognitive thinking levels of students responding to the secondary science topics on 'Water and Solution'.
- To report on the pilot studies conducted with the procedures to establish reliability and validity of HOTTIS using ITEMAN software as an example of ICT tool for item analysis.

### Literature Review

*Higher-order thinking* (HOT) requires students to manipulate information and ideas in ways that *transform* their meaning and implications. This transformation occurs when students combine *facts* and ideas in order to *explain* what they understand or comprehend, to *analyse*, *apply*, *synthesise*, generalise, hypothesise or arrive at some conclusion or interpretation. Such thinking may be referred to as higher cognitive thinking levels (Bloom, Hasting & Madaus, 1971). HOT may also involve integrative thinking including synthesis, creative/critical thinking, metacognition (e.g. monitoring, *evaluating* and revising own thinking) and decision making (based on information gathered, e.g. good/bad points, emotions and organization of thinking). Other types of HOT include logic, perception (and lateral thinking), intuition (e.g. aesthetic, complex judgment, and internalized processes) as well as metaphorical thinking or use of

analogies. There is also thinking concerned with description (such as verbal fluency), and thinking concerned with action (with getting things done) (Bloom, et al., 1971; De Bono, 1985; Sale, 2008; Sill, 2001; Torrance, 1974) e.g. problem-solving activity in PBL.

Literature revealed that numerous studies were conducted to evaluate students' performance in problem-solving classrooms and motivation towards science related studies (e.g. issues about water and environmental sustainability) using psychometric tests, e.g. by Samsudin, Osman and Halim (2007), Tan, Ng, Ch'ng and Teoh (2007), Toh, Yeap, Ng and Isma (2007), and so forth. According to Carter (2005), psychometric tests can be broadly divided into two main categories, i.e. tests of 'maximum' performance and potential in a number of areas, such as ability or aptitude; and tests of 'typical' performance, such as personality or interest. '*Ability*' is defined as 'what the student is able to demonstrate in the present', '*achievement*' is 'what the students have accomplished in the past' and '*aptitude*' means 'how quickly or easily the student will be able to learn in the future'. Students' abilities/performances can be measured separately, or combined to give an assessment of overall general ability. Often tests are constructed so that they relate to a specific job or skills and assess things such as perceptual speed or mechanical reasoning. Examples of 'abilities tests' are knowledge-based *attainment* tests, general intelligence or IQ tests (Eysenck, 1979), and *aptitude* tests that test the ability to use knowledge. It is important to take note that in order to assess specific types of thinking in problem-solving activities, the researcher needs to first define the types of thinking to be assessed, then identify appropriate sources of performance evidence, also produce a valid and practical scoring system as suggested by Sale (2008).

In building a performance test, it is essential to: (1) outline content areas to be covered to ensure content validity; (2) try the test out to obtain data for an item analysis in terms of item difficulty and discriminability. The researcher could also consider that sometimes it does not hurt to retain an extremely easy item. Since it is uniformly easy, it may have motivational value, and because of its uniform ease for both low and high scorers, it will not affect the relative position of scores. For a *performance test*, the establishment of *content validity* (a non-statistical concept) and the use of *item analysis* are usually *sufficient* for test construction. Attempts to establish forms of validity other than content validity (e.g. construct or concurrent validity) are usually unnecessary for *tests of performance*, although it is useful to establish their reliability (Tuckman, 1972).

Item analysis is a procedure to increase the reliability and validity of a test by separately evaluating each test item to determine whether or not that item discriminates in the same way the overall test is intended to discriminate. There are many ways to conduct item analysis, either using formula to calculate manually [as suggested by e.g. Tuckman (1972) and Isaac and Michael (1981)] or using ICT tool such as ITEMAN software. Item analysis program or sometimes being referred to as 'Item and Test Analysis Program' (ITEMAN) is a true classical item analysis software veteran with an origin dating back to the 1970s. Using this software, the output is given in flax text files. ITEMAN has been used for classical item and test analysis for many years and its use remains widespread (Nelson, 2008; Qaqish, 2008).

The most basic level of Item Response Theory (IRT, also known as latent trait theory, strong true score theory or modern mental test theory) that is based on the idea that the probability of getting an item correct is a function of a latent trait or ability to provide explanation on the process of item analysis. According to IRT, a person with higher intelligence would be more likely to correctly respond to a given item on an intelligence test (Wikipedia, 2008). After item analysis, the researchers should be able to see, for example, given a test of academic aptitude or achievement, will students in the top half of the test score distribution also tend to get a particular item correct and students in the bottom half of this distribution tend to miss this same item (Isaac & Michael, 1981). If there are evidences of the aforementioned, such an item has favorable discriminating power and contributes to the overall reliability and validity of the test. This leads to the importance of analyzing the 'difficulty' of the item and its 'discriminability'.

