

LOST: OUR INTELLIGENCE? WHY?¹

QUINN McNEMAR

Stanford University

THE Greeks had a word for it, but the Romans had a word with better survival properties. Regardless of the word, what is now called intelligence has been talked about for at least 2,000 years. And as long as 2,000 years before the advent of attempts to measure intelligence, there seems to have been recognition of the fact that individuals differ in intellectual ability.

The earlier attempts at measuring were based on either of two quite distinct conceptions: the Galton-Cattell idea that intellectual ability manifests itself in simple, discrimination functioning, and the Binet notion that cognitive ability reflects itself in more complex functioning. The Binet concept proved to be more fruitful, and by 1925 there was on the market, in addition to various versions of the Binet scale, a flood of group tests of so-called general intelligence.

A few words about definition may be in order. First, it might be claimed that no definition is required because all intelligent people know what intelligence is—it is the thing that the other guy lacks. Second, the fact that tests of general intelligence based on differing definitions tend to intercorrelate about as highly as their respective reliabilities permit indicates that, despite the diversity of definitions, the same function or process is being measured—definitions can be more confusing than enlightening. Third, that confusion might have been anticipated is evident from a recent reexamination of the problem of definition by Miles (1957). This British chappie found himself struggling with the awful fact that the word “definition” itself has 12 definitions. Perhaps the resolution of this problem should be assigned to the newly formed Division of Philosophical Psychology, or maybe the problem should be forgotten since psychologists seem to have lost the concept of general intelligence.

Why has the concept been abandoned? Was it replaced by something else? By something better?

¹ Address of the President to the Seventy-Second Annual Convention of the American Psychological Association, Los Angeles, September 5, 1964.

Must we admit that the millions who have been tested on general intelligence tests were measured for a nonexistent function? If it is possible that the notion of general intelligence is not lost but merely gone astray, in what corners of what psychological fields should we search for it?

REASONS FOR DISCARDING THE IDEA OF GENERAL INTELLIGENCE

Apparently one reason why concepts are either discarded or modified beyond recognition is that too much is claimed for them. Among the supposed strikes against general intelligence are the following: the earlier false claims about IQ constancy; prediction failures in individual cases; unfounded claims that something innate was being measured by the tests; equally unfounded assertions that nothing but cultural effects were involved; the bugaboo that IQ tests reflect middle-class values; the notion that an IQ standing fosters undesirable expectations regarding school achievement; the idea that IQ differences are incompatible with democracy and lead to educational determinism; and, finally, the great stress on general intelligence caused us to ignore other possible abilities.

This last point leads us right into the problem of factor analysis. Spearman died in battle defending his theory of *g*. Under pressure he reluctantly conceded that factors other than *g* might exist, and he frequently said, in effect, I told you so as long ago as 1906. Actually, Spearman was on the run before the invention of modern factor analysis, but it was not until Thurstone's (1938) first major application of his centroid factor method that Spearman's *g* became, seemingly, nonexistent. Thurstone said, “We have not found the general factor of Spearman” and “We cannot report any general common factor in the battery of fifty-six tests [p. vii].” As anticipated by some, Spearman was not prone to admit defeat. He reworked Thurstone's data and a *g* was found, plus some group factors. He charged that Thurstone's rotational process had submerged the general factor.

American factorists found Thurstone convincing. The description of abilities in terms of seven primaries was an attractive package. The so-called primaries were more amenable to specific definition than the old hodgepodge called general intelligence. Despite the fact that Thurstone was able to replicate his findings on samples from two other populations, thus giving credence to his method and results, there were a couple of events that led to some turbulence in his seven-dimension rarified atmosphere. The first of these was a minor study, by one of his own students, based on the intercorrelations of 1916 Stanford-Binet items, in which the *g* refused to be rotated out. But rather than admit that this might be some kind of general intelligence, the author renamed it "maturational level." Incidentally, this illustrates the first cardinal Principle of Psychological Progress: *Give new names to old things.*

The second disturber of the neat little set of primaries, sans a *g*, resulted when Thurstone took the next logical step, that of constructing tests to measure the primaries. It was found that the primaries were themselves intercorrelated whereas it had, at the time, been expected and hoped that they would be independent. The Thurstones (1941, p. 26) readily admitted that a general factor was needed to explain the interrelatedness of the primaries. This eventually led to the idea of oblique axes, which axes were regarded as representing the primaries as first-order factors, whereas the general factor pervading the primaries was dubbed a second-order factor. It began to look as though Spearman was being revisited, except for the little matter of labeling: anything called second-order could not possibly be regarded as of much importance. Furthermore, it could always be said that, in the ability domain, it is less difficult to attribute psychological meaningfulness to first-order than to second-order factors, so why pay much attention to the latter? Thus it was easy for most American factorists to drop the concept of general intelligence and to advocate that tests thereof, despite their proven usefulness over the years, should be replaced by tests of the primaries. Hence the emergence of differential aptitude batteries, about which more later.

Meanwhile, our British cousins did not tag along with the factor methods preferred on this side of the Atlantic. After all, it is possible to use factor methods that permit a sizable general factor, if

such exists, to emerge as the very first factor. Being first, it is, presto, the most important, as indeed it is as a factor explaining, for the starting battery as a whole, more variance for more tests than attributable to any American-style primary factor. The methods preferred by the British also yield group factors, apt to bear the same name as the primaries, but of attenuated importance. Apparently the British are skeptical of the multitude of ability factors being "discovered" in America. The structure of intellect that requires 120 factors may very well lead the British, and some of the rest of us, to regard our fractionization and fragmentation of ability, into more and more factors of less and less importance, as indicative of scatterbrainedness. This statement presumes that intellectual abilities are brain centered.

