DOCUMENT RESUME

ED 306 233 TM 012 990

AUTHOR	Terwilliger, James S.; And Others
TITLE	Content Validity of the GRE Physics Test as Indexed by Commonly Employed Textbooks.
SPONS AGENCY	Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE	Mar 89
GRANT	R-86-0017
NOTE	19p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, March 1989).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Research/Technical (143)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	College Entrance Examinations; College Students; *Content Validity; *Course Content; Higher Education; *Physics; Professors; *Science Curriculum; Test Items; *Test Validity; *Textbook Content; Textbook Evaluation
IDENTIFIERS	*Graduate Record Examinations

ABSTRACT

The Graduate Record Examination (GRE) in Physics was analyzed as a specific indicator of learning outcomes in that undergraduate major. Textbooks were selected as a measure of curriculum content. A representative sample of colleges and universities (n=59) responded to questions about physics courses and textbooks and the existence of a graduate program in physics. Two panels of experts, 16 physics professors in all, rated the extent to which each text taught students what they needed to know to perform well on the GRE Physics Test. Subfields of physics that were evaluated included mechanics, electricity and magnetism, thermodynamics and statistical mechanics, and modern and quantum physics. The panel prepared two indices: (1) R1, how well a text "covered" the GRE; and (2) R2, how well the GRE "covered" the textbook. The interrater reliability of R1 ratings was good, but the reliability of R2 ratings was not as good. Values from the R1 index indicated that the curriculum generally prepared students well in mathematics. About 60% o the relevant material in electricity and magnetism was covered by the textbooks. For the other subfields, ratings (R1) were higher for schools with graduate programs in physics; this difference was more striking for modern and quantum physics. Many items from commonly used texts were not well represented on the GRE, but it could not be determined whether this is a real mismatch or the result of the random selection process. Five data tables and two figures present the study data. (SLD)

Content Validity of the GRE Physics Test as Indexed by Commonly Employed Textbooks

James S. Terwilliger
J. Woods Halley
Patricia Heller

University of Minnesota

Paper presented at annual meetings of the National Council on Measureme... in Education, San Francisco, March, 1989.

Introduction

This paper describes activities conducted as part of a project sponsored under contract with the Office of Educational Research and Improvement (OERI). The initial request for proposals issued by OERI (#R-86-0017) contained the following general description of the goals of the work.

> Each study will focus on a single discipline or field and determine how indicator(s) can be constructed so as to account for summative undergraduate learning (of content, methods, assumptions, etc.) in that field, and to help answer the question of what it means to be educated in a discipline.

There were 19 responses to the RFP representing a diverse cross-section of disciplines. Five proposals were funded. Besides our project on physics the other four studies represented the disciplines of biology, chemistry, computer science, and mechanical engineering.

The RFP listed a series of specific tasks to be accomplished by each contractor, including the review and analysis of existing instruments and methods for assessing "summative learning" in the discipline. This paper will focus on an analysis of the Graduate Record Examination in Physics as a specific existing indicator of learning outcomes in that undergraduate major.

The Undergraduate Curriculum in Physics

The definitive source of information concerning the undergraduate physics curriculum in U.S. colleges is provided by the Committee on Professional Concerns and Undergraduate Education of the American Association of Physics Teachers (AAPT). The committee has prepared a booklet, AAPT Guidelines for the Review of Baccalaureate Physics Programs which is available through AAPT. The most recent version was published in September, 1987.

The section entitled "Curriculum" suggests that the undergraduate curriculum begin with

> "an elementary course that has at least five subsections: Mechanics, Waves, Heat and Thermodynamics, Electricity and

> Magnetism, and Optics. ----- A time commitment of at least two semesters is required to teach the five standard subsections; a time frame of three semesters is a better choice if Modern Physics is to be included."

With respect to more advanced undergraduate courses that a physics major ought to take, the Guidelines state, "There should be a vigorous, advanced treatment of topics in Mechanics, Electricity and Magnetism, Thermodynamics and Statistical Mechanics, Optics, Quantum Physics, and Experimental Physics." More specifically, the Guidelines present a detailed listing of topics suggested for inclusion in each of these areas. Figure 1 presents the detailed listing of topics recommended for inclusion in each of these six advanced courses.

Insert Figure 1 about here

The Graduate Record Examination Physics Test

The Physics Test is described as follows in Practicing to Take the GRE Physics Test published by Educational Testing Service in 1986.

