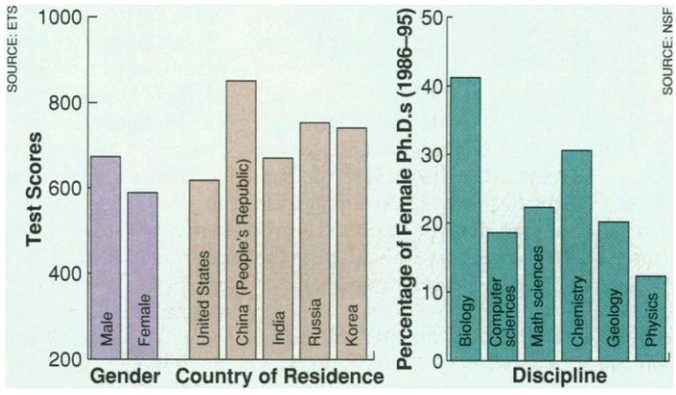
How Not to Pick a Physicist?

Standardized tests are under attack again—this time from physicists who say their field's graduate-school exam overlooks real physics talent and worsens the field's gender disparity

Who gets to do physics? Largely, those who begin their careers by winning admission to graduate programs at top universities. Who gets admitted? In the United States, for the most part, undergraduates with solid grades and warm recommendations who earn a spectacular score on an approximately 3-hour, 100-question, multiple-choice test called the Graduate Record Examination (GRE) Subject Test in physics. While all three factors carry weight, says Howard Georgi of Harvard University, reliance on the GRE score is "very seductive for the admissions committee." So, who gets to do physics can boil down to who blows away the physics GRE.

But an increasingly vocal group of physicists, including Georgi, is calling for what many an overwhelmed admissions officer may find quixotic at best: either that the test's rigid, multiple-choice format be modified—an unlikely prospect in the near future—or that the physics GRE be dropped entirely as a criterion. "I believe that the [test] in its present form ... should be abandoned," says Neal Abraham of Bryn Mawr College, who has served on the committee of physicists that advises the Educational Testing Service (ETS), the Princeton, New Jersey, organization that develops and administers the physics GRE.

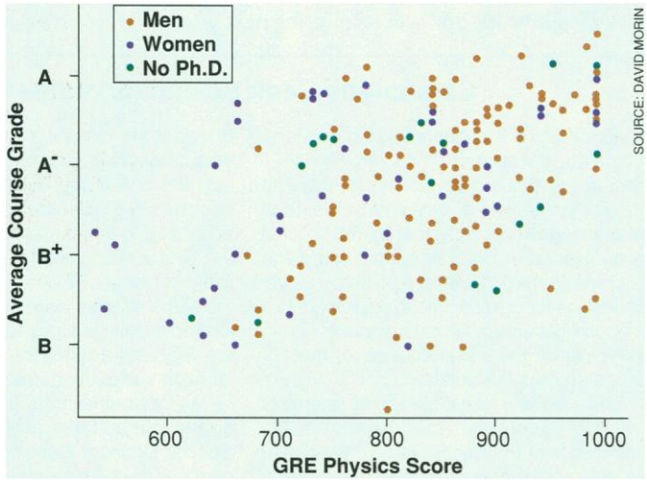
Standardized tests are often under fire, but these critics say that in relying on the GRE, physics admissions committees are on especially shaky ground. They cite a correlation between a student's scores and future performance that is among the weakest for any field in which the GRE is given; the anomalously high physics GRE scores that are logged in some parts of the world (see box); and the suspicion that an undue reliance on the exam could be partly to blame for physics' vanishingly small numbers of women and minorities—both of whom score sharply lower, on average, than other candidates. "As presently constituted, I think it's quite possible that [the test] does more harm than good," said Georgi during a meeting of the American Physical Society (APS) in Indianapolis last May, where a special session was devoted to concerns about the test.



Disparities. Mean scores on the physics GRE exam from 1992 to 1995 *(left)* varied dramatically by country of residence and gender, perhaps contributing to the small representation of women in physics.

The complaints resonate among women scientists and faculty at liberal arts colleges, whose students don't do as well on the test as do those from research universities. "Some of the most vocal complainers about this test are faculty from the elite undergraduate institutions that don't have research programs," says J. Woods Halley, a physicist at the University of Minnesota. And while the discussion of these issues is loudest in physics at the moment, it also cuts across disciplines. At last December's National Science Foundation–sponsored Women and Science conference, for example, the issue of eliminating standardized tests as a criterion for graduate school entry came up repeatedly, according to Judy Franz, executive officer for the APS, who chaired a session. Still, says psychologist Joshua Aronson of the University of Texas (UT), Austin, while "the problem is real in all the programs, the people who are thinking about it most closely are the physicists."

