Contemporary Educational Research Quarterly
September, 2010, Vol.18 No.3, pp. 165-206

Using Revised Bloom's Taxonomy to Analyze Reading Comprehension Questions on the SAET and the DRET

Wei-Hua Lan* Chiou-Lan Chern**

Abstract

Based on the Revised Bloom's Taxonomy, this study aimed to investigate the cognitive process levels and knowledge types measured on the English reading comprehension tests of college entrance examinations administered from 2002 to 2006 in Taiwan. A descriptive analysis was conducted to examine the similarities and differences of the content and cognitive skills intended to be assessed between the two tests, hoping to serve as a reference for English teachers while helping learners develop the needed cognitive skills in reading and test preparation.

Results showed that for both tests, four major levels in the Revised Bloom's Taxonomy (Remember, Understand, Apply, and Analyze) along with eight sub-levels, and three types of knowledge (Factual, Conceptual, and Procedural) along with three subtypes were identified, with a total of five major question types and nine subtypes

* Wei-Hua Lan, Part-Time Lecturer, Division of Preparatory Programs for Overseas Chinese Students, National Taiwan Normal University
E-mail: lanflor93@yahoo.com.tw
** Chiou-Lan Chern, Professor, Department of English, National Taiwan Normal University
E-mail: clchern@ntnu.edu.tw
Manuscript recieved: Feb. 25, 2010; Modified: Jul. 14, 2010; Accepted: Aug. 30, 2010

of questions. Items on Remember Factual Knowledge and Understand Factual Knowledge, which belong to lower cognitive levels, were the majority in the two tests. Few items were found at higher levels of Apply and Analyze. The major differences between the SAET and the DRET were the frequency, occurrence, and distribution of items testing different cognitive sub-skills and knowledge subtypes. It was found that Executing/Apply items were more favored in the SAET, whereas the DRET had more items on Inferring (a subtype under Understand category).

Keywords: test item analysis, reading comprehension, Revised Bloom's Taxonomy, SAET, DRET

《當代教育研究》季刊
第十八卷第三期，2010 年 9 月，頁 165-206

以布魯姆認知分類修訂版分析大學學科能力測驗及指定科目考試英文閱讀測驗

藍偉華* 陳秋蘭**

摘 要

本研究旨在以布魯姆認知分類（修訂版）與內容分析法，將臺灣地區近五年（2002-2006年）大學學科能力測驗與指定科目考試中，英文閱讀測驗試題中的欲測試認知層次與知識型態進行分類，並利用描述統計探討兩種考試之試題的異同，以協助教師及受試者了解兩種考試中之英文閱讀測驗的內涵，進而提供教師在進行閱讀教學時，培養學生建立閱讀和應試應具備的基礎認知技巧之參考。

研究結果顯示，試題的認知層次可分爲四種（即知識、理解、應用、分析）及八種次層次，而內容則分爲三種知識類型（即事實、概念、程序）及三種次類型，共可歸納出五種主要題型及九種次要題型。在這兩種考試中，最常考的是「記憶事實性知識」及「理解事實性知識」這兩類低階認知層次的試

* 藍偉華，國立臺灣師範大學僑生先修部兼任講師
電子郵件：lanflor93@yahoo.com.tw
** 陳秋蘭，國立臺灣師範大學英語學系教授
電子郵件：clchern@ntnu.edu.tw
投稿日期：2010年2月25日；修正日期：2010年7月14日；接受日期：2010年8月30日

題，只有少數題目的認知歷程達「應用」和「分析」兩高層次。此二種考試最大的差異在於次要題型的出現頻率和分布。學科能力測驗含括較多的「執行」（或稱爲「應用」）試題，指定科目考試則包含較多的「推論」問題。

關鍵詞： 試題分析、閱讀測驗、布魯姆認知分類修訂版、大學學科能力英語測驗、大學指定科目英語科考試

Introduction

English reading plays a crucial role in secondary and higher education in the EFL (English as a Foreign Language) context in Taiwan because it is one of the major inputs students receive while learning English and is a major skill measured in nationwide examinations. College entrance examination is a high stake examination for senior high school students and teachers because it determines students' choice of university. College Entrance Examination Center (henceforth CEEC) in Taiwan administers two examinations annually—Scholastic Achievement Test (SAT) and Department Required Test (DRT). In the first stage of screening, students who take the SAT and meet the qualifications set by a particular university are recommended by their high schools to the university. SAT aims to assess students' basic scholastic knowledge and abilities. DRT, implemented in 2002, is designed to identify students who possess the ability in certain subject areas required by a particular university. Usually, a university requires the DRT test results of three to six subjects. Based on the requirements of their preferred schools, students then choose the subjects they want to take the test for. The English achievement test in the second stage is thus comparatively more difficult than that of the first stage and aims at discriminating more proficient students in English.

On both types of English tests, reading comprehension is a major component in measuring students' reading ability. Generally, a reading comprehension section contains a series of passages, each followed by three to five multiple-choice questions, which are assumed to tap different reading or thinking skills. Whether or not the reading comprehension questions truly assess an examinee's reading ability

is a question to be explored. The CEEC makes annual reports on item analysis of both SAET (Scholastic Achievement English Test) and DRET (Department Required English Test), yet studies on the qualitative analysis of reading comprehension questions have not been sufficiently conducted. The statistical analyses conducted by CEEC revealed that most of the items had good statistical values, but test items with good statistical values might not necessarily be good items (Huang, 1994). Therefore, it is necessary to explore what reading skills test items actually measure.

Recently, studies that examined some qualitative factors such as the reading skills measured on each test item and the factors affecting students' reading test performances in both SAETs and DRETs have been conducted (Hsu, 2005; Lu, 2002). However, the scheme they used for analysis, i.e., Mo's (1987) does not clearly show the cognitive process levels that each test item aims to measure.

The present study, viewing reading as a cognitive activity (Bernhardt, 1991), attempts to apply another framework, the Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), to analyze the reading comprehension test items on both SAETs and DRETs. Bloom's original classification includes six major categories in the Cognitive Domain: Knowledge, Comprehension, Application, Analysis, Synthesis, and Evaluation. Except for the Application category, each of these categories contains subcategories. However, several weaknesses and practical limitations could be found in the original taxonomy. One of the major problems is the cumulative hierarchy structure (Furst, 1994). For example, some demands for the Knowledge categories are more complex than certain demands for Analysis or Evaluation levels, and Evaluation is not more complex than Synthesis (which involves evaluation) (Kreitzer & Madaus, 1994). Consequently, Anderson and Krathwohl in 2001

proposed a revised version. The major difference between the two versions is that the revised taxonomy includes two dimensions: knowledge and cognitive process dimensions. There were four knowledge types (Factual, Conceptual, Procedural, and Metacognitive Knowledge) and six cognitive process levels (Remember, Understand, Apply, Analyze, Evaluate, and Create). Thus, it provides a clearer framework for test designers to know what to test and how to test it.

