Sidel, Mark, "University Enrollment in the People's Republic of China, 1977-1981: the examination model returns," Comparative Education, Vol. 18, No. 3, 1982, pp. 257-69.

Streeten, Paul, "The Limits of Development Research," in Social Sciences and Public Policy in the Developing World, Lexington, Mass., Lexington Books, 1982, pp. 21-56.

DOCUMENT RESUME

ED 236 166 TM 830 627

AUTHOR Hunter, John E.
TITLE The Dimensionality of the General Aptitude Test Battery (GATB) and the Dominance of General Factors over Specific Factors in the Prediction of Job Performance for the U.S. Employment Service.
INSTITUTION Michigan State Dept. of Labor, Detroit. Michigan Employment Security Commission.
SPONS AGENCY Employment and Training Administration (DOL), Washington, D.C.
REPORT NO USES-TRR-44
PUB DATE 83
NOTE 54p.; Written under contract to the Northern Test Development Field Center.
PUB TYPE Reports - Research/Technical (143)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Aptitude Tests; Career Counseling; *Factor Structure; Job Performance; Job Placement; Predictor Variables; *Test Construction; Testing Programs; *Test Validity; *Vocational Aptitude
IDENTIFIERS Employment Service; *General Aptitude Test Battery; *General Factor (Intelligence)

ABSTRACT

The structure of this report is as follows: First, specific aptitude theory and general ability theory, the two theories of the relation between ability and job performance, are presented and differentiated. Second, there is a discussion of problems in the current use of the General Aptitude Test Battery (GATB) and of problems in the use of multiple cutoff scores based on small sample data. Third, there is a discussion of the dimensionality of the GATB in traditional factor analytic terms. That is, there is an analysis of the correlations between specific aptitude scores over persons. This culminates in a breakdown of each of the nine specific aptitudes and the three general ability composite scores in terms of general factor variance, specific factor variance, and error variance. Fourth, there are the data on the correlation of aptitude validity coefficients across jobs. These data tend to support the general ability theory. Further evidence for the general ability theory is presented in connection with spatial aptitude. The practical implication of these findings is that the search for applications of specific aptitudes to the prediction of job performance will require either very large sample size studies (N=1,000) for particular jobs or the identification of special job families. (PN)

USES TEST RESEARCH REPORT NO. 44

THE DIMENSIONALITY OF THE GENERAL APTITUDE TEST BATTERY (GATB) AND THE DOMINANCE OF GENERAL FACTORS OVER SPECIFIC FACTORS IN THE PREDICTION OF JOB PERFORMANCE FOR THE U. S. EMPLOYMENT SERVICE

DIVISION OF COUNSELING AND TEST DEVELOPMENT
EMPLOYMENT AND TRAINING ADMINISTRATION
U. S. DEPARTMENT OF LABOR
WASHINGTON, D. C. 20213

1983

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

LIST OF TABLES (CONT.)

6

ACKNOWLEDEGMENT

The United States Employment Service conducts a test research program for developing testing tools useful in vocational counseling and placement.

The purpose of this series of reports is to provide results of significant test research projects as they are completed. These reports will be of interest to users of USES tests and to test research personnel in State agencies and other organizations.

The paper cumulates the findings of validity studies done by hundreds of analysts working for the U.S. Employment Service over a 45 year span. Special thanks go to Ron Boese who did all the computer runs underlying this work at the field center in North Carolina.

This report was written by Dr. John E. Hunter, Michigan State University, under contract to the Northern Test Development Field Center, Michigan Employment Security Commission, Detroit, Michigan. The report was prepared for printing by staff of the Western Test Development Field Center, Utah Department of Employment Security.

ABSTRACT

The correlations between the 9 aptitudes measured by the GATB show that the aptitudes break into three clusters which define three general factors: A cognitive (thought, reasoning, learning) factor with G, V, N; a perceptual factor with S, P, Q; and a psychomotor factor with K, F, M (see Table 2 for explanation of the aptitudes). Correlations between aptitude validity coefficients over 515 jobs show the same structure. Correlations between aptitude validity coefficients for aptitudes measuring the same general factor are high enough to show that virtually all the validity of the individual aptitudes is due to the validity of the general factors rather than the specific factors. Thus the unit weight hypothesis is satisfied within clusters and there is little information lost in rescoring the GATB in terms of three composite scores: GVN, SPQ, and KFM. The perceptual factor is predicted almost perfectly by the cognitive and psychomotor factors and hence the perceptual composite SPQ will contribute little to the predictive power of the battery.

INTRODUCTION

The U. S. Employment Service has completed 515 validation studies over the last 40 years. These studies cover a sampling of all the jobs in the U. S. economy. All studies use the same test battery, the General Aptitude Test Battery (GATB). This report covers the first phase of the application of validity generalization methodology to this data base. The question asked in this phase was: Should the GATB be scored in terms of the nine specific aptitudes measured or should it be scored in terms of three general factors as implied by factor analysis? The data to be presented here shows that the answer is to score in terms of general abilities rather than specific aptitudes.

The central question for this report has long been an issue in employment research. The issue can be couched in terms of two theories. The specific aptitude theory argues that job performance can be best predicted by matching the content of the test items to the content of job materials. For example, the specific aptitude theory would predict that a clerical job working with words would be best predicted by verbal aptitude rather than general cognitive ability and that a clerical job working with numbers would be best predicted by numerical aptitude rather than general cognitive ability. The general ability theory argues that jobs are learned as aptitudes in their own right. This learning is governed by general cognitive ability and hence general cognitive ability will be a better predictor than specific aptitudes.

These two theories can be stated in the form of path models. The path models show that the theories can be tested against one another using a statistic from meta-analysis (quantitative methods of cumulating results across studies). If data are available on the validity of specific aptitudes for a wide variety of jobs, then the two theories predict different patterns of validity across jobs. In particular, if the validity coefficients for two specific aptitudes are correlated across jobs, then the specific aptitude theory predicts a very low correlation while the general ability theory predicts a perfect correlation. This differential prediction can be tested in the U.S. Employment Service data base. The test requires the control of the effects of sampling error, but formulas from meta-analysis are available which do just that.

The analysis of the data for the GATB shows that the classic theories are an oversimplification of reality. The two theories pit one general factor--intelligence--against specific aptitudes. Analysis of the GATB suggests not one but three general abilities: Cognitive ability, perceptual ability, and psychomotor ability. Cognitive ability refers to concepts such as thought, planning, learning, memory. The distinction between cognitive ability and perceptual ability means that cognitive ability is a more narrow concept than general intelligence. However, the correlation between cognitive and perceptual ability is .88 and hence the distinction between cognitive ability and intelligence is quantitatively small.

Most theorists have not considered psychomotor ability in the same context as mental ability. Thus there is nothing controversial about the use of two general factors for cognitive and psychomotor ability. However, the simultaneous presence of mental and psychomotor abilities in the GATB brought out the distinction between the cognitive and perceptual abilities. The perceptual aptitudes are distinct in that they are highly correlated with both the cognitive aptitudes and the psychomotor aptitudes. The three factors span only two dimensions; perceptual ability is redundant on the other two abilities. That is, if all three general abilities are perfectly measured, then general perceptual ability can be perfectly predicted from cognitive and psychomotor ability. Thus, if the general abilities were perfectly measured, then perceptual ability would be superfluous for the prediction of job performance. For the imperfect measures created by using composite scores from the GATB, the multiple correlation of the perceptual composite is only .80. Thus research has used all three composite scores. However, there is only one very narrow job classification (industrial setup work) where the perceptual composite score added to prediction (Hunter, Note 2).

The structure of this report is as follows: First, the two theories of the relation between ability and job performance are presented and differentiated. Second, there is a discussion of problems in the current use of the GATB; problems in the use of multiple cutoff scores based on small sample data. Third, there is a discussion of the dimensionality of the GATB in traditional factor analytic terms. That is, there is an analysis of the correlations between specific aptitude scores over persons. This culminates in a breakdown of each of the nine specific aptitudes and the three general ability composite scores in terms of general factor variance, specific factor variance, and error variance. Fourth, there is the data on the correlation of aptitude validity coefficients across jobs. This data tends to support the general ability theory. Further evidence for the general ability theory is presented in connection with spatial aptitude. The practical implication of these findings is that the search for applications of specific aptitudes to the prediction of job performance will either require very large sample size studies (N=1000) for particular jobs or the identification of special job families. Severe limitations on the use of current methods of job analysis for the identification of such job families have been found in subsequent research (Hunter, Note 2).

Two Theories of Ability and Job Performance

In the first third of this century, research was focused on one general cognitive ability: intelligence. The research findings are presented in a schematic path diagram in Figure 1. Researchers first noted that intelligence could account for the high correlations between educational achievement tests. Then intelligence was found to account for high correlations between all types of problem solving such as analogies, number series, induction, deduction, etc. Finally intelligence was shown to be related to job performance in both complex and simple jobs.

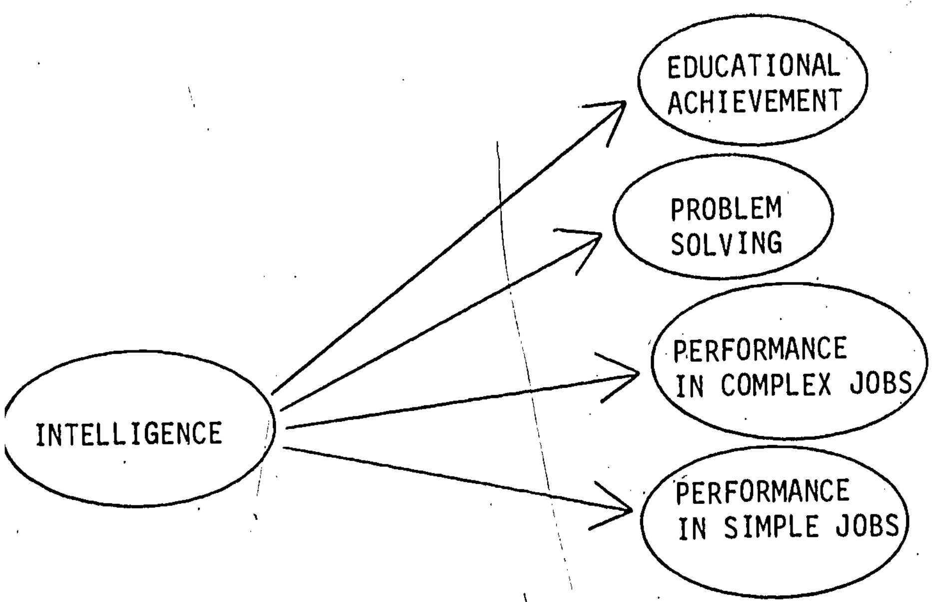


Figure 1. The findings of research on ability in the first third of the century.