In practically all areas of psychological research the demonstration of trivially small minutia is doomed to failure because of random errors. Not so if your technique is factor analysis, despite its being based on the correlation coefficient—that slipperiest of all statistical measures. By some magic, hypotheses are tested without significance tests. This happy situation permits me to announce a Principle of Psychological Regress: *Use statistical techniques that lack inferential power.* This will not inhibit your power of subjective inference and consequently will progress you right back to the good old days when there was no strangling stat or sticky stix to make your insignificant data insignificant.

It may be a long time before we have an ivory tower, strictly scientific resolution of the issue as to whether a scheme involving primary abilities plus a deemphasized *g* is preferable to one involving an emphasized *g* plus group factors. With bigger and better computers we will have bigger, though not necessarily better, factor-analytic studies, but it seems unlikely that such further studies will, in and of themselves, settle the issue under discussion. Until such time as some genius resolves the broader question, so ably discussed by Lee Cronbach in 1957, of the place, if any, of correlational method in a science that aspires to be experimental, we may have to turn to the criterion of social usefulness as a basis for judging whether it is wise to discard general intelligence. Like it or not, much of our heritage in this area is that earlier workers, from Binet on, had as their motiva-

tion the solution of social problems, and currently many in the area have a similar motivation.

THE BEARING OF SOCIAL USEFULNESS

In practice, if you believe that the concept of general intelligence has outlived its usefulness, you may choose from among several differential, or multiple, aptitude batteries, which will provide measures of some of the so-called primary mental abilities. If you happen to believe that there is something to general ability, you can find tests to use. The novice looking for these latter tests may have to alert himself to the first Principle of Psychological Progress—the test labels may have changed from “general intelligence” to “general classification” or “scholastic aptitude.” If you enjoy riding the fence, you might become a devotee of the practice of the College Board, and others, and measure just two abilities: Verbal and Quantitative.

This is certainly not the place to review the voluminous literature that amply demonstrates the practical utility of tests of general intelligence. Nor is it the place to catalog the misuses of the Stanford-Binet for purposes which Terman never claimed for it, or the misuses of the Wechsler scales for purposes which Wechsler *has* claimed for his scales. Neither the Binet nor the Wechsler provides a factorially pure, unidimensional measure of a *g*. The current Stanford-Binet was in reality constructed too early to benefit from the implication of factor analysis for test purity, whereas the Wechsler scales were based on the impossible premise that 10 or 11 subtests can simultaneously provide diagnostic subscores and a meaningful total score. Of the many group tests that appeared between 1920 and 1945 it can be said that few, if any, provide unidimensional measures of general intelligence. The chief difficulty is that most of them lead to a total score based on a mixture of verbal and mathematical material. Thus, with two main sources of variance, marked qualitative differences can exist for quantitatively similar total scores. The College Board—Educational Testing Service people have justifiably refrained from giving a total score involving verbal plus math, but there are those who question the usefulness of the Board's math score and there are those who criticize the Educational Testing Service for failing to change over to a differential aptitude battery.

Let us next turn to a somewhat more detailed

examination of the various so-called multiple aptitude batteries. What and who influenced whom in the development of these batteries is difficult to disentangle. At the risk of oversimplification, it might be said that two prime influences operated.

First, the early factor studies by the Thurstones, by Holzinger, and by Guilford are the progenitors of the Science Research Associates' Primary Mental Abilities (PMA) Test, the Holzinger-Crowder Unifactor Tests, the Guilford-Zimmerman Aptitude Survey, and the Segel-Raskin Multiple Aptitude Tests (MAT).

The second influence, which seems to have emerged from testing experience in the Armed Services during World War II, is the job-element approach, an approach which may or may not differ from the old job-analysis method. For whatever jobs you are dealing with, you study the activities involved in order to decide what aptitudes are called for. Whether or not these aptitudes have been previously isolated by factor analysis is totally irrelevant. It is hoped that some jobs will have aptitudes in common so that the needed number of tests will be less than the number of jobs. The one battery that is built on this approach is the Flanagan Aptitude Classification Tests, a battery that just happens to have the catchy abbreviation, FACT. If we cannot muster any facts in psychology, we can at least have FACT scores!

A cross between testing for factorially defined abilities and job-element derived aptitudes is apparently involved in the General Aptitude Test Battery (GATB) of the United States Employment Service and the Differential Aptitudes Tests (DAT) of the Psychological Corporation, since in both batteries some of the tests seem to have sprung from factor-analysis results and some tests seem to have been thrown in as possible predictors of specific performances.

It is not our purpose to rank order the seven above-mentioned multitest batteries, but a few remarks may be relevant as background for the sequel. Apparently the Employment Service's GATB was made available (but not put on the commercial market) with the idea that there would be a continuing program of validities studies—the accumulation is now impressive. For the DAT of the Psychological Corporation there is an overabundance of data on validity, collected and analyzed prior to marketing the test. Both the Science Research Associates' PMA and FACT were

made available without backing for the claimed usefulness of the tests. Belatedly, that is, 6 years after its appearance, some evidence on the predictive validity of FACT has been reported. Validity information for the other three batteries is not entirely lacking, though far from ample. Some will have noted that that fuzzy dodge called factor validity is being ignored here.

