tests spread over the range .70 to .85. The story on predictive composites is essentially the same as for the Elementary battery.

Primary 2, for grades 2 and 3, requires no reading. A dictated test asks the pupil to judge whether a word such as "suspect" does or does not apply to a picture (man behind bars). Other tasks have to do with computation, everyday information, figure series and vocabulary. A picture-story task displays two silhouette pictures that start a story, plus four additional pictures. The pupil is to mark 3 beneath the picture that fits as third in the sequence, and 4 beneath the correct fourth panel. This subtest falls below the usual excellence of execution: the task is complex, the drawings are tiny, and responses other than the keyed ones can be defended.

For Primary 2 the reliability of GCSS holds up, but concurrent correlations with achievement subtests drop to the .50's and .60's. At this level the attempt to interpret the RCPS-MCPS difference is abandoned, though the option of administering half the test to get either composite remains.

Primary 1 is similar. There are more pictorial tests, and numerical tasks are replaced with one on quantitative concepts. Primary 1 is more nearly a pictorial test of verbal and general abilities than are the achievement-related tests at later levels. It includes a maddening perceptual task calling for selection of similar *kanji*. In general, the demands upon visual acuity and attentiveness seem large for school beginners. A considerable concession to children's interests is made at this level, however. Reliability over a short interval falls off to .80. At this level in particular one wants the missing information on stability over several months, and on predictive validity.

EVALUATION. The manual and technical report are disappointing. The developers obviously made an effort to give honest advice, to qualify statements, and to present extensive evidence. Outright conflict with the *Test Standards* is rare. Yet the manual and technical report seem unlikely to give users a sound understanding of the test. The facts I have found salient in assessing the test are located in widely scattered places, sometimes deeply buried. Some of the needed statistics are not reported even though the data were in hand. There was massive data collection, yet no resources were allocated to one-year follow-up studies. Tabular data are amassed to the point of throwing dust in the reader's eyes. It is agreed, for example, that alternate-forms reliabilities are pertinent, and internal-consistency coefficients less suitable. Why precede the two tables of relevant alternate-forms coefficients with *three* tables of K-R 20 coefficients? (Worse, the K-R 20 values are computed by a lazy-man's formula that probably inflates the coefficients.)

In the discussion of norms the information overload is even greater. The three-stage sampling plan appears to have been very well designed. The technical report offers three tables by way of evidence on the adequacy of the norms. The first table defends the representativeness of the first-stage selection of 2,000 school districts, but leaves us with no direct report on the 69 systems actually used in the norms. The second table, presented without comment, is a regional breakdown of the actual sample. The careful reader discovers that pupils in the Southeast region are overrepresented by 50 percent, with corresponding deficits in Midwest and West. Since the Southeast usually has conspicuously different score distributions, this suggests that the norms are more in error than the developers intended. How much difference does this make? What went wrong in the sampling? Why is the unrepresentativeness not discussed? The third table displays, for each grade, the percentage of standardization cases from various sizes of school system, and shows that each of these distributions matches the distribution for the U.S. population for all grades pooled. But is the U.S. distribution uniform in all grades? One suspects not. In sum, despite a flood of technical information, our confidence in the norms has to rest on our confidence in the authors.

ALP is a carefully made test. Preparation of items is good on the whole. The technical characteristics are entirely adequate, despite the faults noted. Probably ALP will forecast end-of-year grades as well as other tests do, at least in grades beyond the second. Yet it is hard to see what niche it will fit into. It cannot supplant the achievement test a good testing program uses early in the school year. It will do less than a good test of nonverbal reasoning to shake up the school by detecting the many pupils who can reason well but who have not mastered basic educational skills. ALP is conservative in conception, identifying the child's achievement to date with his "potential" and emphasizing pre-

diction to the exclusion of diagnosis and prescription.

ARTHUR R. JENSEN, *Professor of Educational Psychology; and Research Psychologist, Institute of Human Learning; University of California, Berkeley, California.*

The ALP is the result of a major effort to produce a series of group mental tests for school use which can fully compete with the best and most widely used group intelligence tests now available. To this reviewer, the effort appears to have been successful.

The various subtests can be classified by inspection of their item contents as Word Relational Concepts, Number Concepts, and Figure Concepts, although, as we shall see, the factorial complexity of the tests is not as great as all the diverse subtest labels might suggest.

The time required for administration is reasonable and comparable to other good tests. The tests are given in two or three sittings.

The manual for administration is excellent and contains all the information one could wish to find in a test manual.