In the middle third of the century, there was much more work on the development of specialized kinds of cognitive tests. This led to the discovery of a more differentiated relationship between abilities which has been sketched in abbreviated form in Figure 2. Correlations between tests showed them to cluster. The clusters in test correlations can be explained by postulating cognitive abilities that are more general than the skills assessed in specific tests but less general than intelligence. These intermediate abilities will be referred to here as "aptitudes."

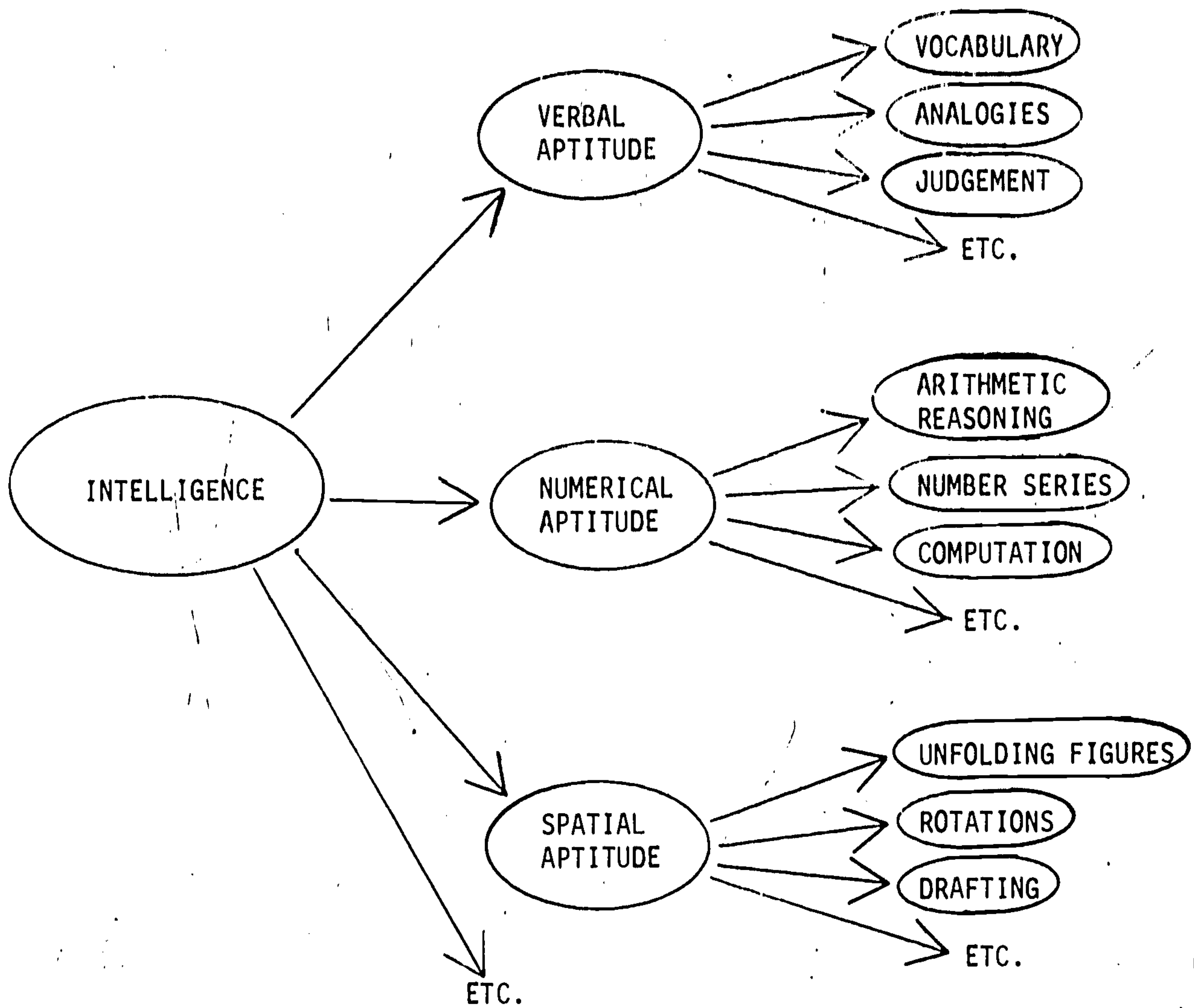


Figure 2. The research findings on general ability and specific aptitudes of the middle third of the century.

The path model in Figure 2 is established by a process of two stage factor analysis. First, the correlation matrix between tests is subjected to confirmatory factory analysis. The factor for each test cluster is an aptitude. Second, the correlation matrix between aptitudes is subjected to confirmatory factor analysis. If there were just one factor required to account for the correlations between aptitudes, then that factor would be intelligence. Most second order factor analyses have found good fit for one general factor (McNemar, 1964; Humphreys, 1962, 1979).

Some authors have confused the issue of general ability versus specific aptitudes, arguing that there can only be one or the other. For example, Horn (Note 3) argues that since factor analysis persistently finds three factors underlying the tests in the Wechlser, it is not possible to have a unitary concept of intelligence. Yet Horn never subjects the correlations between the three factors to second order factor analysis, though these factors are highly correlated. Part of the problem may be a misunderstanding as to the causal ordering in his path diagrams. He postulates each factor as explaining the correlation between the tests in the corresponding cluster. This means that the causal arrows run from factor to test. Horn's Figure 1 shows the causal arrows running from test to factor. Thus although Horn notes the use of intelligence to explain the correlations between abilities, he does not note that the correlations between his primary factors are not explained in his model. The point of Figure 2 is not that there is only one general factor. Rather Figure 2 notes that two levels of factors are present in the data: aptitudes which explain the fine grain clustering in cognitive tests, and a general (second order factor) ability which explains the correlations between the aptitudes.

The factor structure of the GATB to be presented below departs from the single factor intelligence model of Figure 2 in that three general factors are postulated. However, this departure is due to the presence of psychomotor aptitudes and a corresponding psychomotor general ability. Psychomotor aptitudes are not present in the usual study of mental abilities. Furthermore, it is the presence of the psychomotor aptitudes which permits the distinction between the general cognitive ability factor and a perceptual ability factor which are correlated .88. Without the psychomotor tests, this fine distinction would be lost. This would be especially true in data sets based on sample sizes less than 24,000. Without the psychomotor tests, a one factor model would fit the data very closely.

Specific Aptitude Theory and General Ability Theory

Two theories of job performance have grown up in response to the development of specific aptitudes. Some theorists have argued that job performance will be best predicted by specific aptitudes. Some have argued that general intelligence will predict performance better than specific aptitudes. Because of the massive sampling error in particular studies, it is possible to cite studies on both sides of the issue. The debate can be sharpened by expressing both theories as path diagrams. Figure 3 shows such path diagrams for two jobs and two specific aptitudes. These path diagrams lead to a method of distinguishing the theories using meta-analysis which will be developed in the next section.

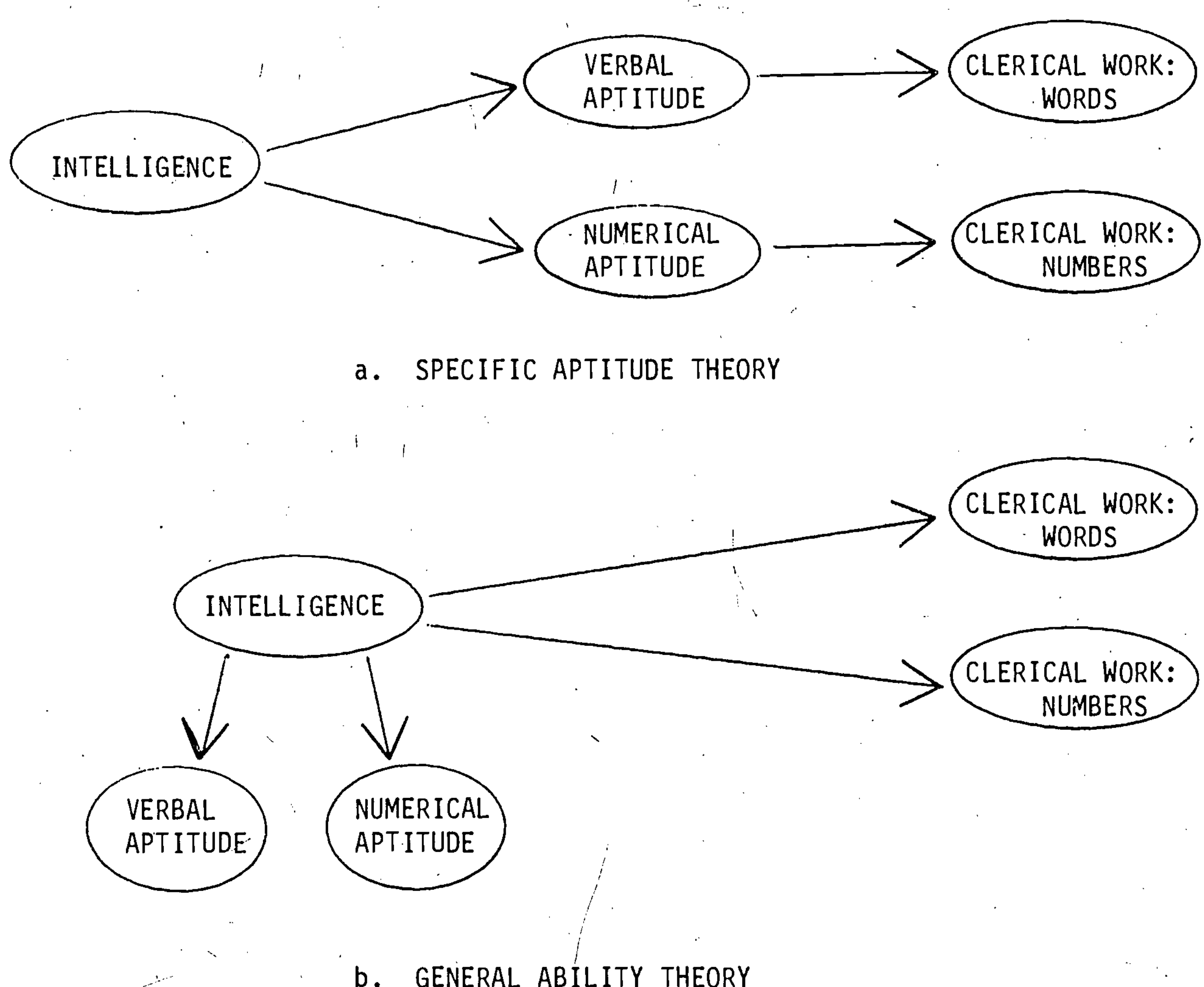


Figure 3. An abbreviated path analytic presentation of two theories relating ability to job performance: specific aptitude theory and general ability theory.

