

DOCUMENT RESUME

ED 467 813

TM 034 357

AUTHOR Wang, Xiang Bo
TITLE Understanding Psychological Processes That Underlie Test Takers' Choices of Constructed Response Items. Law School Admission Council Computerized Testing Report. LSAC Research Report Series.
INSTITUTION Law School Admission Council, Princeton, NJ.
REPORT NO LSAC-R-97-05
PUB DATE 1999-05-00
NOTE 14p.
PUB TYPE Reports - Research (143)
EDRS PRICE EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS Advanced Placement; Chemistry; *Constructed Response; Difficulty Level; Essay Tests; *High School Students; High Schools; *Psychological Patterns; Questionnaires; *Student Attitudes; *Test Items
IDENTIFIERS *Choice Behavior

ABSTRACT

It has been found repeatedly that when test takers are allowed to choose a subset of constructed response (CR) items to answer on a test, they tend to choose differently and often perform lower on more popularly chosen items. This study investigated the psychological factors that influence test takers' choices. Using an experience that incorporated a mini Advanced Placement Chemistry Examination and a related questionnaire, the study reveals a series of psychological processes that consistently influence test takers' choice of CR items. Study participants, 618 Hawaiian students, virtually replicated the choice patterns of their 1989 national counterparts. Findings show that students' perceptions of item difficulty predicted the choice combinations and choice popularities. Essays perceived as easier were chosen more frequently, even though they might not have been truly easier. Students also tended to associate familiarity with easiness. In addition, test items with contents that reflected similar curricular instruction or exposure tended to be chosen together more often. The findings have implications for CR item pretesting, test construction, and other application possibilities for performance oriented tests. (Contains 4 figures, 4 tables, and 13 references.) (Author/SLD)

TM

ED 467 813

U.S. DEPARTMENT OF EDUCATION
 Office of Educational Research and Improvement
 EDUCATIONAL RESOURCES INFORMATION
 CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

PERMISSION TO REPRODUCE AND
 DISSEMINATE THIS MATERIAL HAS
 BEEN GRANTED BY

J. VASELECK

TO THE EDUCATIONAL RESOURCES
 INFORMATION CENTER (ERIC)

1

■ **Understanding Psychological Processes that Underlie Test Takers' Choices of Constructed Response Items**

Xiang Bo Wang
Law School Admission Council

■ **Law School Admission Council
 Computerized Testing Report 97-05
 May 1999**



A Publication of the Law School Admission Council

The Law School Admission Council is a nonprofit corporation that provides services to the legal education community. Its members are 196 law schools in the United States and Canada.

Copyright© 1999 by Law School Admission Council, Inc.

All rights reserved. This book may not be reproduced or transmitted, in whole or in part, by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission of the publisher. For information, write: Communications, Law School Admission Council, Box 40, 661 Penn Street, Newtown, PA 18940-0040.

LSAT® and the Law Services logo are registered marks of the Law School Admission Council, Inc.

This study is published and distributed by the Law School Admission Council (LSAC). The opinions and conclusions contained in this report are those of the author and do not necessarily reflect the position or policy of the Law School Admission Council.

Table of Contents

Executive Summary	1
Abstract	1
Introduction	1
Research Instrument, Subjects, and Methodology	2
Analyses and Results	4
<i>Component I: Comparing the Chemistry Ability Between the Hawaii Participants and National AP Chemistry Population</i>	<i>4</i>
<i>Component II: Comparing Essay Choices Between Hawaii Students and National AP Chemistry Test Takers</i>	<i>5</i>
<i>Component III: Relationship Between Students' Perceptions of Essay Difficulty and Their Choices</i>	<i>6</i>
<i>Component IV: Relationship Between Students' Perceptions of Content Dimensionality and Their Choices</i>	<i>7</i>
Conclusion and Discussion	9
References	10

Executive Summary

In an earlier study by Wang, it was reported that the 18,462 test takers of the 1989 Advanced Placement Chemistry Examination, when asked to select three essay questions to write on from among five possible choices, chose in a seemingly diverse fashion. Their average scores on the more popularly chosen essays were lower than those on the less popularly chosen counterparts. Although such findings have been confirmed by other studies, the causes for the disparity between the popularity of choices and performance on those choices is not known. As a continuation from this earlier study, the purpose of this research is to uncover the psychological processes that influence test takers' choices.

This present study, based on 618 students in Hawaii, an experiment that incorporated a mini Advanced Placement Chemistry Test, and a related questionnaire that tapped into students' perception of item difficulty and similarity, has revealed three major findings concerning how and why test takers chose the constructed response (CR) items as they did. First, by asking the 618 Hawaii students to choose three of the same five essay questions of the 1989 AP Chemistry Examination, this study found that this independent group of Hawaii students virtually replicated the entire choice pattern of their 1989 national counterparts. This indicates an inherent psychological process underlying what appears to be haphazard choice patterns.