PUBLISHER'S CLAIMS VS. REALITY. In the present climate of popular criticism of intelligence tests, which has culminated in their being banned in some school systems, the publishers of ALP seem to have gone all-out to make their product appear to be something other than what it actually is, namely, a good intelligence test. The publisher's blurb, for example, claims the "ALP is more than just a new measure of scholastic aptitude—it is a totally new concept for assessing school ability." The title of the test itself is misleading. There is nothing "analytic" about the scores it yields; there are no profiles or special diagnostic ratios. And the concept of "potential" is quite meaningless with respect to scores on any psychological tests. Some readers are apt to construe the test's title as suggesting a measure of innate ability—a claim not made by the publishers and for which no appropriate evidence is available. Moreover, the test is not best regarded as a measure of "learning" ability. If "learning ability" means rate of acquisition of skills or knowledge, independent of initial status, then the ALP, to the extent that it is like other intelligence tests, is not a measure of "learning ability." The ALP is very much like most other intelligence tests, and ample research has shown that intelligence test scores have relatively low correlations with learning measures or "gain scores" which are independent of initial status.

WHAT ALP MEASURES. Actually, the ALP measures nothing new, nothing different from what is measured by, say, the *Lorge-Thorndike Intelligence Tests*. It is old wine in a new bottle. But the old wine is excellent and the new bottle is indeed attractive. As anyone who has had much experience in factor analyzing a large variety of mental ability tests can readily see from casual inspection of the various ALP subtests, the old wine is nothing other than Spearman's *g*, the general intelligence factor. The various subtests would be somewhat differentially loaded on what Cattell calls "fluid" and "crystallized" *g*, and they can also be grouped in a way that would correspond quite closely to the verbal and nonverbal parts of the Lorge-Thorndike. Through factor analytic experience with many kinds of tests one comes readily to recognize the types of items most heavily loaded with *g*. It would be extremely difficult to make up more different kinds of *g*-loaded items than the authors have succeeded in composing—*g*-loaded items appropriate in difficulty and interest for every grade level from 1 to 12. This heavy *g* characteristic of the ALP tests is most interesting in view of its authors' emphasizing in the manual for administration that the ALP "was *not* developed within a specific theoretical framework concerning the nature of mental ability or intelligence. Thus, the tests were designed to measure neither a single, general ability factor nor to provide factorially 'pure' measures of somewhat discrete mental functions. Tests appearing in each battery were selected solely from the standpoint of their contribution to the prediction of academic success."

FACTORIAL COMPOSITION. To determine the factorial complexity of the ALP, this reviewer performed a factor analysis separately on Forms A and B of each of the five batteries, using the matrix of subtest intercorrelations for each battery provided by the publisher. (Specifically, a principal components analysis was done, followed by a varimax rotation of the components having Eigenvalues greater than 1.) Factorially, Forms A and B are very equivalent. A general factor accounts for between 42 and 59 percent of the total variance in each battery. With few exceptions the various subtests are highly similar in the *g* loadings, which are almost uniformly high (i.e., .70 to .80). In two of the batteries (Primary 1 and Elemen-

tary) no second factor emerged. And in no battery did more than two factors emerge. Rotation of these factors simply divides the variance between verbal and nonverbal factors. In brief, the reviewer's analysis indicates that the variance in these tests is mainly attributable to a large general factor (g), accounting on the average for about 52 percent of the variance, and to two small group factors, verbal and nonverbal (or numerical-spatial), together accounting for about 10 to 13 percent of the variance. The remaining 30 to 35 percent of the variance is attributable to measurement error (less than 10 percent) and to factors specific to the various subtests (about 20 to 25 percent). We can estimate that the total score on the ALP correlates with g between .65 and .77 for the various batteries, with an average correlation of about .72. In this respect, then, the ALP closely resembles most other good tests of general intelligence, which means that it measures, more than anything else, the subject's ability to see complex or abstract relationships, or in Spearman's words, "the eduction of relations and correlates." Whether we like it or not, this is the ability which, more than any other, enters into scholastic achievement under the instructional conditions of present-day schools and as assessed by the traditional criteria of school grades and achievement test scores.

