Three stages mark the development of relevance in testing: relevance to the educational program, relevance to the individual student's past accomplishments, and relevance to the student's future accomplishments. If the College Entrance Examination Board should have to follow paths such as these, it has ahead of it the enormous job of transforming drastically the nature of its activities. Three possible stages are envisaged in the future development of the Board's tests: (1) for the immediate future, an extension of the recent trend toward the diversity of programs and of tests within programs (this being the stage of the multiplex external program); (2) a reduction of emphasis on external examinations and increased reliance on the record compiled by the student in his own school (this being the stage of the school-based program); and, (3) a system, eventually, of examinations in which each student is presented with the individual questions most pertinent to his past preparation and to his responses to test questions earlier in the sequence (this being the stage of the student-based program). It is held that the examination program should match both students and educational programs in diversity. (JM)
Relevance in Testing

Wm. W. Turnbull
Educational Testing Service, 1967

College Entrance Examination Board, 1967
Relevance in Testing
William W. Turnbull

Few agencies have survived since the turn of the century in America unless they have had a notable capacity for change. This proposition is true of agencies working in the field of education, as in other fields. Sir Eric Ashby has noted that, in order to survive, educational institutions must be both sufficiently stable and sufficiently responsive — not an easy job! The College Entrance Examination Board, which has been a notably stable institution over the period since 1900, has at the same time been no stranger to change. The existence of this Commission demonstrates that the 1960's, and indeed the second half of the century, are likely to see the Board continuing as an institution responsive to changes in the social and institutional context in which it operates.

The work of the Board has not been merely the mirror of change. It has also served as a powerful change agent in its own right. In the Educational Testing Service Annual Report for 1961-62, it was pointed out "... that, during the lifetime of the SAT, more opportunities for higher education have been opened up for more students than ever before in this country. In the 1930's, the availability of the SAT provided the additional method needed for identifying able students sufficiently well to justify the award of very large national and regional

scholarships. Since then, the SAT and other tests similar to it have contributed to the effective selection of scholarship students in many colleges throughout the country.

"Another interesting development is the change that has occurred in the undergraduate bodies of colleges that have used the SAT over the last two or three decades. In the Ivy League colleges, for instance, the undergraduate body of the 1920's was a homogeneous one with respect to socio-economic background and a heterogeneous one with respect to intellectual ability. Today the picture is almost reversed -- undergraduates in these colleges come from widely varying socio-economic backgrounds and possess a generally high intellectual ability.

"A third development worth noting is the fact that at colleges where the SAT has been used as part of the admissions process for the last twenty or thirty years, the academic failure of enrolled students has been reduced to a minimal level.

"Surely more than coincidence is involved in the fact that these developments occurred during the lifetime of the SAT. Obviously, many other elements have also contributed to expanding opportunities for higher education, to better identification and encouragement of able students, to better guidance, and to reduction of the academic failure rate at many colleges. The SAT, however, has played its part. And the net effect has been the lifting of many of the earlier restrictions to higher education in this country."

These points are worth developing a little further. It has frequently been said that ten or twenty years ago the most selective eastern private colleges and universities drew 80 per cent of their freshmen from 150 or 200 secondary schools. Now, to fill 80 per cent of their first
year places, they draw from upwards of 500 schools. This movement has, of course, been fueled by the strong desire of institutions of higher education to broaden the composition of their student bodies. It has been made possible, however, by the nationwide availability of examinations designed consciously to allow a good student anywhere to show to advantage, and to be considered favorably, whether or not his school is known to the college.

For a time, perhaps, the influx of students from distant schools into eastern prestige colleges, made possible in part by use of the testing system, posed some problems for the independent secondary schools that could no longer be confident that their own students would find ready admission to the colleges that would surely have accepted them in an earlier day. In the past few years, however, the independent school graduates have accustomed themselves to applying for admission to universities farther afield. Now it is these students whose credentials may have to be attested by the test scores. In the competitive grading system of a good private secondary school, their rank in class may be undistinguished; they frequently must count on high test scores to convince a faraway registrar of their real academic distinction.