What is the underlying psychological process? By asking the Hawaii students to rate the difficulty of essay items, this study found, through unidimensional scaling analyses, that students' perceptions of item difficulty can completely predict the choice combinations and choice popularities of the essay items. Essays perceived as easier by the students were chosen more frequently, even though they might not be truly easier. The students tended to associate familiarity with easiness. That is, the more familiar the items, the easier students viewed them to be.

Moreover, by asking the Hawaii students to evaluate the similarities of the essay items, this study revealed that the students' perception of essay likeness could further explain the choice of essay combinations. Test items whose contents reflected similar curricular instruction or exposure tended to be chosen together more often.

Abstract

It has been found repeatedly that when test takers are allowed to choose a subset of constructed response (CR) items to answer on a test, they tend to choose differently and often perform lower on more popularly chosen items. The purpose of this study is to find the psychological factors that influence test takers' choices. Using an experiment that incorporates a mini Advanced Placement Chemistry Exam and a related questionnaire, this study has revealed a series of psychological processes that consistently influence test takers' choices of CR items. The findings from this study offer a number of suggestions regarding CR item pretesting, test construction, and other application possibilities for performance-oriented tests.

Introduction

Although it is increasingly popular for many performance-oriented tests to contain constructed response (CR) items, a subset of which can be chosen by test takers to answer, it is only recently that some disturbing facts about the consequences of such choices on test performance have come to light. Substantial numbers of test takers perform poorly on the items they choose (Fremer, Jackson, & McPeck, 1968; Pomplum, Morgan, & Nellikunnel, 1992). Test takers of different gender and ethnic backgrounds seem to choose differently, which results in score biases to their disadvantage (Wainer & Thissen, 1993, 1994). In order to alleviate such biases, it has been advocated that scores of differentially chosen CR items be compared and equated (Wainer, Wang, & Thissen, 1994). The equating theory, methodology, and some untestable assumptions have been investigated and explicated (See Thissen, Wainer, & Wang, 1994; Wainer & Thissen, 1992, 1993, 1994; Wang, Wainer, & Thissen, 1995).

Based on systematic analyses of Part D of the 1989 Advanced Placement (AP) Chemistry Exam (The College Board, 1990) this author (Wang, 1996) reported five findings: (1) The five essays in Part D were chosen in dramatically different ways; (2) The more frequently chosen essays belonged to the core-chemistry content, while the least frequently chosen item addressed a highly specialized chemistry topic; (3) Test takers tended to score

lower on the more popularly chosen core-chemistry items than on the noncore-chemistry items; (4) The order in which the essays were presented seemed to have a significant effect on test-taker choice patterns—the test takers who chose items selectively performed significantly better than did those that chose items sequentially; (5) Except for extremely low-ability test takers, all test takers seemed to choose in a similar way.

Table 1 summarizes the information for Findings 1–3, which forms the basis of this present study. For the sake of brevity, evidence that substantiates the other findings presented in the earlier study (Wang, 1996) are omitted in this paper.

TABLE 1

Relationship between rank of essay choice preference and mean performance

Rank	1	2	3	4	5	6	7	8	9	10
Essay Choice	5,6,8	5,7,8	5,6,7	5,8,9	6,7,8	7,8,9	5,7,9	5,6,9	6,8,9	6,7,9
Choice Frequency	5,227	4,198	2,555	1,707	1,392	898	753	457	407	121
Mean Scores	7.10	8.14	4.57	9.89	7.18	9.72	8.37	7.85	8.75	7.12

Although curricular and instructional explanations were offered for the above seemingly contradictory findings through a survey of AP chemistry teachers in the state of Hawaii, there are still several unanswered questions: Why did the national test takers choose the five essays so differently? Can the diverse choice patterns of the national AP chemistry population be replicated by some other independent sample? What role did test takers' perceptions of the difficulty and dimensionality of these essays play in their choices? As a sequel to the earlier study (Wang, 1996), the purpose of this research is to find answers to these questions with emphasis on the psychological processes underlying test takers' choices.

Research Instrument, Subjects, and Methodology

For research purposes, this study used the "Advanced Placement Chemistry Survey and Test Kit" (the Kit hereafter) which was answered by approximately 680 students in Hawaii. This Kit consisted of four parts. Part 1—a general information survey—sought some demographic information on participants, such as age, gender, interest and length of chemistry study, career choices, and so on.