'Difficulty' refers to the extent to which a test item can be responded to correctly by any student. It differs from 'discriminability' in that it does not include any differentiation between students (Tuckman, 1972). An item's difficulty index ( $p$ ) is usually computed for cognitive measures e.g. aptitude and achievement tests. "It is simply a tally for each item of the number of subjects who answered the item correctly, divided by the total number of subjects taking the test" (Borg & Gall, 1989, p.275). It is not a property of the item alone but a function of three things, namely, the complexity of the item, the ability of the pupils in



responding to it, and the nature of instruction preceding the item's use. It was found that a good spread in test results can be obtained if the difficulty index of the test items ( $p$ ) range from 0.20 to 0.80, with a mean value of about 0.50 to 0.60. For example, if the item shows difficulty index ( $P$ ) less than 0.20, it is considered easy item as fewer students got them wrong. For most norm-referenced achievement measures the ideal item difficulty level is 0.50, i.e. half the subjects respond correctly. An exception is when the test has been designed to discriminate among subjects at a given cutoff point (Bloom, et al., 1971; Borg & Gall, 1989; Tuckman, 1972). Of course, the researchers can have a certain degree of flexibility in setting the cutoffs for discriminability and difficulty; in choosing to exercise this flexibility, they should understand what these values mean and what they are doing to their test as they change the items (Tuckman, 1972). Another important determinant of difficulty is the set of possibilities for solving the problem as seen by the problem solver (Eysenck, 1982).

'Discriminability' refers to the extent to which a test item is responded to correctly by those students possessing more of the quality being measured and incorrectly by those students possessing less of this quality (Fyffe, 1971; Robison, 1974; Tuckman, 1972). The item discrimination index ( $D$ ) is meant to communicate the power of an item in separating the more from the less capable on some latent attribute (Wood, 1988 in Keeves, 1988). It describes the extent to which a particular test item is able to differentiate the higher scoring pupils from the lower scoring pupils. The discrimination index ( $D$ ) is the difference between the fraction of the upper group (top 27% higher achievers) answering the item correctly and the fraction of lower group (bottom 27% low achievers) answering it correctly. The index can range from -1.00 to +1.00. Irrespective of the method used, the maximum possible value for  $D$  is +1.0, i.e. all the good students get the question right and all the poor students get it wrong, and vice versa when there is a minimum value of -1.0. If all the students get the question right or if they get it wrong, then the value for  $D$  is 0. Items with discrimination index of 0.35 or more indicate good items that discriminate well. Whereas the discrimination index between 0.30 and 0.39 are reasonable, between 0.20 and 0.29 are acceptable or marginal. Items with  $D$  below 0.19 must be rejected as being not suitable (discriminates poorly). Operationally, the usual procedure is to compute a correlation, either biserial or point biserial, between success and failure on the item and score on a measure that is considered to represent the latent *attribute*. This is typically and conveniently the total score on the test to which the item belongs. The idea is that the higher the correlation between candidates' scores on an item and their scores on the test, the more effective the item is in separating them. But a negative value for  $D$  indicates that there is something wrong with the question and it should either be rewritten or abandoned (Fairbrother, 1988; Wood in Keeves, 1988).

### Methodology

This section will elaborate on the research activities that include the construction and analysis of items in "Higher Order Thinking Test in Science" (HOTTIS) as well as the pilot studies conducted with the implementation of item analysis using ICT tool to establish reliability and validity of the instrument.

### Constructing a paper-and-pencil performance test

HOTTIS was constructed as a performance test to evaluate higher cognitive thinking levels of Form 2 students' (ages about 13 or 14) mastery of basic General Science knowledge and their responses to the items in 'Water and Solution'. The following is a brief illustration of the process in devising HOTTIS.

The researcher first identified those performances to be measured. A list of instructional objectives for the subject area were first constructed. Then all those performance capabilities that students should possess if successfully taught were outlined. On the basis of this content outline, the researcher then wrote performance items and develop scoring keys [refer Table 1(a)]. The use of *content outline* [refer Table 1(b)] should establish *content validity* in that test items based on the outline assess or sample the mastery of that content which they are intended to test. (Tuckman, 1972). More items were initially generated in each specific content area. The tests items were constructed by a team of 3 content experts with more than 10 years of teaching experience in science. References were made from relevant sources such as the Form 2 Science textbooks and past-year PMR questions (Hobson, 2007; Lau, 2006; Ben & Leow, 2006; Ooi & Voon, 2007; Tan & Hasroni, 2007). In these items the specific examples used in teaching the content were modified to make possible the *measurement of transferability of skills* rather than rote learning, also the language used is clear and can be understood by the secondary students. As such, before the administration of HOTTIS to a pilot group, the test items were read by a science teacher who had given advice that the test should include easier items as what had been done in most of the school tests.