The rejection of the GREs is still a minority view. Without the GRE, says Halley, admissions officers would have no objective measures. Declares Peter Meyers, director of graduate studies in physics at Princeton University, "I wouldn't want to drop [the exam] at all." But the criticisms are having an effect, both at ETS and on physics admissions committees. The makers and users of these tests are trying hard to ensure that they are used as recommended—as one data point among many, not an absolute criterion. They are also trying to improve the tests themselves so that they measure something closer to innate physics talent.



Poor correlation? Average course grade versus GRE score for graduate students in the Harvard physics department.

Measuring the measure

The current physics GRE, says Thomas Griffy, a physicist who is associate dean for graduate studies at UT Austin, "tests a student's ability to do simple problems quickly"—a key skill for a well-rounded physicist, says Griffy. Even physicists who defend the test concede that gauging creativity must be left to other measures. "I don't think that is the role of a bulk-delivered and -graded test," says Meyers. But admissions committees nonetheless accord the test a heavy, sometimes overwhelming, weight.

"When your admissions committee gets in their little room and closes the door, they can do whatever they want," says Jennifer Siders, a physicist at Los Alamos National Laboratory who recently received her Ph.D. at UT Austin, where she served as a student member of the admissions committee.

In China, an Enigma in Test Scores

In a field as obsessed with numbers as physics is, here are two whose disparity can hardly be overlooked: 618 and 851. The first is the mean score earned by U.S. residents between 1992 and 1995 on the Graduate Record Examination (GRE) subject test for physics, a standardized test required for admission to most physics graduate programs in the United States. The second is the mean score earned over the same period by residents of the People's Republic of China (PRC), who account for about 6% of Ph.D.s awarded in all fields by U.S. universities each year. On a test that has possible scores of between 200 and 990, says Howard Georgi, a physicist at Harvard University, "the difference between the PRC and everybody else is incredibly dramatic."

The disparity also underscores the caution with which GRE scores have to be approached, say faculty at physics graduate programs (see main text). One factor in the high scores, they say, has to be the huge pool of highly motivated, well-trained students in the PRC. But physics faculty also say that Chinese students' overwhelming advantage on the test doesn't seem to be reflected in other measures of physics ability. An educational system focused on exam-taking and the existence of poorly regulated "coaching" classes for the physics GRE may have inflated the PRC scores, say some observers. And *Science* has learned of another factor that may have played some role in the past: widespread security breaches, which culminated in October 1992, when exam booklets were widely leaked to students in the PRC before the physics exam was given.

No one denies that dozens of top-notch physicists have emerged from the PRC in the 1990s. J. Woods Halley, a physicist at the University of Minnesota who has served on the physics GRE advisory committee, suspects that the effort to identify "unfair" advantages is driven in part by nationalistic bias—"a certain feeling that [PRC students] can't be that good." He points out that "these are very, very able students out of a huge pool." Xueqiao Xu, a physicist at the Lawrence Livermore National Laboratory, suggests that the Chinese educational system may also foster strength in physics, since students in the PRC gain "a very strong mathematics and physics background" as early as elementary school.

But while "there are a lot of good physics students in the PRC," says Georgi, his experience with graduate admissions suggests that those students "are certainly nowhere near as good at physics as they are at taking the GRE subject test." The GRE scores, says Jack Mochel, professor and associate head of physics at the University of Illinois at Urbana-Champaign, "are no indication of how [PRC students] will do in graduate school."

Coaching for the test may explain part of the discrepancy, says Neal Abraham, a physicist at Bryn Mawr College who has also served on the physics GRE committee. And in a recent essay published on the Internet, he cited another factor: "Chinese students report that books of prior exams and exam questions are compiled by test takers and are available for study"—testimony that one former student from the vicinity of Xian province confirmed to *Science*. While American physics students also benefit from "practice" multiple-choice exams, the actual questions from old GREs are not made available. Since some questions are reused, extensive use of such materials could give test-takers an unfair advantage.