Designed to compare the chemistry ability of the Hawaii participants with the 1989 national AP chemistry exam test takers, Part 2 was a mini AP chemistry test. The 12 multiple-choice (MC) items in Part 2 were carefully selected from Part A of the 1989 AP Chemistry Exam to mimic its distribution of item content, difficulty, and discrimination, as well as its test information. Only 12 MC items were used due to the limited time available to the Hawaii students to complete the four parts of the Kit. These 12 MC items, along with 8 MC items in Part 3, appeared satisfactory for the purpose of this study.

Titled "AP Chemistry Multiple Choice Item Comparison and Performance," Part 3 was composed of four pairs of MC items of varying difficulty, content, and discrimination, accompanied by a series of questions. The direct relevance of Part 3 to this paper is that its 8 MC items were used along with the 12 items in Part 2 to compare the ability distributions of Hawaii students and national test takers.

Part 4, "AP Chemistry Essay Problem Comparisons" consisted of the same five essays that constituted Part D of the 1989 AP Chemistry Exam. A set of comparison questions were presented in this part. The purpose of Part 4 was to verify whether or not Hawaii participants would independently replicate the national choice patterns of the five CR items that were found in the 1989 AP chemistry data. A positive finding would strengthen the hypothesis that a systematic influence did underlie the seemingly divergent choices of the 1989 national AP chemistry test takers.

With the support of a wide range of chemistry teachers and students in Hawaii, the Kit was administered to over 680 students in Hawaii. However, data on only 618 students were used in this study due to incomplete responses and information. As summarized in Table 2, the 618 students participating in this study represented four academic levels of chemistry study: (1) one upper-division class of 15 students in the Department of Chemistry, University of Hawaii (UH), who had completed at least four semesters of college chemistry; (2) two

lower-division chemistry classes of 33 students in the Chemistry Department, who had completed their first year of college chemistry; (3) thirteen high school AP Chemistry classes of 237 students, a majority of the registered AP Chemistry population throughout the state of Hawaii in 1992; and (4) eleven high school general chemistry classes of 333 students.

TABLE 2
Summary of demographic information on survey subjects

Category	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Gender				
Male	262	42.4	262	42.4
Female	314	50.8	576	93.2
Unidentified	42	6.8	618	100.0
Ethnicity				
Japanese	211	34.1	211	34.1
Chinese	139	22.5	350	56.6
Caucasian		12.1	425	68.7
Filipino		12.0	499	80.7
Hawaiian or Part-Hawaiian	38	6.2	537	86.9
Portuguese	6	1.0	543	87.9
Black	3	0.5	546	88.4
Samoan	1	0.2	547	88.6
Other Asians	47	7.6	594	96.2
Others	21	3.4	615	99.6
Identified	3	0.4	618	100.0
Student				
AP Chemistry	237	38.3	237	38.3
Level				
College	48	7.8	285	46.1
Non-AP	333	53.9	618	100.0
Total Participating Hawaii Students	618			

Although the participation was voluntary, the Hawaii high school chemistry students seemed to perform more carefully on the Kit than did the UH chemistry students, probably because AP chemistry was more relevant to the former than the latter. A certain number of students, mostly the UH lower division chemistry students, could not completely finish the survey Kit because of time constraints or academic pressures. A small number of students supplied uniform answers on their answer sheets, such as bubbling all "B" options, and so on. To preclude the undue influence of incomplete or random data, those that did not complete Part C of the Kit, or those that supplied uniform answers, were eliminated from this study. As a result, 60 subjects were deleted from the total sample pool.

Table 2 also shows that the 618 Hawaii students participating in this study represented over 10 ethnic groups, with 211 and 139 students from Japanese and Chinese family backgrounds, respectively. Of the 618 students, 314 were female, and 262 were male, while 42 were unidentified.

Analysis methodology of this study employed both uni- and multidimensional scaling methods to the comparison data to reveal participants' perceptions of item difficulty, similarity, and dimensionality. Item-response theory was used for the calibration and construction of the mini AP chemistry test.

Analyses and Results

Analyses are divided into four components. Component I describes the ability comparability between the Hawaii participants and the national AP chemistry population. Component II compares the similarities of the choices between the Hawaii and national groups, and Component III assesses students' perception of essay difficulty and the effect of difficulty on essay choices. Component IV investigates the dimensionality of the five essays and its relationship to the choices.

Component I: Comparing the Chemistry Ability Between the Hawaii Participants and National AP Chemistry Population

Twenty MC items were selected from the original 75 MC items of Section I of the 1989 AP Chemistry Examination and built into Parts 3 and 4 of the Kit. Altogether, 618 students responded to the 20 items. Table 3 summarizes their performance on the 20 items in reference to the national norm.