After careful consideration with the use of Table of Specification [refer Table 1(a)], the first draft of HOTTIS consisted of 30 selected multiple choice items. HOTTIS was made as 'achievement' or knowledge-based attainment test for the evaluation of higher levels of cognitive thinking skills, i.e. 'comprehension, application, analysis, synthesis and evaluation' according to Bloom's taxonomy of cognitive domain (Bloom, et al., 1971). The content of HOTTIS was made compatible to the cognitive levels of learning and thinking skills (Fairbrother, 1988) by ensuring that the content tested in the items was taken from what the students had learnt from Chapter 5 of the Form 2 topic on 'Water and Solution'. However, it was later found from content analysis that these items were mostly designed to assess more on the basic levels of 'comprehension and application' but less in other higher levels based on Bloom's taxonomy of cognitive domains. The researcher had later included two more subjective questions (with multiple items allotted in sub-sections) that provided focus mainly on evaluating 'analysis, synthesis and evaluation'. As such, the edited and final versions of HOTTIS (after pilot study 1 and 2) consists of two sections [refer Table 1(a)].

The final version of the test was planned to take about 1 ½ hours for students to complete. **Section A** consists of 25 items [i.e. multiple-choice question (MCQ) ranges from 'A to D' with either single choice or combination of choices responses] that contributed to 50 percent of the total marks or test scores. Whereas **Section B** consists of 2 items (with questions in essay type of format) requiring open-ended responses by students and contributed to 50 percent of the total marks. It aims at mainly evaluating students' HOTS at the levels of 'analysis, synthesis and evaluation' based on Bloom's taxonomy of cognitive domains.

**Table 1(a)** Table of Specification (TOS) for Questions and Cognitive Levels of Learning for 30 Items (Pilot Study 1) and 25 Items (Pilot Study 2) for "Higher Order Thinking Test in Science" (HOTTIS) on Form 2 topic 'Water and Solution'

Higher Cognitive Thinking Level	MCQ (Question number, with two marks each)		Essay Questions (Question number and marks)		Total Marks	
	Pilot study 1	Pilot study 2	Pilot study 1	Pilot study 2	Pilot study 1	Pilot study 2
Comprehension	1, 2, 5, 12, 13, 16, 20	1, 2, 3, 9, 10, 12, 16	-	-	14	14
Application	7, 9, 11, 14, 15, 18, 19, 23, 26, 30	5, 7, 8, 11, 13, 14, 18, 21, 25	-	-	20	18
Analysis	3, 4, 6, 8,	4, 6, 19	1 (20 marks)	1 (25 marks)	10+	6+
Synthesis	10, 21, 24,	20, 23, 24	And	and	10+	6+
Evaluation	25, 28, 29, 17, 22, 27	15, 17, 22	2 (20 marks)	2 (25 marks)	6+	6+
<b>Total</b>	<b>30</b> (60% of total score)	<b>25</b> (50% of total score)	<b>2</b> (40% of total score)	<b>2</b> (50% of total score)	60%+40% =100%	50%+50% =100%

**Table 1(b)** Content Outline for HOTTIS on Form 2 topic 'Water and Solution'

Main learning or content areas	Sub-topics or specific content area
Physical characteristics of water	Physical properties of water; Effect of impurities on water
Composition of water	Electrolysis of water
Evaporation of water	Evaporation and boiling; Factors affecting evaporation; Applications of evaporation
Solution and solubility	Solute, solvent and solution; Types of solution/suspension; Solubilities of substances; Factors affecting solubility and rate of dissolving; Water as a solvent; Organic solvents
Acids and alkalis	Properties of dilute acids; Properties of alkalis; Acidic and alkaline substances in daily life; Uses of acids and alkalis; Neutralization and its application



---

<i>Methods of water purification</i>	Natural sources of water; Reasons for purification; Methods of purification
<i>The water supply system</i>	Treatment of drinking water; Ways of saving water
<i>Preserving water quality</i>	Water pollutants and effects on living things; Controlling water pollution; Preserving the quality of water

---

### Research activities with data collection and analysis

Pilot studies of HOTTIS were conducted twice in May and July 2008 with the procedures to establish its reliability and validity. More detailed information on how ITEMAN software was used for item analysis will be provided in the subsequent section. The output of item analysis is summarized in Table 2.