This study, aimed at providing a better understanding of the content and construct of the colleague entrance examinations in Taiwan, explores the following questions:

1.What cognitive process levels and knowledge types in the Revised Bloom's Taxonomy were measured in the SAET and the DRET reading comprehension sections?

2.How did the SAET and the DRET differ in the cognitive process levels and knowledge types they measure?

Literature Review

Application of Bloom's Taxonomy

Bloom's Taxonomy has been widely applied in testing and evaluating across different subject matters and various kinds of tests, either state-wide, nationwide, or classroom assessments (Airasian, 1994; Alderson & Lukmani, 1989; Aviles, 1999; Chen, 2004; Masters, Hulsmeyer, Pike, Leichty, Miller, & Verst, 2001; Squire, 2001). However, within the extensive literature on various disciplines, comparatively little research has focused on the application of the taxonomy in EFL reading instruction or assessment (Adams-Smith, 1981; Alderson & Lukmani, 1989; Costin

& Shen, 1986; Surjosuseno & Watts, 1999).

Alderson and Lukmani (1989) investigated whether test items that intended to measure certain skills indeed tested those skills. Nine teachers in the institute for English Language Education at Lancaster University were given four tasks: (1) reading through 41 items as they were taking the test; (2) writing down what they think each item was supposed to test in their own words; (3) classifying each question or sub-question into lower order, middle order, or higher order; and (4) identifying which of the skills were being tested by those questions. In the last task, questions were classified into eight skills described by the Bombay University Communication Skills Group — Recognition of words, Identification, Discrimination, Analysis, Interpretation, Inference, Synthesis, and Evaluation. These skills closely followed the original Bloom's Taxonomy. Results indicated that raters had agreement on only 14 out of 41 items in classification. The possible reason for this low consensus might be the fact that it was hard to really know how individuals arrive at the answer to a question and different people arrive at answers in various ways.

Gierl (1997) also questioned the adequacy of using Bloom's Taxonomy as a model to guide writers to construct items that measured the cognitive processes they hoped to be applied by students. In his study, 30 Grade-7 students (divided evenly into high and low achievers) were asked to think aloud as they solved problems on a mathematics achievement test. Both multiple-choice test items and students' think-aloud protocols were then classified based on Bloom's Taxonomy (only at the three lowest levels of Knowledge, Comprehension, and Application). The overall match between the responses expected by the item writers and the responses observed from

the students was only 53.7% (56% match in the high achievers' group and 50% match in the low achievers' group). Gierl stated that the cognitive process levels in Bloom's Taxonomy did describe students' thinking process while answering questions, yet maintained that it was difficult to judge which question tested those specific skill because readers could arrive at an answer in different ways (e.g., some items expected to be solved by knowledge processes were solved via comprehension processes instead). Alderson (1996), showing similar concerns, argued that despite this problem, defining what to test and then trying to test it would likely lead to a better test item.

Although the results in Gierl's (1997) study did not satisfactorily demonstrate a complete match between test takers' thinking processes and what the test items intended to measure, it should not be taken as conclusive because the pool of population was restricted to the thirty Grade-7 students taking math exams. Gierl's study, however, draws our attention that a coding scheme like Bloom's Taxonomy does show the test writers that the items, at least half of the time, are indeed testing what the items intend to measure.

Masters et al. (2001) analyzed 2,913 multiple-choice questions randomly selected from 17 test-banks of accompanying selected nursing textbooks. Questions were evaluated on thirty generally accepted guidelines for writing multiple-choice questions, the cognitive levels defined by the original taxonomy, and distribution of correct answers. Result showed that most of the questions were written at lower cognitive levels, i.e., 47.3% of the questions were at the Knowledge level, 24.8% were at the Comprehension level; 21% were at Application level, and only 6.5% were at the Analysis level. They reported that the result was somewhat surprising

because most of the textbooks reviewed were intended for upper division courses. In addition, a large amount of NCLEX-RN (National Council Licensure Examination— Registered Nurse, a computer-adaptive test of entry-level nursing competence) questions were written at the Application and Analysis levels, not at the two lowest levels emphasized in those textbooks. A harmful effect of this great discrepancy in learning goals, the assessment tools, and the state-wide examinations might affect students' performances on standardized examinations.

Squire (2001) analyzed the cognitive levels of agricultural science tests in senior secondary schools in Botswana. The materials analyzed were 628 questions taken from the Senior Cambridge Overseas School Certificate (COSC) Agriculture Paper 1 during 1989 to 1998. Data were analyzed by comparing the questions in each section of the examination papers to the sample questions (e.g., what is the capital of France?) and questions with characteristic words (e.g., questions at Knowledge level usually contained key words such as what, who, when). Results indicated that a great amount of questions were at knowledge level of Bloom's original taxonomy. Very few or even no questions could be found at higher cognitive levels of Application, Analysis, Synthesis, and Evaluation. Surprisingly, the essay type items in those tests were at the two lowest levels as well, which contradicted to the belief about the possibility of using constructed-response test items (involving constructing one's own answer) to assess higher level thinking (Buckles & Siegfried, 2006; Simkin & Kuechler, 2005).

Chen (2004) adopted the revised taxonomy to examine the knowledge types and cognitive levels of computer science test in technical college entrance examination from 2001 to 2004 in Taiwan. Three raters participated in the analysis to

ensure the inter-rater consistency of the classification of test items, and together they developed a subject matter table and a table of example questions at all six cognitive process levels gathered from previous computer science literature as principles to classify test items into the two dimensions. Results revealed that firstly, certain types of knowledge are associated with certain cognitive processes. For instance, factual knowledge was related to remembering whereas procedural knowledge to applying. Secondly, similar to previous studies, most of the test items (44% to 77%) measured only lower-level thinking that required students to remember factual information. No item was found at Evaluate and Create levels, possibly due to the constraint of the multiple-choice questions in which test takers are always forced to choose one correct answer. This view is consistent with Buckles and Siegfried (2006), who found that multiple-choice questions can measure elements of in-depth understanding when being carefully designed, and maintained that Synthesis and Evaluation levels could not be accurately measured since the creativity or originality could not be simply tested via multiple-choice questions.

Item Analysis of SAETs and DRETs

Two major studies (Hsu, 2005; Lu, 2002) were conducted on the SAET and the DRET reading comprehension item analysis using Mo's (1987) classification. Mo (1987) proposed that a reading test should include questions about the organization of the text and questions of textual comprehension. He excluded skills such as reading speed, habit, and pleasure that were unrelated to the text structure and then classified reading skills into six main categories: (1) identifying the main idea, (2) finding specific details mentioned in the passage, (3) finding implications and

drawing inferences and conclusions from the text, (4) recognizing style and tone, (5) determining the special techniques used by the author to achieve his effect, and (6) determining the meaning of strange words or phrases as used in the test.