Wernimont and Campbell (1968) argue for specific aptitude theory. They believe that the test which will best predict job performance is that test which is most similar to the job in terms of the behaviors sampled. For example, a clerk running a posting machine works with numbers; thus numerical aptitude should be a better predictor of job performance than general intelligence for posting. On the other hand, a filing clerk works with words; thus verbal aptitude should be a better predictor for filing. Implicit in the Wernimont and Campbell argument is a further argument. Intelligence is correlated with performance in a job only because it is correlated with the relevant specific aptitude.

Figure 3a presents specific aptitude theory for two jobs in the form of a path diagram. In this diagram, there is a direct causal effect of verbal aptitude on the filing job and a direct causal effect of numerical aptitude on the posting job. However, general intelligence is correlated with job performance only indirectly because it has a causal effect on each specific aptitude. Specific aptitude theory can be viewed as an extension of the hierarchical model of Figure 2 in which job performance is viewed as a specific skill entered at the bottom of the hierarchy. The causal arrow could be interpreted as a developmental theory in which the job performance skill is a further differentiation of the specific aptitude to which it is linked.

Humphreys (1979) has argued for general ability theory. His experience with large sample military studies is that it is very hard to find large differences in the pattern of aptitudes in multiple regression equations for different jobs. For example, the patterns are identical for pilot and navigator even though the pilot is dealing with instruments and dials and a three dimensional panorama while the navigator is dealing with maps and numbers. If there are few differences in patterns, then the implication is that the prediction of job performance is being made on the basis of general intelligence rather than the specific aptitudes (Jensen, Note 4).

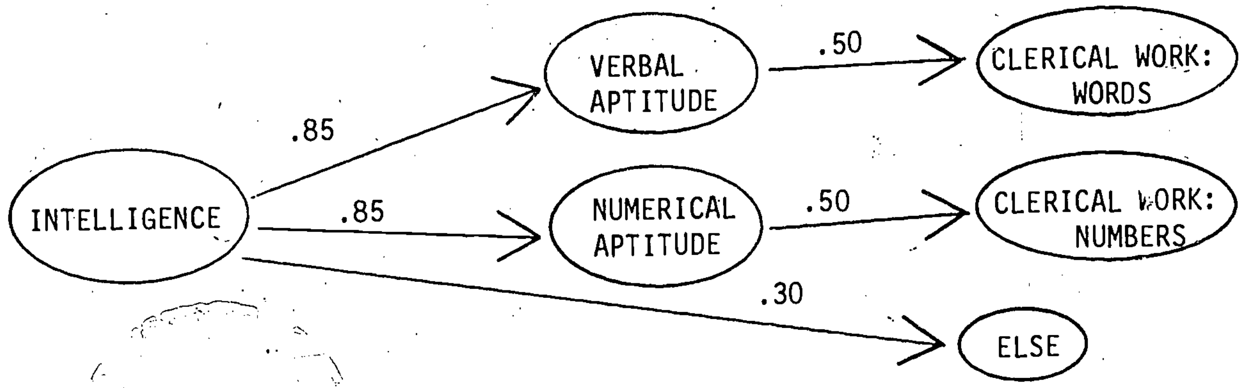
The general ability theory can be derived from the following reasoning. The skills represented in numerical ability matter little in the mastery of the posting machine. Instead, the running of the posting machine is learned as a new skill in its own right. The learning and forgetting which take place during the learning of the posting machine are governed by general intelligence.

Figure 3b presents a path diagram for general ability theory for two jobs and the two specific aptitudes used in the path diagram for the specific aptitude theory (i.e., Figure 3a). According to this diagram, specific aptitudes and job performance are correlated only because both are causally effected by intelligence. That is, the correlation between specific aptitudes and job performance is the "spurious" effect of joint causation by a common antecedent variable. The path diagram in Figure 3b can be viewed as an extension of the hierarchical model of Figure 2 in which each job performance is entered as an aptitude in its own right.

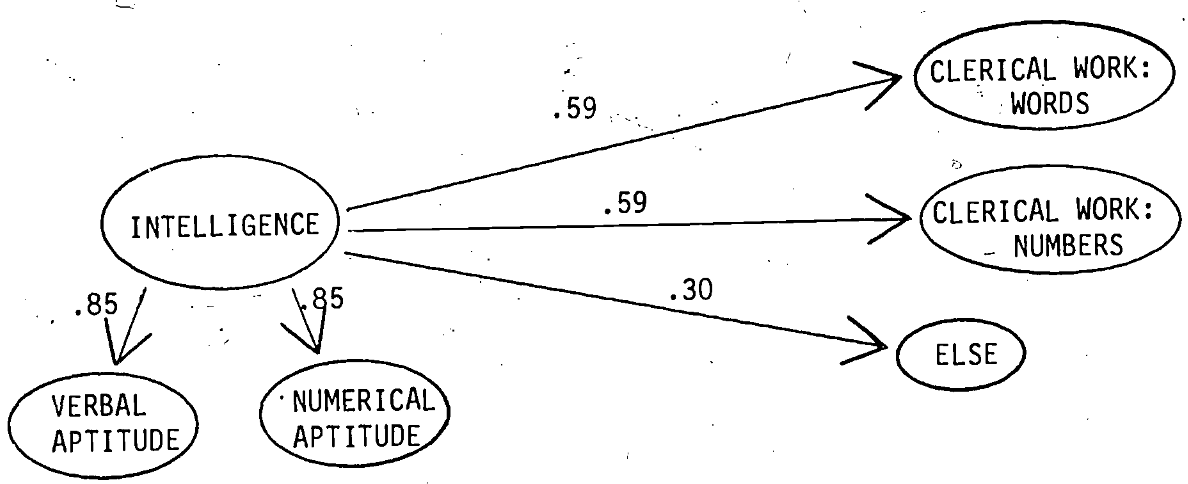
Empirical Differentiation of the Two Theories

The direct test of the path models in Figure 3 would require very large sample data using exactly the same specific aptitudes to predict each job. The U.S. Employment Service data base of 515 validation studies offers a wide variety of jobs with exactly the same specific aptitudes used in each study. However, the average sample size for these validation studies is N = 75 which is too few for the multiple regression required for direct path analysis. The purpose of this section is to derive a method of pitting the two theories against each other in a way that can be applied to the Employment Service data base. The statistic that will be shown to work is meta-analytic: the correlation between aptitude validity coefficients across jobs. The effectiveness of this procedure will be shown in a simplified model in which there are only three job types. The derivation will proceed in three steps. First, the path models of Figure 3 will be fully quantified using assumed but realistic numbers. Second, the quantified path models will be used to generate the validity of each aptitude for each job; i.e., the models will generate predicted correlations between each aptitude and job performance for each of the three job types. Six such sets of aptitude validities will be used to simulate the Employment Service data base of 515 sets. The validity coefficients will then be correlated across jobs. These correlations will be shown to be very different for the two models. This theoretical difference in correlation of aptitude validity across jobs can then be used to test the models against the Employment Service data.

The path models of Figure 3 will be quantified by making several assumptions that are common to both models and approximately true for existing data. The two specific aptitudes considered will be verbal and numerical aptitude. They will be assumed to be correlated .72 which implies that each correlates .85 with intelligence. The correlation of verbal aptitude with performance in clerical jobs dealing with words is assumed to be .50 while the correlation of numerical aptitude with performance in clerical jobs dealing with numbers is assumed to be .50. Finally, there is a third job category of "else." The validity of intelligence for job performance in the else category is assumed to be .30. The quantified path models for each theory are presented in Figure 4.



a. SPECIFIC APTITUDE THEORY



b. GENERAL ABILITY THEORY

Figure 4. A specific quantification of two theories of ability and job performance designed to show how they can be pitted against one another.

17

Table 1

Hypothetical data derived from Figure 4 which shows how to differentiate between two theories of ability and job performance on the basis of correlations between aptitude validity coefficients across jobs (V = verbal aptitude, N = numerical aptitude, G = intelligence)

Aptitude validity coefficients for 6 jobs:

		SPECIFIC APTITUDE THEORY			GENERAL ABILITY THEORY		
Job	Job Type	V	N	G	V	N	G
1	Words	.50	.36	.42	.50	.50	.59
2	Numbers	.36	.50	.42	.50	.50	.59
3	Else	.26	.26	.30	.26	.26	.30
4	Words	.50	.36	.42	.50	.50	.59
5	Numbers	.36	.50	.42	.50	.50	.59
6	Else	.26	.26	.30	.26	.26	.30

Correlations between validity coefficients across jobs:

	SPECIFIC APTITUDE THEORY			GENERAL ABILITY THEORY		
	V	N	G	V	N	G
Verbal Apt.	1.00	.33	.81	1.00	1.00	1.00
Numerical Apt.	.33	1.00	.81	1.00	1.00	1.00
Intelligence	.81	.81	1.00	1.00	1.00	1.00

The path models in Figure 4 generate the validity of each aptitude or ability in the prediction of job performance of each of the three job types. Consider then six hypothetical jobs as shown in Table 1. Given the job type, we can use Figure 4 to generate the correlation between job performance and verbal aptitude, numerical aptitude, and intelligence. There will be a different set of validities predicted by each of the two models. For a job with words, specific aptitude theory predicts a validity of .50 for verbal aptitude and a validity of .36 for numerical aptitude. This pattern is reversed for a job with numbers; the validity is .36 for verbal aptitude and .50 for numerical aptitude. The validity of intelligence is an intermediate .42 for both jobs.

The pattern is strikingly different for general ability theory. That theory predicts that the validity for each specific aptitude will be .50 for both jobs. Intelligence is predicted to correlate higher than either on both jobs at r = .59.

To bring out the difference between the two theories, each set of validity coefficients can be correlated across jobs. That is, the correlations for each job are treated as if they were scores for that job. The columns of correlations are then correlated across jobs as if they were scores. These correlations of aptitude validity across jobs are shown in Table 1. Specific aptitude theory predicts that the correlation between verbal aptitude validity and numerical aptitude validity across jobs will be .33. The correlation predicted by general ability theory is 1.00. The difference between a low correlation and a perfect correlation is very great.

