"Productivity" of Undergraduate Institutions

New analyses show that a college's output of doctors of philosophy depends largely on its input of students.

Alexander W. Astin

The role played by the undergraduate college in stimulating its students to go on for the Ph.D. degree has been the subject of considerable controversy (*1*). Although the original conclusions of Knapp and his associates (*2*) on Ph.D. "productivity" have been questioned in several later studies (*3, 4*), their early work was significant in demonstrating one important fact: undergraduate colleges differ widely in the proportion of their graduates who eventually obtain Ph.D. degrees.

Most of the controversy has been concerned with the *interpretation* of these institutional differences in output of Ph.D.'s. Some investigators have concluded that highly productive institutions foster or encourage eventual attainment of this degree among their students more than less productive institutions do; others have suggested that an institution's Ph.D. output may be mainly a function of the kinds of students it recruits in the first place. Holland (*3*), for example, has pointed out that the "high productivity" colleges cited by Knapp *et al.* without exception recruit higher proportions of Merit scholars than the "low productivity" colleges. In a subsequent study, Thistlethwaite (*5, 6*) found that, even after corrections are made for the ability level of the entering students, marked differences in institutional "productivity" remain. In a sample of 36 colleges, he found that these differences were closely related to certain aspects of the college environment or "press." More recently, however, I have shown (*4*) that the output rates of these 36 colleges are related to the percentage of entering students who plan to major in natural science and the percentage who aspire to the Ph.D. degree, as well as to the academic ability of these incoming students. Moreover, when the effects of these two additional student input characteristics were partialed out, the correlations between college "press" and "productivity" either disappeared entirely or were considerably reduced. In fact, a recent study by Holland and me (*7*) demonstrates that much of what has been called the college "press" can be predicted from a knowledge of certain personal and intellectual characteristics of the students enrolling at the institution.

These latter studies highlight the need for a study of college "productivity" which takes into account differences among institutions in their student inputs.

The study described here uses a new method for assessing Ph.D. productivity—one which, it is hoped, corrects for most of the relevant differences in the kinds of students recruited by different colleges. In addition, this method is essentially longitudinal rather than cross-sectional, thereby eliminating the retrospective assumptions necessary in most of the previous work.

To begin with, the term *productivity* was redefined. The proportion of bachelor's degree recipients who eventually obtain the Ph.D. is, strictly speaking, not a measure of a college's productivity; it is, rather, an output rate. This distinction can be made clearer by an analogy from physical science. In the field of mechanics, the productivity or "efficiency" of a machine cannot be determined solely in terms of the energy or work produced. On the contrary, efficiency is defined as a *ratio* between energy output and input. Obviously, machines with identical outputs can have quite different efficiencies, depending upon their inputs. In the same way, then, the efficiency or productivity of the college as a generator of Ph.D.'s must be viewed in terms of a relationship between its student output and its student input.

In previous work the proportion of students who obtain the Ph.D. has been chosen as the measure of the college's student output. If we are to compare this output rate with an input measure, it is necessary to express student input in equivalent units. The general approach to this problem, which is described in the following sections, has been to express student input in terms of an *expected* output (that is, an expected proportion of Ph.D.'s). The "productivity" of a college, then, is the ratio between its actual output and its expected output.

Sample of Institutions

A total of 265 colleges and universities were selected for study. The two criteria used in selecting each institution were (i) that a reliable estimate of the average ability of its entering students was available (*4, 7*), and (ii) that not more than 5 percent of its undergraduates had major fields designated as "unclassifiable" (this latter criterion is explained below). The bias in this method of sampling tended to exclude very small institutions and institutions with extremely low levels of student ability.

Determining Actual Outputs

Actual Ph.D. outputs were calculated from National Research Council tables which give the undergraduate college origins of persons who received Ph.D.'s in 1957, 1958, and 1959 (*8*). The output rate for each institution was determined by dividing the total number of Ph.D.'s obtained by its graduates during the period 1957–59 by the total number of bachelor's degrees awarded during the years 1951, 1952, and 1953 (*9*) (these years represent approximately the median years of graduation from college for the 1957–59 Ph.D. recipients). These actual Ph.D. output rates ranged from 0.000 to 0.229.

The author is affiliated with the National Merit Scholarship Corporation, Evanston, Illinois.

Table 1. Data on Ph.D. and bachelor's degree recipients. To determine the probability of obtaining the Ph.D. degree with a given baccalaureate major field, by sex, values in columns 2 and 4, respectively, are divided by values in columns 3 and 5, to give the values in columns 6 and 7.

