WHAT CAN YOU LEARN FROM A (GOOD) MULTIPLE-CHOICE EXAM?

Lei Bao[a], **Edward F. Redish**[b]

[a] Department of Physics, The Ohio State University, 174 W 18th Ave, Columbus, OH 43210-1106 USA

[b] Department of Physics, University of Maryland, College Park, MD 20742-4111 USA

Abstract

The information that a teacher typically extracts from a multiple-choice exam is limited. Basically, one learns: How many students in my class can answer each question correctly? Careful studies of student thinking [1] demonstrate that student responses may reflect strongly held naïve conceptions and that students may function as if they think about a particular topic using contradictory models (typically their naïve model and the scientific one taught in class). We have developed tools for extracting information about the state of knowledge of a class from multiple-choice exams that goes beyond how many students answered each question correctly. First, a mathematical function, the *concentration factor*, allows one to determine whether a particular question triggers students' naïve models. Second, by treating the students as if they can exist in "mixed states" of knowledge, we create methods of extracting measures of the *state of confusion* of the class. By this we mean how likely the students are to used mixed models. Our method assists in the construction of multiple-choice tests that respond to what is known about the difficulties students bring into classes and we provide ways of extracting more detail about what students have learned than traditional analysis tools.

1. Introduction

Physics teachers and education researchers have long observed that students can appear to reason inconsistently about physical problems.[2] Problems seen as equivalent by experts may not be treated using equivalent reasoning by students. Qualitative research has documented many different clusters of semi-consistent reasoning students use in responding to physics problems. This knowledge has been used in creating attractive distracters for multiple-choice examinations that allow one to examine large populations.[3] The way that students select wrong answers on such tests contains valuable information on student understanding. Traditional analyses of multiple-choice exams focus on the scores, and possibly on the correlation between correct answers chosen by students. Such an analysis often fails to extract the information about the state of a class that a test provides.

Based on the understanding of student learning, we have developed algorithms to conveniently extract and display such information.[4] Our method allows us to analyze the complete student responses rather than just identifying the fraction of the time they are using the correct approach. In this paper, we discuss an analytical method for analyzing the concentration / diversity of student responses to particular multiple-choice questions. This method is both a tool to extract information from a research-based multiple-choice test and a tool to be used in the cyclic process of creating such tests. A method for evaluating and describing the mixed mental state of a class is described in later papers.[5]

2. The Concentration Factor

As we learn from qualitative research into student learning, student responses to problems in many physical contexts can be considered as the result of their applying a small number of mental models. The way in which the students' responses are distributed on research-based multiple-choice questions can yield information on the students' state: for a particular question, highly concentrated responses implies that many students are applying a common model associated with the question; whereas randomly distributed responses often indicate that students are guessing or have less systematic reasoning.

It is convenient to construct a simple measure that gives the information on the distribution of students' responses on one particular question. To do this, we define the concentration factor, $\boldsymbol{C}$, a function of student response that takes a value in [0,1]. Larger values represent more concentrated responses with 1 being a perfectly correlated response and 0 a random response. We want all other situations to generate values between 0 and 1. The function that does this is

$$\boldsymbol{C}=\frac{\sqrt{m}}{\sqrt{m}-1}\times\left(\frac{\sqrt{\sum_{i=1}^{m}n_i^2}}{N}-\frac{1}{\sqrt{m}}\right) \qquad (1)$$

where m represents the number of choices for a particular question, N is the number of students, and n_i is the number of students who select the *i-th* choice.

To study the details of the distribution of student incorrect responses, we define a new variable, Γ, as the *concentration deviation*:

$$\Gamma=\frac{\sqrt{m-1}}{\sqrt{m-1}-1}\times\left(\frac{\sqrt{\sum_{i=1}^{m}n_i^2-S^2}}{(N-S)}-\frac{1}{\sqrt{m-1}}\right) \qquad (2)$$

Γ is independent from the score (number of right answers) S and thus gives a stable measure of the concentration of the incorrect responses.

