When I was in graduate school, I was taught the science of psychometrics. My professors were staunch members of academia who taught us everything they thought we needed to know about testing. From their academic perspective they truly seemed to believe that if we knew about validity, reliability, classical theory, and item-response theory, that was all we needed in order to be psychometricians. What they neglected to tell us—or maybe they simply did not realize it themselves—was that testing never occurs in a purely psychometric environment. They did not tell us about "the three Ps" of testing in the real world: perception, paranoia, and politics.

Overview of the Texas Academic Skills Program™

I will leave the first two Ps to another discussion and address here the third one: politics in testing. I will examine such politics with particular respect to the Texas Academic Skills Program, or TASP®. To really discuss some of these issues, I first need to provide a framework, i.e., a brief overview of the TASP. In 1987, Texas passed a law requiring all entering college freshmen to take a basic skills test of reading, writing, and mathematics—basic skills that entering freshmen should possess. Students must take this test by the time they accumulate nine credit hours in college; otherwise, they are barred from further enrollment until they take the test. If they fail a portion of the test, then remediation is mandatory. They have to participate in remediation until they pass all sections of the test. If they fail a portion of the test, then remediation is mandatory. They have to participate in remediation until they pass all sections of the test. In a university, students cannot take upper-division coursework beyond 60 credit hours until they have passed all sections of the test. In a community college, students cannot graduate with an associate degree or certain certificates until they have passed all sections of the test. It becomes quite evident that this is a serious, high-stakes test. Many events are contingent upon
taking this test and passing it. These consequences might, at first, seem harsh or punitive, but the whole purpose of the TASP is to identify students' weaknesses early and get them help.

Politics and the TASP

In this context, I would like to explore the relationship between the TASP and the third P, politics. In 1989, the year the TASP was implemented, we offered five test administrations: two in each of the fall and spring semesters and one during the summer semester. During the first two years of the testing program, we were deluged with calls and complaints from parents, teachers, students, and even legislators that students had been unable to take the TASP in one of those five test administrations, they had accumulated their nine credit hours, and now they were unable to enroll in the next semester. Being barred from enrollment is a significant problem, and not just for the affected students. The Commissioner of Higher Education came under a great deal of pressure to do something about this problem, and he passed the message along to us: we had to offer a supplemental test administration in January, one month away. In addition, there was some discussion about the possibility of holding another supplemental administration just prior to the fall semester, because students needed an extra chance to take the TASP and comply with the law. We called National Evaluation Systems, Inc. (NES), and asked if it would be possible to add a test administration, to be held in just 30 days. Fortunately NES rose admirably to the challenge, and we actually held our first supplemental test administration that January of 1991. In fact, we now offer two supplemental testing sessions each year, one in January and one in August.

With that kind of safety valve in place, we anticipated that our troubles would be over. Nevertheless, we still had students who were missing the regular and the supplemental test administrations, students who wanted to go to school but who had not taken the test. The numbers were smaller, so we were a little better off, but those students who had missed the test had no difficulty finding someone in authority to whom to complain about the problem. So we decided to add one more regular administration to the year's testing venue, which would bring us to a total of six regular plus two supplemental administrations.
With eight chances a year for students to take the test, we were certain we had the matter resolved. However, it was not long before we again began hearing the same complaints. We were not hearing as many, but we were still hearing the same refrain: students who had not taken the test could now not attend school unless they could take the test on demand.

Computer-Administered Testing—A Solution?

Since we were already interested in developing a computer-adaptive version of the TASP test, both the state and NES began to consider the use of computers to help us with this issue of test availability. We were not naive, however, about the magnitude of the undertaking we were considering. Computer equipment is expensive, test materials must be kept secure, proctors must be present whenever the test is given, written essays still must be scored, and a host of other factors needed to be considered. We narrowed this list of factors down to the four most critical ones:

- It was essential that computer test results be comparable to the paper-and-pencil test results.
- The test materials absolutely had to be kept secure. On such a high-stakes test, we could not allow unfair advantage because some students had found out information that others had not.
- Using the computer had to have as little impact on the examinees as possible, because not all students are computer literate.
- The test had to be available on fairly short notice, as close to on demand as we could get.

How did we then begin to address these factors? To ensure that the results from the computer test would be comparable to those of the paper-and-pencil test, we just literally put the paper-and-pencil test on the computer. In other words, the two tests are virtually identical in content and format; the only difference is one is on paper and the other is on a monitor. We conducted a field test in which students took both versions of the test, alternately. We compared the results and found that there were no discernible differences in the test results. It is worth noting
that we committed ourselves to watch this carefully over time, and that finding has held true even now that we have given the test for a while. Basically there are no differences in the test results.

