Provision of Service on the Study of Using Assessment Data to Enhance Learning and Teaching (English Language Education)

— Analyzing Multiple Choice Items in the 2008 TSA Reading papers for Primary 3, Primary 6 and Secondary 3

By Dr Anthony K. K. Tong
Faculty of Education
University of Hong Kong   February 1st, 2009

I. Introduction

Serving as a piece of deliverable providing expert views and advisory services for the Provision of Service on the Study of Using Assessment Data to Enhance Learning and Teaching, this document reports an analysis into a set of 2008 Territory-wide System Assessment (TSA) reading papers and their facility indices which have been provided to the researcher by the Assessment and Hong Kong Examination and Assessment Authority Section, Education Infrastructure Division, EDB.

The ultimate objective of this work is to identify, from the given data, prospective students' learning problems, and the dimensions in which weaknesses in students' performance are manifest. It is hoped the identification of these learning problems will lead to reliable and accurate estimates of possible causes of prospective students' learning problems, resulting in the development of some school-based diagnostic assessment tools, the outcomes of which will inform future teaching and learning.

The report will first present a study framework for the study mentioned in the report title, present the findings of an analysis of the TSA data, and make recommendations for the development and design of school-based diagnostic assessment tools.

II. The data set and assumption about the level of difficulty

The data and information which inform this analysis include the papers and the item facility indices of all the Multiple Choice items in the Reading papers of the 2008 TSA papers at the level of Primary 3, Primary 6 and Secondary 3. These papers were attended in 2008 by selected groups of students in schools in Hong Kong. Each of these test papers consists of several parts, and each part contains a number of multiple choice items testing the comprehension of a given text. The topic and genre of these texts vary. A summary of the text type and topic is presented in Appendix A. Some of the texts together with their items appear in a paper are used again in a different version of the same level test. For example, the story of "My first pet" given as comic

in 3ERW1 appears also in 3ERW3. Similarly, repeated use of the same texts and items are seen with the papers for Primary 6 and for Secondary 3.

The present study assumes that the facility indices are indicative of students' reading performances in which the weaknesses of their learning are manifest. Ultimately, the present analysis aims to discern the nature of the difficulties the assessed students experienced in their reading comprehension processes.

At the outset, it is noted that in order to determine the level of difficulty of the concerned items based on their facility indices, it is necessary that a cut score line be established. Given that the TSA is a criterion-referenced assessment and norms do not apply, it has been decided that, for the present analysis, the demarcation line to consider an item as difficult for the testees is set at Facility Index (FI)=50. In other words, the discernment of learning problems will focus upon those items which more than half of the testees failed to score.

III. The theoretical orientation

The examination of the TSA reading papers data in this project is ultimately a study of a phenomenon. Hence the study itself is a learning process in its own right. The phenomenographic approach (Marton and Booth, 1997) adopted for this study rests upon the premise that discernment of an object or a phenomenon must take place in its context. Hence the different items and texts in addition to the FI, and the many aspects related to them are all considered as parts of the whole, and are all therefore subjected to scrutiny for discernment. In order to discern the characteristics of testee's learning problems, the following patterns of variation have been attempted:

Contrast: To discern characteristics of X, one has to experience both X and not X. One example of the use of this method would be to compare the items which yield very low Facility Indices with the ones which yielded higher Facility Indices.

Separation: When a dimension of X varies while other dimensions of X remain invariant, the dimension of X could be discerned. Some weaker performance with items flagged by low FIs may have compounded causes for their difficulties. The separation of these causes with the use of some diagnostic assessment tools at a later stage would be informative.

Generalization: When a dimension of X varies while other dimensions of X remain invariant, the dimension of X could be discerned. For example, certain low FI items across the different items in different sets tend to be dependent on factors such as processing load of the items required.

Fusion: When two dimensions are varying simultaneously, the simultaneity of two dimensions of variation could be experienced. This method has been attempted but at this point in time has not resulted in any supported and useful observations.

IV. The proposed framework for the focuses of varied discernments

While it is reckoned that authentic reading is both cognitive and social, it can also be argued that an Information Processing Model of reading (Rayner & Pollatsek, 1989) would yield higher descriptive power to explain the performance of some primary or secondary students who were asked to read texts in a testing situation. In this information model of reading, students, in their response to meet the demand of some reading tasks, would be actively engaged in some highly interactive, cognitive processing, involving the use of their reading skills and strategies and the deployment of many knowledge systems.