That practice peaked in October 1992, according to sources including the Educational Testing Service (ETS) in Princeton, New Jersey, which produces the exam. At the time, "we occasionally reused [entire] test forms," says Jacqueline Briel, associate program director for the GRE, and it turned out that copies of a previous exam using the same form were circulating among students planning to take the test. The leak was discovered when a student complained to ETS, Halley says. "He said this test wasn't fair because he hadn't seen it beforehand as his friends had. This caused real earthquakes at ETS."

In response, the testing service voided those test results and halted physics GRE testing in the PRC for a year while it made changes to improve the test's security, says Briel. The test is now given in China only once a year with a fresh form each time. She believes that uncaught security problems are now rare, noting that "aberrations" in the scores have not recurred.

Even so, she stresses that admissions committees should also rely heavily on other information about a candidate—course grades, recommendations, and any knowledge of a candidate's personal motivations and English skills, for example. Xu agrees. "First you want a high score," he says. "But if you have [enough] time and manpower, interview these candidates." Only by going beyond the numbers, Xu says, can admissions officers identify which candidates are prepared to make the great cultural leap from Chinese academics to the research community in the United States. **–J.G.**

Siders, who spoke at the APS session, recalls that in her first year on the committee, a faculty member began the proceedings by saying, "Here are our candidates; I think we should do the same thing we did last year, and not look at anyone under a 600" on the physics GRE—even though the cutoff was not a stated criterion for entry into the program.

ETS discourages such cutoffs, but "numbers are easier to hang your hat on than something that is more qualitative, like a letter of recommendation," admits J.D. Garcia of the University of Arizona, who chaired the GRE physics committee for 4 years, until last June. And several sources confirm that a number of physics departments use such cutoffs, although directors of graduate programs often become evasive when asked directly about the practice. Adds Pamela Zappardino, executive director of the National Center for Fair and Open Testing in Cambridge, Massachusetts, who also spoke at the APS meeting: "Asking someone to use test scores with caution is in many ways like asking a jury to disregard something they've heard. It's very hard to get it out of your head."

Yet, Georgi, Abraham, and other critics of the test say that the scores mean little. Says Georgi, "You just can't hope to use these tests, or any tests of this kind, to rank top candidates." ETS's own figures bear out the weak link between GRE scores and first-year grades in graduate school: The correlation is the lowest of any field for which ETS provides statistics, with the exception of geology. The numerical correlation is 0.27, which means that just 7% of the variance in those grades could be predicted by scores on the physics GRE.

And David Morin, a recent physics Ph.D. at Harvard, uncovered other dimensions of the problem in an independent study on groups of graduate students who entered his own program in the 1980s and 1990s. As he reported at the APS session, Morin looked for correlations between GRE scores and several measures of later performance, including course grades in graduate school, how long it took the students to complete the program, and professors' assessments of the students' later performance as physicists. The only link

Morin uncovered was with grades—and although, at 0.48, it was much stronger than ETS's own figure, Morin still considers it weak. "For the students who did come here, the GRE physics [test] didn't mean too much," he concludes.

ETS suggests that the correlation with graduate school grades is weak because students who score very low on the test never make it to graduate school—where, presumably, their low grades would have bolstered the correlation. But Georgi argues that the test sharply underpredicts or overpredicts the physics performance of particular groups of students. On the one hand, he said at the APS session, the very highest scores on the test often go to what he calls "idiot savants": students who are good at manipulating symbols but have little understanding of physics.

On the other hand, Georgi noted that some of his most brilliant women students have bombed on the physics GRE—anecdotal support for a gender disparity that shows up as a nearly 100-point differential between male and female test-takers between 1992 and 1995. As a result, many women, including Siders and Zappardino, see the test as a "gatekeeper" contributing to the dismal gender disparity in physics, the largest such gap of any major scientific field. In an independent study in the graduate physics program at UT Austin, Siders found that the number of women admitted fell precipitously—from an average of about 10 to just one or none—when cutoffs based on physics GREs went into effect.

No one claims to understand why women score lower than men. "Women get lower scores on this test even when they are equally good by any other measure," says John Schwarz, a physicist at Caltech. Possible reasons, he says, range from a different test-taking style to a dislike of time pressure. But when Georgi queried one talented woman about her low score, he said at the APS meeting, "She told me that the physics GRE was simply too nerdy to be taken seriously by an intelligent woman." Georgi drew laughter when he paused and said, "I'm not sure what that says about the men."

Handle with care

ETS isn't laughing about these disparities. Individual GRE questions receive close monitoring for sensitivity and bias toward or against any demographic group, and must pass muster with women and minority members of the GRE faculty committee. "I would not tolerate it if I saw any question which had a taint of bias against women or minorities," says Garcia, of the GRE faculty committee. But he allows that the test "does have warts. It does have flaws."