Now to get back to our main theme, to what extent have the seven batteries contributed to the demise of general intelligence? In attempting to answer this, one encounters a paradox: Some test authors want to eat their cake and have it too—they attempt to measure factors and g with the same instrument. This is understandable in a couple of instances. Three of the 15 tests of the Employment Service GATB were included to provide a measure of general intelligence, apparently because the authors still saw some merit in a g and were not committed to the factor schemata. Holzinger and Crowder suggest a weighted score for a measure of g , perhaps because of Holzinger's long-time alignment with Spearman. The real teaser is why Thurstone ever sanctioned, if he did, the summing of Science Research Associates' PMA scores to obtain an IQ. One has the uncomfortable feeling that his publishers wished to g garnish the factor cake to make it more palatable in the market place.

Although Segel says nothing in his 1957 article about a general score from the Segel-Raskin MAT, the test publishers say that, in addition to yielding scores for four factors, it also provides a "Scholastic Potential" score. Perhaps Flanagan has not completely broken with tradition since he states that four tests of the FACT battery measure "General College Aptitude"—a statement made with the same lack of empirical validity as the claim, which should be anxiety producing for those of you who fly a certain airline, that your highly paid pilot shares four of the aptitudes of a plumber!

Apparently, Guilford and Zimmerman and the test people at the Psychological Corporation are willing to stick to the sound principle that a differential test battery cannot provide factor scores that can be summed to obtain a meaningful IQ, or measure of a g .

Parenthetically, it might be said that the California Test of Mental Maturity (CTMM), which, according to the publisher's 1963 catalog, was originally "designed as a group test of intelligence

patterned after the individual Stanford-Binet," serves as an illustration of factor icing a g cake. Some multitest batteries and the CTMM have a Madison Avenue advantage: The advertising claims the measurement of not only factors but also g ; not only g but also factors. This measurement absurdity is all too apt to go unrecognized by many test users, and hence a sales advantage for the aptitude battery that produces both factor scores and an IQ.

Just how successful have the multitest batteries been? Since by far the most extensive social use of tests has been, and continues to be, in the schools, let us look at the evidence of validity studies therein. As indicated previously, little is known about the predictive usefulness of some of the seven batteries discussed above. The DAT of the Psychological Corporation is the only battery for which adequate predictive (and concurrent) validity data, derived from school sources, are available. It is also the battery that has fared best in the hands of the test reviewers; therefore if we allow the case for differential batteries to rest thereon, we will be looking at the best. So, what is the story?

Recall that the hoped-for advantage of a multitest battery over the old-fashioned general intelligence test was that it would have greater predictive power, a power which could manifest itself in higher validity coefficients for specific subject matter and, perhaps, for overall achievement. It was hoped that such a battery would be truly differential in that particular factors (or subtests) would correlate higher with achievement in some areas than in other areas. Presumably each factor (or subtest) should have unique usefulness. If a battery were truly differential, it would be a boon to school guidance personnel.

Now the manual of the DAT of the Psychological Corporation contains a staggering total of 4,096, yes I counted 'em, validity coefficients. With such a large pool to draw from, one could by gracious selection "show" that the DAT is the answer to the prayer of every counselor, male or female, or by malicious selection one could "prove" that the DAT is far worse than any test ever published. The validity coefficients range all the way down to -.37, which is presumably a chance deviation downward from 0, and all the way up to .90, which is likely not a chance deviation downward from unity. But ranges tell us nothing. After a careful

perusal of the 4,096 correlations, it seems safe to summarize DAT validities as follows:

1. Verbal Reasoning (analogies to most of you) is the best single predictor; Language Usage, as represented by a sentence test dealing with grammar and word usage, and admittedly more achievement than aptitude, is a close second.

2. Numerical Ability, as measured by a test of simple arithmetic operations, designed to tap arithmetic reasoning without the usual verbal component, is the best predictor of achievement in school mathematics. It does not, however, correlate as well with grades in science as does Verbal Reasoning.

3. Aside from the Numerical Ability test, the only other test that shows differential power as a predictor is the Spelling test—if you cannot spell you may have trouble learning shorthand.

4. The remaining five tests in the battery simply fail to show compelling evidence that they are good in the differential predictive sense. For the Mechanical Reasoning and the Clerical Speed and Accuracy tests this may be understandable in that little of school curricula for Grades 8 through 12 requires such abilities, but one would expect that Abstract Reasoning and Space Relations would fare better than they seem to.

Such data as we have been able to locate for the other six multitest batteries tend to support these findings on the DAT. Aside from tests of numerical ability having differential value for predicting school grades in math, it seems safe to conclude that the worth of the multitest batteries as differential predictors of achievement in school has not been demonstrated. Incidentally, the fact that the Verbal and Numerical tests stand out as the only two useful predictors tends to provide some support for the Educational Testing Service-College Board practice of providing scores for just these two abilities.

And now we come to a very disturbing aspect of the situation. Those who have constructed and marketed multiple aptitude batteries, and advocated that they be used instead of tests of general intelligence, seem never to have bothered to demonstrate whether or not multitest batteries provide better predictions than the old-fashioned scale of general intelligence. Be it noted that we are not discussing experimental editions of tests. Some may say that insofar as a test publisher provides validity data for a new battery it is not necessary

to show that the validities are, for the given school condition, better than those of other tests. With this one can agree, but only in case no claims are made, explicitly or implicitly, regarding superior merits for the new battery.