To ensure that HOTTIS and the answer scheme are valid to measure thinking skills and academic performances, the content of the paper was validated by a panel consisting of one senior Biology teacher, one science teacher and one science educator. All of them worked independently and were given a set of the test instrument. Instructions and operational definitions were given to ensure that equitable amount of test items were prepared to evaluate various cognitive levels of learning (e.g. 'understanding' to measure 'comprehension' domain, 'explaining' to evaluate 'application' domain, and so forth). The panel was also required to ensure that the content of the test papers relevant to promote thinking skills (especially 'analysis, synthesis and evaluation' levels) that are required for problem-based science learning. The panel was also requested to give comments or advice with suggestions for the amendment or editing of the test items. Based on a minimum of 70% of mutual agreement among the panel members on the *content validity*, the test items were accepted. Some of the test items were also revised and some were reconstructed based on their comments and suggestions of improvement. The cycle of field test and revision were also repeated at least two times in order to develop effective instruments.

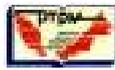
After the revision of the HOTTIS, the tests were pilot-tested the second time before these will be used later for real studies. Total scores on such measures as number of items passed was obtained for each student in the pilot group. Performance on each item was then be compared to total scores using the *item analysis* procedures. The first draft of HOTTIS for 'Water and Solution' with 30 multiple-choice questions (MCQ) items were first pilot-tested with more than 59 students who had already learned the topics in Form 2. The students' responses to the items in the first draft of the test were collected and used in item analysis to compute the *difficulty index* ( $p$ ) and *item discrimination index* ( $D$ ). The second draft contains 25 MCQ items and the instrument was administered among 88 students who had studied the topic in Form 2. The ITEMAN software was used to analyse the items on the test instrument [Refer *Figure 1*] to substitute the item analysis that could be conducted manually.

### Item Analysis Activities via ICT Tool

#### The implementation of Item Analysis and technical features of ITEMAN

The following is a brief description of steps involved in item analysis (pilot 1) using ITEMAN software:

- (1) Prepare a template using SPSS Statistics software to serve as data file for the items to be analyzed (e.g. labeled as 'q1 to q30' in 'variable' column) and the responses by students (e.g. labeled as '1 to 50' in 'id' column), both of which stored in 'Data View'.
- (2) Under the 'Variable View' of the same template, the 'Name' column will include No.1 'id' and No.2 to No.31 'q1 to q30' respectively. The word 'Numeric' is typed under 'Type' column, followed by column width '1' under 'Width' (except the first 'id' with '3') and '0' for all the rows under 'Decimal' column. Under the 'Label' column, except the first written as 'Examinee's ID', the rest of the rows will be keyed in with the answers, e.g. if the answer for the first question is (A), then the 'Label' column will be typed with '1 answer of first question'. Under the 'Values' column, except the first row to be typed



'None', the rest of the rows for 'q1 to q30' will be keyed in with '1=A, 2=B, 3=C, 4=D' if there are four choices. Under 'Missing' column, 'None' to be typed in all rows and for 'Align' column, 'Right' to be typed in all rows. Apart from first row to be typed with '3' under 'Columns' and 'Scale' under 'Measure', the rest of the rows will be typed with '1' under 'Columns' and 'Nominal' under 'Measure'.

- (3) After completing steps (1) and (2), this SPSS data file will be saved as 'Fixed ASCII' (\*.dat)
- (4) Proceed with 'File' → 'Open with' → 'Notepad' → To 'Edit' the data file transferred from SPSS.
- (5) The following is **the** sample output for the data entered in 'Notepad' for 'WaterNSolutionx(edited).dat'

<u>Output appears in the Notepad</u>	<u>Explanation for the function of each row</u>
Δ30ΔNΔ1ΔΔ3	Space, <b>30</b> questions,Space, <b>N</b> ,Space, <b>1</b> ,SpaceSpace, <b>3</b>
143432321312442334421233434342	Answers for q1 to q30
44444444444444444444444444444444	Number of choices (e.g. 1 to 4 labelled for A to D)
YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY	To analyze? 'Yes'
001123332321222432211421231234341	SPSS data collected from the respondents '1 to 59'
002424432211242412334123434241231	.....
.....	.....
059223422221234412343343242344333	.....

- (6) Click on 'ITEMAN3' application software, then follow the instructions to key in files, e.g.
 

Enter the file name of the input file:	<u>D:\1AAITE????~1\WATERN????~1.DAT</u>
Enter the name of the output file:	<u>D:\1AAITE????~1\WATERN????SO.OUT</u>
Do you want to score?	<u>Yes</u>
Enter the name of the score file:	<u>D:\1AAITE????~1\WATERN????SO.SCR</u>
Do you have a key exception file?	<u>No</u>
Do you want to write statistics to an external file?	<u>No</u>
- (7) Then 'Continue' and respond to the following:
 

Type of Correlations: <input checked="" type="checkbox"/> Point-Biserial	Correct Correlations for Spuriousness: <input checked="" type="checkbox"/> No
Ability Grouping: <input checked="" type="checkbox"/> Yes	Express Endorsements (Difficulties) As: <input checked="" type="checkbox"/> Proportions
Subgroup Analysis: <input checked="" type="checkbox"/> No	Score Distributions/Histograms: Interval Width (1)

Then 'Continue' [refer *Appendix F* for the print screen of the use of ITEMAN software for Item Analysis on Higher Order Thinking Test in Science (HOTTIS) to evaluate HOTS]
- (8) Then the output will appear, e.g. (a) Prop Correct (Difficulty) (b) Discrimination (Disc.) Index and (c) Alpha > 0.71 and other details in 'WATERN.SO.OUT' file that can be viewed in Notepad.
- (9) Prepare a table (Refer **Table 2**) to summarize the findings of the items analysis.