> The test consists of about 100 five choice questions, some of which are grouped in sets and based on such materials as diagrams, graphs, experimental data, and descriptions of physical situations.
>
> The emphasis of the test is on the students' firm grasp of fundamental principles and their ability to apply and understanding of them in the solution of problems. Most test questions can be answered on the basis of a mastery of the first three years of undergraduate physics.
>
> The approximate percentages on content topics have been set by the committee of examiners to reflect its judgment about the relative emphases placed on these topics in most undergraduate curricula. These percentages are as follows:

Topic	Percentage of Questions
1. Classical mechanics	18
2. Fundamentals of electromagnetism, including Maxwell's equations	18
3. Atomic physics	15
4. Physical optics and wave phenomena	10
5. Quantum mechanics	10
6. Special relativity	7
7. Thermodynamics and statistical mechanics	7
8. Laboratory methods	5
9. Advanced topics (Lagrangian and Hamiltonian mechanics, nuclear and particle physics, solid state physics, and miscellaneous)	10

The time limit for the test is 170 minutes so examinees have less than two minutes per item. Based upon the responses of 10,625 examinees who took the test (Form GR8677) between October 1, 1981 and September 30, 1984, the median raw score (corrected for guessing) was 37. The average item difficulty for this group of examinees was .45 with individual item difficulties ranging from .14 to .90.

Procedure

Textbook Survey

The approach taken in this study emphasizes the analysis of the undergraduate curricula at a sample of colleges to determine the degree of consensus across institutions with respect to general program requirements and specific course offerings. As previously noted, the AAPT guidelines suggest a fairly "standard" curriculum. In particular, we wished to identify and analyze commonly employed physics textbooks since the content of the undergraduate curriculum is determined to a large extent by the textbooks in common use.

Textbooks were selected as a measure of the curriculum taught because they provide a basis for inter-institutional comparison which no listing of course titles or descriptions can. Further, we suspected (and our survey bears out) that a relatively small number of textbooks represents the curriculum taught in undergraduate programs throughout the country fairly well. We assume in this procedure that the textbooks assigned describe what is actually taught. We have no systematic check on this assumption, though it is consistent with qualitative impressions and anecdotal information.

To obtain a sample of schools for the survey, we selected 80 schools representative of four categories of schools in the U.S.: 20 highly ranked schools with graduate programs (TG), 20 highly ranked schools without graduate programs (TUG), 20 other randomly selected schools with graduate programs (RG), and 20 other randomly selected schools without graduate programs (RUG).

To choose the highly ranked (TG) schools offering graduate programs, we took the top 20 as listed in the 1982 ranking sponsored by the Conference Board of Associate Research Councils (Jones, et.al., 1982). To choose the 20 "other" (RG) schools offering graduate programs, we made a random selection of 20 schools from the remaining schools offering graduate programs in the American Institute of Physics (AIP) listing of institutions which offer undergraduate degrees in physics (Ellis, 1986). Each school was weighted by the number of physics major graduates in 1985 in making the selection.

To select the 20 highly ranked schools (TUG) not offering physics graduate programs, we used the list of 50 liberal arts colleges participating in the Second National Conference on "The Future of Science at Liberal Arts Colleges" (Carrier, et.al., 19-7) and made a random selection of 20 schools from it, weighting each school by the number of physics major graduates in 1985. (This procedure was forced on us by the fact that no reliable ranking of undergraduate physics programs appears to be available.) Finally, we made a random selection of 20 more schools not offering graduate programs (RUG) from the AIP list, weighting the schools in the same way.

Each of the 80 schools was sent a letter explaining the project and a form asking for 1) the number of physics majors graduated in the preceding year, 2) a list of courses taken by physics majors and the number of physics majors taking the course in 1985-86, and 3) the textbook used in each course listed. Each school which did not respond was telephoned at least once.