Lu (2002) conducted both qualitative and quantitative analyses of the reading comprehension test items on SAETs administered from 1995 to 2002. Qualitatively, she classified items into Mo's six question types and examined textual materials, examinees' passing rates on each question type, test variables that affected those passing rates, and discrimination index; whereas quantitatively, she computed the frequency distribution of question types and the correlation between question types and passing rates. Results indicated that the most common question type was items on details, followed by items on inference, main idea, style/tone, organization, and word meaning.

Hsu (2005), applying the same coding scheme, analyzed reading comprehension test items taken from 2001 JCEE (renamed as DRT in 2002) English test and 2002 to 2004 DRET. The themes of the texts and text variables that accounted for item difficulty were also investigated. Different from Lu's study, Hsu examined the use of words in the chosen texts by comparing them with those on the Word List published by the DRT. Instead of directly computing the passing rates of examinees taking those tests, she examined the performances of 76 Grade-2 students (divided into the high-proficiency group, the middle-proficiency group, and the low-proficiency group) from two high school classes.

A couple of problems arose when Lu (2002) and Hsu (2005) claimed that students performed well in what they termed lower-level reading skills (i.e., determining the meaning of words, finding specific details, and identifying the main

idea) and failed when questions tapped higher-level reading skills. If we examine the scheme of reading skills proposed by Mo (1987) carefully, we can find that the categories cannot really describe either higher or lower cognitive processes involved in answering a reading comprehension question. Beatty (1975) indicated that skills such as finding the main idea could involve either the lowest-level processing (when the main idea is in the topic sentence, which can be easily identified) or higher-level processing (when the main idea is implied, and thus should be classified as the level of apprehension rather than recall). Therefore Mo's category of "Identifying the main idea" can refer to different levels of understanding. In addition, none of the reading skills in Mo's classification describe cognitive skills of applying, analyzing, and creating.

Hence, the present study attempts to adopt another framework, the Revised Bloom's Taxonomy, in which the categories represent the cognitive processes (or can be viewed as cognitive skills) involved in reading, to analyze the reading comprehension multiple-choice questions on both scholastic achievement and department required English tests in Taiwan in the past five years. This study aims to illuminate the content and construct of the reading comprehension section in the SAET and the DRET.

Methodology

Materials

The reading passages and their comprehension questions that appeared in SAETs and DRETs administered from 2002 to 2006 were the data source for analysis. In general, a reading comprehension test consists of three to four passages,

with three to five items constituting a set of questions under each passage. A total of 36 reading passages and 140 comprehension items (i.e., 77 items in the SAETs and 63 items in the DRETs) were collected.

Reading comprehension questions on the 2002 SAET were used for trial item analysis for the raters to familiarize themselves with the coding procedure.

Instrument

Revised Bloom's Taxonomy was used to code the reading comprehension test items. The framework is presented in a two-dimensional table—the Taxonomy Table. The rows and columns of the table contain the knowledge categories (Factual Knowledge, Conceptual Knowledge, Procedural Knowledge, and Metacognitive Knowledge) and cognitive process categories (Remember, Understand, Apply, Analyze, Evaluate, and Create).

For the purpose of coding reading comprehension items, the definition for the category of Remember was modified. In the Revised Bloom's Taxonomy, Remember involves retrieving relevant knowledge from long-term memory through recognizing and recalling. However, in examinations, reading comprehension questions do not require students to retrieve the information in their long term memory in order to answer the questions because the reading passage is always there and retrievable through search reading. Items that aim at asking learners the identifiable information (i.e. facts that are explicitly stated in the text) belong to the category of Recognizing, a subtype of Remember (as defined by Beatty, 1975 in the category of Recall). If the item tests students' memorization rather than comprehension, then the item goes to the category of Recalling.

Moreover, examples for each subtype of knowledge were selected from literature to fit the reading context. For instance, "Knowledge of specific details and elements" (one subtype of Factual Knowledge) refers to the specific factual information, statements, which are explicitly stated in the reading passage. Words, phrases, or referents that have contextualized meaning also belong to this type. If a question asks for the meaning of certain word that has only one definition regardless of the context, this type of knowledge belongs to "Knowledge of terminology," for it does not belong to the specific elements that can be found in the text.

Data Analysis

The data analysis procedures included three phases: (1) trial item analysis, (2) formal item analysis, and (3) analysis of the coded data.

Trial Item Analysis

Besides the researcher herself, two MA students in the TESOL program at National Taiwan Normal University were recruited to serve as raters to ensure the validity of coding results.[1] The criteria for coding, the definitions of the Revised Bloom's Taxonomy (Anderson & Krathwohl, 2001), the modified framework (see Appendix A), and some example questions from tests in other disciplines (Yeh & Lin, 2003) were given to the raters to help them become familiar with the coding framework.

Prior to the entire coding procedure, a trial analysis was conducted for the raters

---

[1] Besides the researcher, two other raters participated in the coding process. They both majored in TESOL at National Taiwan Normal University. One rater is now an English teacher in Shilin High School of Commerce; the other is an English teacher in Shimen Juninor High School in Taipei County.

to familiarize themselves with the coding procedure. A set of coding sheets was given to the raters to categorize each test item into the two dimensions. Both the major and sub knowledge types and cognitive process levels were identified.

Based on its item, the correct answer, and the required information in the reading passage to answer the question, each question was classified into a major and sub category of cognitive process levels and then a major and sub knowledge types.

For example, to categorize this item,

Toscanini thought that Marian Anderson _____.

*(A) had a very rare voice

(B) sang occasionally in public

(C) sang only once in many years

(D) was seldom heard by people (Q 48 from 2002 SAET)

the rater needs to first identify the requested information (as below) to determine which cognitive skill is required to answer this question.

> … In 1955, Anderson became the first black soloist to sing with the Metropolitan Opera of New York City. The famous conductor Toscanini praised her voice as "heard only once in a hundred years." She was a U.S. delegate to the United Nations in 1958 and won the UN peace prize in 1977….

To answer this question, test takers need to contextualize the meaning of the statement "hear only once in a hundred years," which is the comment, a praise, that Toscanini gave to Marian Anderson. Readers need to interpret this sentence based on the context to reach the correct answer, which is option A, "had a very rare voice."

The item will then be first categorized as "Understand (2.1 Interpreting)." (see Appendix A). Furthermore, since the meaning of that statement is a fact in that reading passage— Anderson had a very rare voice—it will thus be classified as "Factual Knowledge (1.1 Knowledge of specific details and elements)." This item will then be regarded as "Understand Factual Knowledge" item as well as "Interpreting specific details and elements" item at the same time.

