

APTITUDE TESTING IN PERSONNEL PROCEDURE A. B.
CRAWFORD

Student personnel activity as a whole, whether conducted by registrar, dean, counselor, admissions or placement officer, has, I take it, the common aim of regarding each student as an individual and of trying to help him get the most out of his college experience and himself. As such it should be regarded not as a separate medium of personal development, but as part of a cooperative effort which, to be effective, must be closely integrated and coextensive with the continuous process of education itself. Its various specialized staff functions exist, in other words, not for their own ends, but to serve the inclusive purposes of education. A paper necessarily limited to a few phases only of personnel work naturally cannot attempt to trace connections with the rest of the field or their bearing upon even broader questions. Yet in considering, however inadequately, any one such topic it is well to bear in mind that what, after all, justifies and gives meaning to any particular technique or function is its relationship to the whole problem. Neither any single aspect thereof, nor personnel efforts in toto, can properly stand alone or claim independent objectives.

For that reason, if the methods used by any college for some specific purpose are to play an ultimately effective part in furthering the larger educational aims of that institution, they must fit in with those aims in general. This may seem self-evident, yet many arguments regarding the merits of this or that curriculum or guidance program lose sight of the fact that procedures appropriate for one institution may be quite out of place in another. Our colleges and universities differ far more than we sometimes realize in respect to the make-up of their student bodies, the objectives, standards, and methods of instruction, their physical and financial resources, and the respective 293

obligations they are called upon to discharge. Consequently, for many of our common problems we cannot expect to find a sure and all-inclusive answer. Personnel procedure need not always seek the one best way.

In discussing certain experiments in the selection and guidance of Yale undergraduates I, therefore, hasten to admit the limited nature of my own experience and data, as well as the possible insularity of my resulting interpretations and points of view. Just because we favor certain measures and methods, validated upon our own student body, does not mean that we regard these as equally appropriate elsewhere. For example, it seems to me a fallacy (and unfortunately rather a common one) to assume that even the most carefully developed tests possess uniform, absolute validity, inherent therein under all conditions, irrespective of the group tested. Therefore, the first point I should like to emphasize is that such procedures, even though successful on one campus, should not be uncritically applied on another without application, as well, of the proverbial grains of salt-and plenty of them.

For six years our office has been trying out various means of judging how well our candidates for admission are likely to succeed in their academic work at Yale. Naturally, other factors than those measured in the class room or examination are important in the composition of a student body. Integrity, special interests, ambition, determination, purpose, the ability to get along with other people-in short, all those things which enter into character or personality all play their part in the process of selection. Judgment and common sense must duly weigh the subjective factors which cannot be dealt with by formula. But, for these very reasons, it becomes all the more important for us to consider how to make our objective, academic data as useful and dependable as we can. Moreover, scholastic criteria in the long run furnish the safest guide to admission and, therefore, probably remain the most important for any institution primarily interested in the effectiveness of its educational efforts.

295 — Three different measures of academic promise, singly or in combination, are being widely used for this purpose. today-the school record, one form or another of entrance examination (whether of the essay or of the objective type) and so-called "intelligence" or "scholastic aptitude" tests. Each serves a valuable purpose, more or less independently of the others. All three represent attempts to test the same essential quality, but they proceed by different means.

I shall not take time to comment further on our analyses of these measures, or the methods, for example, by which we adjust data for different school groups in the light of previous experience in order to measure the records of students from hundreds of schools all over the country so far as possible on a comparable basis. Suffice it to say that we find, as do most other institutions, that a student's relative rank in class throughout the preparatory period is the most reliable of these three single measures. Previous reports of our studies have indicated that the College Board entrance examinations, as measured by comparing students' grades thereon with subsequent performance in college, fall far short of predicting either average Freshman standing, or competence in specific subjects, satisfactorily. That is; the correlations between College Board and Freshman averages in general, or between marks in a particular entrance examination and Freshman grades in the same subject, are surprisingly low. This is especially true of the latter type of comparison. Yet I believe that the College Entrance examinations still perform a distinctly valuable service, at least to such institutions as ours, both in holding schools up to uniformly high standards of preparation and in enabling us to evaluate the output of various schools, *judged as a whole*, by reasonably comparable standards. Unfortunately, and perhaps inevitably, the measurement of specific individual competence by the same process is another matter. In respect either to admission or to effective educational placement of any given student, reliance solely upon the examination results seems, in the light of our several years' study of this question, quite unwarranted.