Indices of GRE and Curriculum Congruence

In the next phase of the study we used the textbook survey data to determine quantitative measures of the extent to which the GRE Physics Test measures outcomes represented by textbooks assigned to physics majors. To explain the measures we imagine the material tested by the GRE to be one set labelled GRE and the material taught from the textbooks of the curriculum as another set labelled CURR as shown in Figure 2. We designed an index to estimate the quantity

$$R1 = GRE \cap CURR/GRE$$

and another index to estimate the quantity

$$R2 = GRE \cap CURR/CURR.$$

Since these two indices represent part/whole relationships each takes on values between zero and one. R1 may be thought of as the extent to which the curriculum teaches students what they need to know to perform well on the GRE. R2 may be said to be the extent to which the GRE tests the material which is taught from the texts of the curriculum. Put another way, R1 measures the extent to which the curriculum "covers" the GRE while R2 is the extent to which the GRE "covers" the curriculum.

Insert Figure 2 about here

To obtain an operational estimate of the correspondence of textbook and GRE content, two sets of expert ratings were collected. First, two physics professors with teaching experience in each of four subfields (Mechanics, Electricity and Magnetism, Thermodynamics/Statistical Mechanics, and Modern Physics/Quantum Mechanics) were provided with copies of GRE items from Form GR8677 and with textbooks our survey revealed to be in most common use. Each of the eight professors was asked to answer the question, "To what extent would GRE question q be appropriate (or too easy) for the final examination in a course taught from textbook t?". Ratings for individual items in each subfield were made on a scale from zero to one in intervals of .1. These ratings were employed in the calculation of the R_1 index.

A second group of eight experts (again, two physics professors in each of the four designated subfields) was used to obtain ratings of the degree to which the textbook content "covers" the GRE. A list of 20 end-of-chapter questions was selected at random from each of the textbooks. The questions and a copy of the GRE (Form GR8677) were provided for each professor. They were asked to answer the question, "If you taught a course to prepare students for the GRE, how appropriate would this question be on a final examination for that course?". The rating scale was the same as previously described for the first group of experts. These ratings were used in calculating the R_2 index.

Specifically, the two desired indices were defined as:

1) $R_1 = 1/N \sum\sum_{qt} F_1(q,t)\ f(t)$

and

2) $R_2 = 1/20 \sum\sum_{qt} F_2(q,t)\ f(t)$

where: N = number of GRE items in a given subfield.

$F_1(q,t)$ = average rating of two experts in the first panel of raters.

$F_2(q,t)$ = average rating of two experts in the second panel of raters.

$f(t)$ = the proportion of the students in our survey sample using textbook t.

A simplified illustration of the calculation of the R_1 index is presented in Table 1. In this hypothetical situation, there are ten GRE items rated by each of the two experts with relation to three textbooks. The steps in calculating R_1 for this set of data are presented in detail in the figure.

Insert Table 1 about here

Results

Textbook Survey

Usable data were obtained from 59 schools, or 74% of the sample. The number of schools responding in each category is shown in Table 2.

With regard to textbook adoptions, we found that a small number of texts in each subfield accounts for the great majority of the students at institutions in our sample. To keep the number of texts for further analysis manageable, we decided to restrict attention to texts used by ten percent or more of the students who were physics majors in the schools responding to our survey. Using this criterion, the number of texts retained and the total percent of students using these texts are as presented in Table 3.

Insert Tables 2 and 3 about here

The total percent exceeds 100 in the Modern/Quantum Physics due to the fact that students frequently take more than one course in that category.

Reliability of Expert Ratings

The correlation between the ratings of the two experts in each subfield for each type of rating was calculated for each textbook. The range of correlations and the median correlation for each subfield and type of rating are given in Table 4.

The results in Table 4 reveal that the reliability of ratings of the degree to which textbooks "cover" the GRE items is quite respectable although there is considerable variation both among subfields and among individual textbooks within a particular subfield.

The reliabilities for the ratings of the degree to which the GRE "covers" the textbook items are substantially lower. Inspection of these ratings shows that, on average, the ratings are much lower than are the first set of ratings. Indeed, the average rating for the 20 items randomly selected from textbooks was less than .20 (on a scale from zero to one) for over 40 percent of the ratings. At least half of the textbook items were rated zero over one third of the time. Distributions with such extreme skews and small variabilities obviously place constraints upon correlations among ratings.

A plausible explanation for such low ratings of the textbook items is that, being randomly selected, such items frequently deal with concepts, methods, and problems which are not of central importance when judged in the context of the larger subfield. Obviously, such items are not of the type that are likely to be selected for inclusion on the GRE!