With the factor of test comparability under control, we turned our attention to the matter of test security. We needed to implement a system for keeping test materials secure while they are being used on computers at multiple locations around the state. NES devised a number of approaches. I will not detail all of the innovations, but I will mention a few of the highlights. The test is administered on stand-alone networks using dedicated direct analog phone lines, and we encrypt, or encode, the data. No other traffic is permitted on these lines. The file server at each of the 12 sites is in a heavy, locked, fireproof cabinet. There are no keys to that cabinet on site; NES retains the keys. These file servers are also located in secure areas, so they are doubly secure. Each student sits in a carrel with only a monitor. The actual test resides on the file server itself. There are no disk drives, no way to remove or alter test information whatsoever. Each system is password protected, and no student can take the test unless he or she is registered for it. Security cameras monitor each carrel, and a digital photograph is taken of each student when he or she arrives for the test, so we know who took the test. That picture is later printed on the score report as an additional security measure. Finally, a student cannot take the computerized test more frequently than once every 120 days, because the computerized test is not meant to be the primary test-taking mode. It is intended as the emergency backup, in the event a student is going to be barred from going to school if he or she cannot take the test. There are a number of other features as well, but these examples demonstrate why we are satisfied that we have adequately addressed the security factor surrounding this high-stakes test.

We next turned to our third factor, that of ensuring that a computer-based test would have as little impact on the examinees as possible. Texas has a very diverse population in almost every sense of the word. Not all of the students arriving at the steps of higher education are computer literate. As a matter of fact, there are many who have never worked with a computer and even a fair number who have never seen a computer. How could we, then, use a computer to fill our void in
The Politics of Standardized Testing

test taking and yet keep it from being a barrier to some students or an advantage for others? Again we brainstormed with NES, and we came up with some interesting solutions. When a student takes the test, there is no keyboard—it is all touch screen. For example, when the monitor displays "Do you want the next item?", the examinee can touch that text on the screen and get the next item. It would be difficult to develop an approach that is more user friendly and unintimidating to the non-computer-literate than touch-screen technology. We also minimized the computer aspect in that the computer version is, for all intents and purposes, identical to the paper-and-pencil version. Format, everything, is the same. Examinees can go from item to item or from section to section of the test, forward or backward. They can change answers, which is critical. In a lot of computer tests it is impossible to change answers after they have been indicated. On the computerized TASP test, the examinee can change answers freely until he or she is ready to stop and have the test scored. At that point there is a three-step safeguard to ensure that examinees, especially those who are not facile with computers, do not accidentally have the test scored prematurely. Also, the examinees sit in deep carrels with sides that extend behind them, so testing conditions are optimal. It is quiet, there are minimal distractions, and the monitor is a large, 20-inch screen, all of which helps alleviate any anxiety examinees might be experiencing due to the computer. We feel that all of these factors dramatically equalize the test-taking experience for computer-proficient and computer-uncertain examinees alike.

Finally, we had to address the need for examinees to be able to register on short notice for the test. While the examinees wanted on-demand testing, we were interested in "close-to-on-demand" testing but not "on-demand," precisely, because on-demand testing presents certain security hazards we simply could not risk with this high-stakes test. We needed the shortest workable time frame, so once again we turned to NES to deal with the challenge. The result is that we now have 12 sites for the computer-adapted TASP in Texas. Examinees can register for the test by telephone as late as 4 P.M. the day before they want to take the test. In most cases, then, an examinee can take the test with one day's notice. During certain peak times the sites may get filled, so in those cases examinees can usually be tested within the next day or two. So within one, two, or three days, virtually all students who want to take this test have access to it.
So far we have been able to test everyone who has wanted to test, with only a very few exceptions. In a few cases we have had to ask the institutions to provide more testing time in their blocks. From our perspective, the ability to anticipate periods of peak use is one of the less-controllable aspects of this computer-administered test. The institution provides the space and the proctors. We can only accommodate examinees according to what the institutions will provide. However, our experience thus far has been that when student demand increases, most of the time the institutions have been willing to add one more block in the afternoon or evening or whatever it takes to get the students tested.

To summarize our progress to date: We have a computerized test. It is operational. It meets all the specifications we set. We naturally experienced some problems, but not a lot. We have certainly learned a great deal. We have also learned—or more accurately, our suspicion was reinforced—that ultimately there is nothing we can do about students who wait until the absolute last minute to do everything. We still have students who wait until not even the computer-administered test can rescue them. Obviously it would be unrealistic to imagine that we could ever wholly eradicate issues concerning test availability, so our benchmark for judging the success of our program is that we receive very few complaints from legislators these days. As one college president recently said, and I quote, "We tested 156 students using the computerized test prior to the last term. And while that's not a huge number in terms of overall testing, it represents 156 potential complaints that I didn't have to deal with." Of course, from our perspective at the state level, we did not have to deal with those complaints either. I think this is a fairly good indication of success.

It would be similarly unrealistic to imagine that the real world of testing would ever be devoid of politics. What I would like to envision is that we continue to devise creative solutions, exemplified in our response to political pressure for increased testing availability, and that we work with legislators, using creativity and logic to achieve a common goal, which is to use testing as a tool to ensure that our students acquire a solid educational foundation.