It is far from clear that tests of general intelligence have been outmoded by the multitest batteries as the more useful predictors of school achievement. Indeed, one can use the vast accumulation of data on the validity of the Psychological Corporation's DAT to show that better predictions are possible via old-fashioned general intelligence tests. Consider the fact that a combination of the tests Verbal Reasoning (analogies) and Numerical Ability would be, in terms of content, very similar to many group tests of general intelligence. Consider also that an equally weighted combination of these two tests correlates in the mid-80s with the Otis S-A, Higher Form. Then, when you turn to a careful study of the empirical validities, as reported in the DAT manual, you will not be surprised at the outcome of the application of a little arithmetic, which leads to the definite conclusion that a simple unweighted combination of the Verbal Reasoning and Numerical Ability tests predicts as well as or, in most instances, better than any subtest taken singly, or in the differential sense.

The manual for the DAT contains the following statement (Bennett, Seashore, & Wesman, 1952):

Apparently the *Verbal Reasoning* and *Numerical Ability* tests can serve most purposes for which a general mental ability test is usually given in addition to providing differential clues useful to the counselor. Hence, the use of the so-called *intelligence test* is apparently unnecessary where the *Differential Aptitude Tests* are already being used [p. 71].

Anyone who disagrees with this quotation could, with better justification, say that an intelligence test can serve nearly all, if not all, the purposes for which a multiple aptitude battery is given in the schools because the former, in general, is a better predictor and because, as we saw earlier, the differential clues are too fragmentary to be of use to the counselor. And there is a bonus: one classroom period of testing, compared to six periods. A second bonus: much less costly. A third bonus: fewer scores to confuse the already confused minds of most school counselors.

Thus, we come to the conclusion that general intelligence has not been lost in the trend to test

more and more abilities; it was merely misplaced by a misplaced emphasis on a hope that a lot of us, including the speaker, once entertained, a hope that in turn was based on a misplaced faith in factor analysis: the hope that factors, when and if measured, would find great usefulness in the affairs of society. By the criterion of social usefulness, the multiple aptitude batteries have been found wanting. Now, I have no desire to furnish ammunition for those test critics who would have us stop all testing merely because they find a trivially faulty item in a standardized test. At a time when there is shouting about the tyranny of the testers and the brass of the brain watchers, at a time when school people are showing resentment at the disruption caused by too many national testing programs, at a time when federal and state legislators are all too willing to write legislation that places restrictions on the use of tests, and at a time when both majorities and minorities are being denied the benefits of test-based guidance because certain well-intentioned persons fail to realize that scores for the underprivileged minorities are useful indices of *immediate*, or present, functioning—at a time when all these and other forces are operating to throw out the tests, it is high time for the profession to establish a bureau of standards to test the tests instead of coasting down a road that is tinged with some of the trappings of Madison Avenue. Better to have informed internal control than ignorant, hostile, external control.

INTELLIGENCE ELSEWHERE?

Aside from the near loss of the idea that progress in school may depend on general intelligence, one wonders whether intelligence has come to be regarded as unimportant in other areas.

Any of you who have money invested in stocks and wish some reassurance regarding the intelligence level of business and industry managers should read Edwin Ghiselli's (1963) Bingham Lecture. His summary of his own work indicates that the average intelligence of those in the upper and middle management levels falls at the ninety-sixth percentile of the population. Thomas Harrell (1961) came to a similar conclusion. Furthermore, management level is correlated with intelligence—you can be too dumb to succeed as a manager. Also you can be too bright to be a managerial success! Now it must be admitted that little, if anything, is known about whether man-

agement success might be better predicted by measures of factor-analytic defined abilities. On this you are free to guess—most of you will have already guessed my guess.

A one-by-one cataloguing of what we know or do not know about what abilities contribute to success within various occupational and professional groups would merely add to the dullness of this presentation, so let us turn to some of the more esoteric fields of psychology to see whether the concept of general intelligence has or has had any relevance. One such field, and a very broad one, is creativity. Anyone who peeks over the fence into this field is apt to be astonished at the visible chaos. The definition of creativity is confounded by the diversity of subareas within the field, the criterion problems are far from licked, and so little is known about the creative process that measuring instruments are, seemingly, chosen on a trial-and-error basis.

We might presume that the role, if any, of general intelligence in creativity would increase as we pass from art to music, to architecture, to literature and drama, to science. Your presumption about the ordering may be different and more nearly correct. I would like to discuss briefly the extremes of my ordering.

At the risk of being called a heretic and a has-been statistician, I would like first to resort to the single case of a painter, examples of whose works were reproduced in color recently by Desmond Morris (1962) in a journal called *Portfolio and Art News Annual*. To my uncultured eye these paintings have the general appearance of the so-called school of modern art, and the running comment on the paintings involves what I must presume is the jargon of contemporary art critics: talk about self-rewarding activities, compositional control, calligraphic differentiation, thematic variation, optimum heterogeneity, and universal imagery. Since authors may use pen names, I would guess that this painter is using "Congon" as a "brush" name. Supposedly by now some of you will be guessing that this single case is of interest in the context of this paper because of Congon's IQ. Well, because of this painter's underprivileged cultural background, no test scores are available. Congon, despite striking contribution to art, happens to be a chimp. Aside from a rather obvious conclusion, one wonders what would emerge from a blind (as to source) analysis of Congon's paintings by the

personality boys. We might even tell them that ~~Congon was breast fed.~~

Without in any way implying that creativity in the arts is unimportant, we hasten on to scientific creativity, a specific area in which it seems likely, because of the Sputnik-inspired spurt of interest, that we can learn something of the role, if any, of general intelligence. But immediately we encounter skulls that have been cracked on the criterion problem.