TABLE 3
Comparing AP chemistry performance between Hawaii students and 1989 AP chemistry test takers

Student Category	Student Number	Minimum Score	Maximum Score	Mean	SD	KR-20	SEM
Hawaii AP Chemistry	237	1	20	8.72	4.41	.83	2.04
Hawaii non-AP Chemistry	333	0	16	4.96	2.11	.26	1.92
University of Hawaii Upper Division	15	6	20	12.36	4.09	.83	2.04
University of Hawaii Lower Division	33	2	15	5.73	1.98	.52	1.98
Total Hawaii Sample	618	0	20	7.94	3.78	.76	1.97
National AP Sample	1,000	0	20	8.42	3.66	.76	2.10

The AP chemistry students in Hawaii scored about the same as the national norm. As expected, since most of the non-AP students had completed only about one year of general chemistry study by the time they responded to this instrument, their mean performance on this instrument was only half that of their AP counterparts. The mean score for the UH upper-division class students was 12.36. The mean of the UH lower-division chemistry students was 5.73.

In summary, the overall performance of Hawaii students on the 20 MC items was quite similar to that of the national norm. These 20 MC items functioned equally consistently with both Hawaii and national students as shown by 0.76 KR-20 coefficients for both groups. Figure 1 illustrates detailed score distributions for the four major groups of students.

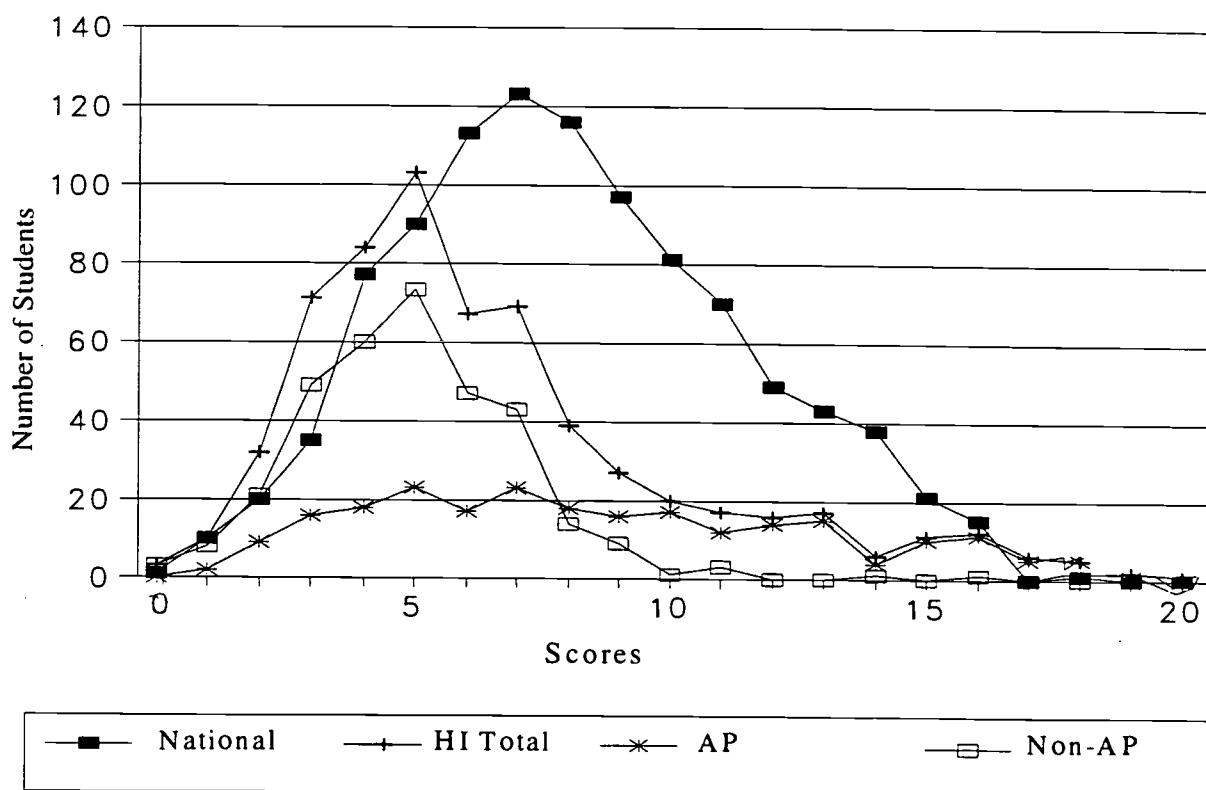


FIGURE 1. Raw score distributions of national, Hawaii total, AP (advanced placement) and non-AP samples

Component II: Comparing Essay Choices Between Hawaii Students and National AP Chemistry Test Takers

Given that the Hawaii students performed so closely to the national norm on average, would they choose the five essays in a similar way as did the 1989 AP Chemistry test takers? Note that the students in Hawaii were asked only to indicate in Part D of the Kit how they would like to choose three of the five essays. Figure 2 reveals strikingly familiar essay choice patterns between the Hawaii students and the national test takers. Based on 554 Hawaii students, the overall choice pattern of the five essays substantially mirrored that of the 18,462 test takers who took the 1989 AP Chemistry Exam. The correlation between the two patterns is .87.