3. Classifying the Response Patterns

A useful method is to combine the concentration factor with scores to form response types. We choose a 3-level coding system with "L" for low, "M" for medium and "H" for high. Based on simulation results, we decided to choose a 3-level coding scheme as defined in table 1.

Score (S)	Level	Concentration (***C***)	Level
0~0.4	L	0~0.2	L
0.4~0.7	M	0.2~0.5	M
0.7~1.0	H	0.5~1.0	H

Table 1. Three-level coding scheme for score and concentration factor

Using the concentration types measured from student data, we can classify the following situations (see table 2):
One-Model: Most of the responses are concentrated on one choice (not necessarily a correct one).
Two-Model: Most of the responses are concentrated on two choices, often one correct and one incorrect.
No-Model: The responses are somewhat evenly distributed among three or more choices.

		Implications of the patterns
One-Model	HH	One correct model
	LH	One dominant incorrect model
Two-Model	LM	Two possible incorrect models
	MM	Two popular models (correct and incorrect)
NoModel	LL	Near random situation

Table 2. This table shows typical response patterns when using the three-level coding system.

4. Concentration Analysis of FCI Data

As an example of the kind of information a concentration analysis can give about an exam and a population, we apply our method to results taken with FCI pre- and post-tests.[6] The data is taken from 14 classes in the introductory semester of a calculus-based physics course at the University of Maryland. The students are mostly engineering majors. Half of the classes were taught with University of Washington-style tutorials and the other half of classes were using traditional instruction.[7]

In table 3, the student responses are grouped into seven categories. The HH and MH types show that students do well on those topics before instruction. The MM type implies that some students are doing well but a significant number of students have a tendency to use a common incorrect model. More interesting results come from the LM and the LH types, which are strong indications for the existence of common incorrect models. The content of the questions suggests that most of the questions with LM and LH types deal with two physics concepts, the Motion-needs-Force relation and Newton's 3rd law.

Types	LL	LM	LH	ML
Questions	15, 24	5, 9, 18, 28	2, 13, 22	3, 7, 21, 26
Types	MM	MH	HH	
Questions	6, 8, 11, 14, 17, 20, 23, 25	12, 16, 29	1, 4, 10, 19, 27	

Table 3. Response types of the FCI Pre-data from both tutorial and traditional classes (778 UMd students).

Table 4 shows the percentage of students selecting the most popular distracters of the questions with LH and LM types of responses. A review of the distracters in the test (original version) confirms that these questions are associated with two naïve models associated with Force-Motion and Newton III.

Force and Motion			Newton's Third Law		
Choice	%	Type	Choice	%	Type
5-c	58%	LM	2-a	66%	LH
9-c	45%	LM	11-d	43%	MM
18-a	63%	LM	13-c	68%	LH
22-c	66%	LH			
28-d	51%	LM			

Table 4. Pre-instruction responses on FCI questions related to the concept of Force-Motion and Newton III (UMd students).

5. Summary

The concentration factor can be useful tool in many both research and instruction. We can use it to facilitate the design of effective multiple-choice questions or use the concentration factor to evaluate student performance and their modeling conditions.

1. L. C. McDermott and E. F. Redish, Resource letter PER-1: Physics Education Research, Am. J. Phys. **67** (1999) 755.
2. I. A. Halloun and D. Hestenes, Common sense concepts about motion, Am. J. Phys. **53**, 1056 (1985).
3. R. R. Hake, Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses, Am. J. Phys. **66**, 64-74 (1998).
4. L. Bao, Dynamics of Student Modeling: A Theory, Algorithms, and Application to Quantum Mechanics, Ph.D. dissertation, U. of Maryland, December 1999.
5. L. Bao and E. F. Redish, Model Analysis: Modeling and Assessing the Dynamics of Student Learning, to be published in PERS to the Am. J. Phys.
6. D. Hestenes, M. Wells, and G. Swackhamer, The Force Concept Inventory, Phys. Teach. **30**, 141-158 (1992).
7. L. C. McDermott et al., *Tutorials in Introductory Physics*, John Wiley & Sons, NY, 1998