"CULTURE-FAIRNESS." In their manual the authors warn that "Pupils who are poorly motivated or who have not had the opportunity to learn the broad, general types of behaviors sampled by the tests should have these limiting factors taken into consideration by teachers and counselors in interpreting the test results. This is particularly true for pupils who have experienced severe cultural deprivation." Does this mean that the ALP has lower validity in predicting the scholastic performance of "culturally deprived" children? We do not know. But we can guess from experience with similar tests that it will predict scholastic achievement (under normal school conditions and assessed by standard tests) as well for the disadvantaged as the majority of children. If it is realized that the *causes* of any particular child's score are unknown (unless specifically investigated) and that the score is simply a statistical predictive device, there need be no concern about the test's "fairness" to all subpopulations of pupils. However, if one should imagine that he is getting at something more "profound" than this, he should be made aware that the test is no more "culture-fair" than most other group tests now on the market. So far as we know, all tests with a high g saturation, whether or not they are called "culture-fair" or look "culture-fair," show very substantial social class differences in performance. The ALP will be no exception. We can expect lower socio-economic status children, on the average, to score about one standard deviation below middle-class children on the ALP.

This reviewer would regard tests like the ALP (and all other standard intelligence tests) as "unfair" only in the sense that they assess such a limited and homogeneous set of abilities. These abilities can be called "intelligence" or g, which is indeed highly correlated with scholastic performance. But g is not the whole spectrum of human ability, nor is it the only ability that can be marshaled for scholastic achievement. One would like to see a broader assessment of children's abilities, especially among groups who are relatively low in g, in hopes of finding other abilities and talents which can serve in the educational process of making schooling rewarding even for children with below average academic aptitude. Although tests like the ALP may give the impression, with all their diverse subtests, that they are getting at a broad assessment of many abilities, they are actually very unidimensional. Unless supplemented by other forms of assessment they tend to rank-order pupils along a single dimension of aptitude. Does the form of school instruction in turn attempt to maximize the correlation between "scholastic aptitude" (as represented by nonscholastic tests of g) and scholastic achievement? Highly homogeneous tests of ability, yielding a single score, may not be the most useful instruments in schools tending toward a diversity of curricula and instructional programs intended to make school beneficial to a wide range of individual differences. Only by inventing additional tests with low loadings on g will we be able to discover possible areas of educationally relevant strengths in those children who are below the average in g-type abilities. Unfortunately, there is no standard test battery one can recommend at present to fill this need.

NORMS. The ALP norms are based on a sample of 165,000 pupils in 75 school systems selected so as to be highly representative of the U.S. school population according to the latest census. Norms for different regions of the U.S.

or for different types of communities (e.g., rural-urban, lower–class—middle–class, etc.) are not provided. The authors suggest that school systems should develop their own local norms, presumably by accumulating large numbers of test results and converting the raw ALP scores to normalized standard scores. A further step would be to determine the regression equation for the school's particular achievement measures as "predicted" from the ALP. Local norms, if properly established and maintained up-to-date, make a good deal of sense, considering the fact that the average level of scholastic achievement in a school or a community is highly related to a host of community characteristics over which the schools themselves have little or no control, such as the educational level of the adult population, home ownership, cost of housing, proportion of native-born whites, rate of unemployment, proportion of professional workers, etc. Comparison of a particular school's or community's scores with overall national norms, though it may serve the purpose of describing one aspect of the school population, is of little or no value in dealing with an individual pupil. On the other hand, there may be some value in comparing an individual's score on an intelligence test with an assessment of his scholastic achievement. This is best done by putting the intelligence and achievement scores on the same scale (e.g., normalized standard scores with the same mean and standard deviation) normed on the same reference population. The authors of the ALP emphasize, correctly, I believe, that the subtests are designed, for the most part, to be "relatively free from specific school-learned skills. The tests ....do assess learned abilities gained from a number of somewhat diffuse sources whose exact nature cannot be clearly specified." The detection of large and reliable discrepancies between a measure of extra-scholastically acquired skills and measures of scholastic achievement can be of diagnostic value, both for individual children and for the means of classrooms and of whole schools. Detection of "underachievers" for special attention is a useful function of ability tests. The ALP can serve this purpose as well as any other intelligence tests on the market. The "cut-off" discrepancy between ALP and achievement scores that would pick out "underachievers" is not specified, but it is a largely arbitrary matter anyway. (It should probably be at least twice the test's standard error of estimate.) The fact that there will be about as many "over-achievers" as "underachievers" belies the test's label, "Analysis of Learning *Potential*," since theoretically no one should be able to exceed his "potential."