It can fairly be said, then, that the Board's testing program has played and continues to play a significant role in broadening and democratizing the basis for the admission of students to a wide array of institutions of higher education in this country. This is by no means a trivial accomplishment. It has been achieved during a critical period in educational history, when we in this country have been engaging in the most dramatic extension of college education to young people of college age that the world has yet seen. It is instructive, however, to look behind the steady upward trend in the proportion of
students continuing into college and to ask about the qualitative differences in the student body introduced by this remarkable change.

As was implied earlier in this paper, during the early part of the century the operative selection factor for much, if not most, college-going was socio-economic. But as we all know, college was seen by many people as the path to upward social mobility for their children -- the avenue to a job with more prestige and better pay for a boy, or to a "better marriage" for a girl. As college came within the economic reach of the children of these families, and as means were developed to recognize their ability to profit from collegiate study, able young people began moving into college places regardless of their family backgrounds.

Still, it is fair to say that the scholastic abilities of the incoming students did not change radically before World War II. In the 1940's, it was frequently pointed out that for every high school graduate who proceeded into higher education, there was another of equal aptitude who did not. Increasingly, over the years, the numbers swelled, but scholastically the new clientele was not unlike the old.

What, then, of the educational programs into which the students were progressing? At all levels, and increasingly in the post-secondary years of education, we have seen changes in the nature of what has been taught. We have labored heroically to modernize the curriculum, especially in the last decade. And we have added courses in vocational and technical subjects, whether through the creation of special institutions or through the incorporation of these studies in the programs of comprehensive high schools or institutions of higher education. But the fact remains that over the period since 1900 for most stu-
dents, the core of the instructional program has remained the traditional academic disciplines. In 1964, it could still be said\(^3\) that the nine literary works most widely studied in public secondary schools in the United States were: Macbeth, Julius Caesar, Our Town, Great Expectations, Hamlet, Red Badge of Courage, Tale of Two Cities, Scarlet Letter, and, inevitably, Silas Marner. And in a paper presented at the 1967 annual meeting of the American Council on Education, Philip Werdell said \(\ldots\) for most students, undergraduate education continues to be an extension of the goals, the requirements, and the methods developed a century ago in response to the industrial revolution.\(^4\)

Whether or not one wants to adopt this viewpoint entirely, it is probably reasonable to assert that during the first half of the present century we have seen an increasingly large but academically quite comparable student body passing through a substantially unchanging array of educational offerings in both school and college.

I raise this matter of historical similarity or continuity, not to deplore it, but to point out the unremarkable fact that stable educational structures beget stable examination programs. More important, if a single educational program, albeit of varied levels of academic challenge from institution to institution, is expected to


serve the great majority of students proceeding from school to college, a single program of examinations can reasonably be expected to serve the function of sorting students into the several rough ability groups that match the levels of institutional demand. In the terms of the title of this paper, a stable examination program, evolved carefully over many years, may well be most highly relevant to a stable educational system.

In such a comparatively stable milieu, it should perhaps come as no surprise to us that a rather stable examination program has proved serviceable in the transition from school to college. The Scholastic Aptitude Test was introduced some 40 years ago. It is about 30 years since achievement examinations were introduced in something approximating their present form.

The objection will immediately be raised that the examinations are different. They have indeed been changed and expanded. We hope they have been improved. We are aware of at least some of their shortcomings, and we are working hard to vary and improve them further. Granting those facts, however, I would return to the assertion that in form, function, and essential content, the tests now in use are recognizably of the same genus as those of a generation or more ago.