The third factor mentioned earlier, which we regularly use in studying a candidate's entrance records, is the Scholastic Aptitude Test. Though, like other so-called "intelligence tests," it, too, can be regarded as a sort of qualifying examination, it looks forward toward a student's educational ° potentiality rather than backward toward more traditional measures of past achievement. While the aptitude it measures is itself to a considerable degree a product of formal education, this method of probing for intellectual capacity may often reveal what other methods, through the process of averaging academic grades, only obscure.

The Scholastic Aptitude Test, developed by a committee of which Professor Carl Brigham, Associate Secretary of the College Entrance Examination Board, is Chairman, has two sections-one verbal, with special significance for such academic subjects as English and history; the other mathematical. We have found that the verbal section predicts individual Freshman grades in **English and history, and** the mathematical section Freshman grades in, mathematics and science, in each instance, about twice as well as do the specific entrance examinations in those subjects.

More valuable than any of these measures-school grades, entrance examination, or the Scholastic Aptitude Test-alone, is an index which is a compound, in due proportion, of all three. Some device of this nature has been used at various institutions-first, I believe, at Minnesota and Princeton- and given various names, such as Academic Index, Bogie, and Predicted Grade or General Prediction. Technically, it is the statistically best weighted combination obtained by the method of multiple correlation. Otherwise, it may be described as the summary of several different kinds of scholastic evidence.

Our General Prediction now correlates, for the whole class, a little better than .70 with Freshman Year averages. As those familiar with correlation coefficients in academic work know, this is reasonably high. It compares for our students with correlations of .57 for school grades alone and of about .45 each for S.A.T. scores and weighted average

on the College Board examinations. Yet, though the latter thus are seen to yield lower correlations in respect to our students than do the adjusted school grades or the General Prediction, both the latter measures are themselves probably dependent to a large degree for their higher validity upon the stabilizing effect of the College Board situation as a whole.

This Prediction, in the majority of cases, is accurate within a range of 3 or 4 points on a scale of 100-i.e., the

chances are that a student whose predicted grade is 78 will not fall below 75 or rise above 81 or 82. About sixty per cent of this year's Freshmen at mid-year conformed to last summer's predictions within that range, and nearly 70 per cent within a range of five points-equivalent to one step on the marking scale, since our individual subject grades are reported on five-point intervals. However, no such measure should be regarded as better than an odds-on bet. The odds are more favorable for a prediction so determined than they are for any other single measure thus far devised. If you gamble on it, you will win more often than you will on any other system-but it still remains a betting system and every so often it will, for one reason or another, go haywire. It should, therefore, be used with common sense and discretion-just as any other test score or grade always should be-and particularly so when the different components entering into the prediction appear markedly at variance with each other.

In view of all the trouble we take in carefully adjusting and weighting school grades, College Board results, and Aptitude Test scores, why is our resulting Prediction still more or less inaccurate for a third of the class? One reason is, of course, that individuals do, for various reasons, actually change in attitude and performance, particularly as they progress from school to college courses. Some have developed more rapidly before entrance than others have. Not all have had equal advantages in preparation for college work. Still others fail to adjust themselves adequately to the new and freer environment of a university.

Then, too, our admissions procedure, on the whole, results in such a highly selected and homogeneous group that prediction of relative accomplishment within that group becomes correspondingly difficult. For example, trial at an Eastern University of two widely used tests of general intelligence (Otis and American Council) indicates that over 90 per cent of the Freshmen there ranked in the upper half of all college students, as measured by the published norms for these tests. That means a reduction of nearly half, in the range of ability group and those of many other colleges on the accredited list. That is one reason why, despite attempts to refine predictive procedures as far as possible, wholly satisfactory correlations with subsequent performance are not obtainable.