SCALES AND SCORES. The ALP provides five types of derived scores, all of which can be found in tables in the Norms-Conversion Booklet. The Index of Learning Potential (ILP) is a normalized standard score with a mean of 50 and a standard deviation of 15. The reference group is based on chronological age, within 2 or 3 month intervals. Although the term IQ is assiduously avoided by the authors, the ILP is essentially a deviation IQ, comparing the individual's standing among others of his age, and thus it has the same meaning that IQ has on the Lorge-Thorndike or any other up-to-date tests of intelligence which provide deviation IQ's. To put the ILP on the same scale as the IQ, one must simply add 50 to the ILP.

The General Composite Standard Score (GCSS) is normed on pupils making normal progress within grade levels limited to an 18-month age range within each grade level, a range that comprises the middle 80 to 90 percent of pupils in any one grade. The GCSS also has a mean of 50 and a SD of 15.

So, if you want to know where a pupil stands with respect to those of similar chronological age, regardless of their grade level, you use the ILP. If you want to see where a pupil stands in relation to others in his grade who are making normal progress (presumably the middle 80 to 90 percent of the age range of children in regular classes), you use the GCSS.

The Composite Prognostic Score (CPS), also with a mean of 50 and a SD of 15, is a score based on a weighted combination of several subtests that correlates maximally with either reading or mathematics achievement. The CPS provides hardly any appreciable gain over the total ILP or GCSS in terms of correlation specifically with achievement in verbal or quantitative curricula.

Finally, the ALP raw scores can be converted to percentile ranks and to stanines.

The most useful score, one might think, would be one which is on the same scale as the scholastic achievement test, so that direct comparison of "intelligence" (i.e., extra-scholastic achievement) and scholastic achievement would be possible. But most standard achievement tests provide normalized T scores with a mean of 50

and a standard deviation of 10. The ILP unfortunately combines the mean of achievement tests (i.e., 50) with the standard deviation of IQ tests (i.e., 15).

Scoring of the ALP by hand (made easy by simple templates) or by machine is possible. IBM, MRC, and Digitek answer sheets are available, and scoring can be done locally or by the publisher's scoring service.

RELIABILITY. The internal consistency reliability (Kuder-Richardson) is very high, ranging from .92 to .97 in different grades. The alternate-forms reliability is also quite satisfactory (.80 to .94).

VALIDITY. The ALP is expressly intended to predict scholastic achievement. Its validity was established by correlating the GCSS scores at each grade level. In the high school grades the validities are nearly as high as reliability will permit. If the authors' chief aim was to "predict" concurrent scholastic achievement with the ALP, it is hard to see how they could have been any more successful. The ILP score was not used in the validation evidence but would probably yield validity coefficients very similar to those for the GCSS.

The ALP correlates .83 with Lorge-Thorndike total IQ (at grade 5) and has correlations between .29 (Mechanical) and .86 (Verbal Reasoning + Numerical Ability) with various parts of the *Differential Aptitude Tests*.

TECHNICAL INFORMATION. In addition to the very complete printed manual that accompanies the test, the publishers have prepared a mimeographed Preliminary Technical Report which contains much more detailed information about the construction and validation of the ALP. It contains information which will be of primary interest to educational researchers and could also serve as a model in courses on the theory and practice of test construction. One of its most useful features is an appendix which gives normalized standard scores, with a mean of 500 and SD of 50, for the entire five batteries. This puts all the five tests, spanning grades 1 to 12, on the same scale, a feature which enhances the test's usefulness for longitudinal studies. The method of scaling the various tests is fully described in this report.

USEFULNESS. If one wants to use a test of "general intelligence," or "IQ," or "scholastic aptitude," the ALP is about as good as any of the current top competitors in the field. Since it correlates almost as highly with tests of scholastic achievement as reliability permits, one might ask, Why use the ALP at all? Why not just measure achievement? Indeed, why not? Unless the school authorities have some special purpose intended which calls for a measure of ability which is not directly based on the subject matter of the curriculum, there would seem to be little justification for the time, bother, and expense involved in getting group-administered intelligence test scores on all pupils in a school. Achievement scores should suffice for most purposes. Other diagnostic measures (including nonscholastic measures of general intelligence) would be called for in those cases where a pupil's scholastic performance is markedly deviant. As previously suggested, an intelligence test used along with achievement tests can spot the underachievers who may then receive further diagnosis to determine the causes of the underachievement. Since the correlation is so high between ALP and achievement, those pupils who show a marked discrepancy between the two scores would warrant special attention, especially if achievement scores are markedly *below* the ALP scores. The ALP scores, reflecting more extra-school influences, could also be used in the same way for comparing the average achievement of whole classes, schools, or communities. The ALP can also serve as a control variable in educational experiments.