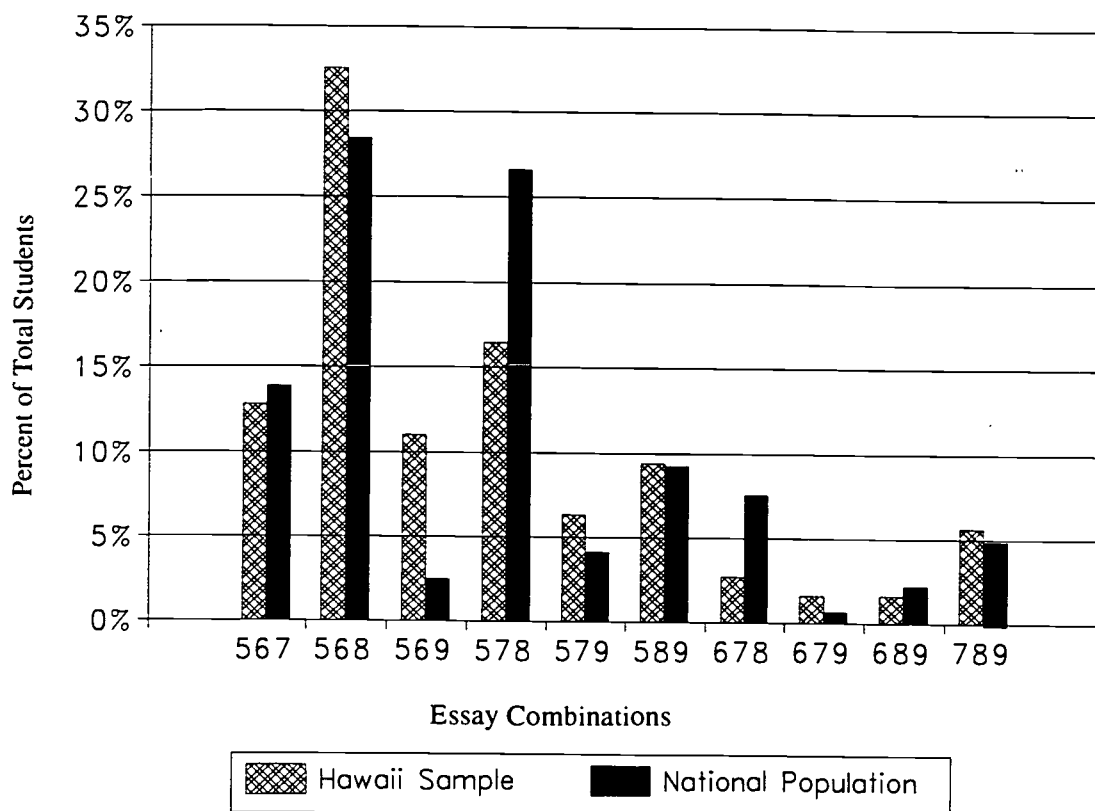


FIGURE 2. Comparison of the overall essay choice pattern of 554 Hawaii and 18,462 national students

Three points should be noted. First, like the 1989 AP Chemistry exam population, the Hawaii students favored essay combination 5, 6, and 8 most. Essay combination 5, 7, and 8 remained the second most popular combination. Second, almost the same proportion of Hawaii students (13%) as the 1989 AP test takers (14%) chose essay combination 5, 6, and 7. Note that essay 7, originally the third essay in the 1989 AP Chemistry Exam, was presented as the fifth essay in the Kit to test the effect of positioning essay 9. In spite of the position change, essay 7 was still preferred over essay 9. There seems to be something inherent about essay combination 5, 6, and 7 that attracts AP chemistry students.

The above findings seem to embody a universal regularity in the way students choose these essays, possibly due to the commonality of the AP chemistry textbooks. Although there is no standardized national AP Chemistry curriculum, all AP chemistry textbooks are quite similar in their curricular emphases, which translates into students' varying familiarity with various subjects and eventually the way they choose items on a test.