Major field	Ph.D. recipients, 1958 (No.)		Bachelor's degree recipients, 1952 (No.)		Probability of obtaining Ph.D.	
	Males	Females	Males	Females	Males	Females
Business	154	10	32,844	5,288	.005	.002
Chemistry	1,005	72	5,717	1,102	.176	.065
Mathematics	253	29	3,389	1,332	.075	.022
Psychology	516	88	3,789	2,839	.136	.031

Actual output rates were also computed separately for two broad Ph.D. fields: natural science (NS) and arts, humanities, and social science (AHSS). Actual outputs in natural science ranged from 0.000 to 0.213; in arts, humanities, and social science they ranged from 0.000 to 0.092.

Determining Expected Outputs

Two basic questions must be answered in determining a college's expected output of Ph.D.'s. First of all, what characteristics of students entering college are predictive of their eventual level of academic attainment; and, secondly, what are the *magnitudes* of the relationships between these characteristics and the eventual level of academic attainment? Fortunately, there are several recent studies which provide evidence for answering these questions. Harmon (*10*) has shown a marked relationship between a person's intelligence and the probability of his obtaining the Ph.D. degree. Several studies conducted at the National Merit Scholarship Corporation (*4, 7*) suggest that, in addition to intelligence, the student's sex and choice of undergraduate major field also affect the probability of his eventually obtaining the Ph.D. On the basis of these studies, the decision was made to compute "expected Ph.D. outputs" for each college on the basis of the intelligence, sex, and major fields of its students.

National Research Council data on the high school and college backgrounds of 1958 Ph.D. recipients (*8*) were used in calculating the expected outputs. These data provided breakdowns on the intelligence, major fields in college, and sex of those people who were awarded Ph.D.'s by universities in the United States during 1958. The probability that an undergraduate of a given sex and a given major field in college would eventually obtain the Ph.D. was estimated as follows: since the median year of graduation from college for these 1958 Ph.D.'s was 1952, U.S. Office of Education tables of "earned degrees" for 1952 (*9*) were consulted to determine the total numbers of bachelor's degrees awarded by all colleges in the United States, by sex, in each major field. By combining these data with the data on sex and undergraduate major fields of the Ph.D. recipients, it was possible to form a two-way table which showed the probability of obtaining a Ph.D. for a person of either sex with a given baccalaureate major field. This procedure is illustrated in Table 1, for four of the undergraduate major fields.

It may be seen from Table 1 that the student's undergraduate major field and sex are major factors affecting his chance of obtaining the Ph.D. Males with undergraduate majors in psychology, for example, are about 68 times more likely to obtain Ph.D.'s than females with undergraduate majors in business. The procedure followed to obtain the probabilities of Table 1 was also followed to obtain separate probabilities of obtaining the doctorate in natural science and in arts, humanities, and social science. For example, a male with an undergraduate major in chemistry has a .168 chance of obtaining a Ph.D. in a field of natural science but only a .008 chance of obtaining one in a field of the arts, humanities, and social science.

The same breakdown, by sex and major field (Table 1, columns 4 and 5), that was obtained for the 1951, 1952, and 1953 graduates was obtained separately for each of the 265 undergraduate institutions. The initial expected Ph.D. output for each college was arrived at by multiplying the number of its graduates in each sex-times-major-field cell by the corresponding probability for that cell (Table 1, columns 6 and 7) and summing the products. Three initial expected outputs (total, NS, and AHSS) for each institution were computed in this manner.

The final step in determining expected Ph.D. outputs was to correct these initial estimated outputs for the ability level of the student body (the method used to determine these expected outputs is similar to that used for determining outputs based on sex and major field). Since the overall probability of obtaining the Ph.D. for all 1952 recipients of the bachelor's degree was .0241 (8011 Ph.D.'s in 1958 per 331,924 graduates in 1952), the objective was to obtain for each college an ability correction factor $X/.0241$, where X is expected output based only on the ability level of its students.

Harmon (*10*) has divided the 1958 Ph.D. recipients into 12 groups by level of intelligence. By estimating the numbers of 1952 bachelor's degree recipients falling into these same 12 I.Q. intervals (*11*), the probability that a graduate in a given I.Q. interval will obtain the Ph.D. was determined. Then, by estimating the proportions of students at a given college A who fall into these I.Q. intervals (*12*), an ability correction factor $X_A/.0241$ for that college was computed as follows:

$$\frac{X_A}{.0241} = \frac{\sum_{i=1} (p_i A_i)}{.0241}$$

where p_i is the probability of obtaining

Table 2. Correlations between expected and actual Ph.D. outputs for six types of institutions.

Type	Institutions (No.)	Correlation between expected and actual outputs: Natural science	Arts, humanities, social science	Total
Technological institutions	14	.99	.69	.98
Men's colleges	19	.44	.76	.66
Men's universities	11	.90	.87	.93
Women's colleges	27	.58	.78	.77
Coeducational colleges	97	.70	.70	.81
Coeducational universities	97	.72	.47	.62
All institutions	265	.86	.62	.78

the Ph.D. for a bachelor's degree recipient in the *i*th I.Q. interval; A_i is the proportion of students of college A that fall in the *i*th I.Q. interval; and .0241 is the median probability of obtaining the Ph.D. for all bachelor's degree recipients.