The comparison of correlations for real data is complicated by the presence of sampling error. Consider the prediction of a perfect correlation made by general ability theory. The effect of sampling error in the validity coefficients is similar to the effect of error of measurement on the correlation between scores on two parallel test forms where the correlation is reduced to the reliability of the tests. For the sample sizes in the Employment Service data, the perfect correlation would be reduced to .47. The correlation of .33 for the specific aptitude theory would be reduced to .16. As it happens, the effect of sampling error is known and there is a formula to correct for its effect (Hunter, Note 1; Hunter, Schmidt, and Jackson, 1982).

There is one study in the literature which has looked at correlations between aptitude validity across jobs. Schmidt and Hunter (1978) correlated aptitude validity coefficients for Army data on 35 jobs gathered together by Helm, Gibson, and Brogden (Note 5). The correlation between verbal and numerical was .98. The correlations between spatial aptitude, mechanical aptitude, and shop mechanics averaged .93. The cross correlations averaged .65. The clerical speed test validity correlated .88 with the verbal numerical cluster and .38 with the spatial cluster. This suggests that the prediction of job performance can be done with two general factors rather than six specific aptitudes: a cognitive factor (underlying verbal and numerical) and a perceptual factor.

This study could be criticized on the basis of the performance measure; all the data analyzed in Schmidt and Hunter used training success. Vineberg and Joyner (Note 6) noted in their review of military studies that training success is consistently the best predictor of later job performance. However, it might still be true that training success is predicted by general ability while later performance is predicted by specific aptitudes. In the U.S. Employment Service studies, 425 out of 515 studies used job proficiency as the performance measure. Thus the study reported here does not have the potential problem of the Army data.

Current Use of the GATB: Multiple Cutoff Problems

For 45 years, the U.S. Employment Service has done validation work using the General Aptitude Test Battery (GATB), a set of 12 tests scored as 9 aptitudes. The scores are combined for prediction using a multiple cutoff procedure in which two, three or four aptitudes are considered. This multiple cutoff procedure has a number of problems which are listed below. The main problem is that all employers in all labor markets are forced to use the same top two-thirds, bottom one-third selection ratio regardless of the percentage of workers to be hired. Many have urged the Employment Service to change to multiple regression.

However, a shift to multiple regression raises its own problems. The Employment Service has the same problems of small sample size which plague other field workers. Most of the 515 validation studies have had fewer than 100 persons. This is a very small sample size for multiple regression with nine predictors. Should prediction using the GATB be based on a modified multiple regression procedure such as unit weights (Schmidt, 1971, 1972) or ridge regression (Price, 1977; Rozeboom, 1979)?

Unit weights or ridge regression are used only if the data satisfy two assumptions: (1) All or most variables are indicators of the same underlying factor, and (2) the validity of prediction is due to the general factor rather than the specific indicators. This paper will consider both assumptions. Examination of the aptitude correlation matrix will confirm the findings of others (Fozard, Nuttall, & Waugh, 1972; Nuttall & Fozard, 1971; Watts & Everitt, 1980) that the GATB aptitudes define three factors; a cognitive (thought or reasoning) factor, a perceptual factor, and a psychomotor factor. However, these three factors are very much correlated; the perceptual speed factor can be almost perfectly predicted from the cognitive and psychomotor factors. Thus the three factors span only two dimensions.

In addition to examining these assumptions, this paper will go further. The Employment Service has done 515 validity studies: 425 with a proficiency measure of performance and 90 predicting training success. In each of these studies, each aptitude has a validity coefficient for job performance in that study. Thus aptitude validity can be correlated over jobs. This correlation matrix shows the same structure: three general factors for the same aptitude clusters with the perceptual factor being predictable from the other two. Furthermore, the correlations within clusters are as high as sampling error will permit, thus showing that the prediction of job performance is due to the general factors rather than the specific factors in the specific aptitudes.

The conclusion of this paper is that for validation purposes, the GATB should be scored in composite scores estimating the three general factors. These composite scores can then be combined using multiple regression. Inclusion of the perceptual factor will add little to the prediction of the cognitive and psychomotor factors.

Multiple Cutoffs and Linearity

The multiple-cutoff procedure is highly nonlinear. Mathematical formulas for this procedure are intractable and there has been little statistical development. No efficient algorithms for finding cutoffs have been developed and there are no known formulas for sampling error or shrinkage. As a result, the Employment Service has been forced to use trial-and-error computational procedures with no formal treatment of sampling error. Instead the Employment Service has had to rely on cross validation with small samples. The sampling error in such cases is such that many valid test combinations have been rejected because of Type II error, i.e., because the validity on the small sample was falsely low.

The multiple-cutoff procedure is only optimal if the relationships between tests and job performance are nonlinear. If job performance is linearly related to test scores, then multiple regression is known to be optimal. Hawk (1970) made a statistical study of over 3,000 test performance relationships in Employment Service data and found "significant" nonlinearity at exactly the chance level. Thus there is no nonlinearity in Employment Service data. This is also true of the employment field as a whole as shown in a review of such studies by Hunter and Schmidt (in press; see also Schmidt, Hunter, McKenzie, & Muldrow, 1979).

Since the relationships between test and job performance is known to be linear, multiple regression will yield better prediction than multiple cutoffs. This does not mean that past work with multiple cutoffs was invalid (cross validation has shown past work to be valid), but merely that shifting to multiple regression will mean better prediction of job performance.

Multiple Cutoff and Apparent Inconsistency Across Studies

Because of the complexities of the trial and error computational procedures used with multiple cutoffs, the Employment Service never uses more than four tests in a prediction equation. Yet cumulative analysis shows that for jobs such as fork lift truck operator, all nine aptitudes are predictive. Thus which two, three or four are chosen for use in a given study is largely a matter of sampling error. Any combination of two, three or four would work just about as well.

This situation has been very much misinterpreted by people who distrust tests. They see the inconsistency in the tests used across studies for the same job and they believe it to mean that tests which are valid at one point in time are not valid at a later point in time. The error in their reasoning is the assumption that if a test is not used in the multiple cutoff procedure, then that test is not valid. They are not aware that the restriction to four tests is an artifact of the computational scheme.

Multiple regression would not have this problem. There is a known formula for the sampling error in a beta weight. Thus even though weights vary considerably from sample to sample, the variation could be shown to be due to sampling error.

Multiple Cutoffs and Incorrect Selection Ratios

An employer uses a valid selection test because workers with high test scores are more productive on the average than workers with lower test scores. If performance is linearly related to test score (as shown by Hawk, 1970, and others reviewed in Hunter & Schmidt, in press), then there is an exact multiplicative relation between the gain in productivity due to selection and the average test score of those hired. Thus optimal productivity of the work force requires hiring the top applicants.

Therefore if an employer has positions for 10 percent of the available applicants, the optimum productivity would be achieved if the employer hired those in the top 10 percent of the test score distribution. However, the Employment Service multiple cutoff procedure identifies only those who are in the top two thirds. Thus instead of hiring the top 10 percent, the employer who works with the Employment Service at present can only hire randomly from among the top 67 percent. The utility equations of Brogden (1946, 1949; see also Cronbach & Gleser, 1965; Hunter & Schmidt, in press; and Schmidt et al., 1979) show that the employer loses over two thirds of the benefits of testing under these conditions.

Furthermore, not all employment offices recommend only those with a high (H) rating. Many offices recommend those with medium (M) ratings as well. In this case the employer is hiring at random from the top 80 percent with loss in utility of nearly 80 percent. Thus the use of multiple cutoff procedures robs the employer of nearly all of the benefits of increased productivity.

Loss in Productivity at a National Level

The previous discussion showed that most employers receive only about one fifth to one third of the potential benefit of fitting people to jobs if current multiple cutoff procedures are used by the Employment Service. However, a national service should also consider the national implications of such a policy.

Hunter and Schmidt (in press) have used standard utility equations and known validity equations to estimate the national impact of using tests to fit people to jobs. According to their figures, if tests were abandoned, there would be a decrease in national productivity of 80 billion dollars a year. If, instead of abandoning tests, everyone were to use the artifactually low cutoff of top two thirds currently used by the Employment Service, then the loss in national productivity would be about 54 to 60 billion dollars. This would have very considerable implications for economic growth and future employment.

In particular, it should be noted that small differences in productivity can lead to grave differences in economic outcome. American businesses are all now in competition with foreign manufacturers. Even a small difference in labor productivity can lead to a difference in price which results in the complete loss of an industry to the foreign competition. Thus wide use of the low cutoff scores currently used by the Employment Service would lead directly to a higher rate of unemployment.

Unit Weights and Ridge Regression

If population correlations were known, then multiple regression would always give optimal prediction. However, if data is available only on a small sample (less than 500), then multiple regression on the sample data will not always give optimal prediction. In a survey of cases from the employment literature, Schmidt (1971, 1972) showed that multiple regression yielded beta weights that varied much more than was optimal. He then recommended a procedure of using unit weights for those predictors that are significantly related to job performance. He showed that unit weights would actually work better in the situations that he surveyed. Later, Price (1977; Rozeboom, 1979) developed ridge regression as an alternative method of estimating beta weights which would vary less than those of multiple regression and hence be more nearly optimal. Should such methods be used by the Employment Service in predicting from the GATB?

It is mathematically obvious that unit weights and ridge regression are not always better than multiple regression. For example, if the predictors are uncorrelated with each other, then unit weights and ridge regression would be a disaster no matter how small the samples. It is now known that unit weights and ridge regression work better than multiple regression only if the predictors are quite closely linked to each other. However, the exact theory of this linkage has not yet been fully explicated.

The one situation which is known to be poor for multiple regression is a situation in which the predictors are all indicators of the same underlying general factor and it is the general factor which predicts job performance rather than the specific factors of the individual predictors. If the specific factors of the predictors are known in advance to be uncorrelated with the criterion, then we can state in advance that optimal prediction will be achieved by weighting the predictors in accordance with how well they correlate with the general factor, i.e., their factor loadings on that factor. These factor loadings are all usually high and hence lead to approximately equal weights (hence the success of ridge regression or unit weights). Schmidt's (1971, 1972) findings are explained by the fact that most contemporary predictors of job performance are all indicators of general intelligence (i.e., school grades, verbal or numerical ability, etc.). If it is general intelligence that is the basic predictor, then Schmidt's findings follow at once from the equations of factor analysis.