To use the methods of variation effectively, there is also the need to identify key foci for discernment. A review of the literature of reading in English as a Second Language has been conducted to provide an array of possible foci for discernment. The varied observations should include at least the following:

(1) Comprehension processing types;
(2) Comprehension (receptive) skill types;
(3) Knowledge types for comprehension; and
(4) Task (question) demands.

It should be noted that these four factors are not mutually exclusive. In fact, they are interrelated and sometimes intertwined. A discussion on these four dimensions is presented below.

(1) Comprehension processing types
Two common processing types which are often discussed in the literature are the Top-down processing and Bottom-up processing (Grabe and Stoller, 2002; Harmer, 2001, 1983; Nuttall, 1986).

> "In top-down processing, we draw on our own intelligence and experience – the predictions we can make, based on the schemata we have acquired – to understand the text. As we saw, this kind of processing is used when we interpret assumptions and draw inferences." (Nuttall, 1996, p. 16)

> "In bottom-up processing, the reader builds up a meaning from the black marks on the page: recognising letters and words, working out sentence structure. (Nuttall, 1996, p. 17).

It is viewed that both top-down and bottom-up processing are essential for effective comprehension and that comprehension is possible only when these two processing types are used interactively and appropriately (Grabe & Stoller, 2002). A reader (or listener) will rely on top-down processing to activate schemata and expectations, enable himself or herself to hypothesize, and predict what sort of information to come. In the bottom-up processing, the reader (or listener) deciphers incoming linguistic codes and cues to inform himself or herself details of the message, to confirm or refute hypotheses, to adjust expectations, and to revise predictions, etc.

The overreliance on top-down background (e.g. not being able to decipher linguistic codes in bottom-up processing) is dangerous because a reader often come with his own biases and beliefs which can urge him or her to impose wrong expectations and misinterpret the text. Relying solely on bottom-up processing, on the other hand, will result in reader having to cope with very heavy processing load as he or she is burdened with a larger quantity of information. This may in turn lead to failure to understand the intentionality of the writer, the context, and the totality of the message, resulting in failure to arrive at a coherent interpretation of the whole.

(2) Comprehension skill types

One type of variations for discernment in this study would be the specific skill types involved in testees' attempts to read and answer the questions. A reader has to be equipped with different skills to make meaning out of a text. J. Harmer (2001), for example, considered that comprehension, reading as well listening, may involve the following receptive skills:

1. Identifying the topic
2. Predicting and guessing
3. Reading for general understanding
4. Reading for specific information
5. Reading for detailed information
6. Interpreting text

It should be noted that while Harmer's inventory is useful for the teaching and learning of reading and listening, some of the skill types which he mentioned are often not included in many standardized reading comprehension tests because these skills may pose threat to marking or scoring reliability. For example, the skills of *predicting and guess* hardly appear

in standardized reading tests because the answers given by candidates may not be convergent, resulting in low marking reliability.

While Harmer (2001) considered skimming under the skill of Reading for general understanding, other writers (e.g. Nuttall 1986) explicitly list skimming and scanning as key reading skills. See Appendix B for the reading skills inventory proposed by Nuttall (1986). Other writers in Reading English as a second language have offered similar inventories of reading skills and strategies (e.g. Hedge 2000; Garbe and Stoller, 2002). It is, however, not always possible to know from their responses, if testees have actually used a certain skill or have not answered in a supposed manner a specific item.

To tap testees' cognitive processing, it is necessary that verbal reports such as think-loud protocols be adopted. (Guass & Mackay, 2000). However, for the present analysis, observations can still be made and hypotheses formed about the specific skill or processing the testees have probably experienced based upon the specific nature of the corresponding cues in the item, the options and the option chosen by the testees. This will be further discussed under the section entitled Task demands.

(3) Knowledge types for comprehension

It is believed that reading comprehension involves the use of many different knowledge systems. Some of these are related to language, and some in the general category of world knowledge. Hedge (2000, p. 189) discussed the knowledge bases for reading comprehension and suggested the following:

    (1) Syntactic knowledge
    (2) Morphological knowledge
    (3) General world knowledge
    (4) Sociocultural knowledge
    (5) Topic knowledge
    (6) Genre knowledge
    (7) Schematic knowledge
    (8) Language knowledge

Because of space, the following discussion will only cover (i) Topic and schematic knowledge; (ii) Genre knowledge; and (iii) Language knowledge, which are considered relevant to the postulation of testees' comprehension difficulties in this investigation. As far as possible, examples taken from the present sets of TSA reading papers will be supplied to explain the specific nature of the dimension under discussion.