Note that if college A's X value (expected output rate as predicted from the ability level of its student body) were 0.0482, its ability correction factor would be 0.0482/.0241, or 2.0. This means that, on the basis of ability alone, we would expect college A's graduates to obtain Ph.D.'s twice as frequently as the typical college graduate. For the 265 institutions, these ability correction factors ranged from 0.46 to 2.84.

To obtain final expected Ph.D. outputs, each institution's initial (from sex and major field) expected output rates —total, NS, and AHSS—were multiplied by the ability correction factor for that institution. The ranges of these final expected Ph.D. outputs for the 265 institutions were as follows: total, from 0.003 to 0.201; NS, from 0.000 to 0.184; and AHSS, from 0.001 to 0.093.

Predicting Actual Output

Table 2 shows the Pearson correlations between expected and actual Ph.D. outputs for each of five types of institution. The substantial size of most of the correlations suggests that much of the variation among undergraduate institutions in Ph.D. output is a function of student input. Technological institutions and men's universities, in particular, appear to have Ph.D. outputs which are highly predictable (.98 and .93, respectively) from student input. The poorest predictions occur in the case of men's colleges and coeducational universities (.66 and .62, respectively).

Figures 1 to 5 show the relationships between actual total Ph.D. output and expected total Ph.D. output for each of the five types of institution. (Coeducational colleges and women's colleges were combined in a single plot, Fig. 4, since differences in correlation coefficients and in regression coefficients between these two categories were not significant.)

In each of these figures the actual regression line is shown; a 45-degree line would represent a productivity ratio (a ratio of actual to expected Ph.D. output) of 1.00. Institutions falling to the left of a 45-degree line would be "overproductive" (that is, their actual output would exceed their expected output), whereas institutions falling to the right of the line would be "underproductive." Of special interest is the suggestion that men's institutions (Figs. 2 and 3) are generally underproductive. Technological institutions, on the other hand, tend to be overproductive.

The slopes of the regression lines in Figs. 1 to 5 were compared by means of an analysis-of-variance test for homogeneity of regression. The mean slope (regression coefficient) for the 265 institutions was .81. The test revealed that two of the subgroups, technological institutions (Fig. 1) and men's colleges (Fig. 2), had slopes which differed significantly ($p < .01$) from .81. In the case of technological institutions, the slope was significantly steeper (1.13), whereas with men's colleges it was significantly flatter (.37). Men's universities also tended to produce a flat-

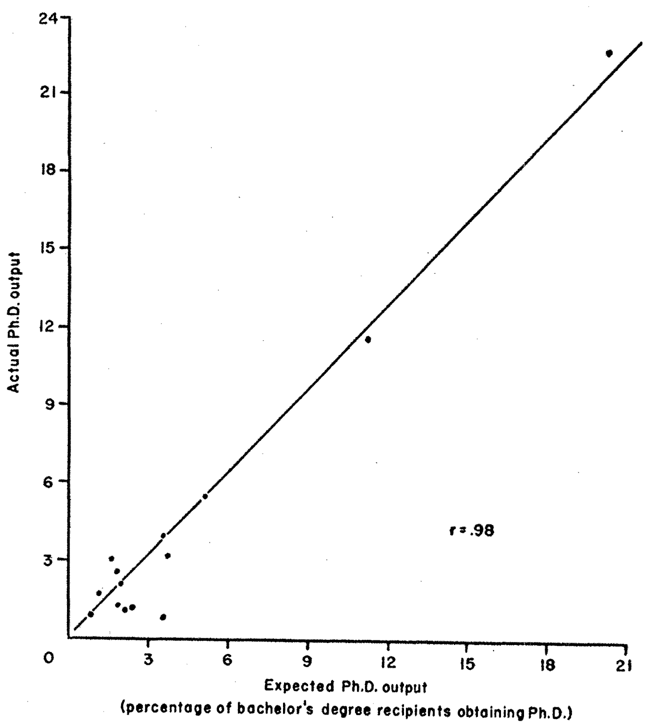


Fig. 1 (above). Relationship between expected and actual Ph.D. outputs for technological institutions ($N=14$).