Is it likely that the GATB would be unidimensional in this sense? The GATB is much broader in coverage than most contemporary predictor sets in the employment domain. The GATB not only contains the cognitive aptitudes (verbal and numerical aptitude) but contains three perceptual aptitudes and three psychomotor aptitudes as well. It is likely then that evidence will suggest that unit weights be used within each of these three sets of aptitudes but not across them. The remainder of this paper will show that this is indeed the case.

Dimensionality of the GATB

The GATB is a set of 12 tests combined to measure nine aptitudes. The test aptitude combinations are shown in Table 2. The aptitudes are deliberately grouped by general factor. The first three aptitudes are the classical cognitive aptitudes (G, V, N, for general, verbal, and numerical aptitude) which assess an applicant's ability at thought, reasoning, and ability to learn. The middle three aptitudes are perceptual aptitudes (S, P, Q, for spatial visualization, form matching or pattern recognition, and name matching or clerical speed, respectively). The final three aptitudes are psychomotor skills (K, F, M, for motor coordination, finger dexterity, and manual dexterity, respectively). The G aptitude is the only aptitude not measured independently from the others. The G aptitude overlaps with verbal, numerical, and spatial ability. If independence of measurement is desired in an analysis, then G would be dropped.

Table 2

The nine aptitudes measured by the GATB and the tests used for each (USES, 1970, p. 40)

Symbol	Name	Test(s)
G	General Intelligence	Vocabulary + Arithmetic Reasoning + Three Dimensional Space
V	Verbal Aptitude	Vocabulary
N	Numerical Aptitude	Computation + Arithmetic Reasoning
S	Spatial Aptitude	Three Dimensional Space
P	Form Perception	Tool Matching + Form Matching
Q	Clerical Perception	Name Comparison
K	Motor Coordination	Mark Making
F	Finger Dexterity	Assemble + Disassemble
M	Manual Dexterity	Place + Turn

The correlations between the nine aptitudes are given in the Employment Service manual (USES, 1970, section III, page 34) for 23,428 workers. These correlations are presented in Table 3. The correlation matrix is blocked off by ability grouping. The diagonal blocks contain the correlations between aptitudes in the same general factor set, while nondiagonal blocks contain correlations between aptitudes in different general factor sets. The average correlation within the diagonal blocks are .79, .53, and .45 for the cognitive, perceptual, and psychomotor blocks, respectively. The average correlation between variables in different sets is .39. The average correlation between cognitive and psychomotor clusters is .26. That is, the cognitive and psychomotor aptitudes are relatively uncorrelated with each other, but both are highly correlated with the perceptual aptitudes.

Table 3

The correlations between aptitudes (USES, 1970, p. 34) and their reliabilities (USES, 1970, p. 269); N = 23,428 for the correlations, decimal omitted

		G	V	N	S	P	Q	K	F	M
Intelligence	G	100	84	86						
Verbal Aptitude	V	84	100	67						
Numerical Aptitude	N	86	67	100						
Spatial Aptitude	S	74	46	51	100	59	39			
Form Perception	P	61	47	58	59	100	65			
Clerical Perception	Q	64	62	66	39	65	100			
Motor Coordination	K	36	37	41	20	45	51	100	37	46
Finger Dexterity	F	25	17	24	29	42	32	37	100	52
Manual Dexterity	M	19	10	21	21	37	26	46	52	100
Reliability		88	85	83	81	79	75	86	76	77

Table 4 presents the results of a confirmatory factor analysis of the correlations in Table 3. Aptitude G was left out of the analysis since it is not defined independently of V, N, or S. Spatial aptitude was left out of the perceptual factor because it is closer to the cognitive factor than are P and Q. The factor analysis was done with communalities using oblique multiple groups analysis (Harman, 1976; Tryon & Bailey, 1970, under the rubric "cluster analysis"; Hunter, 1980; Hunter & Gerbing, 1982). The factor analysis shows only slight departures from the content analysis; spatial aptitude is farther from the psychomotor factor than are the other two perceptual abilities, and motor coordination K is closer to cognitive and perceptual abilities than are the dexterity measures, F and M.

Table 4

Confirmatory factor analysis of the correlations between aptitudes in Table 3; correlations between factors and aptitudes; done using oblique multiple groups factor analysis with communalities; factors defined by the aptitudes listed under them; decimal omitted

		Factors		
		Cognitive	Perceptual	Psychomotor
		VN	PQ	KFM
Intelligence	G	..	..	..
Verbal Aptitude	V	82	68	32
Numerical Aptitude	N	82	77	42
Spatial Aptitude	S	59	61	35
Form Perception	P	64	81	66
Clerical Perception	Q	78	81	54
Motor Coordination	K	48	60	64
Finger Dexterity	F	25	46	67
Manual Dexterity	M	19	45	72

The factors in this analysis are correlated. Table 5a presents the correlations between the factors of Table 4. The cognitive and psychomotor factors are substantially separated, though each has a high correlation with the perceptual factor. Table 5a also contains a multiple regression analysis of the perceptual factor onto the other factors. The multiple correlation is .96, showing that the perceptual factor is essentially perfectly predictable from the cognitive and psychomotor factors. If all factors were perfectly measured, then the perceptual factor could be dropped from the battery.

Table 5. Correlations between the factors of Table 4 and between the corresponding composite scores and associated regression analyses (decimals omitted)

Table 5a. Correlations and regression analysis for the factors of Table 4

Correlations

	Cog	Per	PMo
Cognitive Factor	100	88	46
Perceptual Factor	88	100	75
Psychomotor Factor	46	75	100

Regression of perceptual factor onto the cognitive and psychomotor factors

Beta for cognitive	=	68
Beta for psychomotor	=	44
Multiple correlation	=	96

Table 5b. Correlations and regression analysis for the composite scores which estimate the factors of Table 4 (decimals omitted)

Correlations

		GVN	SPQ	KFM
Cognitive Composite	GVN	100	76	35
Perceptual Composite	SPQ	76	100	51
Psychomotor Composite	KFM	35	51	100

Regression of perceptual composite score onto the cognitive and perceptual composite scores

Beta for cognitive composite	=	66
Beta for psychomotor composite	=	28
Multiple correlation	=	80

In practice, the factors can only be estimated by composite aptitude scores: the cognitive factor by G + V + N, the perceptual factor by S + P + Q, and the psychomotor factor by K + F + M. Table 5b shows the correlations between the actual composite scores. These correlations are similar to but smaller than the correlations between the factors. Table 5b also contains a regression analysis of the perceptual composite onto the other composites. Again the larger beta weight is for the cognitive factor. The multiple correlation is only .80, reflecting the imperfect measurement implicit in the use of composite scores to estimate factor scores.

This analysis is consistent with previous work on the structure of the GATB, most notably Fozard et al. (1972). They used VARIMAX factor analysis and identified three factors: information processing ability (i.e., cognitive ability), pattern analysis capability (i.e., perceptual ability), and manual dexterity (i.e., psychomotor ability). Their analysis differed only in that they put clerical perception in with the cognitive aptitudes. Because they used orthogonal factors, they had no way of knowing how high the correlation is between clerical perception and the psychomotor factor; i.e., they could not see that clerical perception departs markedly from the cognitive abilities in this respect and is hence like the other perceptual abilities.

Although the GATB aptitudes define three general factors, the perceptual factor is dependent on the other factors. In this sense only two dimensions are spanned. The reasons that Fozard et al. (1972) could find three orthogonal factors were (1) they did not use communalities, and (2) they included three other variables in their study (age, education, and socio-economic status).

Since they didn't use communalities, their three factors were orthogonally transformed composite scores and hence the perceptual composite was not completely accounted for.

The analysis of the dimensionality of the GATB shows that the conditions for the use of unit weights or ridge regression are NOT met for the nine aptitudes which define three general factors rather than one. However, it is quite possible that unit weights might be appropriate within the three clusters. There is an additional requirement before the use of three composite scores in place of the nine aptitudes can be recommended. It must be shown that it is the general factors rather than the specific factors which predict job performance. This will be shown in the section on correlations between aptitude validity coefficients.

Causal Relations Between Factors

Linear dependence between variables is a symmetric condition. If $z = x + y$, then $x = z - y$ and $y = z - x$. Thus the fact that the perceptual factor can be predicted from the other two factors does not mean that the dependence would have to be in that direction causally. However, there are two sets of evidence which tend to confirm the dependence of the perceptual factor: negative beta weights for other causal orderings, and the correlations of the three factors with race and age.