(i)   Topic and schematic knowledge

During reading, readers will make use of the cues in text (e.g. lexical items) to help activate expectations and the schema of the topic under discussion. The use of the schemata, frames or scripts will help the reader to understand specifics and details of a message. For example, a poem such as the one in the Primary 6 Reading paper (Page 2 in 6ERW1) entitled "A test" containing words such as "hall", "timer", "read", "question" will activate in a reader's mind certain schema, as for example, the script of an oral test: This script may look like this: The student goes to the teacher inside the school hall; the timer starts ringing; the student is supposed to read aloud a text within a given time; after that reading aloud a text, he or she moves to answering some questions from the teacher. If a candidate fails to activate the right schema when reading this poem, he or she is likely to provide a wrong answer for a test item.

(ii)   Genre knowledge
Genre knowledge can also be considered as a kind of language knowledge depending on how it is categorized. Genre knowledge is also closely related to schematic knowledge because genre or text type by definition involves the use of a repeated pattern, so it can be a kind of schemata. When people talk about a text being a story, a schema will then be activated, and the reader expects that the text will unfold following more or less the pattern of "Orientation-conflict-Resolution-Coda", a pattern which all stories tend to follow. So knowing the specific genre or text type to which a text should belong would help comprehend the information in it. There is evidence in the data sets to show that some students simply lacked the basic knowledge to identify some common text types such as play or conversation. The facility index of Question No. 1 in Part 4, 6ERW3 (FI=45.5), for example, suggests that more than half of the testees did not know that the text, Martin the dreamer, is a conversation taking place in a play; 26% of the testees thought it was a story and 20.8% thought it was a telephone conversation.

Six different text types have been suggested in the English Language Curriculum Guide (Primary 1 - 6) (The curriculum Development Council, 2004, p. 17). They are: Narrative; (2) Information Texts (3) Exchanges (4) Procedural Texts (5) Explanatory Texts and (6) Persuasive Texts.

A review of the TSA 2008 Reading papers for P3, P6 and S3 suggests that the reading tests exploited only four text types: (1) Narrative Texts, (2) Information Texts, (3) Explanatory Texts, and (4) Exchange Texts. The table in Appendix C summarizes the text types used by all the reading papers for this review. It can also be seen here that in fact that while the Primary testees were exposed to more different text types, the Secondary 3 students met with a much narrower range of text type samples.

Appendix C consists of three tables displaying the average FIs of the different parts linking with different texts in the paper sets. It can be seen that certain texts seem to have exerted greater difficulties to the testees. With the Primary 3 paper sets, the

Table of Contents in Part 2B in 3ERW2 and 3ERW3 posed the greatest difficulties to the testees. (Average FI=42.3 and Average FI=48.1). This is followed by the texts of two stories, "The three frogs" and its sequel "Frogs and Princess", with Average FIs=43.5. and 45. 2, and Average FIs=48.6 and 52.4).

With the Primary 6 papers set, the parts which posed the greatest difficulty for the testees are a text of school announcement, which was a multiple text (texts in a text) with average FIs= 47.7 and 47.9).

The only text which calls for concern (FIs below 50) at S3 level is the Explanatory Text of Traditional Chinese wedding, with Average FIs=48.9 and 47.8.

Observations about the relative difficulties of these parts in the papers have prompted the conjecture that certain texts, and possibly testees' own limited knowledge and exposure to these text type or text structure or both, may have created for them difficulties.

(iii)   Language knowledge

Language knowledge consists of many different knowledge systems. In light of the foci of the present analysis, the discussion will cover only (a) Lexical knowledge and (b) Discourse knowledge

(a) Lexical knowledge and lexical barriers
Words are the building block of a language, hence it is only logical that difficult or unfamiliar words in a text will pose challenge to a reader. A reader can of course try to guess the meaning of the unknown word, but even good justified guessing does not ensure getting it right, and hence guessing in turn does not ensure understanding the text; the guess could be wrong. If a text contains too many difficult words, comprehension is obviously a problem even if the reader has good skills to guess word from its context. Nation (2001) has suggested that any text containing more than 3-5% of unknown words will make guessing very difficult, if not impossible. Several items in the present data sets show that students were cued to make guess of word meaning.