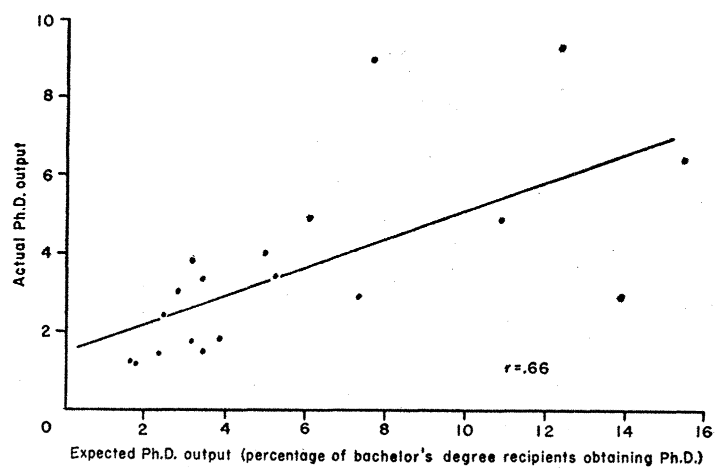


Fig. 2 (above right). Relationship between expected and actual Ph.D. outputs for men's colleges ($N=19$).

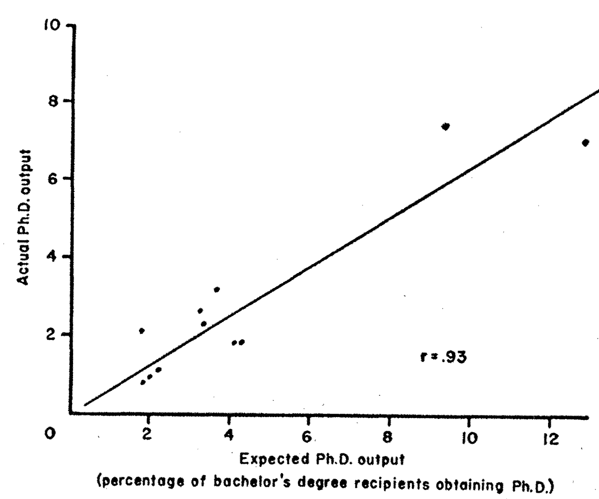


Fig. 3 (right). Relationship between expected and actual Ph.D. outputs for men's universities ($N=11$).

Table 3. Comparison of the productivity of six types of institutions: $\chi^2 = 15.971$; $p < .01$.

Type	Over-productive (ratio > 1.00) (No.)	Under-productive (ratio < 1.00) (No.)
Technological institutions	9	5
Men's colleges	3	16
Men's universities	1	10
Women's colleges	14	13
Coeducational colleges	46	51
Coeducational universities*	48	48
All institutions	121	143

* One coeducational university, which produced a productivity ratio of exactly 1.000, has been omitted.

ter slope (.64), but this did not differ significantly from .81 ($p > .05$).

These differences in slope suggested that there might be significant differences among the types of institutions in their productivity ratios. To test this notion, institutions in each of the original six groups were classified as either "overproductive" (ratios greater than 1.00) or "underproductive" (ratios less than 1.00) (Table 3). It appears from Table 3 that men's institutions—colleges as well as universities—tend to be underproductive. This conclusion was confirmed by the chi-square test of the cells in Table 3, which proved to be significant ($p < .01$). Removal of men's universities from Table 3 does not reduce the chi-square to nonsignificance, and removal of the men's colleges reduces it only to borderline significance ($.10 > p > .05$). There is, then, the strong suggestion that both groups of men's institutions in this sample are significantly underproductive of Ph.D.'s.

It might be argued that the apparent underproductivity of the men's institutions reflected some artifact in the calculation of expected outputs—that is, perhaps the correction for sex was somehow overestimated. But three other findings contradict this interpretation: (i) women's colleges did not tend to be overproductive; (ii) no relationship was found between productivity and the sex ratio of the student bodies at coeducational institutions (this is discussed later); and (iii) the technological institutions, which are attended almost exclusively by men, tend to be somewhat *over*productive.

Table 4 shows results for the overproductive and underproductive colleges arranged by geographic regions (the men's institutions have been omitted, because they are concentrated almost exclusively in the northeastern states). There is some tendency for colleges located in the Middle West and the Middle Atlantic states to be overproductive, but the trend is not significant ($\chi^2 = 6.487$, $p > .05$).

Table 4. Productivity of colleges by geographic region (men's institutions have been omitted): $\chi^2 = 6.487$; $p > .05$.

Region	Over-productive (ratio > 1.00) (No.)	Under-productive (ratio < 1.00) (No.)
New England	8	10
Middle Atlantic	23	18
South	22	29
Middle West	40	26
West	24	34
Total	117	117

To investigate geographic factors in the productivity of men's institutions, the men's universities and colleges (not including technological institutions) were grouped into two regions: Northeast and Other regions (Table 5). The chi-square value in Table 5, with correction for continuity, is significant ($\chi^2 = 4.338$, $p < .05$), suggesting that men's institutions located in the northeast tend to be more underproductive than men's institutions located elsewhere.