Another and still more technical cause for shortcomings in this respect is frequently overlooked. According to published data, the reliability of our own grades is no lower than it is at other institutions, yet it still leaves much to be desired. The raw or uncorrected coefficients of correlation between Freshman marks and those of the Sophomore year vary, for our three undergraduate schools, from .75 to .82. Our correlation between predictions and first year grades, now over .70, is really not much inferior to that between first and second year grades. If, despite the other causes of variation just mentioned, we can predict before entrance the Freshman averages of students we have not yet seen, almost as well as the latter will forecast their Sophomore standing, we are doing about all that can be expected under the circumstances. In other words, scholastic prediction, by tests, examinations, combinations, or other devices cannot be improved much further until our college marking systems themselves become more reliable.

Scholastic prediction derived from a combination of several measures is, therefore, everything considered, a pretty good measure. One of its most useful attributes is the

valid because it can only arise from consistent excellence on *all* counts. To be sure, extra-curricular interests, week-end parties, and other pleasant diversions frequently result in the repudiation of such promise. Still a college ought to know when this is the case. If a man who might stand in the upper tenth of his class and, with little or no effort, is coasting along in the middle, it is a pedagogical crime to regard his performance as satisfactory just because he's not in trouble. Nothing will make every such student work to capacity; but at least you can make him realize, if he

doesn't, that he is not getting away with mental murder wholly unsuspected. Indulgent toleration of continued loafing in any form, with no check-up thereon, is nothing less than a travesty on education.

Although such a measure as the General Prediction indicates which students possess distinctly superior capacity, it does not reveal the particular fields in which their talents severally lie. Nor for that matter is unusual ability, at least in some one subject, by any means limited to those of exceptional *all-around* ability. Taking academic grades as an illustration, one student may obtain marks of around 75 each in English, history, science, and mathematics. Another may barely pass, or even fail, in the first two but do 90 work in the others. Both will have a general average of 75- yet for guidance purposes they present very different characteristics. With predictions, likewise, we should be on the lookout for promise not only among those of highest *general* standing but also among others who may have some distinct aptitude, even though this is not apparent until their average standing is analyzed. The latter, in such cases, may be pulled down by weakness in certain fields which have obscured unusual capacity in others, and which neither the students nor their teachers have fully realized.

The counsel of parents or teachers, both of whom more often than not want everyone else to follow the paths which have proved most interesting to themselves; the effect of school and college curricula which, through

flexibility, emphasize certain subjects to the neglect of others; various accidental or essentially irrelevant influences—all these factors too often obscure students' appreciation of some outstanding capacities which they, all unsuspecting, may have. It is, I feel, a major responsibility of the college to search for such possibilities and to encourage their fruition.

How can this be done? Educational procedure in this respect, until recently, has been relatively laggard. Emphasis in selection, indeed, has been negative rather than positive. A first university enrolments was to raise the bars against those ill-prepared or unfit for college studies. Not nearly so much attention was then paid to the guidance of those admitted as to limitations, in one form or another, of their numbers. Selective admission really started out as selective rejection.

Of late, marked interest in the more effective direction of individual capacity and effort has been manifested by the increased use of placement examinations and test of educational aptitude. Selective placement of entering students in the light of objective evidence is now a recognized essential of thorough-going educational guidance. The usual all-around test of intelligence, though looking distinctly towards the future, is not sufficiently diagnostic to facilitate differential guidance. For example, by breaking down the separate parts of The American Council's Psychological Examination and analyzing their respective significance for certain fields of study, we raised the correlation of scores on this test with strictly academic subjects from .39 to .51 by using only two of its five sections. Since the individual validity of such measures roughly varies as the squares of their correlation coefficients, effectiveness for guidance purposes in this particular situation appears to be nearly doubled by utilization only of those parts of the whole test which show the greatest predictive value for work in liberal arts. Achievement tests, on the other hand, though more dif-

1 are *both* forward-looking and specifically diagnostic. That is what Brigham has already developed in the verbal and mathematical sections of the Scholastic Aptitude Test; and what various other investigators are working towards in respect to different branches of advanced study.

Objective measures of any sort are frequently assailed as unduly factual in nature—as not affording **sufficient scope** for thoughtful organization of facts, their bearing upon each other, and their intelligent application to general questions. There is some ground for these criticisms. As teachers like to point out,

facts to work with. But it is the use one makes of them, rather than command of the isolated facts themselves, which denotes intelligent and educated thinking. Cultivation of the latter, after all, should be the chief aim of higher education.