Fifteen items from 2002 SAET reading comprehension section were used for trial analysis. Each rater conducted his/her analysis independently. Subsequent to three raters' classifications, three sets of coded data were compared to check the inter-rater consistency.

The agreement rate for the trial analysis was only 33.33% due to raters' insufficient understanding of the coding scheme, particularly of the cognitive process dimension. Variations in judgments were discussed among all raters until a consensus was reached. Before the formal analysis, a second discussion on the first coding of items on 2002 SAET was held. The coding results of the 2002 SAET finalized in the discussion will be adopted as data for statistical analysis.

Formal Item Analysis

In the formal item analysis, procedures for coding and discussion in the trial analysis were applied. The resulting consistency rates among the three raters were merely 66.13% in the SAET data sets and 55.56% in the DRET data sets, presumably due to the inadequate rater training. Therefore, to ensure the inter-rater validity and consistency on classification, the data coding results were finalized between two raters who demonstrated higher coding consistency. Also, 20% of the discordant items were checked by another expert who was familiar with Bloom's

Taxonomy to reach a final decision.❷

Analysis of the Coded Data

A Crosstabulation was computed to see the frequency counts and percentages of the major combinations as well as the sub-combinations of the cognitive process and knowledge dimensions occurred in the two examinations. The counts and percentages of each combination were then tabulated across years separately in both tests in order to draw detailed comparisons of the occurrences, frequencies, and distributions of the question types between the SAET and the DRET reading comprehension item designs.

Results and Discussion

Cognitive Skills and Knowledge Types Measured

From the item classification of the 2002 to 2006 SAET and DRET, four major cognitive process levels along with eight sub-levels and three types of knowledge with four subtypes were found. The four main cognitive skills were: Remember, Understand, Apply, and Analyze. As to its sub-skills, of these four cognitive skills identified, only the category of Understand was found to have more than one sub-skill. Five out of the seven sub-processes of Understand in the Revised Bloom's Taxonomy surfaced in both examinations: Interpreting, Summarizing, Inferring, Classifying (only in the DRET), and Explaining (only in the SAET). Other main categories had merely one sub-category identified: Recognizing (under Remember),

❷ Specializing in reading and using Bloom's Taxonomy in test items classification, the expert who helped to finalize the discordant coding results is a professor in National Taiwan Normal University.

Executing (under Apply), and Attributing (under Analyze). As for the knowledge types, Factual Knowledge included one subtype, i.e., Knowledge of specific details and elements; Conceptual Knowledge had two subtype, i.e., Knowledge of principles and generalizations and Knowledge of classification and categories; and Procedural Knowledge had one subtype, i.e., Knowledge of subject-specific skills and algorithms.

These identified cognitive levels and types of knowledge were then recognized to constitute five major combinations: (1) Remember Factual Knowledge, (2) Understand Factual Knowledge, (3) Understand Conceptual Knowledge, (4) Apply Procedural Knowledge, and (5) Analyze Conceptual Knowledge. Nine sub combinations were also identified: (1) Recognizing specific details and elements, (2) Interpreting specific details and elements, (3) Inferring specific details and elements, (4) Classifying into classifications and categories, (5) Summarizing principles and generalizations, (6) Inferring classifications and categories, (7) Explaining principles and generalizations, (8) Executing subject specific skills and algorithms, and (9) Attributing principles and generalizations. These major and sub combinations are referred to as major and sub question types in the present study, which are summarized in Table 1.

Items at the Evaluate and Create levels were not found in the tests analyzed. The reason might be that evaluating and creating are more productive skills that cannot be easily tested through multiple-choice questions. For knowledge types, metacognitive knowledge was not identified. As Anderson and Krathwohl (2001) stated, metacognitive knowledge is hard to be measured using multiple-choice questions; it will be more easily assessed in classroom activities and discussion of

Table 1 Major Types and Subtypes of the Cognitive Skills and Knowledge Tested

Major Types	Subtypes
1. Remember Factual Knowledge	Recognizing specific details and elements
2. Understand Factual Knowledge	Interpreting specific details and elements, Inferring specific details and elements
3. Understand Conceptual Knowledge	Classifying into classification and categories Summarizing principles and generalizationsInferring classifications and categories Explaining principles and generalizations
4. Apply Procedural Knowledge	Executing subject specific skills and algorithms
5. Analyze Conceptual Knowledge	Attributing principles and generalizations

various reading strategies.

A Crosstabulation was conducted to see the frequency counts of each question type in the SAET and the DRET. The results are presented in Table 2.

Table 2 The Crosstabulation of the Cognitive Skills and Knowledge Types Identified on the SAET and the DRET Test Items

The Knowledge Dimension		The Cognitive Process Dimension			
		Remember	Understand	Apply	Analyze
Factual Knowledge	Count	58	45	0	0
	% of Total	100.0%	70.3%	.0%	.0%
Conceptual Knowledge	Count	0	19	0	13
	% of Total	.0%	29.7%	.0%	100.0%
Procedural Knowledge	Count	0	0	5	0
	% of Total	.0%	.0%	100.0%	.0%

Table 2 shows that in general, the SAET and the DRET combined included 58

Remember Factual Knowledge items, followed by 45 items coded under Understand Factual Knowledge, 19 items under Understand Conceptual Knowledge, 13 items under Analyze Conceptual Knowledge, and 5 items under Apply Procedural Knowledge.

As illustrated in Table 2, among test items classified as Remember (58 items in total), all of them were identified as testing factual knowledge. For items under the category of Understand, 70.3% of the items in this category (45 items) were labeled as Factual Knowledge items, 29.7% (19 items) were Conceptual Knowledge items, and none were Procedural Knowledge items. This indicated that most of the Understand items were Understand Factual Knowledge mainly focusing on specific details in the reading passage, while around 30% of the items were Understand Conceptual Knowledge questions that asked students to understand the text as a whole, a general concept of the text such as the main idea. As for items classified as Apply, all of the items were labeled as Procedural Knowledge, while items belonging to Analyze category were classified as Conceptual Knowledge.

This finding is actually predictable and close to our intuition. In the context of reading comprehension test, for example, if we want to test student's knowledge of specific details in a reading passage, we usually would test students' ability to recognize factual information, and a Remember Factual Knowledge item is thus constructed to meet this aim. Likewise, when we intend to assess students' summarizing skill, we would test their knowledge of the general idea of that passage. In line with Chen's (2004) finding in the computer science discipline which revealed this knowledge-and-cognitive association, result in the present study further confirms Anderson and Krathwohl's (2001) proposition that certain types of

knowledge tended to be associated with certain types of cognitive skills.