One elaborate study (C. Taylor, Smith, Ghiselin, & Ellison, 1961), on a sample of 166 physical scientists working at Air Force research centers, came up with 150, yes, believe it or not, 150 criteria of scientific productivity and creativity. By combining some scores and eliminating others, the number of criteria was reduced to 48. A factor analysis of the intercorrelations of the 48 reduced the number to 14 "categories." Apparently the 150 original criterion measures included everything except success at turning on a kitchen faucet, so one need not be surprised at the outcome of the factor analysis. For example, one factor-derived criterion of scientific productivity and creativity is "likableness," another is "status seeking," another is extent of membership in scientific and professional societies—the joiners, no doubt.

The fact that the intercorrelations among the 14 criterion categories, derived from factor analysis, range from $-.08$ to $+.55$, with a median of only $.18$, indicates either criterion complexity or else a whale of a lot of vagueness as to what is meant by productivity and creativity in science. Now this criterion mess emerged from a study of interview results of 166 "physical scientists," but nearly half of these so-called scientists were engineers, and the education of the total group indicates only 2 years of graduate work on the average; so when is a scientist a scientist a scientist?

The next step in this study was to collect data on 107 of these so-called scientists for a whopping total of 130 potential predictors, which, when pitted against 17 criterion measures, produced 2,210 "validity coefficients." The distribution of these, excluding 30 values involving un-cross-validated empirical keys, almost restores one's faith in the random-sampling distribution of correlation coefficients around zero! There were 16 predictors based on aptitude tests, hence 16×17 , or 272, "validities" for this area. Since only 4% of them reach the 5% level, we can do no more than accept

the null hypothesis: Aptitude ain't important in scientific productivity and creativity. The idea that some scientists are more equal in ability than others apparently is not true.

But this criterion-based study did not contribute to my worry about the role of general intelligence—the failure to include a general intelligence measure as a potential predictor may be interpreted as indicating that the authors already had the answer.

Let us turn to another criterion-based study (D. Taylor, 1961). The criterion measures for creativity and productivity were based on the checking by supervisors of statements that had been scaled by Thurstone's equal-appearing interval method. Creativity and productivity, so gauged, correlated .69 with each other on a sample of 103 researchers (electronic scientists and engineers). For this same group, intelligence, as measured by the Terman Concept Mastery Test (CMT), correlated only $.20$ or less with the criteria. Two Psychological Corporation tests and an American Institute for Research test did a little better. Creativity is slightly more predictable than productivity. Insofar as these two criteria are themselves valid, the findings indicate that within a group of research workers, precious little of the variance in creativity, and still less in productivity, can be predicted by the tests.

A third study (MacKinnon, 1962) based on criterion (rated) measures of performance was concerned with the creativity of architects. Although the author reports that *within* a creative sample the correlation is essentially zero between intelligence (CMT) and rated creativity, it is not clear from the context what is meant by "within" sample. If this means within the sample of 40 creative architects selected as the "most creative" in the country, then we indeed have such a drastic restriction in range on the *criterion* variable that little, if any, correlation can be expected for any and all predictors. Now the author says, without presenting any evidence, that "Over the whole range of intelligence and creativity there is, of course, a positive relationship between the two variables [p. 488]." One wonders just what is meant by creativity in architecture as rated either by fellow architects or by editors of architectural journals. If judged creativity reflects engineering-structural innovation, then intelligence would likely be a correlate; if judged creativity depends on new artistic designs, then the intelligence component

would likely be of less importance. It would seem that when the author says we "may have overestimated . . . the role of intelligence in creative achievement [p. 493]," he should have included some marked qualifications as to what type of creativity he had in mind.

That such qualification is indeed necessary is supplied by a finding of still another investigator (Barron, 1963). For a group of highly creative writers it was estimated, by way of the Terman CMT, that their average IQ is about 140, which we interpret as meaning that a high IQ is a necessary, though not sufficient, condition for outstanding success as a writer. On the basis of his own studies and those of other persons, this same investigator suggests that "over the total range of intelligence and creativity a low positive correlation" of .40 probably obtains. This sweeping generalization is for all areas of creativity.

And speaking of sweeping generalizations, consider the suggestion in a 1961 study (Holland, 1961) that "we need to use nonintellectual criteria in the selection of students for scholarships and fellowships [p. 146]." The author did not say so, but presumably he meant in addition to intellectual ability; maybe he did not, since he had previously concluded that "intelligence has little or no relationship to creative performance in arts and science . . . [p. 143]" at the high school level. His data back up this conclusion, as might have been expected when correlations are based on groups restricted in range to the top 1%!

If the foregoing examples of criterion-based studies of creativity seem to indicate that general intelligence is relatively unimportant for creativity, it should be remembered that drastic but unknown or unspecified curtailment of range exists for both ability and criteria. Why do correlational studies under such adverse circumstances?

Next we turn to a few studies of creativity which cannot be criticized because of restriction of range on the criteria—these studies simply avoid this problem by never having actual criterion information. The approach is to claim that certain tests, which typically are scored for novel responses or novel solutions to problems, *are* measures of creativity, with no evidence whatsoever that the tests have predictive validity for nontest, real-life creative performance. This bit of ignorance does not prove to be a handicap to those who think that creativity can be studied without the nuisance of

obtaining criterion measures. We reluctantly accept the test-based criteria solely for the sake of seeing what happens to general intelligence as a part of the picture. Time permits only three examples.