Insert Table 4 about here

Indices of Content Validity of the GRE

R_1 indices were calculated separately by subfield and by type of institution in our survey sample. The results are shown in Table 5A. The R_1 values indicate that the curriculum in Mechanics generally prepares students well for the GRE, regardless of institution type. However, the courses in Electricity and Magnetism seem to be covering only about 60 percent of the relevant material on the GRE. A more detailed analysis reveals that there are a significant number of questions on optics in the GRE while the typical curriculum does not give great emphasis to this topic.

For the two remaining subfields, the R_1 values for institutions with graduate level programs are quite high but the values for institutions with strictly undergraduate programs are much lower. A close inspection of our data suggests that this arises because schools without graduate programs offer or require fewer courses in these two subfields. Also, in the absence of a graduate level curriculum and research program, fewer undergraduate students are likely to take courses in Statistical Mechanics/Thermodynamics and Modern/Quantum physics even when they are available.

Insert Tables 5A and 5B about here

The R_2 indices were also calculated separately by subfield and institution type. These are presented in Table 5B. The most striking feature of these results is the much lower magnitude of the values in relation to the values in Table 5A. About 40 percent of the textbook content in Mechanics and Electricity and Magnetism is represented by the GRE. The figure is less than half of that for the other two fields. Also, there is a uniformity of results across institutions with each subfield. However, these results need to be interpreted with caution in view of the low reliabilities previously noted.

Conclusions

Ratings of GRE items indicate that the commonly used texts in Mechanics do a good job in preparing students for GRE items in that subfield, but texts in Electricity and Magnetism do less well in that regard. In the other two subfields there is a clear difference between institutions with graduate programs and those without graduate programs. The difference is especially striking in

Modern/Quantum Physics, a subfield which represents over one third of the GRE items. This suggests that students who attend strictly undergraduate institutions may be at a substantial disadvantage on the GRE Physics Test.

The ratings of items randomly selected from the most commonly adopted texts suggest that such items are not well represented by items on the GRE. The ratings were so low that the inter-rater reliabilities were generally low and quite variable from one text to another. It is not possible to state whether this results from the random selection process or from a "true" mismatch between the texts and the GRE. To clarify this issue, it would be necessary to use expert judgments in selecting textbook items that are "representative" of outcomes considered to be most important. This is presumably how the GRE items were selected.

References

Carrier, S.C. and Davis-van Atta, D. (1987). "Maintaining America's Scientific Productivity: The Necessity of the Liberal Arts Colleges," Offices of the Provost and Institutional Research, Oberlin College, Oberlin, OH.

Ellis, S.D. (1986). "Data on Physics Enrollment and Degrees," AIP Publication No. R-121.23A, American Institutes of Physics, NY.

Jones, L.V., Lindzey, G. and Coggeshall, P.E. (1982). "An Assessment of Research-Doctorate Programs in the United States: Mathematical and Physical Sciences," National Academy Press, Washington, D.C.

Figure 1

Description of Recommended Content for Advanced Undergraduate Courses for Physics Majors[a]

Mechanics: The mathematical level of the course should require the use of differential equations. Central forces should be studied through at least the development of Kepler's laws. The study of systems of particles should pursue the consequences of the conservation of energy, momentum, and angular momentum; the latter should include the use of the inertia tensor. The analysis of rigid body motion should include the application of Euler's equations. Lagrangian mechanics should be treated in sufficient depth for its application to small oscillations and coupled oscillators.

Electricity and Magnetism: The mathematical level of this course should require the use of field operators and vector integral theorems. The treatment of electrostatics should encompass Coulomb's law; the electrostatic field and potential; the Laplace and Poisson equations; electric dipoles; multipole expansions of potentials; electrostatic energy and force; capacitance; polarization; dielectrics; and the electric displacement field. The topics of electric current, Ohm's law, and the continuity equation will lead to discussions of magnetism, including the magnetic induction field; the Biot-Savart law; Ampere's law; magnetic energy, force, and torque; magnetization; and the magnetic field. Maxwell's equations should be considered essential components of this course and they should be applied to simple geometries (e.g., plane waves in infinite, non-conducting medium). Complex waves could also be introduced. If time permits, relativistic electrodynamics could be briefly considered.