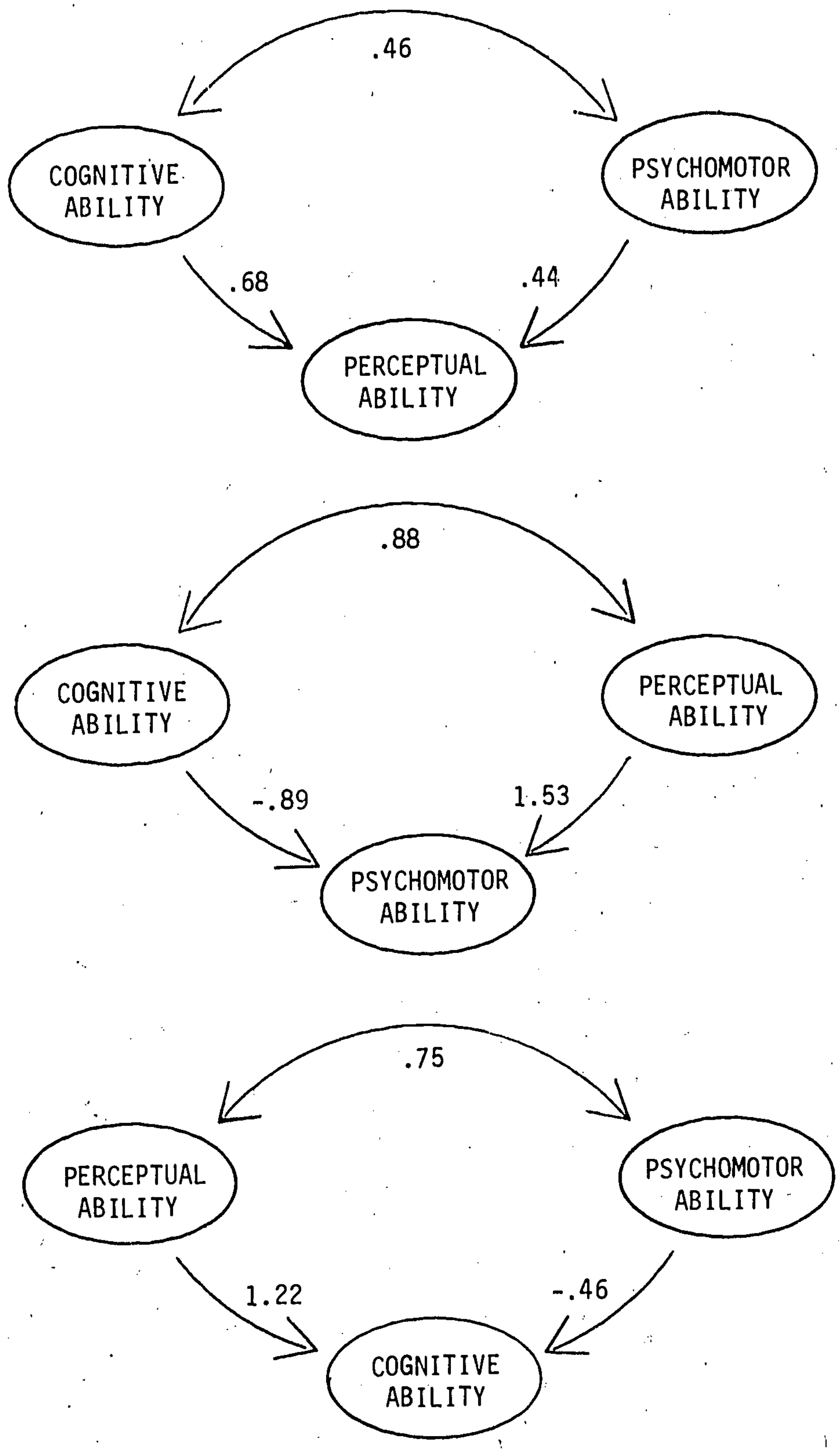


Figure 5. Three alternative path models for the causal relations between the three factors underlying the GATB.

Figure 5 shows three alternate path models for the three factors of the GATB. In each path model, a different ability is assumed to be causally dependent on the other two. The first path diagram assumes that perceptual ability is causally dependent on the other two factors. Both beta weights are positive; .68 for cognitive ability and .44 for psychomotor ability. The curved arrow between cognitive and psychomotor indicates that this model makes no attempt to explain the correlation between them.

The second path model assumes that psychomotor ability is dependent on the other two abilities. If psychomotor ability is taken to be dependent, then there are beta weights of 1.53 for perceptual and -.89 for cognitive ability. That is, cognitive ability is a suppressor variable in the prediction of psychomotor ability from perceptual ability. I can think of no substantive theory which would make such a prediction, though it follows from the assumption that perceptual ability is the dependent variable.

The third path model assumes that cognitive ability is causally dependent on the other two abilities. If cognitive ability is taken to be dependent, then the beta weights are 1.22 for perceptual ability and -.46 for psychomotor ability. Since psychomotor ability is positively correlated with cognitive ability, the negative beta weight is usually interpreted as meaning that psychomotor ability is a suppressor variable for the prediction of cognitive ability from perceptual ability. I know of no substantive hypothesis which would make such a prediction.

Of the three models, only one model has both beta weights positive. Of the three models, only one model has both beta weights less than one. While negative beta weights and weights greater than one are mathematically possible, they occur only under very peculiar conditions. Usually they indicate a reversal of causal order. If they did indicate suppressor variables, then that is usually taken to mean that the variable being predicted is a causal agent of certain test content while the suppressor variable is a second uncorrelated causal agent. This is tantamount to saying that the suppressor model actually assumes the model in which perceptual ability is dependent anyway. Thus by the usual interpretation rules for beta weights, the only acceptable causal model is that in which perceptual ability is dependent on the other two abilities.

Table 6

The correlations between age, race, and the three factors which underlie the GATB (decimals omitted)

	A	R	PM	C	P
Age	100	0	-44	-19	-39
Race	0	100	8	26	19
Psychomotor ability	-44	8	100	46	75
Cognitive ability	-19	26	46	100	88
Perceptual ability	-39	19	75	88	100

Table 6 shows the correlations between age, race, and the three abilities underlying the GATB. There are decrements due to age on all the aptitudes in the GATB. These decrements are greater for psychomotor ability than for perceptual ability; as would be the case if psychomotor ability were an intervening variable between age and perceptual ability. There are racial differences on all aptitudes. These differences are greater on cognitive ability than on perceptual ability; as would be the case if cognitive ability were an intervening variable between race and perceptual ability. That is, the correlations with age and race suggest that cognitive and psychomotor ability are causally intervening variables between age and race on the one hand and perceptual ability on the other hand. This would be consistent with the path model which assumes perceptual ability to be causally dependent on cognitive and psychomotor ability.

The age decrement in psychomotor ability is also much larger than the age decrement on cognitive ability. In fact it is larger in exactly the ratio of the correlation between them. This suggests that psychomotor ability intervenes between age and cognitive ability. That is, the age correlations in Table 6 suggest that psychomotor ability is causally prior to cognitive ability.

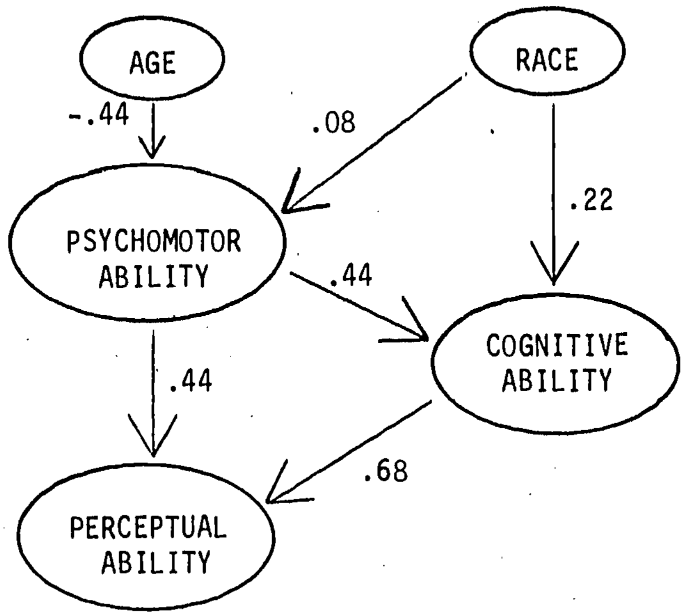


Figure 6. A path model relating age and race to the three abilities which underlie the GATB.

Figure 6 presents a path model which incorporates the causal relations suggested by the previous two paragraphs. According to this model, age produces decrements in psychomotor ability. The causal processes which produce the correlations between psychomotor ability and the other two abilities then transmit those decrements proportionately. According to this model, race is associated with environmental or genetic differences which produce differences in cognitive ability and to a much lesser extent psychomotor ability. The causal processes which connect these abilities to perceptual ability then transmit the racial differences proportionately to perceptual ability.

Most of the correlations in Table 6 are used to estimate the path coefficients in Figure 6. However, there are three residual correlations free to depart from 0. The residual correlation between age and cognitive ability is exactly 0 as predicted by the model. The residual correlation between race and perceptual ability is -.02 which is not 0 but is trivial in magnitude. The residual correlation between age and perceptual ability is -.06 which is small but significantly different from 0 with a sample size over 24,000. Thus there is a small impact of age on perceptual ability above and beyond the impact of age on psychomotor ability.

The close fit of the path model in Figure 6 supports the suggestion that perceptual ability is causally dependent on cognitive and psychomotor ability. The data also suggest that the correlation between cognitive and psychomotor ability stems from causal processes that go from psychomotor ability to cognitive ability. Could it be that videogames will eventually eliminate the small decrements now found in cognitive ability over time?

General and Specific Factors

Consider two aptitudes which are related to the same general factor, say finger and manual dexterity. These aptitudes will not correlate perfectly with each other. First of all, there is random error in their measurement. The reliability of finger dexterity is .76 and of manual dexterity is .77. However, even if the correlation between them is corrected for attenuation, the correlation between true scores will not be perfect. That is, if we correct the correlation .52 (from Table 3) for attenuation we have $r = .52 / \sqrt{(.76)(.77)} = .70$. The gap between .70 and 1.00 is a measure of the extent to which the aptitudes F and M measure specific factors as well as the psychomotor general factor. The specific factor for F is the measure of the extent to which a person's finger dexterity differs from the level that would be predicted from the person's general psychomotor ability. The specific factor for M is the measure of the extent to which the person's manual dexterity differs from the level predicted by the person's general psychomotor ability. The greater the extent to which a given aptitude depends on its specific factor rather than on its general factor, the lower the correlation of that aptitude with other aptitudes measuring the same general factor.

It is the hypothesis of the unit weights model that specific factors will NOT be correlated with job performance. If this hypothesis is true, then the specific factor is also an "error" factor in the sense that it represents unwanted variation that reduces the validity coefficient. However, the specific factor can be thought of as a systematic error factor rather than a random error factor. If the specific factor of an aptitude is in fact irrelevant to the job, then the reliability of the test is not a good index of quality of measurement. Instead we should use an index of the correlation between the test score and the relevant general factor. In terms of squared correlations, this turns out to be the communality of the test rather than its reliability.

It is important to have an exact quantitative statement of the relationships between general factors, specific factors, and error factors. These relations are stated in full in the Appendix. For present purposes, it is sufficient to note that the total variance on an aptitude can be written as the sum of three numbers which can be interpreted as the amount of variance explained by the general factor, the amount of variance explained by the specific factor, and the amount of variance explained by error. That is,

$$\text{Total Variance} = \text{General Factor Variance} + \text{Specific Factor Variance} + \text{Error Variance}$$

For purposes of predicting and correcting correlation coefficients, these are usually expressed in percentage terms. That is,

100 percent = Percent General + Percent Specific + Percent Error

These percentages can be linked to other terminology:

Percent General = Communality

Percent General + Percent Specific = Reliability

Computationally, the reliability is determined from either an internal consistency measure such as coefficient alpha or from a test retest correlation. The communality is determined from the aptitude correlation factor analysis. The specific variance is then obtained by subtracting the communality from the reliability.