The use of low frequency words in text could be another issue. Words which students do not have opportunity to be exposed to would of course pose threats to comprehension. Easily confused words such "hot spots" and "hot sports" in 3ERW2 is one such example. Many students of Primary would know the word "sports", but many of them might not have learned "spots", although the term "hot spots" is used idiomatically by many native speakers of English. (Item 2, Part 2B, 3ERW2)

One type of lexical knowledge which is relevant to the present investigation is testee's awareness of English polyseme, combined with proper use of lexicographical skills. Many English words are polysemous, a word having different

but similar meanings in different contexts. Such words in a dictionary are naturally given several explanations. When a student consults a dictionary for the meaning of a polysemous word in a text, it is important that the testee is aware of the context of the word so that the appropriate explanation can be chosen from several. Item No. 7 in Part 4, 6ERW1 and 6ERW3 for the text of the play Martin the dreamer is one such example. A total of four senses of the word "beg" were given in this item, and testees were asked to choose the correct one for the word in "and the rich and famous will beg to eat my super creamy chocolate cheesecake…". Only 39.6% candidates chose the key, B, "to ask for something very strongly". A total of 27% candidates went for A. "to ask for good or money because you are very poor". Responses to this item show that many students (a) lack good dictionary skills; (b) not knowing how to use context to guess meaning of words; and (c) lack the awareness of polysemous words.

(b) Discourse knowledge
Discourse knowledge here refers to the knowledge base for the ability to interpret a stretch of language longer than a sentence. Hence, discourse knowledge helps to achieve understanding of meanings which is conveyed by language longer than a sentence or a group of words. For example, the answer to Question No. 4, Part 2A, in 3ERW1, a blank-filling item, should be "The frogs jumped all day to get to the hill top." This answer is cued by a long paragraph consisting of three sentences:

> "When the sun came up, the three frogs started to jump. Jump, jump, jump. They only got to the top of the hill when the sun went down." (The three frogs in Part 2A, 3ERW1, p. 6)

This item requiring global processing was attained by only one third of the total candidates (FI=32.6). A total of 30.8% of the candidates chose distractor A, "all morning", and 17.1% opted for "all afternoon", and 18.5% chose "all night".

Discourse knowledge may also take the form of *genre knowledge*. It may refer to the knowledge system to enable a reader to recognise a specific text type, and to look at the appropriate place in a text for a specific piece of information. For example, knowledge about the letter layout will inform the reader where to find out on which date a letter was written. An example of the use of this knowledge type would be to infer 'the exact date of the next visit to a park' from an expression in a letter such as "Dad will take us to the park again tomorrow" when a reader at the same time knows where to find the specific date on which the letter was written (3ERW1, Question No. 8, P. 13).

Another discourse knowledge type would be the ability to identify the referent of a pro-form (e.g. a pronoun such as "it" or "she").

(iv) Task demands

The notion of Task Demands refers to the way in which testee's comprehension processing is cued by an item, which usually appears in the form of either a question or blank filling. The forces contributing to the task demands can come from (a) the specific processing type which is required by the item, such as whether it is a local, global or an inferencing item; (b) an item requiring testee's perusal of more than one text; (c) an item requiring very careful reading of the item stem; and (d) an item with strong attraction from certain distractors. Because of the different conditions a testee is forced into by the specific nature of an item, a reader's information load can increase sharply, resulting in his or her performance errors.

(a) *Local*, *Global* and *Inferencing* items
It has been observed that certain question types are relatively easier to handle than the others. A *Local* question requires a testee to read and understand the meaning conveyed locally by a group of words, or by one or two connected sentences. A typical Local question would be one that asks, for example, "Who will get the postcard?" in which case, the reader would only need to see the name of the addressee on the postcard, "Dear Chris", to determine that Chris is the person to get this postcard (Question No. 2 in 3ERW1, p. 12). A local item is often a lexical item. For example, when a testee read in the given text that a teacher enjoys "going to concert", he or she should know that this would mean the teacher enjoys "listening to music". (6ERW4 part 3 Question 6; FI=25.9)

A *Global* question, which demands the appropriate use of discourse knowledge, requires a reader to interpret a longer stretch of words. A typical global question would be to ask testees to read a poem and then suggest from four options the most suitable title for the poem. For example, Question 8 in Part 2, in 9ER1, testees were asked to provide the poem with a title. It is observed that half of the candidates failed this particular item (FI=49.3).