Component III: Relationship Between Students' Perceptions of Essay Difficulty and Their Choices

In light of the high levels of similarity in choice tendencies involving the five essays, one might wish to discern the cognitive dimensionality underlying the choices. The immediate questions are: "How did the students perceive the difficulties of the five essays? How did their perception influence their choices?"

In Part D of the Kit, Hawaii students were asked to conduct pair-wise comparisons of the relative difficulty levels of the five essays through the question "Which one (essay) seems easier for you?" Such pair-wise comparison data were analyzed through Ranko (Dunn-Rankin, 1983), which carries out variance-stable rank-scaling analysis. The linear plot with scale scores from Ranko is reproduced in Figure 3.

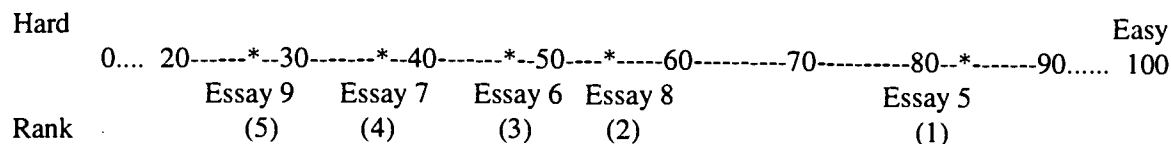


FIGURE 3. *Perceived order of essay difficulty*

Hawaii students ranked essay 5 as the easiest, followed by essays 8, 6, 7, and 9 in that order. Such an order of perceived essay difficulty conforms completely with the popularity of the five essays of the 1989 national chemistry test takers reported in Wang's earlier study (See Figure 2 of Wang, 1996). Using this rank order, one also can predict the popularity of the 10 essay combinations found with the national test-taker population (see Figure 1 of Wang, 1996). For example, from Figure 2 of this study, we know that essays 5, 6, and 8 form the most popular essay combination. These happened to be the three easiest essays on the Ranko scale. The Ranko scale also confirms that essays 5, 8, and 9 and 5, 6, and 7 form the second and third most popular essay combinations. The three essays that form the least popular essay combination 6, 7 and 9 in Figure 2 turn out to be the last three essays in the Ranko scale. Table 4 further summarizes the critical differences among the essays, and all the essays are shown to be significantly different from one another in terms of their difficulty at 0.01 significance level.

TABLE 4
Rank differences

	5	8	6	7	9
5	0				
8	565	0			
6	735	170	0		
7	927	362	192	0	
9	1,108	543	373	181	0

Note. The critical differences are 137 at .05 level and 163 at .01 level.

Component IV: Relationship Between Students' Perceptions of Content Dimensionality and Their Choices

These five essay questions are known to involve the following areas of chemistry knowledge:

- Essay 5: valence, electronic configuration, covalent bonding, molecular geometry
- Essay 6: periodic trends, stability, ionization, energy, properties of halogens, properties of alkali metals
- Essay 7: properties of metals, writing and balancing chemical equations, conservation of mass, double displacement reactions
- Essay 8: rates of reaction, physical behavior of gases, energy changes
- Essay 9: nuclear chemistry

It is clear that these five essay questions differ in content. How did such content diversity affect test takers' choices?

Because not all test takers responded to these five essays, traditional factor analysis based on students' scores cannot be applied here to ascertain the dimensionality of the five essays. In order to offset this information deficit, pair-wise similarity comparisons ("How similar are Problem I and II?") have been incorporated into Part D of the AP Chemistry Survey and Test Kit. Based on the similarity data from the Hawaii students, the dimensionality in terms of content similarity of the five essays is revealed through multidimensional scaling in Figure 4.