We first note that general intelligence has not manifested itself as a correlate of so-called creativity tests in the factor-analytic studies of creativity. The explanation for this is easily found—no measures of general intelligence are used in these studies! When discussing his plans for studying creativity, a certain author (Guilford, 1950) said that "we must look well beyond the boundaries of the IQ if we are to fathom the domain of creativity [p. 448]." He went on to say, the conception "that creative talent is to be accounted for in terms of high intelligence or IQ . . . is not only inadequate but has been largely responsible for lack of progress in the understanding of creative people [p. 454]." With a part of this one can agree, but does it follow that one should prejudice the role of general intelligence as a source of variance in creativity tests or factors derived therefrom? Does the failure to include an IQ test help one learn the extent to which one must go beyond the boundaries of the IQ to fathom creativity? Apparently the author, although willing to predict that the correlations between IQ and the many types of creativity tests "are only moderate or low," was unwilling to include an IQ test for the sake of finding out. However, negation by omission is not very convincing.

That at least one test bearing the label "creativity" is correlated more than moderately with IQ is evidenced by the value of .67 (average for boys and girls) for the carefully chosen sample of 15-year-olds in Project Talent (Shaycoft, Dailey, Orr, Neyman, & Sherman, 1963). This sample-stable r (based on a total N of 7,648) becomes .80 when corrected for attenuation.

In a recent extensive study (Getzels & Jackson, 1962), already extensively criticized, creativity is defined as the sum of scores on five tests (median intercorrelation of only .28). Although the investigators use the sum score for most of their analyses, they do not bother to report the correlation of creativity, so defined, with IQ. From the published report I have ascertained (via the correlation-of-sums formula) that creativity and IQ correlate to the extent of .40 for the total of 533 cases. Now this r of .40 has been greatly attenuated because of three things: first, the usual measurement errors;

second, the cases were highly selected on IQ (mean of 132); third, the IQs are a mixture from the Stanford-Binet, Henmon-Nelson, and Wechsler Intelligence Scale for Children (the use of regression-estimated Binet IQs from the other two scales aggravates rather than improves the mixture). We deduce that intelligence and the creativity tests used here have far more common variance than the authors believe.

Much is made of the finding that the creativity tests tended to correlate higher than did IQ with verbal-content school achievements. Again the IQ comes in for an unfair drubbing because of the same mixture of IQ scores and, what is more pertinent, because of explicit selective curtailment on the IQ variable and only incidental selection on the creativity variable.

Of more importance to the present paper is the analysis, by these same authors, based on a high IQ group and a high creative group, these groups being selected as the top 20% for each variable but excluding those who were in the top 20% on both variables. These two selected groups were then contrasted on total school achievement (and a host of other variables that are of no interest here). The mean IQ for the high IQ group was 150 whereas the mean IQ for the high creative group was 127, yet the achievement means of the two groups were "unexpectedly" equally superior to the school population mean despite the 23-point difference in mean IQ. The authors say that it "is quite surprising" that the high creativity group achieved so well. From this it is concluded that the "creative instruments account for a significant portion of the variance in school achievement [p. 24]," and the subsequent argument implies that creativity is more important for ordinary school achievement than is the IQ. Now anyone who is at all familiar with a three-variate problem will not be "unexpectedly" surprised at the foregoing results—indeed, if the authors had bothered to give the three basic correlations among the three variables (IQ, creativity, and total school achievement) for the entire group, any person versed in simple multivariate analysis could deduce the results. Furthermore, he could deduce a further result (and this one has been overlooked by the critics) which might be unpleasantly surprising to the thesis of these authors: namely, the high IQ and the high creative groups did equally well in school achievement despite an unreported differ-

~~ence in mean creativity that is of the same order as the much stressed difference in IQ.~~² Utilizing the half-blind logic of the authors, one can say that creative ability is not as important as IQ for school achievement—just the opposite of their position.

Now the fact that seven of nine replications of this study confirm the original findings merely indicates that repetition of the same faulty design and false logic will lead to the same false conclusions. The design being used is such that, if two variables are equally correlated with a third, the conclusion will be reached that the two are actually unequally correlated with the third. This is the neatest trick of the decade for supplying educationists with an antidote for the IQ virus. I cannot refrain from saying at this point that, although discouraged, I am still hopeful that people who do statistical studies will first learn a modicum of elementary statistics!

Time does not permit a discussion of other studies in which creativity is defined in terms of test performance instead of being based on actual creativity of the sort prized by society. In summary of this brief on creativity studies, I would like to offer a few dogmatic-sounding observations. First, one need not be surprised at the fact that so-called creativity tests do not yield high correlations with IQ tests—but the correlations are generally far higher than those found in typical studies with range restrictions. I would anticipate that for normalized scores, the uncurtailed scatters for IQ versus creativity tests will be bivariate normal. Second, if we have honest to goodness criterion measures of literary or architectural or scientific creativity, the scatter diagram between IQ and such creativity (not normalized, since it makes sense to expect a skewed distribution for actual creativity) will be triangular in shape for unselected cases. That is, at the high IQ levels there will be a very wide range of creativity, whereas as we go down to average IQ, and on down to lower levels, the scatter for creativity will be less and less. Having a high IQ is not a guarantee of being creative; having a low IQ means creativity is impossible. Third, it remains to be seen whether or not the so-called creativity tests and/or factors derived therefrom have appreciable value as predictors of

² Since this was written, the replication study of Yamamoto (1964) gives data that corroborate this deduction.

actual creative performance. Such tests may or may not yield better predictions than a test of general intelligence. Fourth, as far as I am concerned, to claim factorial validity for creativity tests, along with definitions of creativity in terms of tests, is an unwarranted avoidance of the fundamental problem of validity.