Different from previous studies on test item analyses, the present study identified the sub knowledge types and cognitive processes tested on each item to offer a better understanding of the content and cognitive skills measured. The results showed that a total of nine sub combinations/types of items surfaced on the SAET and the DRET. Items of Remember Factual Knowledge, Apply Procedural Knowledge, and Analyze Conceptual Knowledge each had only one subtype, namely Recognizing specific details and elements, Executing specific skills and algorithms, and Attributing principles and generalizations. There are a number of possible explanations for these major categories to include only one subcomponent. For the Remember category, since all the texts were retrievable through search reading, no Recalling items were needed, since testing students' memorization in reading is not the aim. For the Apply category, it was hard to find an Implementing item because items at this level in the current study refer only to items involving calculation. As for the category of Analyze, one partial explanation of the nonexistence of its sub-skill "Organizing," which aims at testing readers' understanding of textual relationship, might be that it was tested in the Discourse Structure section in the DRET.

The Understand items included six sub question types: Interpreting specific details and elements, Classifying into classifications and categories (only on the DRET), Summarizing principles and generalizations, Inferring specific details and elements, Inferring classifications and categories, and Explaining principles and generalizations (only on the SAET). As can be seen in the cognitive process dimension, five out of seven sub cognitive skills of understanding in Bloom's

Taxonomy were included on the test. The result indicates that the SAET and the DRET in the past five years indeed attempted to measure a variety of sub skills of understanding. Skills such as interpreting, summarizing, and inferring were emphasized in both tests. In turn, it shows that various sub cognitive skills of Understand in reading can be measured via multiple-choice questions.

Costin and Shen (1986), while surveying teachers' views on the most-required cognitive skills and skills that the weak ESL readers lacked, indicated that cognitive skills of knowledge, comprehension, application, and analysis (corresponding to remember, understand, apply, and analyze in the revised taxonomy in this study) were the most essential skills for college students while reading in English. Congruent with his perspective, results of item analysis showed that the SAET and the DRET did include items that measure these four skills in reading. Moreover, the inclusion of these four major skills in reading test is in accord with the testing objectives that the College Entrance Examination Center (CEEC) set for each discipline, especially those for the DRET, in which thinking skills such as understanding the implied meaning, analyzing the data, or applying the knowledge to different contexts are major goals for testing.

Similarities and Differences Between the SAET and the DRET

To investigate the similarities and differences in the frequency and distribution of each type of question between the SAET and the DRET, crosstabulations were employed separately to both examinations. Crosstabulation analyses revealed the frequency counts and percentages of the major and subordinate cognitive processes and knowledge types measured in the SAET and the DRET administered from 2002

to 2006. The results are presented in Table 3.

SAET

As shown in Table 3, in SAETs administered from 2002 to 2006, 35 out of 77 items (45.5%) were categorized as Remember Factual Knowledge, which was the most frequent question type. Understand Factual Knowledge items (28.6%) came second, comprising 20 Interpreting items and two Inferring specific details and elements items. Understand Conceptual Knowledge was the third frequent items (11.7%), including six Summarizing items, two Inferring classifications and categories items, and one Explaining item, followed closely by Analyze Conceptual Knowledge items (seven items, accounting for 9.1%). Only four items were coded as Apply Procedural Knowledge (5.2%). The result of item frequency in the SAET showed that, in general, around half of the items aimed to test students' ability to recognize facts in the articles and almost one third of the items measured the ability to understand specific details. Few items, only 11.7%, tested learners' cognitive skills like summarizing, explaining, or inferring the source of the article; 5.2% of the questions asked students to execute calculation, and 9.1% of the questions measured higher cognitive skills of finding the writer's tone or intention.

Another observation which can be made from Table 3 is that, within the major question type of Understand, Interpreting items apparently stood out, with a proportion specifically higher than others (20 out of 31 items). This was followed by items of Summarizing principles and generalizations (six items), Inferring specific details and elements (two items), Inferring classifications and categories (two items), and Explaining principles and generalizations (one item). These results indicated that even though a diversity of sub-skills of Understanding were tested, there was an

Table 3 Cognitive (Sub)skills and Knowledge (Sub)types Measured on the SAET and the DRET Test Items

Test	The Knowledge Dimension			The Cognitive Process Dimension							
				Remember		Understand				Apply	Analyze
				Recognizing	Interpreting	Classifying	Summarizing	Inferring	Explaining	Executing	Attributing
SAET	Factual	Specific details and elements	Count	35	20		0	2	0	0	0
			% of Total	45.5%	26.0%		.0%	2.6%	.0%	.0%	.0%
	Conceptual	Classifications and categories	Count	0	0		0	2	0	0	0
			% of Total	.0%	.0%		.0%	2.6%	.0%	.0%	.0%
		Principles and generalizations	Count	0	0		6	0	1	0	7
			% of Total	.0%	.0%		7.8%	.0%	1.3%	.0%	9.1%
	Procedural	Subject-specific skills and algorithms	Count	0	0		0	0	0	4	0
			% of Total	.0%	.0%		.0%	.0%	.0%	5.2%	.0%
DRET	Factual	Specific details and elements	Count	23	19	0	0	4		0	0
			% of Total	36.5%	30.2%	.0%	.0%	6.3%		.0%	.0%
	Conceptual	Classifications and categories	Count	0	0	3	0	4		0	0
			% of Total	.0%	.0%	4.8%	.0%	6.3%		.0%	.0%
		Principles and generalizations	Count	0	0	0	3	0		0	6
			% of Total	.0%	.0%	.0%	4.8%	.0%		.0%	9.5%
	Procedural	Subject-specific skills and algorithms	Count	0	0	0	0	0		1	0
			% of Total	.0%	.0%	.0%	.0%	.0%		1.6%	.0%

Note. The percentages were calculated based on the total number of reading comprehension items in each kind of test.

uneven distribution among these subtypes of questions, with a predominant emphasis on the interpreting skill.

DRET

As to the DRET administered from 2002 to 2006 (also shown in Table 3), the most frequent question types were items on Remember Factual Knowledge (36.5%) and Understand Factual Knowledge (36.5%), which included 19 Interpreting and four Inferring specific details and elements items. Understand Conceptual Knowledge items (15.9%), containing three Classifying, three Summarizing, and four Inferring classifications and categories items, came the second, followed by six Analyze Conceptual Knowledge items (9.5%), and one Apply Procedural Knowledge item. The trend was almost the same as that of the SAET. Similar to the SAET, test items in the DRET also focused on measuring students' abilities to identify and to understand facts, with less attention on testing analytical skills.