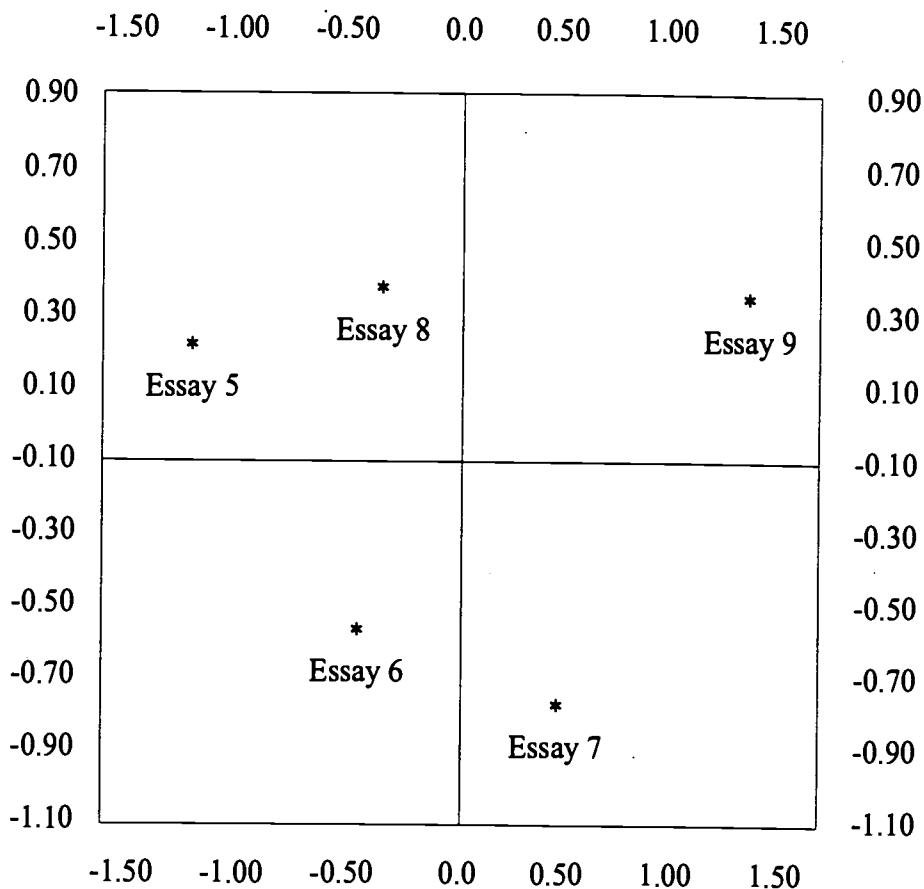


FIGURE 4. *Dimensionality of the five essay questions*

Corresponding to content differences, these five essays are spread out in the four quadrants of the two dimensional space. However, under the seemingly large differences lies a certain vein of commonality. First, if we examine the two halves divided by the vertical line, we see that essays 5, 6, and 8 are in the left half, and essays 7 and 9, in the right half. What do essays 5, 6, and 8 share in common? They reflect the most frequently taught topics and constitute the common core of general chemistry. On the other hand, essay 7 requires more extensive lab experience in addition to classroom lecturing. Since not all students have equal access to laboratories, essay 7 cannot be dealt with as readily by average students. Furthermore, since essay 9 is concerned with the least taught topic of nuclear chemistry, it is further removed from the other four essays.

Figure 4 designates the horizontal dimension as “extent of textbook coverage” with the left half symbolizing “core chemistry” topics and the right half, the “noncore chemistry” topic.

Now, let us try to interpret the top and bottom halves of Figure 4. The Hawaii AP chemistry teachers survey (Wang, 1996) indicates that both essays 6 and 7 tap into deeper and more complex chemistry theories, structures, and lab experiences, while essays 5, 8, and 9 are relatively descriptive and fact-oriented. Therefore, the vertical dimension denotes “complexity of problems” with the upper half standing for the relatively “straightforward” questions, and the lower half representing relatively “complicated” questions.

How did the dimensionality affect test takers’ choices? More specifically, did students tend to choose essays of the same or similar dimension? The answer appears to be *yes*. For example, according to Figure 2, essay combination 5, 6, and 8 was the most frequently chosen combination by both the national population and the Hawaii students. These three essays form the “core chemistry” half of the horizontal dimension. Any essay combination with essays 7 and 9 was usually avoided. essays 6, 7, and 9 were the least favored combination, basically because essays 7 and 9 form the “noncore chemistry” half of the horizontal dimension.

Moreover, as shown in Table 1, we know that essay combination 5, 8, and 9 has the highest mean score. This is probably because these three essays are from the “straightforward” half of the vertical dimension. It also can be seen that any essay combination that includes essays 6 and 7 has a lower mean score. For instance, essay combination 5, 6, and 7 has the lowest mean score of the 10 essay combinations, basically because essays 6 and 7 were from the “complicated” dimension of the essays.

It can be concluded from the above findings that not only were students’ choices influenced by item dimensions, but also their scores. The origin of essay dimensions can be attributed to various factors. In the case of these five essays, it is reasonable to believe that these two dimensions are attributed to the order of textbook presentation and tasks involved to solve the problems. It is argued here that with most subject tests like AP chemistry tests, the predominant mode of textbook presentation must have had a long-lasting effect on how students are taught, which subsequently influences how they will choose on a test.

Conclusion and Discussion

This study has produced three major findings. First, the national choice pattern of the five essays was well replicated by the participants of Hawaii, indicating the existence of a general and consistent influence on test takers’ choices. This consistent influence is attributed to the commonality of AP chemistry curricula and textbooks. The second finding is that students’ perception of essay difficulty, when transformed into a rank order, can accurately predict and account for the popularity of essay choices. The third finding is that students’ perception of content dimensionality of the essays coincides with students’ choices and provides a reasonable explanation for their mean performance on the essay combinations. The second and third findings vividly suggest the links between students’ psychological processes and their choices.