The final gross institutional correlate of productivity to be investigated was the type of control. For this purpose, institutions were classified as being under public, private (nonsectarian), Protestant, or Catholic control. Technological institutions, all of which are nonsectarian, and men's institutions were omitted. Table 6 shows the remaining institutions classified by type of control. It is apparent from Table 6 that private institutions, whether nonsectarian, Protestant, or Catholic, tend to be much less productive than public institutions. The significant chi-square test confirms this conclusion ($\chi^2 = 12.264$, $p < .01$). When public institutions are removed from Table 6, the chi-square drops to nonsignificance ($p > .05$).

Why public institutions should tend to produce more than their expected proportions of Ph.D.'s is difficult to explain. It might be that, because of the relatively high attrition rate among students at certain public institutions, we have underestimated the ability level of the graduates and thereby underestimated the expected outputs of these institutions. However, since we failed to find any relationship between productivity and our estimates of student ability (this is discussed later), this explanation does not appear to be appropriate.

Another possibility is that there are relevant student-input variables other than sex, major field, and intelligence which differentiate students who choose public institutions from students who choose private institutions. Holland (*13*) has shown, for example, that, even among a homogeneous group such as National Merit finalists, students choosing private institutions differ from those choosing public institutions with respect to several personal and socioeconomic variables. Whether these variables are related to later academic attainment is, of course, not known.

College Environments and Productivity

Even though the correlations between expected and actual Ph.D. output were generally high, the fact that most of them were considerably less than unity suggests that the characteristics of different college environments may account for the discrepancies between actual and expected outputs.

In a previous study (*5, 6*) 36 institutions were classified as "most productive" and "least productive" in both natural science and arts, humanities, and social science. Students attending these institutions were asked to rate their faculty on a number of "faculty behavior" items. Since differences were found with respect to several of these items between the faculties at "most productive" and those at "least productive" institutions, it was concluded that "the teacher plays an important role in motivating talented undergraduates to seek advanced degrees" (*5*, p. 75).

In order to evaluate these earlier findings on the basis of the newer productivity rates (rates based on the ratio of actual to expected output), these same 36 institutions were reclassified as overproductive in natural science (actual NS output greater than expected NS output) or as underproductive in natural science (actual NS output less than expected NS output). They were also reclassified as either over- or underproductive in the arts, humanities, and social science. It was found, after re-

Table 5. Productivity of men's institutions in the northeastern states and in other regions: χ^2 (with correction for continuity) $= 4.338$; $p < .05$.

Region	Over-productive (ratio > 1.00) (No.)	Under-productive (ratio < 1.00) (No.)
Northeast	0	18
Other regions	4	8
Total	4	26

classifying these 36 institutions, that in no instance were the differences between members of pairs of over- and underproductive groups significant ($p = .05$) with respect to the 24 items (10 for natural science, 14 for arts, humanities, and social science) previously reported to be related to productivity. As a matter of fact, even the *direction* of the over-under discrimination was reversed in 14 of the 24 comparisons. It would thus appear that the conclusions of this earlier work with regard to faculty "influence" are no longer tenable.

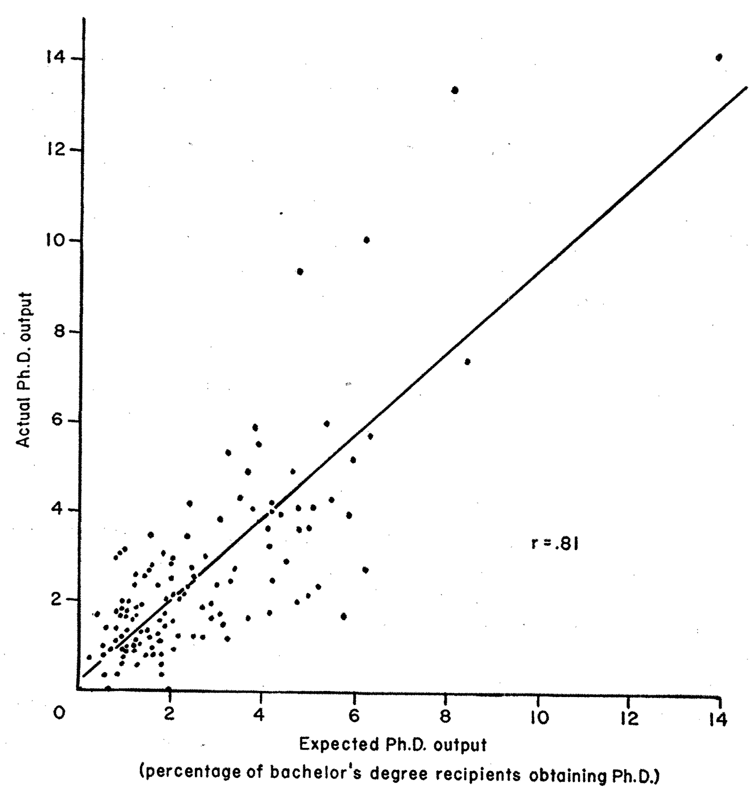
Next, an attempt was made to relate differences in productivity ratios to a variety of characteristics of the colleges. From the entire sample of 265 institutions, the most overproductive and the most underproductive institutions were identified. These two extreme groups were then examined to find pairs of institutions (one overproductive and one underproductive) which had similar *expected* Ph.D. outputs. Matches were made only *within* subgroups previously found to show differences in productivity—that is, men's colleges were matched only with other men's colleges, public institutions only with other public institutions, technological institutions only with other technological institutions, and so on. Within these limits it was possible to identify 35 matched pairs of institutions. The 35 overproductive institutions were then compared with the 35 underproductive institutions with respect to 25 institutional characteristics by means of t ratios for correlated means. These 25 characteristics included special measures of the college "environment" (*7*), several indexes of the college's financial status (tuition, endowment, research funds, and so on) (*4*), and such miscellaneous characteristics as ability level of the student body, rate of growth, size of library, and faculty-student ratio.