For the major question type of Understand, the DRET included five sub question types, with Interpreting items took up the largest proportion (19 items), followed by items of Classifying (three items), Summarizing (three items), Inferring specific details and elements (four items), and Inferring classifications and categories (four items). Obviously, similar to the SAET, the DRET also focused on testing students' ability to interpret textual meaning in context more than other sub-skills of understanding.

Similarities

In general, the SAET and the DRET did not exhibit considerable differences in their inclusions of major and sub question types. According to Table 3, these tests both contained major questions on Remember Factual Knowledge, Understand

Factual Knowledge, Understand Conceptual Knowledge, Apply Procedural Knowledge, and Analyze Conceptual Knowledge in the past five years. Of the five subtypes of questions under the Understand category, both tests measured students' skills of interpreting, summarizing, and inferring.

In terms of frequency, both tests showed a similar pattern: most of the items were Remember Factual Knowledge, followed by Understand Factual Knowledge, Understand Conceptual Knowledge, Analyze Conceptual Knowledge, and Apply Procedural Knowledge. These findings are in line with studies on test item analyses that adopted either the original or the Revised Bloom's Taxonomy in different disciplines. All of the studies showed a consistent finding that cognitive skills of remembering, understanding, applying, and analyzing were found most commonly tested on multiple-choice questions, with items on Remember/Knowledge and Understand/Comprehension being the most frequent (Chen, 2004; Masters et al. 2001; Squire, 2001). This indicates that most of the test items were merely testing the two lowest levels of Bloom's Taxonomy, and higher cognitive skills were rarely tested. A possible explanation for this phenomenon might be that items of Remember and Understand are more easily constructed in multiple-choice formats, yet items of higher level cognitive skills such as evaluating and creating are productive in nature and therefore not easily tested in a multiple-choice format.

Based on close examination on the frequency of the sub question types, Table 3 shows that these two types of questions took up a great proportion in both tests—Recognizing specific details and elements and Interpreting specific details and elements. They together constituted a substantially higher proportion, accounted for over 66%, i.e., 71.5% for SAET vs. 66.7% for DRET. Recognizing and Interpreting

questions in the present study are more language bound that aim at testing readers' bottom-up processing skills in reading. Remembering/Recognizing items require students to find out specific facts such as events, location, and reason stated explicitly in the reading passages; whereas Understanding/Interpreting items measure students' ability to understand sentential, vocabulary, or pronoun meanings in context. The SAET and the DRET had a great number of items of these two types. This suggests that both tests highly emphasized examinees' abilities to locate details, understand specific information, or interpret sentential or lexical meanings in context, which belong to micro-level comprehension. This concentration on assessing the micro-level understanding in the SAET and the DRET would likely lead the EFL teachers or learners to believe that lower-level reading skills are much more important than the higher ones.

As Alderson and Lukmani (1989) stated, if measuring language ability rather than cognitive skills was the goal in a reading test, it would be better to design a test with more lower-level questions to increase the content validity. From what was found in the present study, it was apparent that the SAET and the DRET aimed more at tapping students' language abilities rather than their cognitive skills. Higher cognitive level items occupied less than 10% in either test, and those items were designed mostly for "identifying the author's tone." The other two analytical skills, i.e., differentiating and organizing under the Analyze category in the revised taxonomy, were not tested, neither were even higher level thinking skills like evaluating or creating.

Whether this overemphasis on lower-level reading skills contradicts with the testing goals set by the CEEC or whether higher cognitive skills like Evaluate and

Create should be tested in the college entrances examinations require in-depth investigations on the objectives for English reading comprehension test in the future.

Differences

To explore the differences between the two kinds of tests, the frequency of question subtypes were further examined year by year.

As can be seen from Table 4, in the SAET, an average of 45% of the total was Recognizing items, accounting for the largest proportion among the item types every year. The proportion of the Recognizing items in the DRET, on the other hand, fluctuated broadly from 33.3% in 2002 to 53.3% in 2003, 45% in 2004 and dropped to 18.2% in 2005 and 27.3% in 2006, according to the amount of emphasis given to the Interpreting items. For the category of Understand, the DRET had slightly more Inferring items than the SAET did. As to the category of Apply/Executing, the SAET included more items than the DRET. This finding indicated that certain types of questions were preferred by these two tests. Executing questions were more common in SAETs (occurred in 2002, 2003 and 2005), while more Inferring questions asking for either inferential facts or concepts were favored in DRETs (occurred nearly every year, except 2005). The second difference was the occurrence of certain types of items. Of the sub question types, Explaining items occurred in the SAET in 2002; whereas Classifying items appeared in the DRET in 2005 only. The last difference between the two examinations was the distribution of the cognitive sub-skills and knowledge subtypes through years. If we examine the distribution of those cognitive sub-skills and knowledge subtypes by year, as shown in Table 4, the SAET consisted of at least three question subtypes each year, ranging from three to seven among these five years: five types in 2002, seven in 2003, three in 2004, six in 2005, and

Table 4 Question Subtypes in the SAET and the DRET by Year

			Question Type								
			Recognizing specific details and elements	Interpreting Specific details and elements	Classifying classifications and categories	Summarizing principles and generalizations	Inferring specific details and elements	Inferring classifications and categories	Explaining principles and generalizations	Executing subject specific skills and algorithms	Attributing principles and generalizations
SAET	2002	Court	7	5	0	0	0	0	1	1	1
		% of Total	46.7%	33.3%	.0%	.0%	.0%	.0%	6.7%	6.7%	6.7%
	2003	Court	6	2	0	3	1	1	0	1	1
		% of Total	40.0%	13.3%	.0%	20.0%	6.7%	6.7%	.0%	6.7%	6.7%
	2004	Court	7	6	0	2	0	0	0	0	0
		% of Total	46.7%	40.0%	.0%	13.3%	.0%	.0%	.0%	.0%	.0%
	2005	Court	6	3	0	0	1	1	0	2	3
		% of Total	37.5%	18.8%	.0%	.0%	6.3%	6.3%	.0%	12.5%	18.6%
	2006	Court	9	4	0	1	0	0	0	0	2
		% of Total	56.3%	25.0%	.0%	6.3%	.0%	.0%	.0%	.0%	12.5%
DRET	2002	Court	5	5	0	0	1	1	0	0	3
		% of Total	33.3%	33.3%	.0%	.0%	6.7%	6.7%	.0%	.0%	20.0%
	2003	Court	8	3	0	1	1	1	0	0	1
		% of Total	53.3%	20.0%	.0%	6.7%	6.7%	6.7%	.0%	.0%	6.7%
	2004	Court	5	3	0	1	1	1	0	0	0
		% of Total	45.5%	27.3%	.0%	9.1%	9.1%	9.1%	.0%	.0%	.0%
	2005	Court	2	4	3	1	0	0	0	1	0
		% of Total	18.2%	36.4%	27.3%	9.1%	.0%	.0%	.0%	9.1%	.0%
	2006	Court	3	4	0	0	1	1	0	0	2
		% of Total	27.3%	36.4%	.0%	.0%	9.1%	9.1%	.0%	.0%	18.2%

Note. The percentages were calculated based on the total number of reading comprehension items in each year.

four in 2006.