The third type of question is called *Inferencing* question. With this type of question, the correct answer is often not explicitly given anywhere in the text. The testee has to infer to fill the gap of reasoning to arrive at the right response. An example of this is Item No. 8, Part 2, in 6ERW1, with an FI=49.2. This blanking-filling item aims to elicit the response of "John will arrive at home at about 5:15 p.m. (Option D: "5:15 p.m." (49.2%). In fact, the last entry of time on the schedule is 4:30 pm, the time for the students to arrive at school upon completion of the visit. It should be noted that the key of 5:15 pm, however, appears nowhere in the text, which is a letter to parents by a school, informing them the proposed schedule for a visit to Stanley.

(b) Perusal of multiple texts

Some items (questions) are more demanding in that they require the testee to process *multiple texts* (e.g. several smaller texts in a larger text) to determine which text to

read on first before they identify the specific detail for the item. This point has already been discussed in the section entitled Genre knowledge.

These items involving testee's chains of decisions would raise their risks of making mistakes in their responses. For example, Question No. 8 in Part 3, 6ERW3 requires that candidates should determine which one of the six given texts to read first before they decide on the details to answer the question. An example for this type of item would be Item No. 8 in Part 3, 6ERW3. The text is a multiple text consisting of four information texts (Posters providing information about four different cultural events). The question asks" How many shows are there for the play?" The FI for this item is a particularly low, FI at 14.8, meaning that more than 85% of the candidates failed to score. The key is A, "5 shows". In view of the very low FI and that lots of 35% of the candidates chose "3 shows" (cued by My Message) and "2 shows" (cued by summer concert), there is reason to believe that many of them simply could not decide which one of the four texts to go to in order to figure out the answer.

Anther type of item which requires students to peruse more than one text is exemplified by Item No. 7 (FI=33.4) in Part 3, 6ERW4. This is an item about a multiple text — a school announcement about their three new teachers in the new school year. The question asks: Who has been teaching the least number of years? The answer is D. Thomas Cheung. Although the cue to Thomas Cheung being the key is quite obvious in the text — he "joins the school this year" and "it is his first year of teaching". Wrong responses to this item suggest that perhaps students were not used to processing several texts to get to the answer.

Candidates being made to process both some comic strips and the texts or language in these strips were also being asked to process multiple texts. It would have been a reasonable assumption that many Primary 3 students should appreciate the use of comic strips. However, the FIs of this particular item seem to show that many testees dealt with these items in the most careless manner. Item No. 1 in 3ERW1 (FI=45.9), for example, asks "How many people live in Jack's house?". The cue to the key is a picture of a 3-storey house with three persons standing on the three storeys, each of them saying in a speech bubble that they were living with one other person on the same floor. Hence the answer to the item should be D, "6 people". However, more than half of the candidates offered a wrong answer, and among these mistaken testees, the vast majority (Option A, 37.1%) suggested "3 people" as the answer, suggesting that these testees answered the item by simply counting the number of figures in the pictures.

Two possible explanations proposed for this response pattern are that either these testees got into a panic because they were not used to handling comic strips when reading, or they had carried with themselves the assumption that if it was a picture for comprehension, processing the texts in the speech bubbles in it would not be necessary. Further investigation into the causes of comprehension failure is thus warranted.

(c) Very strong distractor(s)

On type of task demands is the *force of the distractors*. The condition in which a student has to deal with reading multiple choice items is actually unique to testing. With authentic reading behaviours a reader does not have to deal with distractors. It is also in such a condition that a reader is subjected to very strong distortive forces coming from MC distractors which would skew his or her understanding of a text. Item No. 6, in Part 1, 9ER3 for the Poem Caterpillar-Butterfly, with a low FI=28.7 exemplifies the power of a strong distractor. The item asks: "What does the writer want to see in the last stanza?" The last stanza reads like this: "And I stand here,/Breathing dirty city air,/Hoping to see brightly coloured wings,/All the wondrous beauty nature brings.". The key to the item is Option C, "Beautiful things" (28.7%). The distribution of the options are: A. "More trees" (9%); B. "a clean city" (56%); D. "more people" (4.8%). The hypothesis arises from this observation is that B was a very powerful distractor attracting the choices from more than half of the total candidates. Because of the word "dirty" in the line "Breathing dirty city air", these candidates were inclined to the belief "the writer wanted to see a clean city".