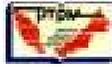
**Figure 1** Summary of item analysis procedures using ITEMAN software



Table 2

*Difficulty and Discrimination Indices of first (n<sub>1</sub>=30) and second (n<sub>2</sub>=25) piloted items in HOTTIS for 'Water and Solution'*

Item No.	Cognitive Level	Thinking Skills	First pilot study with 30 items		Item No.	Cognitive Level and/or Thinking Skills	Second pilot study with 25 items	
			Difficulty Index (p)	Item Discrimination Index (D)			p	D
1	Comprehension	Understanding	0.86	0.16	1	Comprehension (Understanding)	0.88	0.22
2	Comprehension	Sequencing	<u>0.15*</u>	<u>-0.11*</u>	2	Comprehension (Sequencing)	0.30	0.28
3	Analysis	Differentiating	0.34@	0.77	-	-	-	-
4	Analysis	Categorizing	0.44@	0.43	-	-	-	-
5	Comprehension	Understanding	0.81	0.44	3	Comprehension (Understanding)	0.84	0.32
6	Analysis	Categorizing	0.59	0.55	4	Analysis (Categorizing)	0.69	0.58
7	Application	Explaining	0.59	0.72	5	Application (Explaining)	0.66	0.57
8	Analysis	Classifying	0.41	0.20	6	Analysis (Classifying)	0.45	0.53
9	Application	Explaining	0.69	0.61	7	Application (Explaining)	0.71	0.43
10	Synthesis	Estimating	<u>0.15**</u>	<u>-0.17**</u>	-	-	-	-
11	Application	Solving	<u>0.20*</u>	<u>-0.04*</u>	8	Application (Solving)	0.38	0.30
12	Comprehension	Defining	0.68	0.56	9	Comprehension (Defining)	0.71	0.37
13	Comprehension	Defining	0.69	0.78	10	Comprehension (Defining)	0.69	0.61
14	Application	Clarifying	<u>0.42*</u>	<u>0.14*</u>	11	Application (Clarifying)	0.47	0.59
15	Application	Calculating	0.69@	0.55	-	-	-	-
16	Comprehension	Identifying	0.46	0.37	12	Comprehension (Identifying)	0.50	0.54
17	Evaluation	Making judgement	0.42	0.42	15	Evaluation (Making judgement)	0.50	0.39
18	Application	Classifying	<u>0.25*</u>	<u>0.07*</u>	13	Application (Classifying)	0.31	0.17
19	Application	Explaining	0.34	0.48	14	Application (Explaining)	0.39	0.51
20	Comprehension	Defining	0.63	0.37	16	Comprehension (Defining)	0.66	0.28
21	Synthesis	Making inference	0.47@	0.55	-	-	-	-
22	Evaluation	Making judgement	0.73	0.72	17	Evaluation (Making judgement)	0.76	0.44
23	Application	Solving	0.56	0.66	18	Application (Solving)	0.55	0.51
24	Analysis	Classifying	<u>0.10*</u>	<u>-0.11*</u>	19	Analysis (Classifying)	0.31	0.26
25	Synthesis	Interpreting	<u>0.17*</u>	<u>0.06*</u>	20	Synthesis (Interpreting)	0.29	0.44
26	Application	Categorizing	0.27	0.25	21	Application (Categorizing)	0.43	0.42
27	Evaluation	Making judgement	0.53	0.31	22	Evaluation (Making judgement)	0.52	0.46



28	Synthesis	Making inference	0.66	0.38	23	Synthesis (Making inference)	0.69	0.57
29	Synthesis	Making inference	0.42	0.94	24	Synthesis (Making inference)	0.56	0.72
30	Application	Explaining	<u>0.15</u> *	<u>0.01</u> *	25	Application (Explaining)	0.34	0.36
<b>Note</b>	Number of items ( $n_x$ ) :		<b><u>Pilot study 1</u></b>		<b><u>Pilot study 2</u></b>			
	Number of examinees ( $N_x$ ) :		$n_1 = 30$		$n_2 = 25$			
	Alpha ( $\alpha_x$ ) :		$N_1 = 59$		$N_2 = 80$			
	Decisions made for pilot study :		$\alpha_1 = 0.747$		$\alpha_2 = 0.826$			
			**Item is deleted due to poor p or D		All items were accepted and used in the subsequent studies as $0.2 \leq p \leq 0.8$ and $D > 0.16$			
			* Item is refined and adapted for pilot study 2					
			@ Item is deleted based on the suggestion by teachers.					