It was found that for only two of the 25 college characteristics were differences between the overproductive and the underproductive institutions significant ($p < .05$). Overproductive institutions were shown to have higher faculty-student ratios (fewer students per faculty member) and to charge *lower* tuition fees. It should be stressed that these results may well have occurred by chance (it is not improbable that, from sampling fluctuation alone, two of 25 t ratios will exceed the .05 level).

Table 6. Productivity by type of institutional control (men's institutions and technological institutions have been omitted): $\chi^2 = 12.264$; $p < .01$.

Type of control	Over-productive (ratio > 1.00) (No.)	Under-productive (ratio < 1.00) (No.)
Public	60	37
Private (nonsectarian)	14	27
Protestant	25	37
Catholic	8	11
Total	107	112

A careful inspection of the most overproductive and the most underproductive institutions from the original sample of 265 revealed several small, homogeneous groups of institutions which are of some interest. For example, among the most overproductive institutions were four colleges located in New York City: Brooklyn College, City College of New York, Queens College, and Yeshiva. Also, all three institutions located in the state of Utah (Brigham Young, Utah, and Utah State) were highly overproductive. It is difficult to determine whether the overproductivity



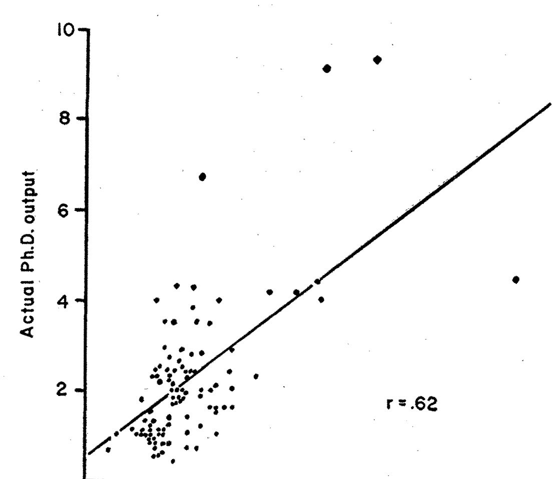


Fig. 4 (left). Relationship between expected and actual Ph.D. outputs for coeducational and women's colleges ($N = 124$).

Fig. 5 (above). Relationship between expected and actual Ph.D. outputs for coeducational universities ($N = 97$).

of these groups can be attributed to the effects of the institutions themselves. Certain ethnic or religious characteristics of the students entering these institutions may be important factors in the colleges' productivity. That is, there may be motivational factors associated with certain cultural or religious groups in addition to the factors of sex, college majors, and intelligence level.

A rather unexpected finding was the uniform underproductivity of the eastern men's colleges reported earlier. Six of the eight most underproductive (Dartmouth, Hamilton, Harvard, Princeton, Wesleyan, and Williams) were among Knapp and Greenbaum's original 50 "top-ranking" institutions in "productivity." Why these institutions should produce consistently fewer than their expected numbers of doctorates is difficult to explain. However, when we examine overproductive institutions (Antioch, California Institute of Technology, Oberlin, Reed, Swarthmore, Wabash) which have expected outputs similar to those of the underproductive six, some possible interpretations are suggested. In the first place, it may be that graduates of eastern men's colleges are more frequently and earlier recruited into business and professional careers, so that many are diverted from the pursuit of graduate academic degrees. (This interpretation appears especially tenable when one realizes that representatives of business and industry are probably much more aggressive in recruiting the graduates of high-prestige eastern men's colleges than in recruiting graduates of other institutions.) Secondly, there may be uncontrolled factors in student input—factors not reflected in sex, major field, and intelligence—which contribute to these discrepancies. For example, the attainment of the Ph.D. may not represent the same degree of personal achievement or social mobility to the typical student at these eastern men's colleges that it does to the student of comparable ability enrolling at other institutions.