(d) Not reading the stem carefully

Another type of task demand posed onto the testee is the particular condition specified by or embedded in the *stem of the question*. Careless readers may miss processing certain important specific detail in the stem, leading to suggesting a wrong answer. For example, Item No. 2 in Part 3, 9ER3 (FI=46.7), for the text Traditional Chinese wedding requires testee's careful reading of the words in the stem. The question asked:

"What do the bride and groom do the night before the ceremony?" The options and the distribution of responses are:

> A. They light candles. (21.2%)
> B. They see each other. (13.7%)
> C. They stay in different places. (the key: 46.7%)
> D. They comb each other's hair. (14.6%)

The corresponding text for the item gives the cue to the answer rather explicitly: "The ceremony is performed for both the bride and the groom, *in their own homes*, as they are not allowed to see each other before the wedding day."

Fewer than half of the candidates scored this item. A review of the wrong options would suggest the conjecture that those who opted for "light candles" (21.2%), or "comb each other" (14.6%), which are both acts DURING the ceremony, did not actually read carefully enough the stem and hence missed seeing that they were required to talk about things BEFORE the ceremony, and not DURING.

V. Findings from the analyses

Item-by-Item Analyses have also been conducted to examine the papers sets, the texts and their items, in addition to a review of all the facility indices by text and by paper. Specific observations have been made to individual items which are considered low (FIs=50). Discernment of the possible causes of the comprehension difficulties were carried out in relation to the corresponding texts and tasks. A summary of this item-by-item analyses are presented in Appendix D. A summary of the FIs of items by paper/section/text is presented in the EXCEL file in Appendix E.

At a general level, the researcher would like to propose the following overall observations:

(A) Primary 3: A review of the most difficult items experienced by this testee group tends to suggest that:

1. Many students did not have the proper genre knowledge, particularly knowledge about table of contents, and postcard.

2. Many of them could not handle well information from comics, either comics as a text to process (My first pet) or comics accompanying a text (the three frogs and Frogs and the Princess).

3. Many lacked the ability to deal with inferencing items.

(B) Primary 6: A review of the most difficult items experienced by this testee group tends to suggest that:

1. Many students did not have the strategies needed to deal with multiple text (texts in a text) or to make cross-text comparison.

2. Many could not handle well inferencing questions, especially when they did not have any lexical clues to rely upon.

3. Students tended to lack the skills and knowledge to deal with the referent of pronouns, possibly suggesting a lack of reading practices at discourse level.

4. Many of them lacked dictionary skills, something which they should begin to pick up and develop for their upcoming secondary language learning.

5. Some of them seemed to possess a very small vocabulary. They did not even know words like "magazine", listed as a word among the first 1,720 English words in term of frequency by Collins COBUILD (2001).

(C) Secondary 3: A review of the most difficult items experienced by this testee group tends to suggest that:

1. Many did not have the skills to infer based on surface linguistic cues.

2. Many testees were not able to guess word meaning from context.

3. They would need to develop their ability to interpret language at discourse level.

4. Some of them showed carelessness in their reading behaviours, not being sensitive to details in texts.

VI. Recommendations for further investigations:

Many of the observations made in this report are reasoned and grounded on some well-received understandings in the literature as, for example, the information processing load theory, reading English as a second or foreign language. However, further work is called for and the following are suggested.

(a) Students' cognitive processing

It is suggested that further investigations be conducted with students in schools to test and verify some of the key hypotheses raised in this analysis.

(b) Classroom teaching and learning

It is also important that we need to better our understanding about how reading lessons are actually conducted in schools. This researcher has formed the impression, from his own classroom observations in both primary and secondary schools and from this analysis, that while many students in schools are given ample opportunities to practise top-down processing, e.g. to do predictions and guesses based on book covers, comics and illustrations and half-told stories, etc., it is likely that there will have to be more work on how bottom-up processing could be developed in our students.

(c) The design of task, text and items for test

The TSA papers have used some texts and their items repeatedly in different versions of the paper in the same level years. Because of this, with many of the items, two sets of facility indices were generated. It is good to see that the very same items used for the same texts have generated very similar indices, validating the reliability of the tests. On the other hand, it is suggested that future test writers give more thoughts to the issue of

authenticity of texts and tasks. It has been observed that authenticity was an issue with some of the texts used in the papers. The multiple text for Part 3 in 6ERW2 is considered problematic, for example. Against common practice, it was observed that the category labels of the four texts in text, [PLAY, DANCE, MUSIC, ART], were vertically printed. This obviously created problems for most of the Primary 6 readers in the test, as it actually puzzled the researcher himself for some time when he tried to tackle Item No. 8 (FI=16.8). The crux of the problem is that in real life, one can hardly see Posters or Information texts such as this one anywhere in City Hall or the Hong Kong Cultural Centre, with category label printed vertically.