The recently renewed interest in "gifted" children, along with the flurry of creativity studies, has led to a reexamination of methods for identifying the gifted. It has long been recognized that identification in terms of high IQ is too narrow—those gifted in such areas as art and music would be overlooked. The argument against the IQ is now (Torrance, 1962) being reinforced by the claim that the selection of the top 20% on IQ would mean the exclusion of 70% of the top 20% on tested creativity. This startling statistic, which implies a correlation of only .24 between IQ and creativity, is being used to advocate the use of creativity tests for identifying the gifted. Be it noted that these creativity tests will also miss those gifted in art and music.

We are being told that it is important "to identify creative talent early in life," hence you need not be surprised that the search goes down to the kindergarten level, with claims of successful identification. The creativity tests are presumed to be better for this purpose than the IQ tests because of the failure of the IQ to be constant, an argument that completely overlooks the fact that the IQ does have some constancy whereas absolutely nothing is known about the stability of standings on creativity tests. The IQ tests, known to be imperfectly valid as predictors of outstanding achievement in life, are to be replaced by the creativity tests, known to be of unknown validity as predictors. Anyway, progress, defined as change, is in the offing.

The IQ is being linked with *learning* as an outmoded educational objective; the new objective involves an emphasis on *thinking*. Somehow or other creativity, not general intelligence, is being associated with thinking. The horrible idea of underachievers and overachievers, in terms of expectancies based on the IQ, will be abolished. But no thought is given to the fact that the use of creativity tests will simply define a new crop of under- and overachievers.

In an apparent zeal to rid us of general intelligence, it is argued that measured creativity is sig-

nificantly related to ordinary school achievement. Maybe so, but never, never does one find complete data reported as to the relative sizes of validity coefficients. And, as we have seen, the technique being used will show that equal coefficients are unequal. Why not the full facts, free of fantasy?

An additional difficulty is not being faced by those who would replace IQ tests by creativity tests, or creative-thinking tests. The factor-analytic studies indicate either no, or a trivially small, general creativity factor in these tests, yet these self-characterized "bold, adventurous" reformers (see Torrance, 1963) do not hesitate to advocate a total score which is nearly devoid of meaning. Changing the curriculum to the teaching of creativity and creative thinking will not overcome this measurement difficulty. Again, I express the hope that the IQ is replaced by something better rather than by something worse.

There are other areas, such as reasoning, problem solving, and concept formation, in which one might expect to find some consideration of intelligence as an aspect. One might also expect that investigators of thinking would have something to say about individual differences in thinking being dependent upon intelligence, but for some unintelligent reason these people seem never to mention intelligence. Surely, it cannot be inferred that thinking about thinking does not involve intelligence!

IN CONCLUSION

It has been the thesis of this paper that the concept of general intelligence, despite being maligned by a few, regarded as a second-order function by some, and discarded or ignored by others, still has a rightful place in the science of psychology and in the practical affairs of man. It has not been argued that the nature of general intelligence is well understood. Much, however, has been written about its nature. Over 40 years ago (Intelligence, 1921a, 1921b), an editor secured and published the reasoned views of 13 well-known test psychologists. Later, Spearman set forth his speculations about the nature of *g*. Prior to these, Binet had, of course, given much thought to the problem.

More recent discussions exist. Hebb (1949) has considered the problem from the viewpoint of neurology and brain functioning. Cyril Burt (1955), always a vociferous defender of the concept of general intelligence, has reviewed the evidence for a *g*

and restated the idea, dreadful to some, that intelligence is innate. Perhaps it was inevitable that Raymond Cattell (1963), who has camped with the general intelligence contingent, should gaze into his crystal n -dimensional factor ball and find evidence for crystallized as opposed to fluid general intelligence. Joseph McVicker Hunt's (1961) book on *Intelligence and Experience* is in large part devoted to questions pertaining to the nature of intelligence.

By far the most provocative recent discussion that I have encountered is the closely reasoned 44-page paper by Keith Hayes (1962). He puts forth a motivational-experiential theory of intelligence. In essence, he presumes that there are hereditary differences in motivation. "Experience-producing drives" and environmental differences produce differences in experience, which in turn, by way of learning, lead to differences in ability. Therefore, differences called intellectual are nothing more than acquired abilities. I think that Hayes has ignored the possibility of individual differences in learning ability, but if such a formulation leads to experimental manipulation of variables, we may eventually make progress in an area that has too long been dominated by ever increasing fractionization by factor analysis, with little thought as to how the fractured parts get put together into a functioning whole.

Abilities, or capacities, or aptitudes, or intellectual skills, or whatever you choose to call them, are measured in terms of response products to standardized stimulus situations. The *stimulus* is presented to an *organism* which by some *process* comes up with a *response*; thus any attempt to theorize and/or study intellect in terms of a simple stimulus-response (S-R) paradigm seems doomed to failure unless drastically modified and complicated by the insertion of O for organism and P for process.

There have been thousands of researches on the multitudinous variations from organism to organism, and the results fill books on individual differences. These studies can be roughly classified into two types. First, those that ascertain the inter-correlations among scaled response products to various stimulus situations, known as tests, have to do with the structure of intellect; and whether the resulting factors are anything more than dimensions for describing individual differences need not concern us here. The second type of study

seeks the nontest correlates of test performance, and whether or not any of the found correlates can be regarded as explaining individual differences is not of interest here. Both types of studies certainly force one to stress the overwhelming diversity exhibited among the organisms.