When the function of an examination program is to spread individuals along a single dimension, a unitary set of tests fulfills the function of revealing individual differences. The jingle fallacy that standardized testing leads to standardized students has been surprisingly long-lived. It is equivalent to suggesting that the standardized conditions of a race such as the mile will somehow lead to uniform performance on the part of the contestants. Every passing year reminds me more poignantly that this is not likely to be the case. The more rigorous
the conditions of the trial, the more dependably are in-
dividual differences -- including all the frailties of the
human condition -- revealed. Standardized tests have, in
fact, been the single most powerful stimulus to the rec-
ognition of individual differences in this century.

Let us, however, pursue the sports analogy a bit further.
A single contest, like the mile event, may serve reasonably
well as an index of the respective running abilities of a
large group of athletes. If we are setting up a 5-minute
test of this athletic talent, a well-standardized mile
race would perhaps work fairly well. But suppose we were
trying to select a team, not just for track, but for a
meet that would include pole vaulting, shot-putting, and
perhaps even swimming. Immediately we would want a set of
tests that would reveal accomplishments of much greater
variety.

In this example, if we were selecting contestants for
the decathlon, we might assign all aspirants a broad vari-
ety of athletic tasks to perform and choose those who were
most proficient overall. But suppose we were faced with
some hopeful swimmers, some discus throwers, and so on.
The indicated strategy would be to test each person's
ability to perform well in just those events he wanted to
enter, or thought he might want to enter. In short, we
would tailor the tests to the abilities and plans of the
individual.

I have made the argument that in the past we have had a
high degree of continuity in both the college applicants
and the college program; that, to oversimplify, we hav-
had runners turning out for track. But is this stability
likely to continue? In fact, there are many signs to indi-
cate that the continuity is vanishing rapidly, and that
we are facing a future characterized by a high degree of
diversity in both the clientele and in the programs avail-
able to them. The conclusion I want to suggest is that the examination program should undergo changes of equal scope and in the same direction: that it should be as diverse as the students and the educational programs if it is to be relevant to the new situation.

In order to develop some idea of the extent of the diversity that lies ahead, let's look first at the composition of the student body entering, or aspiring to enter, higher education. Here I'd like to cite a few figures taken from just two of the several studies of the topic. Wolfle in his data describing the situation in 1953, estimated that 35 per cent of all high school graduates were proceeding to college. The Project TALENT survey showed that in 1960, 43 per cent of all high school students were going on.

The increase over this period will surprise no one. What I find much more illuminating, however, is a breakdown of these overall percentages to show the proportion of high school graduates at various ability levels going on to college. Let me read you the figures as I have derived them from these two studies.

<table>
<thead>
<tr>
<th>Wolfle</th>
<th>TALENT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1953</td>
<td>1960</td>
</tr>
<tr>
<td>Lowest Quarter</td>
<td>20</td>
</tr>
<tr>
<td>Third Quarter</td>
<td>32</td>
</tr>
<tr>
<td>Second Quarter</td>
<td>38</td>
</tr>
<tr>
<td>Top Quarter</td>
<td>48</td>
</tr>
</tbody>
</table>


These data were gathered in somewhat different ways, and I do not wish to imply that the comparisons can be made with any great precision. But the gross differences are large and revealing. In the third and fourth quarters of the high school ability range, the 1953 and 1960 data were essentially the same. At the top of the range, they were strikingly different. In the highest quarter, in 1953 only 48 per cent continued into college, while in 1960, the figure was 80 per cent. In SAT terms, these are students who scored in the area of, say, 450 or better.

If one looks at only the top 10 per cent academically of the high school graduates, in 1953 about 54 per cent of them were going on, and in 1960 about 89 per cent. These are students who would be expected to score above perhaps 550 on the SAT.

This is a sharp and encouraging difference at the high end of the academic scale. I have no doubt that if I had the figures for 1967, they would be changed again in the upward direction. But obviously, they cannot have moved up again as dramatically for these able students: already, in 1960, at the 80 per cent level of continuation for the top quarter of the class, we were approaching what may be a practical saturation level.