References

Anderson, N. (1999). *Exploring second language reading*. Boston: Heinle & Heinle.

Cohen, A. (1998). *Strategies in language and using a second language.* New York: Longman.

Harmer, J. (1983). *The practice of English language teaching*. Harlow: Longman.

Harmer, J. (2001). *The practice of English language teaching*. Harlow: Longman.

Hedge, T. (2000). *Teaching and learning in the language classroom.* Oxford: O.U.P.

Gass, M. S. & Mackey, A. (2000). *Stimulated recall methodology in second language research.* New Jersey, USA: Lawrence Erlbaum.

Grabe, W. & Stoller, F. L. (2002). *Teaching and researching reading*. Harlow: Longman.

Marton, F. & Booth, S. (1997). *Learning and awareness*. New Jersey, USA: Lawrence Erlbaum.

Nation, P. (2001). *Learning vocabulary in another language*. Cambridge: Cambridge University Press.

Nuttall, C. (1996). *Teaching reading skills in a foreign language. (New edition)* Oxford: Macmillan Heinemann.

Rayner, K. & Pollatsek, A. (1989). *The psychology of reading*. NJ, USA: Lawrence Erlbaum Associates.

Appendix A

Table: Text types and topic of the texts in the 2008 Reading papers in P3, P6 and S3*

Categories of Text Types	Examples of Text Types	Primary 3	Primary 6	Secondary 3
Narrative Texts:	Cartoons and comics	My first Pet		
	Rhymes or Poems	At the farm	The Test; Climbed a hill	Caterpillar-butterfly
	Stories	Three Frogs; Frogs and Princess		
Information Texts	Charts			
	Notices	School Notices (6 activities)	Halloween party; Notice to parents (Stanley)	
	Tables	Table of contents		
	Dictionaries		Entry of "beg" as a polyseme	
	Announcements		1. Four cultural Events; 2. Three new teachers in school	
	Book Review*			Book Review of The way home
	Book Cover	Happy Island		
Exchanges	Conversations		Martin the dreamer	
	Postcard	Postcard from Peter (Water Park)		
Explanatory Texts	Explanations of how and why		Are you fit? (2 charts: bar and pie)	HK - City of Life; Cinemas in HK; Traditional Chinese wedding
Procedural Texts	Directions Instructions			
Persuasive Texts	Discussion			
	Expositions			

[Note: It should be noted that the classification of text types is not unique. The key rhetorical functions deployed in a text have been considered when categorizing. For example, most book reviews in the real world would contain value judgment and evaluation by the review writers. However, the book reviews in the present tests set (9ER1, p. 1) are very descriptive and do not contain any evaluative statements. They are therefore placed in the category of Information Texts. Book cover is also placed under Information Texts. On the other hand, some writers would consider the three S3 texts, HK - City of Life, Cinemas in HK and Traditional Chinese Wedding, as exemplifications of Information texts, rather than explain. This research has placed them under explanatory because the rhetorical functions in the texts tend to focus on explaining why, rather than just telling something.]

Appendix B: Reading skills proposed by Nuttall (1996)

Nuttall (1996) suggests the teaching and training of the following sub-skills:

Word attack skills:
      Grammatical clues
      Structural clues: morphology
      Inference from context
Text attack skills:
      Understanding syntax
      Recognising and interpreting cohesive devices
          interpreting elliptical expressions
          interpreting lexical cohesion
      Interpreting discourse markers
          markers that signal the sequence of events
          markers that signal the discourse organisation
          markers that signal the writer's of view
      Recognising functional value
          independent functions: e.g. defining, classifying, generalizing, naming, describing, reporting, speculating, predicting
          Text-dependent functions: e.g. asserting, exemplifying, explaining, reinforcing, hypothesizing, commenting, concluding
      Recognising text organisation
            Rhetorical organisation
            Organisation of paragraphs into texts
            Organisation of sentences into paragraphs
            Text diagrams
      Recognising the presupposition underlying the text
      Recognising implications and making inferences
      Prediction

Appendix C

Table 1: No. of Items, Parts in paper, Text type, Topic, and Average Facility Indices in the Primary 3 Reading Paper, 2008 TSA