Table 7

The breakdown in general, specific, and error factor variance for each of the nine aptitudes and for the three composite scores (decimals omitted)

		General Factor Variance			Specific Factor Variance	Error Factor Variance
		Cog	Per	PMo		
Intelligence	G	79			13	8
Verbal Aptitude	V	67			18	15
Numerical Aptitude	N	67			16	17
Spatial Aptitude	S		37		44	19
Form Perception	P		65		14	21
Clerical Perception	Q		65		10	25
Motor Coordination	K			41	45	14
Finger Dexterity	F			45	31	24
Manual Dexterity	M			52	25	23
Cognitive Composite	GVN	80			12	8
Perceptual Composite	SPQ		79		11	10
Psychomotor Composite	KFM			75	16	9

Table 7 shows the breakdown into general, specific, and error variance for each of the GATB aptitudes and for the composite scores which estimate the general factors. The key fact brought out by this table is that it would be very unwise to try to avoid composite scores by just using one of its component aptitudes. For example it would be very convenient if we could use just K to estimate the psychomotor general factor. But Table 7 shows that the relevant measure of the quality of K as an estimate of the psychomotor general factor is not its reliability of .77, but its communality of .52 which is quite poor. Even the composite scores are poorer than we might like. The communality of the KFM psychomotor composite is only .75 even though its reliability is .91. In general, the communalities of the composite scores are actually somewhat lower than the reliabilities of single aptitudes even though the composite scores are based on about three times as many responses.

The Dimensions of Validity

Correlations Between Aptitude Validities Across Jobs

Suppose that an aptitude is correlated with job performance on some job. That aptitude could be correlated because its general factor is a valid predictor of job performance or it could be correlated because its specific factor is correlated with job performance on that job (or both). Consider the unit weights hypothesis; assume the specific factors are irrelevant to job performance. Then an aptitude will have high validity for a given job only if its general factor has a high validity for that job. But if the general factor has a high validity, then so will the other aptitudes which reflect that same general factor. Thus according to the unit weights hypothesis, all the aptitudes which measure the same general factor will either have high validity or low validity together for each given job. Thus if we correlate validities across jobs, then there should be a high degree of correlation between the validities for aptitudes which measure the same general factor. This argument is spelled out mathematically in the Appendix.

Table 8

Correlations between validity coefficients

across 515 jobs (decimals omitted)

		G	V	N	S	P	Q	K	F	M
Intelligence	G	100	80	81						
Verbal Aptitude	V	80	100	61						
Numerical Aptitude	N	81	61	100						
Spatial Aptitude	S	67	32	40	100	53	30			
Form Perception	P	45	30	48	53	100	57			
Clerical Perception	Q	57	54	63	30	57	100			
Motor Coordination	K	19	16	24	8	41	40	100	46	56
Finger Dexterity	F	9	1	15	26	45	23	46	100	62
Manual Dexterity	M	-2	-7	9	14	36	19	56	62	100
Reliabilities		54	47	47	47	46	44	45	53	52

Table 8 presents the correlations between aptitude validity coefficients across 515 validation studies done by the U.S. Employment Service. The structure of the correlations of validities across jobs is an exact mirror of the structure of the correlations between aptitude scores across persons. The high correlations are those between aptitudes measuring the same general factor, though the correlation between spatial aptitude and clerical perception is again lower than would be expected. The similarity of structure can also be seen in the close match between the factor analyses. Table 9 presents the confirmatory factor analysis of the correlations between validities across jobs using the same specifications as those used for Table 4; communalities, with G left out of the analysis, and with S left out of the specification of the perceptual factor.

Table 9

Confirmatory factor analysis of the correlations between aptitude validity coefficients across jobs in Table 8; numbers shown are correlations between factors and aptitude validity coefficients (decimals omitted); method is oblique multiple groups factor analysis with communalities; factors defined by the aptitude-validities listed under them

		Factors		
		Cognitive	Perceptual	Psychomotor
		VN	PQ	KFM
Intelligence	G	..	..	..
Verbal Aptitude	V	78	57	5
Numerical Aptitude	N	78	74	22
Spatial Aptitude	S	46	55	22
Form Perception	P	50	75	55
Clerical Perception	Q	75	75	37
Motor Coordination	K	26	54	70
Finger Dexterity	F	10	45	73
Manual Dexterity	M	1	36	78

Table 10. Correlations between the factors of Table 9 and between the corresponding composite validity coefficients (decimals omitted) and the associated regression analyses

Table 10a. Correlations and regression analysis for the factors of Table 9

Correlations

		VN	PQ	KFM
Cognitive Factor	VN	100	83	17
Perceptual Factor	PQ	83	100	61
Psychomotor Factor	KFM	17	61	100

Regression of perceptual factor onto the cognitive and psychomotor factor

Beta for cognitive	=	.75
Beta for psychomotor	=	.48
Multiple correlation	=	.96

Table 10b. Correlations regression analysis for the composite score validity coefficients

Correlations

		GVN	SPQ	KFM
Cognitive Composite Validity	GVN	100	65	8
Perceptual Composite Validity	SPQ	65	100	38
Psychomotor Composite Validity	KFM	8	38	100

Regression of perceptual composite onto the cognitive and psychomotor composite validity

Beta for cognitive composite validity	=	.62
Beta for psychomotor composite validity	=	.33
Multiple correlation	=	.73

The factors in Table 9 are correlated and the correlations are shown in Table 10a. Again there is a relatively low correlation of .17 between the cognitive and psychomotor factors, but high correlations between the perceptual factor and the others. The regression analysis in Table 10a shows that there is a multiple correlation of .96 for perceptual onto cognitive and psychomotor factors. Thus the validity of the perceptual factor is almost perfectly predictable from the validities of the cognitive and psychomotor factors.

Validities can be computed for the composite scores GVN, SPQ, and KFM which estimate the general factors. These validities can be correlated across jobs. The validity correlations across jobs are shown in Table 10b. These correlations are lower but follow the same pattern. The multiple correlation of the validities for the perceptual composite SPQ onto the others GVN and KFM is only .73 reflecting the impact of sampling error on the validities involved.

The structure of the matrix of correlations between aptitude validities across jobs mirrors the matrix of correlations between aptitudes across people. This is the prediction of the unit weights hypothesis that specific factors are largely irrelevant. But are the correlations high enough? Hunter (Note 1) developed a formula for the expected variance of sampling error across studies and converted that into a "reliability" formula for use in correcting correlations across studies for attenuation due to sampling error. These "reliabilities" are shown as the last row in Table 8. According to the unit weights hypothesis, the correlation between any two aptitudes measuring the same general factor should be the average of their respective "reliabilities" (actually the geometric average). Verbal and numerical should correlate about .47 (versus the actual .61); spatial, form perception, and clerical perception should correlate about .475, .455, and .45 (versus actual correlations of .53, .30, and .57); motor coordination, finger dexterity, and motor dexterity should correlate about .49, .485, and .525 (versus actual correlations of .46, .56, and .62). If these correlations are corrected for sampling error in a manner analogous to correction for attentuation, then the average corrected correlation is 1.09. This confirms the prediction of the general ability theory over the specific aptitude theory. The general ability theory of job performance predicts a correlation of 1.00 between aptitude validities within the same general ability cluster. The specific aptitude theory predicted a low correlation.

The structure of the correlations between aptitude validities across jobs confirms the unit weights hypothesis; the specific factors are largely irrelevant to the prediction of job performance. Therefore, the aptitudes should be rescored in terms of unit weight composite scores before prediction is begun. Since there are three general factors present in the GATB, there should be three composites: GVN, SPQ, and KFM to estimate the cognitive, perceptual, and psychomotor factors, respectively.

This does not mean that the specific factors are always irrelevant. The model does not fit exactly. Moreover, the test for fit used here does not distinguish between perfect fit and near perfect fit. Thus if specific factors were occasionally relevant, these few deviations from the model would not be detected by the present procedure. However, the identification of those few jobs where specific factors are relevant would be a prodigious task. One must distinguish between the few times when the specific factor is truly correlated with job performance and the many times when the specific factor is falsely correlated with job performance by sampling error. This cannot be done on a statistical basis without eliminating sampling error, i.e., without using a very large sample (at least 500) for the validation study for a given job.

In practical terms we have the following conclusion: the composite scores can be validated for job families using validity generalization (Schmidt, Hunter, Pearlman, & Shane, 1979). If specific factors are ever relevant, then they must be validated for given jobs. If criterion related validity for the specific factor is to be shown, then the sample size for that job must be at least 500. Finally it should be noted here that a given test or aptitude is valid if either its general or its specific factor is valid. Thus test validity as defined in the various guidelines requires only validation of the general factor underlying the test which can be done with validity generalization.

Further Work on Spatial Aptitude

At the time that this research was done, most analysts in the U.S. Employment Service were staunch advocates of the specific aptitude theory of ability and job performance. They were particularly convinced that spatial aptitude will be valid in jobs different from those where the other perceptual aptitudes are valid. This hypothesis can be tested using two sources of job analysis information directly relevant to spatial aptitude. The corresponding analysis will be described below and disconfirms the beliefs of those advocating specific aptitude theory. However, in the course of still later research using an entirely different job analysis, we did find two narrow job families where spatial aptitude is more relevant than the perceptual factor.

Most Employment Service validation studies are conducted across organizations, usually across states. There is a thorough job analysis done at each participating organization to ensure that the job there is the same as at other organizations in the study. As part of that job analysis, analysts are asked to rate the job in terms of the relevance of each of the nine aptitudes on the GATB. A consensus of these reports is presented as part of the validation study writeup. Thus for each of the 515 validation studies, there is a consensus rating for the relevance of spatial aptitude for the job in question. When the data was broken down using these ratings, no job family was found where spatial aptitude was more valid than the perceptual factor.

For every job in the Dictionary of Occupational Titles (1977), there have been several dimensional job analyses; one job analysis is relevant to spatial aptitude. Every job is rated in terms of the expected aptitude level of people in that job. Analysts based this judgment on the idea that if an aptitude is relevant to the job, then those who survive on the job will be higher on that aptitude than those who did not. For each of the 515 validation studies, this "expected means" rating was obtained for spatial aptitude. When the validity data were broken down using these ratings, there was no job family found for which spatial aptitude was more valid than the perceptual factor.

Table 11

The average validity of selected aptitudes and aptitude composites for job families determined by ratings of the relevance of spatial aptitude; low category numbers mean high relevance for spatial aptitude

Spatial Relevance Rating	Number of Studies	Average Validity		
		Perceptual Composite	Spatial Aptitude	Other Perceptual Aptitudes
4	43	.29	.28	.24
5	80	.26	.21	.24
6	92	.25	.18	.24
7	192	.23	.14	.23
8	8	.14	.05	.17

The analysts' job analysis ratings and the expected means ratings are similar in their meaning. However, they are not perfectly correlated. This suggests that neither rating is perfectly reliable. Thus there is the possibility that the specific aptitude prediction failed because the rating error weakened the effect to below visibility. Averaging the two ratings would be the equivalent of doubling test length in increasing the reliability of the rating. The average ratings could then be used to stratify the validation studies. Table 11 presents the average validity for each category with more than one study. The category labels in Table 11 are defined so that a small label indicates high relevance.