Thermodynamics/Statistical Mechanics. A thorough grounding in the concepts of temperature, work, specific heat, compressibility, and entropy should result from this course. The laws of thermodynamics, from the zeroth to the third should be discussed, with a thorough discussion of the import of the second law. The four thermodynamic generating functions (internal energy, enthalpy, Helmholtz function, and Gibbs function), together with Maxwell's relations, should be used to solve practical problems such as gas laws, engines, radiation, and phase transitions. The kinetic theory of gases, partition functions, and, as possible, ensembles are covered. Here is the student's first exposure to Maxwell-Boltzmann, Fermi-Dirac, and Bose-Einstein statistics

Optics Both geometrical and physical optics should be included. Enough time should be spent on thin lenses and mirrors to provide an understanding of simple optical systems and such concepts as magnification, entrance and exit pupils, and stops. The treatment of physical optics should include a discussion of two-beam and multiple-beam interference, diffraction at apertures, and the application of those principles to simple interferometers, double-slit diffraction, the diffraction grating, and diffraction-limit resolution Polarization and reflection should also be included. As time permits, there should be selective coverage of thick lenses, lens aberration, lens design, vision, color, ray tracing, birefrigence, spectroscopy, scattering, transfer functions, radiometry, and photometry. Some mention of lasers, holography, fiber optics, gradient-index optics, phase conjugation, and optical computing would tie the course to current technological development

Quantum Physics: The historical foundations of quantum physics, blackbody radiation, Compton scattering, the Davisson-Germer experiment, and the Bohr-Sommerfeld model of the atom should be established (if not previously given the student in a modern physics course). The quantum physics course should include in-depth applications of the Schrödinger equation to one-dimensional problems such as the square-well potential, barrier scattering and tunneling, and the harmonic oscillator. The treatment of quantized angular momentum should include some elementary work with operator methods and commutators. Three-dimensional problems should, at a minimum, describe the hydrogen atom and should include relativistic corrections. If this course is extended into a second semester there should be applications of essential quantum concepts to major fields of contemporary physics, e.g., multiple particle wavefunctions vis-a-vis elementary quark models, shell theory applied to nuclear models, group theory, and matrix methods applicable to the theory of solids.

Experimental Physics. The goal of this laboratory course is to give the student experience with real-world apparatus such as lasers, high field magnets, detectors, radioactive sources, vacuum equipment, and sophisticated electronics (at the level of lock-in amplifiers and multichannel scalers). The schedule should be a blend of classic experiments illustrating concepts from electricity and magnetism and quantum physics (the Franck-Hertz experiment, Zeeman effect with ions, measurement of the speed of light, etc.) as well as experiments designed to convey the flavor of contemporary experimental physics. Examples of the latter are experiments on tunnel junctions, angular correlation of gamma rays, nuclear decay spectroscopy, and magnetic resonance spectroscopy. Special attention should be given to written communication of scientific information (see the *AIP Style Manual*).

[a] Taken from the 1987 AAPT Guidelines for the Review of Baccalaureate Physics Programs.

Figure 2

Representation of the Relationship Between the Content of the GRE Physics Test and the Content of the Undergraduate Curriculum in Physics