Discussion

It is clear from these results that the actual Ph.D. outputs of a variety of undergraduate institutions can be predicted with substantial accuracy from a knowledge of the abilities, major fields, and sex ratios of the student bodies. This fact, together with the failure to find any consistent relationships between a college's productivity and specific characteristics of its environment, suggests that colleges do not differ appreciably in the extent to which they stimulate their students to seek higher academic attainment or inhibit them from seeking such attainment. (Exceptions to this are, of course, the public institution and the eastern men's college, but here again it may be that we have not yet recognized and made allowance for all the relevant differences in student input at institutions of these types.)

Table 7. Reliability of the actual Ph.D. output rates as compared with the expected-actual correlations in six types of institutions.

Type	Correlation of 1957–59 actual Ph.D. output with: Expected output*	Actual output, 1951–56 (reliability)
Technological institutions	.98	.97
Men's universities	.93	.95
Coeducational colleges	.81	.87
All institutions	.78	.85
Women's colleges	.77	.84
Men's colleges	.66	.81
Coeducational universities	.62	.73

* From Table 2, column 5.

If we assume, then, that Ph.D. output is primarily a matter of student input, rather than of special college influence, how can we account for the varying degree of success with which actual output can be predicted from expected output from one type of college to another (Table 2)? One possibility lies in the relative *reliability* of the actual output rates used (for some colleges, particularly those with very low output rates, a difference of only one or two actual Ph.D.'s can drastically alter the actual output rates). If the reliability of these actual output rates is indeed a factor in the success with which output can be predicted, we would expect to find the most reliable output rates in those subgroups of institutions where the correlation between expected and actual output is highest.

To explore the validity of this expectation, the 1957–59 output rates used in this study were correlated with the 1951–56 output rates used in previous studies (*4–6*). The reliability estimates are presented in Table 7, along with the original correlations between expected and actual outputs. It is evident from Table 7 that the more reliable the actual output rate is, the more accurately it can be predicted from expected output rates (the rank correlation between the two sets of correlations in Table 7 is 1.00). It seems safe to conclude that differences in the reliability of the actual rates contribute to differences in the degree of success with which predictions can be made from expected rates (*14*).

Other factors which tend to reduce the expected-actual correlations are errors in the expected output rates (*15*) and, as mentioned previously, the possibility that there are uncontrolled student-input variables. When these two factors, as well as the evidence on the reliability of the output rates, are taken into consideration, the correlations shown in Table 2 can be regarded as conservative estimates of the true relationship between student input and Ph.D. output.

These results suggest that it might be well to consider abandoning "Ph.D. productivity" as a useful criterion in comparing or evaluating the effects of the undergraduate institution on the student's motivation. This is not to say, of course, that attainment of the Ph.D. is an unimportant variable, but simply that there appears to be little reliable residual variance in productivity which one might attribute specifically to the influence of the college. Moreover, there are several other criteria of later student achievement (for example, teaching effectiveness and research competence) which may prove to be more dependent on the influence of the college than Ph.D. productivity is.

It should be stressed that we are speaking here of the *gross* influence of institutions. It is probable that *within* a given institution there are certain professors, student associates, and particular experiences which encourage the individual student to attain the Ph.D., or inhibit him from attaining it. When the overall productivity of the institution is evaluated, these experiences could well counterbalance one another and thereby not be reflected in a gross output measure. An intensive study of these intracollege influences on the student may prove to be much more rewarding than a study of gross influence as a means of learning what specific educational practices affect the student's motivation to seek graduate training.

As an alternative to an experimental study of the effects of colleges (such a study is, of course, not feasible since students cannot be randomly assigned to colleges), the method used in this study would appear to be useful in arriving at inferences about effects of the college on the student. One of the drawbacks of methods used previously (*4–6*) is that the corrections for student input were based on correlations in which the institution was used as the unit of sampling. When real differences in institutional effects are assumed to be present, this correlational procedure tends to partial out the institutional effect along with the input variables ("to throw out the baby with the bath water," so to speak). In contrast, in the approach used in this study an actuarial method is used for weighting student-input characteristics, whereby students are pooled without respect to the institutions they attend. An actuarial table of this nature can accommodate any number of input variables (sex $\times$ major field $\times$ intelligence $\times \ldots n$), as long as there is a sufficient number of students in each cell to produce stable probabilities. Note that, under these conditions, institutional effects (if they exist) would not be partialed out. The major problem in this (or any other) method is, of course, that of insuring that we have included in the actuarial table all input variables which might reasonably be expected to relate to output.