Let me say emphatically, although it may interrupt the train of logic for a moment, that we cannot afford to call 80 per cent a near-saturation figure and go on. The 20 per cent of the top quarter of the class who were not continuing to college in 1960 were a vitally important group in many ways. As Ralph Berdie's studies and others have been showing for years, the academically talented students who fail to go to college usually come from working-class

7. See, for example, Berdie, Ralph F. After High School, What? (Minneapolis: University of Minnesota Press, 1953)
families. What happens to their education is of tremendous significance to the structure of American society.

For the moment, however, I would like to return to the comparative figures to make a different point. That is, that in the years after the war to 1960, we succeeded in expanding markedly the proportion of the top quarter of the class going on to college, to the point where we had very nearly reached the limit for them. Automatically, this means that all of the further increase in the proportion of students going beyond high school must come from the second, third and fourth ability groups. This is not a new thought, but I think the cold statistics of the SAT scale may serve to drive it home: these students are largely from the score range in the two and three hundreds. We can emphatically lay to rest the old assertion that for every high school graduate who proceeds into higher education, there is another of equal academic ability who does not.

To look at the student body along the narrow dimensions of academic talent is, of course, grossly inadequate. For the students newly represented on college rolls, skills and aptitudes of quite different orders are probably the pertinent dimensions of comparison. It is symptomatic of our problem that we do not have the data to show systematically the ways in which the college-going population is changing with respect to dimensions other than scholastic aptitude.

Having looked at the incoming student body and its increasing heterogeneity, let us turn to the institutions prepared to receive these students. Clearly, in education we are moving away from the relatively uniform academic program of earlier decades to a much more diversified assortment of offerings. At the higher education level, the community college in particular offers a ready example of
an institution that has accepted just this responsibility. It provides the academic equivalent of the hammer throw as well as providing the cross-country run.

In these circumstances, the day when a single entrance measure or an array of traditional academic measures was an adequate yardstick for all candidates has vanished forever, if indeed that day ever existed. The academic dimension is relevant to only a fraction of the tasks to be performed. We are at a point where we need to tailor the entrance measures to the particular abilities and aspirations of the individual students, rejecting the concept that all students should "prove themselves" on the same set of examinations.

If your Greek mythology is in good repair, you will recall that Procrustes was the gentleman -- or villain -- with the bed to which all wayfarers were fitted. I suspect it may have revealed individual differences superbly along one dimension. Proteus, by contrast, showed infinite flexibility by assuming whatever form might be most fitting in the instant circumstances.

I am suggesting, in this paper, that for a well-defined and homogeneous form of education, the Procrustean model of examinations may be most relevant. As education varies its forms, however, a Protean examination program may be needed.

I do not wish to suggest that the College Board Admissions Testing Program constitutes an intellectual bed of Procrustes. For several decades, in fact, each student has been free to choose those subjects in which he would be examined, subject to some constraints laid on by his intended college. In more recent years, the Board has moved much farther and much faster to meet the individual's needs for a personal choice of examinations relevant to the educational options more widely available to him.
Consider the advent of the Advanced Placement Examinations. One may agree that the use of these measures has grown because of the increased variance in the demand level of secondary school programs, or that the causality has run the other way and that the availability of Advanced Placement Examinations has encouraged the schools to stretch their instruction. In fact, it has undoubtedly worked both ways. In any case, we see here an educationally important innovation, permitting the advanced student to be examined in a fashion relevant to his own preparation.

By now, it is probably apparent that I welcome the advent of Proteus. In this circumstance, you may ask where I think he is going to lead us, or should lead us.

I see three possible stages in the future development of the Board's tests: (1) For the immediate future, an extension of the recent trend toward the diversity of programs and of tests within programs. I will call this the stage of the multiplex external program.

(2) Next, a reduction of emphasis on external examinations and increased reliance on the record compiled by the student in his own school. I will call this the stage of the school-based program.

(3) Eventually, a system of examinations in which each student is presented with the individual questions most pertinent to his past preparation and to his responses to test questions earlier in the sequence. I will call this the stage of the student-based program.