The DRET had five question subtypes on average each year, except 2003 when six question types emerged. This showed that the DRET consistently covered various question subtypes each year whereas the SAET fluctuated in its variety of question types. Overall, the major differences between the two tests were frequency, occurrence, and distribution of the cognitive sub-skills and knowledge subtypes through years.

To sum up, these findings showed that the cognitive skills and knowledge types tested in both examinations were relatively fixed. Five major question types accompanying nine subtypes surfaced in the SAET and the DRET. Recognizing items (known as Remember items) and Interpreting items (subtypes of Understand Factual Knowledge items) were in the majority, which suggested that both kinds of tests emphasized measuring examinees' micro-level comprehension of a text. A few items asked students to understand the passage as a whole through summarizing, classifying, establishing a cause-and-effect model, or inferring, yet fewer items measured higher cognitive skills such as applying and analyzing. Whether even higher cognitive skills like Evaluate and Create in reading should be tested or whether these two types of skills are considered essential in senior high English reading instruction becomes an issue to be explored by the CEEC while setting test specification, and by the MOE (Ministry of Education) in setting national curriculum objectives for English subject.

Conclusion

This study was undertaken to investigate what cognitive process levels and

knowledge types were tested on the reading comprehension test items of the SAET and the DRET administered from 2002 to 2006. The results of item analysis on the SAET and the DRET can have a positive washback effect to EFL reading instruction in senior high schools. The major result shows that from both tests, five major question types were identified. They were (1) Remember Factual Knowledge (also known as its subtype, recognizing), (2) Understand Factual Knowledge, (3) Understand Conceptual Knowledge, (4) Apply Procedural Knowledge (also known as its subtype, executing), and (5) Analyze Conceptual Knowledge (also known as its subtype, attributing). Under these major classifications, six sub question types were identified under the Understand category: interpreting, classifying (only in the DRET), summarizing, inferring specific details, inferring classifications, and explaining (only in the SAET). There was a tendency that the Apply/Executing items were more frequently found in the SAET whereas the DRET had more Inferring items, showing that certain types of questions were favored in specific kind of test. These skills are hence suggested to be taught or assessed in the classroom. Although the analysis reveals a diversity of questions types, items on Remember and Understand Factual Knowledge, which are closely related to the bottom-up processing, were still the majority in the two tests throughout the five years. It seems that students' English ability, rather than higher cognitive skills, is the test goal in college entrance examination. Teachers then need to help learners build up their lexical and syntactical competence as well as the ability to use contextual clues to interpret words or sentences in context.

In both tests, few items were found at higher cognitive levels of Apply and Analyze, yet they seem to be constructible with a slight revision of items or texts.

For example, to construct a differentiating item (under the category of Analyze), which require readers to make comparison of two things through discriminating relevant from irrelevant information and to find out what elements distinguish those two objects being compared (Anderson & Krathwohl, 2001), possible ways of asking a differentiating question are: "According to the passage, the two objects differ primarily in…?" or "What makes this object different from the others?" etc.

For even higher cognitive skills of Evaluate and Create, they can probably be tested in open-ended questions in the writing section on the college entrance examinations. For example, to test learners' cognitive ability of evaluating, test takers can be asked to give judgments or critiques in terms of writer's argument or to evaluate how far the author has achieved their argumentation. As for the skill of creating, examinees can be asked to produce a solution or to give their opinions for a problematic issue or condition. However, although it is possible to pose questions at higher cognitive process levels, we probably need to go back to the curriculum objectives set by the MOE to see if they are truly required for senior high students while learning English.

There are of course some limitations in this study. First, item analysis in the present study can only predict the cognitive process levels and knowledge types attempted to be tested on each item. To know if test takers truly apply certain cognitive skills while answering a question, methods such as think aloud or interview can be adopted. Second, the relatively small number of items for analysis each year is not sufficient for statistical multi-comparisons. There were only 11 to 16 test items on the reading comprehension section each year; they were classified into at least three types of questions, which made each category contain less than two

items sometimes.

The present study has attempted to classify reading comprehension items on the SAET and the DRET. It is suggested that the Revised Bloom's Taxonomy be adopted to analyze reading comprehension questions in senior high school English textbooks. It can also be used to analyze teachers' questions asked in class to see whether these questions indeed match the basic cognitive skills in reading. Lastly, the new Senior High Curriculum in 2010 specifies that students' critical thinking abilities in English and creativity, which are higher cognitive skills mentioned in the Revised Taxonomy, need to be acquired. It would be interesting to see how the new guidelines on those abilities can be tested in the future.

References

Adams-Smith, D. (1981). Levels of questioning: Teaching creative thinking through ESL. *Forum, 19*(1), 15-21.

Airasian, P. (1994). The impact of the taxonomy on testing and evaluation. In I. L. W. Andersion & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 82-102). Chicago, IL: The National Society for the Study of Education.

Alderson, J. C., & Lukmani, Y. (1989). Cognition and reading: Cognitive levels as embodied in test questions. *Reading in a Foreign Language, 5*(2), 253-270.

Alderson, J. C. (1996). *The testing of reading*. In C. Nuttall (Ed.), *Teaching reading skills in a foreign language* (pp. 212-228). London: Heinemann.

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's educational objectives*. New York: Longman.

Aviles, C. B. (1999). Understanding and testing for "critical thinking" with Bloom's taxonomy of educational objectives. Washington, DC: ERIC Clearinghouse. (ERIC SO032219)

Beatty, R. J. (1975). Reading comprehension skills and Bloom's taxonomy. *Reading World, 15*(2), 101-108.

Bernhardt, E. B. (1991). *Reading development in a second language: Theoretical, empirical and classroom perspectives*. Norwood, NJ: Ablex.

Buckles, S., & Siegfried, J. J. (2006). Using multiple-choice questions to evaluate in-depth learning of economics. *Journal of Economic Education, 37*(1), 48-57.

Chen, H. J. (2004). An analysis of computer science exam questions using revised Bloom's taxonomy. Unpublished master's thesis, National Taiwan Normal University, Taipei.

Costin, B., & Shen, W. H. (1986). Cognition and comprehension: A study of the need for incorporation of Bloom's taxonomy of education objectives in English as a second language remedial reading programs. Unpublished M. Phil. Thesis, Chinese University of Hong Kong, Hong Kong. (ED297595)

Furst, E. (1994). Bloom's taxonomy: Philosophical and educational issues. In L. W. Anderson & L. A. Sosniak (Eds.), *Bloom's Taxonomy: A forty year retrospective* (pp. 28-40). Chicago, IL: The University of Chicago Press.