Summary

A new method for evaluating the "Ph.D. productivity" of undergraduate institutions was applied to a sample of 265 institutions. It was found that a college's actual Ph.D. output can be predicted relatively accurately from an "expected" output based on the sex, major fields, and intelligence level of its students. Public institutions were found to be significantly overproductive, and eastern men's colleges and universities were found to be significantly underproductive. Previous findings indicating that the faculty had a causative effect on productivity were not confirmed. These results suggest that Ph.D. productivity may not be a sensitive measure of the effectiveness of undergraduate institutions.

References and Notes

1. The study discussed in this article is part of the research program of the National Merit Scholarship Corporation and was supported by a grant from the National Science Foundation. I am indebted to Nancy Bailey for her assistance in all phases of this project, and to A. V. Astin for suggesting the physical-science analogy.
2. R. H. Knapp and H. B. Goodrich, *Origins of American Scientists* (Univ. of Chicago Press, Chicago, 1952); R. H. Knapp and J. J. Greenbaum, *The Younger American Scholar: His Collegiate Origins* (Univ. of Chicago Press, Chicago, 1953).
3. J. L. Holland, *Science* **126**, 433 (1957).
4. A. W. Astin, *J. Educ. Psychol.* **52**, 173 (1961); ——— and J. L. Holland, *Coll. and Univ.* **37**, 113 (1962).
5. D. L. Thistlethwaite, *Science* **130**, 71 (1959).
6. ———, *J. Educ. Psychol.* **50**, 183 (1959).
7. A. W. Astin and J. L. Holland, *ibid.* **52**, 308 (1961).
8. I am indebted to Lindsey Harmon of the National Academy of Sciences–National Research Council for making available, prior to publication, tables of 1957–59 Ph.D.'s and for providing detailed data on the backgrounds of 1958 Ph.D. recipients.
9. "Earned degrees conferred by higher educational institutions," 1950–51; 1951–52; 1952–53, *U. S. Office Educ. Circs. 333, 360,* and *390* (Washington, D.C., 1952–54). It would be more appropriate to base an "input" measure on the major fields of the students as they *enter* college, rather than on their final baccalaureate field, but no such data are generally available. However, a recent study by J. A. Davis and N. Bradburn, *Great Aspirations: Career Plans of June 1961 College Graduates* (National Opinion Research Center, Chicago, 1961) does provide data on the initial and final major fields of the 1961 graduating classes at 81 of the 265 institutions used in the present study. To test the relationship between initial (entering) and final (degree) fields of the students at these institutions, two "expected" Ph.D. outputs were computed: one based on the initial major fields of the students, and one based on their final major fields. Despite the fact that nearly half of these students actually changed fields between their freshman and senior years, the correlation between the two expected Ph.D. outputs was .97. This finding suggests strongly that expected Ph.D. outputs based on the initial major fields of the students would not be appreciably different from expected outputs based on their final major fields. (I am indebted to James A. Davis and Norman Bradburn of the National Opinion Research Center for furnishing the data for this analysis.)
10. L. R. Harmon, *Science* **133**, 679 (1961).
11. The average intelligence quotient (AGCT) of college graduates has been estimated as 121 by D. Wolfle [*America's Resources of Specialized Talent* (Harper, New York, 1954)]. If we assume this mean and a standard deviation of 20, however, we find that a graduate in the highest level of intelligence (160+) is *less* likely to get a Ph.D. than a graduate in the next-to-highest interval (150–159). Since assuming the same mean with a smaller standard deviation tended to produce similar improbabilities at other points on the distribution, we assumed a mean of 115 and a standard deviation of 20. Under these conditions, the relationship between I.Q. and probability of obtaining the Ph.D. for college graduates shows a consistently positively accelerated function similar to the one for the general population reported by Harmon (*10*).
12. The ability estimates for the student bodies at the 265 institutions were mean National Merit SQT (NMSQT) scores of samples of entering students in 1959. Several recent studies (*4, 7*) have shown that these means are highly correlated with mean Scholastic Aptitude Test (SAT) scores given to the *entire* entering student body at several of these institutions in 1956. Through a series of transformations (NMSQT → SAT → AGCT I.Q.), these mean NMSQT scores were converted into equivalent mean I.Q. scores. By assuming an I.Q. standard deviation of 20 and a normally shaped distribution, it was then possible to estimate the proportions of students at a given college falling into each I.Q. interval.
13. J. L. Holland, *Coll. and Univ.* **35**, 11 (1959).
14. We have not applied the "correction for attenuation" here, because there is no way to be sure that these "test-retest" reliability estimates reflect only random error in actual output rates.
15. Since many of the 1957–59 Ph.D. recipients obtained their bachelor's degrees in years other than 1951–53, any changes in the size, major fields, or sex ratio of a college's student body before or after these three years might have affected its actual Ph.D. rate. Such changes, however, would not have been reflected in the expected Ph.D. rate and would therefore represent sources of error in the expected-actual correlation.