The above findings offer at least two avenues to better implement CR item choices on a test. The first useful application is to pretest how likely test takers would be to choose a set of CR items on a test. It is well-known to testing agencies that most CR items are difficult to pretest because of logistical difficulties in scoring them and the high risk of test-security breaches. Yet, this paper shows that using a small number of potential test takers to rate difficulty and similarity levels can offer fairly accurate estimates of how test takers perceive items in terms of difficulty and dimensionality, and of how likely a candidate would be to choose them. This method would minimize pretesting costs and test-security risks.

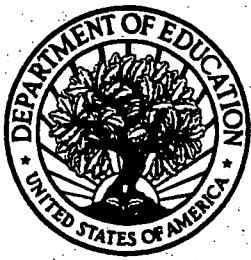
The second possible application is to minimize the potential differences among CR items. The principle of test equity demands a fair and equal chance of success for each test taker. Yet, allowing test takers to choose among a set of CR items potentially different in difficulty, content, and dimensionality would easily jeopardize such a principle. Controlling content similarities is certainly one solution to this possible inequity. However, it is well known that content-similar items frequently produce psychologically different dimensions. The technology used in this study can certainly help reveal such psychological dimensionality differences to further improve the quality of tests in general.

This paper shows that students’ perceptions of item difficulty and dimensionality can account for their choices, but it does not claim that such perceptions can accurately predict the actual difficulty of a test item. According to Wang (1996), test takers on the national level performed lower on the more frequently chosen items. Why was there a negative relationship between familiar items and performance? The investigation of this paradox will be reported in a separate study. It suffices to say that there appeared to be a negative interaction between the scoring rubrics and the curricular emphases on the essay items. More specifically, more commonly taught chemistry items might have been scored more stringently than less commonly taught essay items. Through an experiment with MC items requiring straightforward “right” or “wrong” scoring without scoring rubrics, test takers do score higher on the items they choose, as long as they understand them.

This study does have its shortcomings. The author would have preferred a more standard and controlled fashion of delivery for the Kit. More reliable and complete responses might have been obtained if students responded to the Kit under some mandatory condition. However, best efforts were made to collect the data, given all the resources available to the author, and the results have been instructive. It is hoped that this study will stimulate more research to help increase the accuracy, reliability, and validity of tests that involve CR item choices.

References

- The College Board. (1990). *The 1989 Advanced Placement Examination in Chemistry and its Grading*. New York: Author.
- Dunn-Rankin, P. (1983). *Scaling methods*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Fremer, J., Jackson, R., & McPeck, M. (1968). *Review of the psychometric characteristics of the Advanced Placement tests in Chemistry, American History, and French* (Unpublished Tech. Memorandum). Princeton, NJ: Educational Testing Service.
- Masterton, W. L., Slowinski, E. J., & Stanitski, C. L. (1985). *Chemical Principles*, (6th edition). Philadelphia: Saunders College Publishing.
- Pomplum, M., Morgan, R., & Nellikunnel, A. (1992). *Choice in advanced placement tests* (Unpublished Statistical Report No. SR-92-51). Princeton, NJ: Educational Testing Service.
- Thissen, D., Wainer, H., & Wang, X-B. (1994). Are tests comprising both multiple-choice and free-response items necessarily less unidimensional than multiple-choice tests? Analysis of two tests. *Journal of Educational Measurement*, 31, 113-123.
- Wainer, H., & Thissen, D. (1992). *Choosing a test*. (Tech. Report No. 92-25). Princeton, NJ: Educational Testing Service.
- Wainer, H., & Thissen, D. (1993). Combining multiple choice and constructed response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, 6, 103-118.
- Wainer, H., & Thissen, D. (1994). One test taker choice in educational testing. *Review of Educational Research*, 64, 159-195.
- Wainer, H., Wang, X-B., & Thissen, D. (1994). How well can we compare scores on test forms that are constructed by test takers' choice? *Journal of Educational Measurement*, 31, 183-199.
- Wang, X. B., Wainer, H., & Thissen, D. (1995). On the viability of some untested assumptions in equating exams that allow test taker choice. *Applied Measurement in Education*, 8(3), 211-225.
- Wang, X. B. (1996, April). *Investigating allowing test takers to choose constructive response items on a test*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Wittrock, M., & Baker, E., (1992). *Testing and cognition*. Englewood Cliffs, NJ: Prentice Hall.



*U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)*



NOTICE

Reproduction Basis

- This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.
- This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").