Year of Paper	Part	Text Type	Topic	No. of Items		Average of Facility Indices (%)	No of items with FIs which are below 50	No. of items with FIs which are below 40
				Sub-total	Total			
3ERW1						50.5		
	Part 1	Dialogue accompanied by pictures	My first pet	8		59.7	3	1
	Part 2A	Story	The three frogs	5		43.5	4	3
	Part 2B	Story	Frogs & Princess	7		45.4	4	2
	Part 3	Letter from Peter	Water Park in Japan	8	28	50.2	4	2
3ERW2						52.9		
	Part 1	Notice from school (Multiple)		8		60.1	1	1
	Part 2A	Book cover	Happy Island	4		57.7	1	1
	Part 2B	Table of contents	Various topics	7		42.3	5	4
	Part 3	Postcard from Peter	Water Park in Japan	8	27	52.7	4	2
3ERW3						57.5		
	Part 1	Dialogue accompanied by pictures	My first pet	8		66.5	1	0
	Part 2A	Book cover	Happy Island	4		62.4	1	1
	Part 2B	Table of contents	Various topics	7		48.1	4	4
	Part 3	Poem	At the farm	7	26	53.7	3	0
3ERW4						56.3		
	Part 1	Notice from school (Multiple)	Various events	8		67.7	2	0
	Part 2A	Story	The three frogs	5		48.6	4	1
	Part 2B	Story	Frogs & princess	7		52.4	1	1
	Part 3	Poem	At the farm	7	27	52.8	2	0

Table 2: No. of Items, Parts in paper, Text type, Topic, and Average Facility Indices in the Primary 6 Reading Paper, 2008 TSA

Year of Paper	Part	Text Type	Topic	No. of Items		Average of Facility Indices (%)	No of items with FIs which are below 50	No. of items with FIs which are below 40
				Sub-total Total	Total			
6ERW1						55.0		
	Part 1	Poem	The test	7		53.7	2	0
	Part 2	Announcement to Parents	A visit to Stanley	8		68.2	1	0
	Part 3	School Announcement (multiple)	Three new teachers in the year	8		47.7	5	4
	Part 4	Narrative: play	Martin the dreamer	11	34	55.0	5	3
6ERW2						56.2		
	Part 1	Poem	The test	7		56.4	0	1
	Part 2	Poster	Halloween at HK Fun Park	8		70.4	0	0
	Part 3	Events announcement (Multiple)	Play-Dance-Music-Art	9		53.0	3	2
	Part 4	Explain/Inform	Are you fit?	8	32	56.2	2	0
6ERW3						59.7		
	Part 1	Poem	Climbed a hill	7		59.7	2	1
	Part 2	Announcement to parents	A school visit to Stanley	8		69.2	0	0
	Part 3	Events announcement (Multiple)	Play-Dance-Music-Art	9		50.4	3	2
	Part 4	Narrative: Play	Play: Martin the dreamer	11	35	55.8	4	2
6ERW4						59.3		
	Part 1	Poem	Climbed a hill	7		61.2	3	2
	Part 2	Poster	Halloween in HK Fun Park	8		70.8	1	0
	Part 3	School announcements (multiple)	Three new teachers	8		47.9	4	4
	Part 4	Explain/Inform	Are you fit?	8	31	57.6	1	0

Table 3: No. of Items, Parts in paper, Text type, Topic, and Average Facility Indices in the Secondary 3 Reading Paper, 2008 TSA

Year of Paper	Part	Text Type	Topic	No. of Items		Average of Facility Indices (%)	No of items with FIs which are below 50	No. of items with FIs which are below 40
				Sub-total	Total			
9ER1						58.4		
	Part 1	Book Reviews (Two texts)	About "The way home"	8		71.1	0	0
	Part 2	Poem	Caterpillar	8		51.7	4	2
	Part 3	Explain	Traditional Chinese wedding	5	21	48.9	3	0
9ER2						61.5		
	Part 1	Book Reviews (Two texts)	About "The way home"	8		71.4	0	0
	Part 2	Explain	HK - City of life	6		67.2	0	0
	Part 3	Explain	Cinemas in HK	5	19	61.5	1	0
9ER3						60.3		
	Part 1	Poem	Caterpillar	8		51.6	4	2
	Part 2	Explain	HK - City of Life	6		66.2	0	0
	Part 3	Explain	Traditional Chinese wedding	5	19	47.8	3	0