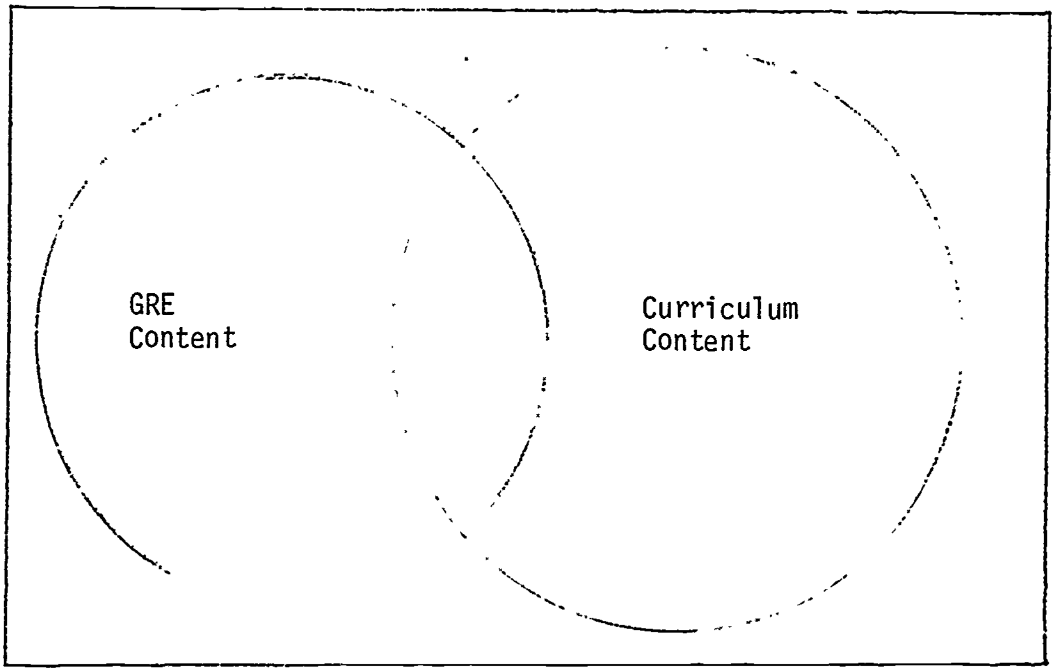

Table 1

Illustration of the Computation of R_1 for a Hypothetical Setting

	Expert (E)	t_1 E_1	t_1 E_2	$F_{(q,1)}$	t_2 E_1	t_2 E_2	$F_{(q,2)}$	t_3 E_1	t_3 E_2	$F_{(q,3)}$	$\Sigma F_{q,t} f(t)$
GRE Item (q)	1	.7	.8	.75	.4	.3	.35	.8	.7	.75	.515
	2	.5	.7	.60	.6	.7	.65	.4	.3	.35	.510
	3	.6	.5	.55	.3	.2	.25	.3	.4	.35	.335
	4	.8	.9	.85	.5	.5	.50	.7	.6	.65	.585
	5	.9	.8	.85	.8	.7	.75	1.0	.9	.95	.745
	6	1.0	1.0	1.00	.9	1.0	.95	.9	1.0	.95	.870
	7	.5	.7	.60	.6	.7	.65	.7	.6	.65	.570
	8	.6	.5	.55	.1	.2	.15	.6	.5	.55	.335
	9	.8	.7	.75	.3	.6	.70	.9	.7	.80	.665
	10	.7	.8	.75	.6	.5	.55	.8	.6	.70	.585
	Mean	.71	.74	.725	.56	.54	.550	.71	.63	.670	5.715 (Sum)

$f(t_1) = .30$ $f(t_2) = .40$ $f(t_3) = .20$

$$R_1 = \frac{1}{10} (.5715) = .5715 \quad \text{(From Eqn 1, p. 6)}$$

Alternatively,

$$R_1 = .30\ (.725) + .40\ (.550) + .20\ (.670) = .5715$$

Table 2

Summary of Survey Responses

Category of School	# of Schools	# Phys Grads, '86
TG	14	575
RG	13	222
TUG	19	208
RUG	13	78

Table 3

Summary of Textbook Usage

Curriculum Category	# Texts Used By 10% or More	Total % Students Using These Texts
Mechanics	4	92
Electricity/Magnetism	5	75
Thermodynamics/ Statistical Mechanics	2	80
Modern/Quantum	5	117

Table 4

Summary of Correlations for Expert Ratings

Subfield

Type of Rating	Mechanics		Electricity/ Magnetism		Thermodynamics/ Statistical Mech		Modern/Quantum	
	Range	Mdn	Range	Mdn	Range	Mdn	Range	Mdn
GRE Items	.51 to .86	.66	.63 to 1.00	.80	.84 to .87	.80	.46 to .87	.71
Textbook Items	-.13 to .58	.32	-.23 to .44	.34	.07 to .45	.26	.31 to .64	.42

Table 5A

R_1 Values by Subfield and Institution Type

Subfield	Type of Institution TG	TUG	RG	RUG
Mechanics	0.79	0.76	0.84	0.84
Electricity and Magnetism	0.63	0.58	0.59	0.61
Statistical Mechanics and Thermodynamics	0.77	0.61	0.82	0.63
Modern/Quantum	0.82	0.50	0.85	0.47

Table 5B

R_2 Values by Subfield and Institution Type

Subfield	Type of Institution TG	TUG	RG	RUG
Mechanics	0.44	0.38	0.48	0.45
Electricity and Magnetism	0.44	0.42	0.42	0.40
Statistical Mechanics and Thermodynamics	0.17	0.15	0.16	0.18
Modern/Quantum	0.24	0.14	0.25	0.14