It is the purpose of educational aptitude tests to measure the individual's capacity for learning in this very vital sense, and particularly to discover the subjects of study in which an individual is likely to develop this capacity most effectively. In addition to finding out what students have already *acquired*, further inquiry as to how they may effectively *apply* such knowledge in new ways, and towards more advanced studies, is also necessary. School work is largely concerned with laying a foundation, with providing the tools of higher education. The imagination, judgment, and interpretation which will guide the use of those tools in carving out the higher education itself; the cultivation of that thoughtful approach to evernew problems which is the mark of a first-rate mind—these abilities should particularly be developed by the increased scope and added maturity of college experience. Testing adaptability to these new and enlarged educational opportunities, and the power for intellectual growth, is a particular function of those tests of specific educational aptitude—that is, not just of academic intelligence in a *general* sense, nor yet of past achievement alone in a particular sense; but of *differential aptitude* for this or that special field of future

endeavor. Such tests are intended to be direction finders by which an individual who does have markedly greater promise for a particular field of study-say science, or mathematics, or engineering, or literature-than for others, may take his bearings and set his course.

A boy in our present Freshman Class had planned to enter Harvard but flunked his comprehensive English examination badly. That; to Harvard, seems to be a fatal, unforgivable fault. His score on the verbal section of the Scholastic Aptitude Test confirmed his weakness in the English field. He ranked on this within the lowest two per cent of the thousands taking College Boards. But his score on the mathematical section ranked him in the upper five per cent of the same group.

On the relatively few mathematical and scientific subjects which the traditional school curriculum includes, he made grades over 90. Refused by Harvard, he decided to make the best of a bad job and apply to Yale, with the intention of electing work in the Scientific School. Our Board of Admissions, taking into account his marked promise for such a program, waived his English failure and the Dean of Freshmen assigned him to one of our best English instructors, with a note of comment on the case, and a prayer. Special attention and effort are enabling him thus far barely to get by in his English (for, even with us, scientists are supposed to be reasonably literate) ; but in his professional studies-mathematical and scientific-he had three A's at mid-year. I think that's one time this year we put something over on Harvard.

Although such cases are exceptional, they are less rare than one might suppose. About six per cent of our present Freshman Class show differences almost as significant-that is, their individual scores on the verbal and mathematical aptitude tests vary by two standard deviations or more. Nearly 70 per cent of an entire class are grouped within a corresponding range of plus or minus one standard deviation from the mean. Roughly, therefore, this means that the differential aptitudes of fifty Freshmen, in respect

AMERICAN ASSOCIATION OF COLLEGIATE REGISTRARS
303 to their promise for literary versus scientific studies, show a spread in one direction or the other equivalent to almost 70 per cent variation in their respective percentile ranking on these two measures. In addition, 60 more scored at least two standard deviations above the mean in one or the other of these tests, or in General Predictions-that is, according to at least one criterion they ranked exceedingly high either in all-around, or in some specific, promise. Thus 110 out of a class of 830-nearly one man out of every seven-evidenced, at entrance, aptitudes distinctly meriting special consideration.

In an effort to improve the facilities for effective educational guidance of these individuals, and others with less striking but still important differences in potential ability, our *office* at Yale for the last five years has conducted investigations in this field. This has been done through the medium of an experimental group of Yale Freshmen, from 130 to 150 in number each year, carefully chosen to be representative of the entire class in respect to background, preparation, academic intelligence as measured by the Scholastic Aptitude Test, and our General Prediction of expected classroom performance.

By this means we have tried out some 30 tests, and revisions thereof, in relation to later performance in different branches of study. Besides those intended to measure certain educational aptitudes, we have experimented with widely used intelligence tests, with the General Culture Examination of the Cooperative Test Service, with such personality inventories as Thurstone's and Bernreuter's, and with Strong's very useful Vocational Interest Blank. In this paper I can touch only briefly upon some aspects of the aptitude test investigations. I trust it is clear that by this term "educational aptitude" we mean relative promise for some particular branch of learning, irrespective of the amount of exposure to that field which a student may already have had.