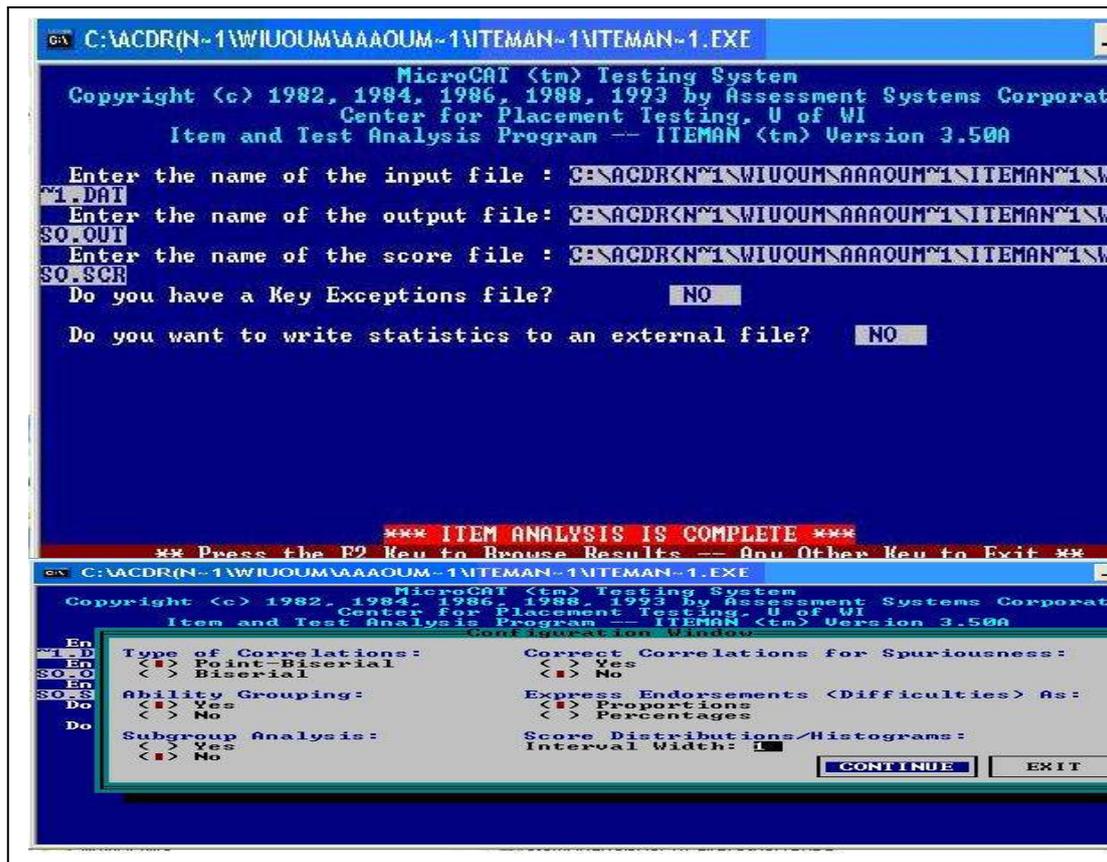


Figure 2 Printscreens pictures of item analysis using ITEMAN software

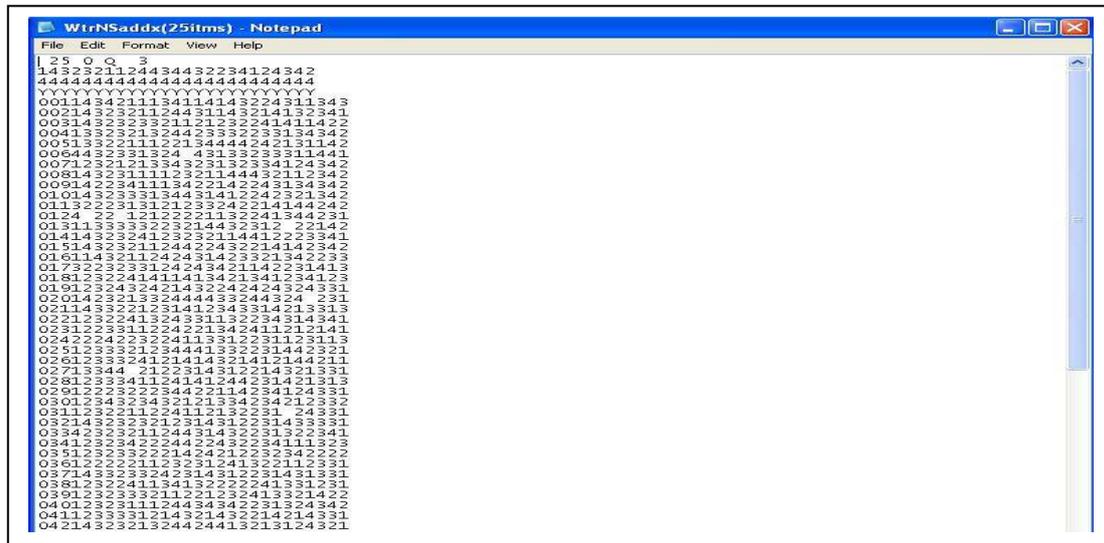


Figure 3 Printscreens pictures of sample input ITEMAN DAT file transferred to notepad from pilot study 2

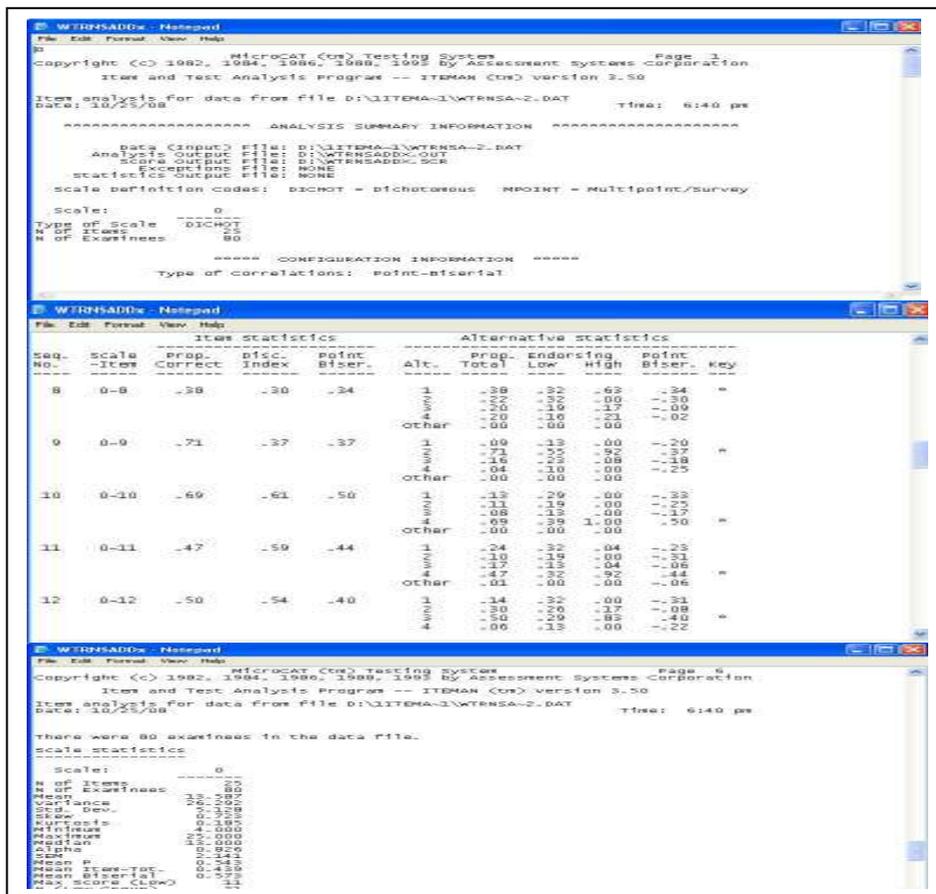


Figure 4 Printscreen pictures of sample item analysis output from pilot study 2 using ITEMAN software

### Conclusion

Although ITEMAN is useful to analyze objective data for paper-and-pencil test, some see its interface as dated and it may be found not so practical among teachers due to the inaccessibility of software and the need to follow lengthy steps. Alternatively item analysis formula could also be prepared in the MS-Excel which can be more easily accessible or Excel-based LERTAP software that is easy to operate. Our increasingly information-based societies is very much dependent on the availability of ICT that change the ways how we learn, do research, manage our activities and reach out to others. The technologies fuelling advances like these will continue to move forward at breakneck speeds and it is worth exploring various ICT tools to optimize learning and assessment techniques.

### Acknowledgement

The first author wish to acknowledge the Ministry of Education Malaysia for granting the research pass to conduct studies in local schools; the Principal, teachers and students involved in the pilot studies of the HOTTIS instrument. Special thanks also to the input given by Ms. Li Xuanxi on the use of ITEMAN software in item analysis of this study.