SUMMARY. From both technical and practical standpoints, the *Analysis of Learning Potential* (despite its title being a misnomer for "intelligence test") is an excellent battery of five group tests, covering grades 1 to 12, for measuring general intelligence or scholastic aptitude by means of test materials for the most part not specifically taught in school. Its correlations with tests of scholastic achievement are exceptionally high. The ALP appears to be fully competitive with the best group intelligence tests currently available.

[335]

★**Boehm Test of Basic Concepts.** Grades kgn–2; 1969–70, c1967–69; BTBC; Form A ('69, 16 pages in 2 booklets); manual ('70, 22 pages); $5.90 per 30 tests; 50¢ per manual; $1 per specimen set; postage extra; Spanish edition available; (30–40) minutes in 2 sessions; Ann E. Boehm; Psychological Corporation. *

REFERENCE

1. BOEHM, ANN ELIZABETH. *The Development of Comparative Concepts in Primary School Children.* Doctor's thesis, Columbia University (New York, N.Y.), 1966. (*DA* 27: 4109B)

BOYD R. MCCANDLESS, *Professor of Education and Psychology and Director, Educational Psychology, Emory University, Atlanta, Georgia.*

The purpose of the test, as stated by the author, is "to assess beginning school children's knowledge of frequently used basic concepts widely but sometimes mistakenly assumed to be familiar to children at their time of entry into kindergarten or first grade." The reviewer considers this statement accurate, modest, and realistic.

The test was inspired by the author's awareness that many children beginning school do not comprehend many of the printed or spoken instructions taken as "givens" by most teachers, and by her assumption (well supported by survey data) that deficits at the beginning of school are cumulative over time. She hopes to provide an instrument to pinpoint these deficits, lead the way to remedying them, and thus prevent irrelevant interference with school progress.

The initial item content was empirically and apparently somewhat subjectively determined by inspection of curriculum materials, together with checks to see what concepts were difficult or unfamiliar to substantial numbers of children.

The test was finally narrowed to 50 items placed in two test booklets to facilitate administration in two sessions to children in grades K, 1, 2, and 3. Items proved so easy for third graders that the final form of the test includes norms for only grades K through 2, but the test is too easy to be of great value for first graders from middle or higher socioeconomic levels or for second graders of any social class.

Two waves of preliminary testing were conducted before the final form of the test was set up. Item selection was made according to conservative and acceptable principles—e.g., sampling of a range of concepts, point biserial correlations exceeding .30, "even rises of percent-passing values across age levels," and "normal distribution of percent-passing values, centered around .50 for the kindergarten pupils."

Testing with two booklets, each including 25 questions to be answered by making X's on pictures, requires 15 to 20 minutes per booklet. Instructions to teachers state that groups of 8 to 12 may be tested. Although the reviewer has not tried the test with kindergartners, he is skeptical that it can be feasibly administered to groups of this size unless there is a generous supply of proctors.

The booklets are made up of black line drawings on white (Booklet 1) or buff (Booklet 2). For the most part, the drawings are clear, though a few seem ambiguous. The people in the illustrations are appropriately integrated racially. Scoring instructions are clear and the mechanics are about as simple as is possible when working with test protocols for children in kindergarten and grades 1 and 2.

The standardization sample came from five cities, one western, one south-midwestern, one southeastern, and two northeastern. School personnel in each city were asked to administer the test within one high, one middle, and one low socioeconomic school. A disproportionate number of the low socioeconomic class children come from the southeastern city school system. The test author makes no pretense to having a representative U.S. normative sample, but has sampled widely in reasonably representative school systems.

No validity evidence other than face validity is presented, although the face validity is convincing enough. As anyone familiar with kindergarten and first grade instruction will realize, the items tap concepts that children need to know. The author represents the BTBC modestly as a screening device and a guide for instruction. Used thoughtfully, it can be quite useful to teachers. A section in the manual devoted to interpretation and use of the results in instruction is very practical.

CHARLES D. SMOCK, *Professor of Psychology, The University of Georgia, Athens, Georgia.*

Children enter school with a variety of experiential backgrounds and variation in knowledge of the physical and social environments. Current interest in cognitive developmental theory and enrichment of the environments of "disadvantaged" children has increased concern for adequate assessment of their intellectual level upon entering school. Also, curriculum development specialists have found it necessary to modify the typical first grade curriculum in order to create effective learning conditions for these children. Of particular importance is the fact that both the available curriculum materials and "readiness" tests assume a set of fundamental concepts which many disadvantaged children do not yet understand; for example, the under-