Gierl, M. J. (1997). Comparing cognitive representations of test developer and students on a mathematics test with Bloom's taxonomy. *Journal of Educational Research, 91*(1), 26-32.

Hsu, W. L. (2005). An analysis of the reading comprehension questions in the JCEE English test. Unpublished master thesis, National Kaohsiung Normal University, Kaohsiung.

Huang, T. S. (1994). *A qualitative analysis of the JCEE English tests*. Taipei: Crane.

Kreitzer, A. E., & Madaus, G. F. (1994). Empirical investigations of theherarchical structure of the taxonomy. In L. W. Andersion & L. A. Sosniak (Eds.), *Bloom's taxonomy: A forty-year retrospective* (pp. 64-81). Chicago, IL: The National Society for the Study of Education.

Lu, J. J. (2002). An analysis of the reading comprehension test given in the English subject ability test in Taiwan and its pedagogical implications. Unpublished master thesis, National Chengchi University, Taipei.

Masters, J. C., Hulsmeyer, B. S., Pike, M. E., Leichty, K., Miller, M. T., & Verst, A. L. (2001). Assessment of multiple-choice questions in selected test banks accompanying text books used in nursing education. *Journal of Nursing Education, 40*(1), 25-32.

Mo, C. C. (1987). A study of English reading comprehension and general guidelines for testing reading. *Journal of National Chengchi University, 55*, 173-206.

Simkin, M. G., & Kuechler, W. L. (2005). Multiple-choice tests and student understanding: What is the connection? *Decision Science—The Journal of Innovative Education, 3*(1), 73-98.

Squire, P. J. (2001). Cognitive levels of testing agricultural science in senior secondary schools in Botswana. *Education, 118*(1), 100-110.

Surjosuseno, T. T., & Watts, V. (1999). Using Bloom's Taxonomy to teach critical reading in English as a foreign language classes. *Queensland Journal of Educational Research, 15*(2), 227-244.

Yeh, L. C., & Lin, S. P. (2003). Exploring the revision of Bloom's educational objectives. *Journal of Educational Research, 105*, 94-106.

Appendix A

Cognitive Process Dimension

1. A test item is categorized as *Remember,* retrieving relevant knowledge from long-term memory, when it requires examinees

- to identity the information stated explicitly, especially when it includes key words like "the main reason is that…" or the "the main idea is …"

(1.1 Recognizing)

- to retrieve relevant knowledge from long-term memory

(1.2 Recalling)

2. A test item is categorized as *Understand*, determining the meaning of instructional messages, including oral, written and graphic communication, when it requires examinees

- to understand the sentence(s) through changing from one form of representation to another (e.g., words to pictures, or words to words)

(2.1 Interpreting)

- to find specific examples of a concept or principle

(2.2 Exemplifying)

- to determine that something belongs to a category

(2.3 Classifying)

- to abstract a general theme or major point(s)

(2.4 Summarizing)

- to draw a logical conclusion from presented information

(2.5 Inferring)

- to detect correspondences between two ideas, objects, and the like

(2.6 Comparing)

- to construct a cause-and-effect model for a system

(2.7 Explaining)

3. A test item is categorized as *Apply*, carrying out or using a procedure in a given situation, when it requires examinees

- to apply a procedure to a familiar task

(3.1 Executing)

- to apply a procedure to an unfamiliar task

(3.2 Implementing)

4. A test item is categorized as *Analyze*, breaking material into its constituent parts and detecting how the parts relate to one another and to an overall structure or purpose, when it requires examinees

- to distinguish relevant from irrelevant parts or important from unimportant parts of the text

(4.1 Differentiating)

- to determine how elements fit or function within a structure

(4.2 Organizing)

- to determine a point of view, bias, values, or intent underlying presented material

(4.3 Attributing)

5. A test item is categorized as *Evaluate*, making judgments based on criteria and standards, when it requires examinees

- to detect inconsistencies or fallacies within a process or product; to determine whether a process or product has internal consistency; to detect the effectiveness of a procedure as it is being implemented

(5.1 Checking)

- to detect inconsistencies between a product and external criteria, to determine whether a product has external consistency; to detect the appropriateness of a procedure for a given problem

(5.2 Critiquing)

6. A test item is categorized as *Create*, putting elements together to form a novel, coherent whole or make an original product, when it requires examinees

- to come up with alternative hypotheses based on criteria

(6.1 Generating)

- to devise a procedure for accomplishing some task

(6.2 Planning)

- to invent a product

(6.3 Producing)

Knowledge Dimension

A. test item is categorized as *Factual Knowledge*, the basic elements students must know to be acquainted with a discipline or solve problems in it, when it measures examinees' knowledge of

- definition of words, phrases, etc.

(Aa. Knowledge of terminology)

- specific and explicitly stated details or elements, propositional information inferences (answering questions beginning with who, when, what), propositional explanatory inferences (answering questions beginning with why, how), etc.

(Ab. Knowledge of specific details and elements)

B. A test item is categorized as *Conceptual Knowledge*, the interrelationships among the basic elements within a larger structure that enable them to function together, when it measures examinees' knowledge of

- the genre, or the parts of sentences (e.g. nouns, verbs, adjectives)

(Ba. Knowledge of classifications and categories)

- main ideas, major generalizations, major principles, implications, etc.

(Bb. Knowledge of principles and generalizations)

- structure of a sentence/paragraph/passage, or interrelationships among principles or theories, etc.

(Bc. Knowledge of theories, models, and structures)

C. A test item is categorized as *Procedural Knowledge*, how to do something, methods of inquiry, and criteria for using skills, algorithms, techniques, and methods, when it measures examinees' knowledge of

- skills used to determine word meaning based on structural analysis, the skills of solving a mathematic problem, etc.

(Ca. Knowledge of subject-specific skills and algorithms)

- various methods of literary criticism, etc.

(Cb. Knowledge of subject-specific techniques and methods)

- the criteria for determining which of several types of essays to write, which skills to apply, etc.

(Cc. Knowledge of criteria for determining when to use appropriate procedures)

D. A test item is categorized as *Metacognitive Knowledge*, knowledge of cognition in general as well as awareness and knowledge of one's own cognition, when it measures examinees' knowledge of

- elaborating, outlining, planning, comprehension-monitoring strategies, etc.

(Da. Strategic knowledge)

- elaboration strategies like summarizing and paraphrasing can result in deeper levels of comprehension, etc.

(Db. Knowledge about cognitive tasks)

- strategies used in certain situations, goals, personal interests, etc.

(Dc. Self-knowledge)