### References

- Assessment Systems Corporation (2006). *User's Manual for the ITEMAN Conventional Item Analysis Program*. Manual for the 32-bit Windows Version 3.6. St.Paul,Minnesota. [URL: [www.assess.com](http://www.assess.com)]
- Ben, M.E., & Leow, C.S. (2006). *Science Form 2 Questions and Answers based on New Exam Format*. Pustaka Yakin Pelajar Sdn. Bhd.
- Bloom, Hasting & Madaus (1971). *Handbook of formative and summative evaluation of student learning*. New York: McGraw Hill.



- Borg, W.R., & Gall, M.D. (1989). *Educational research: An introduction*. Fifth edition. NY: Longman Inc.
- Carter, Philip (2005). *The complete book of intelligence tests: The IQ workout series*. England: John Wiley & Sons Ltd. [URL: <http://www.scribd.com/doc/3150576/Complete-book-of-intelligence-test>]
- De Bono, Edward (1985). *Tactics: The art and science of success*. UK: William Collins.
- Eysenck, H.J. (1979). *The structure and measurement of intelligence*. With contributions by David W. Fulker. NY: Springer-Verlag Berlin Heidelberg.
- Eysenck, H.J. (Ed.) (1982). *A model for intelligence*. NY: Springer-Verlag Berlin Heidelberg.
- Fairbrother, R.W. (1988). *Methods of assessment*. Longman group UK Ltd.
- Fyffe, D.W. (1971). *The development of test items for the integrated science processes: Formulating hypotheses and defining operationally*. Unpublished doctoral dissertation. Michigan State Univ.
- Hobson, A. (2007). *Physics: Concepts and connections*. 4<sup>th</sup> Edition. New Jersey: Pearson Prentice Hall.
- Isaac, S. & Michael, W.B. (1981). *Handbook in research and evaluation for education and behavioural sciences*. Second edition. San Diego, California: EdITS Publishers.
- Keeves, J.P. (Ed.) (1988). *Educational research, methodology and measurement: An international handbook*. Oxford: Pergamon Press.
- Lau, H.Y. (2006). *Science Facts in Graphics: Total Revision for Form 1, 2 and 3 with latest PMR Format*. Petaling Jaya: Sasbadi Sdn. Bhd.
- Nelson, L. (2008). ITEMAN and Lertap 5. Curtin University of Technology. [URL: [www.lertap.curtin.edu.au](http://www.lertap.curtin.edu.au)]
- Ooi, S.T., & Voon, F.L. (2007). *Science sample copy for PMR*. Oxford Fajar Sdn. Bhd.
- Qaqish, B. (2008). *Developing Multiple Choice Tests for Social Work Trainings*.
- Robison, R.W. (1974). *The development of items which assess the processes of controlling variables and interpreting data*. Unpublished doctoral dissertation. Michigan State University.
- Sale, D. (2008). *Assessing specific types of thinking in problem-based learning activities*. [URL: <http://pbl.tp.edu.sg/Others/default.aspx>, articles related to PBL]
- Samsudin, M.A., Osman, K. & Halim, L. (2007). *Can Problem-based Learning create a science culture in the classroom?* Published in CoSMEd 2007 conference proceedings. RECSAM.
- Sill, David J. (2001). Integrative thinking, synthesis and creativity in interdisciplinary studies. *The Journal of General Education*, Volume 50, Number 4, 2001, pp.288-311. [URL: [http://muse.jhu.edu/login?uri=/journals/journal\\_of\\_general\\_education/v050/50.4sill.html](http://muse.jhu.edu/login?uri=/journals/journal_of_general_education/v050/50.4sill.html)]
- Tan, K.A., Ng, K.T., Ch'ng Y.S., & Teoh, B.T. (2007). *Redefining mathematics classroom incorporating global project/problem-based learning programme*. Paper presented in CoSMEd'07 conference, 13<sup>th</sup>-16<sup>th</sup> Nov.
- Tan, S.H.A., & Hasroni, H. (2007). *Sample Copy of Science for PMR (Koleksi Kertas Peperiksaan Sebenar)*. Shah Alam: Cerdik Publications Sdn. Bhd.
- Toh, L., Yeap, C.H., Ng, K.T., & Isma, F. (2007). *Cross-curricular teaching incorporating human values: A collaborative research by reflective practitioners*. Paper presented and published in the Proceedings (refereed) of CoSMEd 2007 (13<sup>th</sup> to 16<sup>th</sup> November) conference. SEAMEO RECSAM.
- Torrance, E. Paul (1974). *Torrance Tests of Creative Thinking*. Ginn and Company (Xerox Corporation).
- Tuckman, B.W. (1972). *Conducting educational research*. U.S.: Harcourt Brace Jovanovich, Inc.
- Wikipedia (2008). *Item Response Theory (IRT)*. [URL: [http://en.wikipedia.org/wiki/Item\\_response\\_theory](http://en.wikipedia.org/wiki/Item_response_theory)]