The respective aptitudes, in education as in the vocations, to a considerable degree cut across the usual cate-

gories because they are broader in nature than can be represented by the labels ordinarily classifying divisions either of the curriculum or of occupations. Thus, besides English itself, certain subjects (for example history or sociology) so largely involve the intelligent use of English that, from an aptitude angle, the common factor of verbal facility is seemingly more important than are the differences in content of their respective courses. Thinking through the medium of verbal symbols is a characteristic of most

other hand, the mathematician deals in symbols of quite different nature. A single formula may compress into one line what several pages could not so clearly convey to him in words. A sigma, or an integral sign, or many another notation, is a sort of shorthand character by which the scientist or mathematician can express or understand the whole of some complicated, abstract relationship. He uses a language of his own and people seem to differ in their facility to acquire that language—just as they have long been observed to differ in ability to pick up a foreign tongue. Like the people in Webster's cartoons, we just don't all speak the same language.

What the nature of such aptitudes is, whether they are innate or acquired, I do not believe anybody knows. But by the time some students reach college age, these differences have developed in them to a significant degree. We must realize that the mental process often works more effectively with certain kinds of material, or uses certain symbols with greater facility, than applies, even for the same mind, to others.

One of these special aptitudes of particular importance is that for visualizing in three dimensions—a sort of spatial intelligence. This capacity seems, like musical or literary talent, to develop early and naturally, or else not so well and only with great difficulty. I speak with some feeling because for several years we have been devising tests for this spatial sense which leave me a jittering maniac. I can make them up but I can't do them myself for sour apples.

This spatial sense is essential to the architect, and to the engineer. It offers one example of relatively close correspondence between educational and vocational aptitude, because the training for those fields is more professionally specialized, almost from the beginning of college, than is the case for most others. The studies of Professors C. R. Mann at the Missouri School of Mines and Metallurgy, C. J. McCauley at the University of Arizona, and John W. Cox, of University College, London, represent exceedingly

valuable contributions to this problem.

Successful engineers and many other scientists usually combine with spatial sense an aptitude for mathematics. The traditional school curriculum is less likely to discover or develop promise of that nature, than capacity for so-called academic or liberal studies. This is because the preparatory course is naturally concerned with general subjects, most of which are essentially verbal in nature. Our investigations indicate that students not infrequently have an aptitude and interest for engineering which, before entrance as Freshmen, they had not much chance to discover.

Therefore, using a combination of school records and appropriate sections of the Scholastic Aptitude Test, we now calculate not only the General Prediction used by the Board of Admissions in selecting candidates for entrance, but two differential predictions, one for academic and the other for scientific promise. This year, the former correlated .65 with Freshman grades in English and history (our academic criterion) and the latter .64 with grades in mathematics and science. Carrying this project further, we obtained, from a combination of measures, a multiple correlation of nearly .70 with Freshman science, mechanical drawing, and mathematics, the most important prerequisites for advanced training in the engineering field. The battery yielding this result consists of McCauley's and Mann's Spatial and Brigham's Mathematical Aptitude Tests. Thus, by extending the method of combining several measures from general to specialized prediction, we can differentiate a student's respective aptitude for one or the other of these

broad divisions of upper class study with reasonable success -almost as well, that is, before entrance by this means, as we can later on from his Freshman grades.

Our work thus far has chiefly been experimental, and actual use of tests for guidance purposes has awaited further analysis of their suitability to our situation. Compared with the correlations which school and entrance examination grades alone yield with the same criteria, however, these aptitude tests give much the more satisfactory results. For example, College Board grades, on the average, correlate about .30 with Freshman work in corresponding subjects of study, as against .65 for the best aptitude test batteries. The latter, while still leaving much to be desired so far as individual reliability is concerned, therefore, appear to be three or four times more effective for this particular purpose than the time-honored entrance examinations.

So far, the undergraduate fields for which such special aptitudes or educational talents have been identified seem to be, (1) academic, literary, or other largely verbal subjects, (2) pure science, (3) mathematics, (4) subjects involving spatial (three dimensional) problems, (5) engineering (a combination largely of the two foregoing), and (6) foreign languages. Analogous tests for still other fields are also being developed-notably, on a higher level, in the measurement of aptitudes for law and medicine. We have been working for several years on a test of this nature now used by our own Law School; while allied investigations in respect to legal promise have been made at Columbia, Wisconsin, Minnesota, and elsewhere. The Moss Medical Aptitude Test, officially adopted in 1930 by the Association of American Medical Colleges, is now in general use as part of the machinery for admission to the study of that profession. Schools, too, are increasingly interesting themselves in differential prediction of future accomplishment, several of them at present working directly with various new tests, at the preparatory level, of

ceed in driving these measures, the more valuable, for guidance purposes, they should become.