Let's examine each of the three proposed stages briefly.

1. Multiplex external program

The Board has already moved from the original "simplex" system of achievement examinations only — which was
never so simple! — through the "duplex" system of scholastic aptitude and achievement, substantially into the beginning of a new "multiplex" system. The program is already made up of many tests of aptitude and achievement.

Two recent developments extend the options significantly. One is the College-Level Examination Program, designed especially with the object of allowing each student to pursue his own educational course in his own way and demonstrate his competence at the time of his choosing. Another is the Comparative Guidance and Placement Program, which marks the first major move of the Board to recognize, and accommodate to, the varied talent requirements of two-year institutions. Increasingly, then, the College Board program is preparing to meet the student on his own terms. The Board should, I believe, broaden still further the aptitudes, interests, and varieties of attainment for which it provides measures. These should be seen as extending rather than as replacing the present program.

The multiplex program seems to me to present the best hope of reaching the severely disadvantaged students — the potential college group that now and in the near future has no real prospect of demonstrating competence on traditional academic measures. At the other extreme, it offers the possibility of meeting the needs of the extremely capable students, for whom the regular admissions program may not be entirely adequate, and of whom we cannot lose sight in our desire to recognize the proper claims of the educationally underprivileged.

A more highly diversified examination system should serve the placement purpose well, just as the Advanced Placement Examinations now provide a better match than the regular Achievement Tests for some students. Perhaps more important, it should permit more confident guidance of students in relation to a wide spectrum of possible insti-
tutions, curricula and courses. It will permit colleges to provide student profiles covering a broader array of characteristics, including non-academic aspects of the college climate. It will be fiendishly difficult to manage!

By managing, I do not mean primarily the administrative activities of the Board and ETS, complicated as these activities may be in a multiplex system. I am thinking more about the problems posed to the student and counselor, who must find their way to the tests most relevant to them. The examining agencies have a responsibility to make this process as easy as possible, and this is a matter in which we should invest a substantial research effort. But it would be easy in dwelling on the difficulties to obscure the gains. Students are managing, with a little help, to find their way now among the PSAT, SAT, Achievements, Advanced Placement and the rest. With reasonable added help, they will, I suspect, find their way in a more extended spread of examination offerings. If you have a good road map and a little advice, you can find your way in some very strange territory.

Without expanding further on the multiplex external program, I'd like to move on to the second proposed stage.

2. The school-based program

To my mind, the ideal system is one in which the information needed not only for instruction and guidance, but also for college admission and placement, is gathered, organized, and used in the school. By the time a person reaches college-going age, he has obviously built up a very substantial record of accomplishment, differentiated according to the abilities and interests he has already displayed in school — that is, those that are relevant to him. The one-day external examination should, under the circum-
stances, be redundant.

It seems to me that we should work toward a system for collecting, on a regular basis, information about each child's performance as he passes through the school system, a record of his accomplishments in terms of grades and standardized tests, his interests, extra-curricular activities, and so on. We need to find a way of organizing this information efficiently, of expressing it in an unambiguous language that can be communicated to other people and manipulated rationally, of storing this information and retrieving it rapidly for use in pupil guidance within and between educational levels, including grades 12 and 13. As everyone here is well aware, we are rapidly developing the means of doing just that. The data bank movement is proceeding apace, and communication both at a given educational level and between levels seems likely to improve with improving technology. The gap between grades 12 and 13 seems likely to shrink to the dimensions of the present gap between junior and senior high. In these circumstances, the formidable mechanisms of national examinations may come to seem cumbersome indeed -- vestiges of an earlier era, when college-going was the exception rather than the rule.

I suspect that such a school-based system of information storage and retrieval will tend to grow up first within large educational aggregates that look upon themselves as being in some way homogeneous. Typically, the unit may be a statewide system within which a data bank would be organized and maintained. Moreover, one might well expect that state or community colleges, which tend to draw most students from a known area and which emphasize guidance and placement above selection, would be the first to find such a system as good as, and probably much better than, one relying heavily on external examinations. If the signs
are not misleading me, in a few years that description will fit the colleges to which the overwhelming majority of students will be going.