In an effort to minimize them, says Jacqueline Briel, associate program director for the GRE, ETS has tentatively scheduled a so-called FAME conference—for fairness, access, multiculturalism, and equity—for the winter of 1997 to explore possible reasons for the demographic disparities. The testing service is also looking at ways to base bulk-graded exams on something other than multiple-choice questions—say, by using computers to grade a student's approach to a problem, rather than just the final answer. But Minnesota's Halley notes that "it's daunting to teach a machine to measure that," and any solution is still probably far off.

In the meantime, Briel stresses, the tests should be applied cautiously: They are "one source of information to be used in conjunction with other criteria"—grades, recommendations, essays, and any record of other activities, such as published research. Adds James Stith, a physicist at Ohio State University: "What we have to do is convince our colleagues on these committees that this is not a good single indicator. For those who use it as a data point" among many others, says Stith, "I don't see that it does any harm."

Georgi, however, insists, "We should either fix it or seriously consider getting rid of it." As the ongoing, and sometimes agonizing, discussion shows, neither side in the debate is likely to let the issue rest.

–James Glanz

PLANETARY SCIENCE

The Moon Looms Large in Japan's Plans

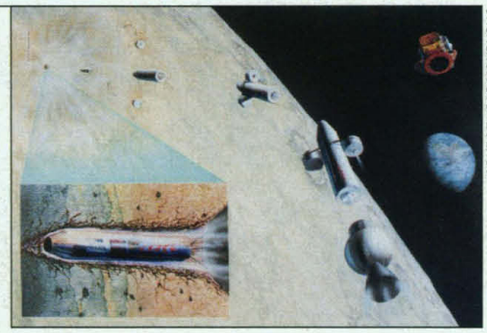
KYOTO—Moon viewing is an autumn tradition in Japan in which friends gather to gaze at the harvest moon while penning poetry and sipping sake. The moon is also getting a lot of attention from Japan's space scientists: Last week, at a major international meeting here, Japanese researchers and officials provided fresh details of an ambitious lunar exploration effort that holds a high priority in the nation's space program.

Japan is pressing ahead with its lunar program at a time when most other nations are focusing their space efforts elsewhere—a fact that drew some envy from the 150 researchers in fields ranging from astrophysics to civil engineering who gathered here for the Second International Lunar Workshop, held simultaneously with the first International Lunar Exploration Working Group. "I wish we had done this years ago," says Michael Duke of the Lunar and Planetary Institute in Houston. Hitoshi Mizutani, a planetary scientist at the Institute for Space and Astronautical Science (ISAS) and chair of the workshop session, acknowledged that "Japan has taken on a leading role in lunar activities." In particular, Japan is backing two of the three international missions now scheduled to land on or orbit the moon in the next several years.

The first of these missions, the LUNAR-A mission, lifts off next summer. An ISAS project, the satellite will orbit the moon and release three penetrators that will slam into the moon's surface. The penetrators, traveling at 250 to 300 meters per second, are designed to burrow 2 meters into the ground in the first attempt to drop such penetrators from an orbiting spacecraft. Two of the 90-cm-long penetrators, each housing a seismometer and a heat probe, will be placed on the near side of the moon and one will be placed on the far side. They will transmit seismic and heat-flow data to the orbiter when it passes overhead, about once every 15 days for a year. The orbiter will then relay the data to Earth, helping scientists to better understand some geological forces that are easier to monitor on the moon than on its larger sister. "It will capture a part of the Earth-moon history that is lacking on the Earth," says Carlé Pieters, a planetary scientist at Brown University in Providence, Rhode Island.

A few months later, NASA will launch its Lunar Prospector mission, a 223-kg package of remote-sensing instruments placed into lunar orbit. The 1-year mission hopes to plot the distribution of elements such as uranium, iron, and silicon; search for ice at the lunar poles; and study the moon's magnetic and near-side gravitational characteristics. The data will add greatly to knowledge of the composition of the moon and identify potentially valuable resources.

The most ambitious of the three missions is Japan's Selenological and Engineering Explorer (SELENE), planned for launch in 2003. A joint mission with Japan's National Space Development Agency (NASDA) that is



SOURCE: ISAS/MONBUSHO

New digs. Japan's LUNAR-A satellite will launch penetrators to burrow into the moon's surface.