In this respect it is my hunch that proper educational guidance is more important and more meaningful, so far as cultivation of the individual's highest powers are concerned, than is vocational guidance. Not all people have abilities more markedly educable in one direction than in another in fact most do not. But for those who do, even if it is only in one field, that fact itself is of primary significance. For the others, guidance on the basis of their interests, motives, and opportunities, of course, should not be neglected; but, by definition, they are not the ones for whom the choice of this or that field may make as much *difference*, either for them or for the world, as may the choice of students with more marked ability of at least some particular nature. The most important decisions affecting a career, for example, are not made in senior year on the basis of whether an able man should enter advertising, banking, law, or the steel business. He can probably succeed in any of these or many other callings of equal importance. The significant decision **affecting his future occurred in the choice** of his general field of study. If someone with a real flair for chemistry or historical scholarship or engineering or languages or medicine or art goes so far in his college course without discovering such a talent that he is deflected from developing it or handicapped therein, then he and society alike are the losers.

Encouragement of specialization, whether in college or in later professional work, is naturally more appropriate with the student displaying at least some distinct educational aptitude than for the one who may be a good, steady, all-around man, without, however, any real scholarly talent. Educability of a special sort is certainly dependent upon intellectual promise. Intelligent guidance, it seems to me, becomes proportionately more important as we deal with progressively more able persons. The scholar does not always prove the most useful being in the large sense, the best citizen, or the one who achieves the best-rounded and most worthwhile total development. Yet the direction of a

more average individual's studies and career into this' or that particular channel is probably a less significant problem-just because he is a sort of normal all-around person, and not one with marked aptitude for a particular field than is the guidance of one with really superior capacities. It is a curiosity of our educational procedure that not infrequently our institutions of higher learning devote more effort to retaining in the traditional college situation people who perhaps do not belong there in the first place, than to encouraging the most intelligent development of their best material. This remark does not connote lack of sympathy for the academically weaker group, but simply the feeling that they should be handled in a different way than is appropriate for the abler students, and particularly that the superior promise of the latter should not be sacrificed, or their educational opportunities diluted, in the interest of their inferiors.

These considerations, and recent progress in the study of aptitudes, all serve to illustrate the tie-up mentioned earlier between specialized psychological procedures like testing, the use of resulting data in the personnel function of counselling, and the bearing of both upon education as a whole.. Measurement, guidance, teaching, curriculum planning, vocational placement-none of these can any more keep growing by themselves alone, than can the vitals cut out of a living organism. Further attacks along the lines I have so roughly sketched, if worth pushing at all, need coordinated efforts along the whole front. Thus far the development of testing and guidance techniques has outstripped that of measuring actual scholastic achievement in college. In other words, our aptitude, intelligence, and other predictive measures have been carried further, and made more scientific and actually more stable *per se*, than are the marking systems towards which they are pointing. What we need in this general cause more than anything else at the moment is increased reliability of college grades. Efforts of the Cooperative Test Service and of the American Council on Education's Committees on Testing and on Personnel

309 Methods; The Carnegie Corporation Pennsylvania Survey; furtherance by the Educational Records Bureau of standardized achievement tests and cumulative records of academic progress-all these emphasize a growing and insistent demand for advance in this direction, and suggest means for its attainment.

So long as pseudo-accurate marks continue to be naively entered and trustfully dealt with by those responsible for assigning and recording them, there is not much chance for further progress in aptitude testing, either general or specific. Presumably some such formal measures of classroom accomplishment-some means of ranking students relatively to each other-are still regarded as administrative and pedagogical necessities. But if they are precise only in appearance and not in fact, their value is specious and slight. Consequently those responsible for the administration of marks and interested in dependable prediction of scholastic success in these terms, must strive to make them more meaningful and valid than they usually are today. I suggest, therefore, that the next step in this game of measurement is up to the registrars and deans. They occupy a post of the greatest strategic importance in this fight. Let them give the testers and forecasters something to shoot at that isn't itself doing a continual shimmy!