The problem will be more difficult for the highly selective institutions, which, in Richard Pearson’s phrase, are heavily in interstate commerce. They have to be able to place their faith in the common meaning of records from different areas of the country and therefore will be most concerned with the adequacy of any system for calibrating school records that would succeed the common currency now provided by the Board scale. They may well have a need for an external testing program for some time after other institutions have been able to dispense with it. It is possible, however, that they will be able to rely on and compare scores from any one of several large educational regions, and require external examinations only for students from other areas of the country.

I believe the conditions necessary for a workable school-based system are within reach and that we should take steps to hasten their attainment.

What are those conditions? First, that there be an adequate supply of good tests and other measures of performance available to and used by the schools. This condition is well on its way towards being met. The missing element is a way to express the results of both standardized tests and school performance in terms meaningful to post-secondary education, in a language at least as well understood, let’s say, as the College Board scale.

The second condition is that school guidance be strong enough to bear the weight of a school-based system. And while guidance may still be far from perfect, it is improving rapidly. In the best systems the guidance program is already perfectly capable of operating a school-based system exceedingly well and in a few years, it may well
be quite generally adequate to this complex task.

The third condition is that college admissions officers place increased reliance on the school guidance officers. This condition is likely to develop as a consequence of demonstrably better guidance, especially as one looks at the within-state movement of students.

I see the school-based system as evolving out of the multiplex external exam system -- or, if you will, as absorbing the multiplex exam system. When we have developed a wide variety of good exams, oriented toward the most diverse set of post-secondary opportunities, designed explicitly for guidance and placement, I think the schools will want to use them early for their own purposes, just as they have wanted to use the PSAT. When a common scaling procedure allows the colleges to interpret the results, I think they will be ready to accept the scores on such tests without the requirement of a separate, validating set of scores from an externally administered program.

One benefit I see flowing from such a system is that the emphasis on a long-term record in which test results are embedded will reduce the pressure that now builds up around the single, dramatic moment of the external test. In guidance and also in college admissions, to paraphrase the title of a talk made some years ago by B. Alden Thresher, the testorama will give way to the silent search.

I will take time in this paper only to foreshadow the third stage of providing tests that have maximum relevance to the individual student: viz.,

3. The student-based system

The provision of whole examinations geared to the preparation level of the individual student is a long step forward, since it makes the test contingent on the student's
background. The next step should be to provide examinations in which the individual questions are contingent on the student's responses to previous questions. If you will permit the computer to raise its ugly tapes, I would like to put forward the prospect of an examination in which each examinee's sequence of questions is determined by his response to items earlier in the sequence. The questions will be selected to provide the individual student with the best opportunity to display his own profile of talent and accomplishment, without wasting time on tasks either well below or well beyond his level of developed ability along any one line.

Looking further down this same path, it is not hard to foresee a time when such tailor-made tests will be part and parcel of the school's instructional sequence; when the results will be accumulated and displayed regularly as a basis for instruction and guidance; and when the pertinent elements of the record will be banked as a basis for such major choice points as those involving selection of a new institution.

The three stages I have somewhat artificially created represent, to me, steps toward relevance in testing: relevance to the educational program and to the individual student -- his past accomplishments, and his future plans -- in an age when diversity and change rather than uniformity and stability will increasingly dominate the educational scene. If the College Board should choose to follow paths such as these, it has ahead of it the enormous job of transforming drastically the nature of its activities, not to mention those of my own institution. Such a transformation is never easy to contemplate. Even the best and most selfless institutions inevitably tend to crystallize around their procedures rather than around their purposes. But the Board has always seen that the real business
of educational institutions is to find the best solutions to educational problems rather than to perpetuate the solutions of the moment. That, Mr. Chairman, is what I assume this Commission is all about.