But these studies of individual differences never come to grips with the *process*, or operation, by which a given organism achieves an intellectual response. Indeed, it is difficult to see how the available individual difference data can be used even as a starting point for generating a theory as to the process nature of general intelligence or of any other specified ability.

As a basis for a little speculation, let us conceive of a highly hypothetical situation in which the two members of a pair of identical twins, with identical experiences, find themselves cast up on an uninhabited tropical island. Let us assume that they are at the supergenius level, far beyond that of your favorite man of genius. Let us also assume that, though highly educated in the sciences, they have been fortunate enough to have had zero exposure to psychology. In addition, we presume that, being highly involved and abstracted in the pursuit of science, they have never noticed what we call individual differences in abilities.

A quick exploration of the island assures them that food is plentiful, that shelter is available, and that clothing is not a necessity. To allay the boredom that they foresee as an eternity in this laborless heaven, they decide to spend their time in the further pursuit of science, but the lack of the wherewithal for constructing gadgets rules out any research in the physical sciences. Having had a college course in Bugs and Bites they proceed to study the life of the island's insects, then the habits of the birds, and the antics of a couple of monkeys. The manner in which the monkeys adjust to the environment leads them to set up some trial situations for more systematic observation. Needless to say, the monkeys show evidence of what we call learning and what we call problem solving.

Eventually they decide that attempting to out-wit each other might be more fun than being out-witted by the monkeys, so they begin to cook up and use games and problems for this purpose. This activity leads each to speculate and introspect about how problems are invented and how solved. Then by cleverly designed experiments, preceded of course by theory, they set forth highly developed

laws and principles about what we call reasoning and problem solving. Incidentally, they switch back and forth between the roles of experimenter and subject, there being no college sophomores available. They continue for years the study of their own mental operations, constantly on the alert for new phenomena to investigate.

And now with apologies to the ancient Greeks, who did have some ideas along these lines, we leave with you the 64-million drachma question: Will our two identical supergeniuses, being totally unaware of individual differences, ever hit upon and develop a concept of intelligence?

REFERENCES

- BARRON, F. *Creativity and psychological health*. Princeton, N. J.: Van Nostrand, 1963.
- BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. *Differential Aptitude Tests, manual*. (2nd ed.) New York: Psychological Corporation, 1952.
- BURT, C. L. The evidence for the concept of intelligence. *Brit. J. educ. Psychol.*, 1955, **25**, 158-177.
- CATTELL, R. B. Theory of fluid and crystallized intelligence: A critical experiment. *J. educ. Psychol.*, 1963, **54**, 1-22.
- CRONBACH, L. J. The two disciplines of scientific psychology. *Amer. Psychologist*, 1957, **12**, 671-684.
- GETZELS, J. W., & JACKSON, P. W. *Creativity and intelligence*. New York: Wiley, 1962.
- GHISELLI, E. E. Managerial talent. *Amer. Psychologist*, 1963, **18**, 631-642.
- GUILFORD, J. P. Creativity. *Amer. Psychologist*, 1950, **5**, 444-454.
- HARRELL, T. H. *Manager's performance and personality*. Cincinnati, O.: South-Western, 1961.
- HAYES, K. J. Genes, drives, and intellect. *Psychol. Rep.*, 1962, **10**, 299-342.
- HEBB, D. O. *The organization of behavior*. New York: Wiley, 1949.
- HOLLAND, J. L. Creative and academic performance among talented adolescents. *J. educ. Psychol.*, 1961, **52**, 136-147.
- HUNT, J. McV. *Intelligence and experience*. New York: Ronald Press, 1961.
- Intelligence and its measurement: A symposium. *J. educ. Psychol.*, 1921, **12**, 123-147. (a)
- Intelligence and its measurement: A symposium. *J. educ. Psychol.*, 1921, **12**, 195-216. (b)
- MACKINNON, D. W. The nature and nurture of creative talent. *Amer. Psychologist*, 1962, **17**, 484-495.
- MILES, T. R. On defining intelligence. *Brit. J. educ. Psychol.*, 1957, **27**, 153-165.
- MORRIS, D. The biology of art. *Portfolio art News Annu.*, 1962, No. 6, 52-63, 122-124.
- SEGEL, D. The multiple aptitude tests. *Personnel guid. J.*, 1957, **35**, 424-432.
- SHAYCOFT, M. F., DAILEY, J. T., ORR, D. B., NEYMAN, C. A., JR., & SHERMAN, S. E. *Project Talent: Studies of a complete age group—age 15*. Pittsburgh: University of Pittsburgh, 1963. (Mimeo)
- TAYLOR, C. W., SMITH, W. R., GHISELIN, B., & ELLISON, R. Explorations in the measurement and predictions of contributions of one sample of scientists. *USAF ASD tech. Rep.*, 1961, No. 61-96.
- TAYLOR, D. W. Variables related to creativity and productivity among men in two research laboratories. In C. W. Taylor & F. Barron (Eds.), *Scientific creativity*. New York: Wiley, 1961.
- THURSTONE, L. L. *Primary mental abilities*. Chicago: Univer. Chicago Press, 1938.
- THURSTONE, L. L., & THURSTONE, T. G. *Factorial studies of intelligence*. Chicago: Univer. Chicago Press, 1941.
- TORRANCE, E. P. *Guiding creative talent*. Englewood Cliffs, N. J.: Prentice-Hall, 1962.
- TORRANCE, E. P. *Education and the creative potential*. Minneapolis: Univer. Minnesota Press, 1963.
- YAMAMOTO, K. Role of creative thinking and intelligence in high school achievement. *Psychol. Rec.*, 1964, **14**, 783-789.