The average validities shown in Table 11 show that job analysts were indeed sensitive to the relevance of spatial aptitude. As relevance ratings go down, the mean validity of spatial aptitude goes down. However, ratings of the relevance of spatial aptitude also predict the relevance of the other perceptual aptitudes. They, too, are less valid for jobs with low spatial relevance ratings. Thus the validity of the perceptual composite score as a whole varies with ratings of spatial aptitude relevance. In all categories, even the first, the average validity of spatial aptitude was less than that of the perceptual composite score.

Table 11 shows that there are some jobs where the specific factor for spatial aptitude predicts job performance. The relative range of variation in mean validity is .28/.05 = 5.60 for spatial aptitude and only .29/.14 = 2.07 for the perceptual composite. However, the spatial relevance ratings by job analysts do not differentiate between the relevance of the general perceptual factor and the relevance of the spatial specific factor sharply enough to detect the small set of jobs where there is an empirical payoff for the use of the spatial aptitude rather than the perceptual composite.

During later work on job analysis schemes, two job families turned up where spatial aptitude was higher validity than the perceptual composite. The Guide to Occupational Exploration was developed by the Department of Labor (Droege & Padgett, 1979; Droege & Hawk, 1977) as a job classification system based on Holland's (1973) theory of interests and occupational choice. There are 11 basic interest categories which are then further subdivided. The first two categories are Artistic and Scientific jobs. For these categories there are 11 validity studies in the U.S. Employment Service data base. The average validity is .26 for spatial aptitude, .16 for the other perceptual aptitudes (i.e., the PQ composite), and .22 for the perceptual composite. After suitable correction for attenuation and range restriction (using the findings of Hunter, Note 2), the multiple regression equation for this job family shows expected job performance (EJP) to be EJP = .40 GVN + .17 S with a multiple correlation of .52.

Spatial aptitude is the most likely of the cognitive and perceptual aptitudes to show specific validity. First, the specific factor variance for spatial aptitude is 44 percent whereas the next highest specific factor variance for perceptual and cognitive aptitudes is 18 percent for verbal aptitude. Second, the one departure from the general factor model is the correlation between spatial and clerical aptitude. Yet the number of jobs in which the specific factor is relevant is very small and even there the increment is marginal: a correlation for spatial of .26 versus a correlation for the perceptual composite of .22. Furthermore, in both scientific and artistic work, there is heavy use made of the spatial principles learned in school, i.e., heavy use of formal geometry. Thus in these jobs, the conditions for the validity of the specific factor are met: Use is made of the specific learning associated with the specific aptitude.

CONCLUSION

There are two theories relating ability to job performance. Specific aptitude theory asserts that job performance will be best predicted by that aptitude which is most similar to the job in test content. Specific aptitude theory asserts that intelligence is only indirectly correlated with job performance because it is correlated with the relevant specific aptitude. General ability theory asserts that job performance is learned as a new aptitude in its own right. Specific aptitudes will be valid only indirectly because they are correlated with general intelligence. The findings of 515 validation studies carried out by the U.S. Employment Service clearly disconfirm the specific aptitude theory and show almost perfect fit for the general ability theory. The only known departure is for artistic and scientific jobs where spatial aptitude is more valid than general perceptual ability.

The departure from perfect fit shown by spatial aptitude may be of theoretical significance but is of no practical significance. Prediction even in the artistic and scientific jobs is only trivially higher using spatial aptitude than the perceptual composite.

The analysis of the GATB showed three factors (cognitive, perceptual, and psychomotor ability) spanning two dimensions. The cognitive factor is slightly different from the classic concept of general intelligence in that it is distinguished from the perceptual factor. The correlations between the factors and additional data on age and race show the perceptual factor to be perfectly predictable from the other factors and causally dependent on them.

REFERENCE NOTES

1. Hunter, J. E. Cumulating results across studies: Correction for sampling error, a proposed moratorium on the statistical significance test, and a critique of publishing practices regarding multiple regression, canonical correlation, MANOVA, and factor analysis. Invited address, American Psychological Association, 1979. (Author's address: Department of Psychology, Michigan State University, East Lansing, Michigan, 48824).

2. Hunter, J. E. Test validation for 12,000 jobs: An application of job classification and validity generalization analysis to the General Aptitude Test Battery. Department of Psychology, Michigan State University, written for the U.S. Employment Service, July 13, 1982.

3. Horn, J. L. Statistical weakness of a unitary construct of intelligence. American Psychological Association symposium, September 5, 1979.

4. Jensen, A. R. Test validity: G versus the specificity doctrine. Invited address, Division 14, American Psychological Association, August 26, 1981.

5. Helm, W. E., Gibson, W. A., & Brogden, H. E. An empirical test of shrinkage problems in personnel classification research. Personnel Board Technical Research Note 84, October 1957.

6. Vineberg, R., & Joyner, J. N. Prediction of job performance: Review of military studies. HumRRO: Human Resources Research Organization, 1981.

REFERENCES

Brogden, H. E. On the interpretation of the correlation coefficient as a measure of predictive efficiency. Journal of Educational Psychology, 1946, 37, 65-76.

Brogden, H. E. When testing pays off. Personnel Psychology, 1949, 2, 171-183.

Cronbach, L. J., & Gleser, G. C. Psychological tests and personnel decisions. Urbana: University of Illinois Press, 1965.

Droege, R. C., & Hawk, J. Development of a U.S. Employment Service Interest Inventory. Journal of Employment Counseling, 1977, 14, 65-71.

Droege, R. C., & Padgett, A. Development of an interest oriented occupational classification system. Vocational Guidance Quarterly, 1979, 27, 302-310.

Fozard, J. L., Nuttall, R. L., & Waugh, N. C. Age-related differences in mental performance. Aging and Human Development, 1972, 3, 19-24.

Harman, H. H. Modern factor analysis. Chicago: University of Chicago Press, 1976.

Hawk, J. Linearity of criterion-GATB aptitude relationships. Measurements and Evaluation in Guidance, 1970, 2, 249-251.

Holland, J. L. Making vocational choices: A theory of careers. Englewood Cliffs, NJ: Prentice Hall, 1973.

Humphreys, L. G. The organization of human abilities. American Psychologist, 1962, 17, 475-483.

Humphreys, L. G. The construct of general intelligence. Intelligence, 1979, 3, 105-120.

Hunter, J. E. Factor analysis. In P. Monge & J. Cappella (Eds.), Multivariate techniques in human communication research. New York: Academic Press, 1980.

Hunter, J. E., & Gerbing, D. W. Unidimensional measurement, second order factor analysis, and causal models. In B. M. Staw & L. L. Cummings (Eds.), Research in Organizational Behavior (Vol. 4). Greenwich, CT: JAI, 1982.

Hunter, J. E., & Schmidt, F. L. Fitting people to jobs: The impact of personnel selection on national productivity. In E. A. Fleishman (Ed.), Human Performance and Productivity, in press.

Hunter, J. E., & Schmidt, F. L., & Jackson, G. B. Meta-analyses: Cumulating research findings across studies. Beverly Hills, CA: Sage Publications, 1982.

APPENDIX

GENERAL AND SPECIFIC FACTORS

Basic Definitions

For simplicity, the definitions will be developed in terms of a concrete example. Consider the dexterity measures F and M which are both measures of the same general psychomotor factor. Let the true scores of F and M be denoted T_1 and T_2 respectively, and let the errors be denoted e_1 and e_2. If all variables are expressed in standard scores, then we can write the traditional equations of reliability theory as:

$$F = \rho_1 T_1 + \gamma_1 e_1$$

$$M = \rho_2 T_2 + \gamma_2 e_2$$

Since true scores and errors are uncorrelated, we have

$$\rho_1 = r_{FT_1} \text{ and } \gamma_1^2 = 1 - \rho_1^2$$

$$\rho_2 = r_{MT_2} \text{ and } \gamma_1^2 = 1 - \rho_2^2$$

By convention ρ_1^2 is called the reliability of F and ρ_2^2 is the reliability of M.

The specific aptitude true scores T_1 and T_2 are related to the psychomotor general factor in the same way that the observed scores are related to true scores. If we denote the psychomotor general factor by GM, then the regression of the true scores onto GM is given by

$$T_1 = \lambda_1 GM + \omega_1 S_1$$

$$T_2 = \lambda_2 GM + \omega_2 S_2$$

where S_1 is the error in predicting the true score for F from the psychomotor general factor GM, and S_2 is the error in predicting the true score for M from the psychomotor general factor GM. If all factors are expressed in standard scores, then:

$$\lambda_1 = r_{T_1 GM} \text{ and } \omega_1^2 = 1 - \lambda_1^2$$

$$\lambda_2 = r_{T_2 GM} \text{ and } \omega_2^2 = 1 - \lambda_2^2$$

The errors of prediction (as opposed to the random errors of measurement e_1 and e_2) S_1 and S_2 are called the specific factors for F and M respectively.

There is an analogy between specific factors and errors of measurement: Specific factors are to the general factor as errors are to true scores (i.e., deviations from an expected uniformity). However the analogy ends there. Errors of measurement stem from random processes in single item responses and will not correlate with any later measurement including job performance. On the other hand, the specific factor represents those aspects of learning or interest or special ability which differentiate between performance in different tasks of the same general kind. These specific factors might correlate with performance in certain jobs which rely on exactly these same skills.

To quantify these relationships, let us combine the two prediction processes into one equation. That is, we substitute the equation relating the true score to the general factor into the equation relating the true score to the observed score.

$$F = \rho_1 T_1 + \gamma_1 e_1 = \rho_1(\lambda_1 GM + \omega_1 S_1) + \gamma_1 e_1$$

$$= \rho_1\lambda_1 GM + \rho_1\omega_1 S_1 + \gamma_1 e_1$$

$$M = \rho_2\lambda_2 GM + \rho_2\omega_2 S_2 + \gamma_2 e_2$$

If we define new parameters by $\alpha_i = \rho_i\lambda_i$ and $\beta_i = \rho_i\omega_i$, then we can write these equations more simply as:

$$F = \alpha_1 GM + \beta_1 S_1 + \gamma_1 e_1$$

$$M = \alpha_2 GM + \beta_2 S_2 + \gamma_2 e_2$$

All of the parameters can be defined in te[illegible] correlations and several of them have special names. Let us spell out thes[illegible]ions for F:

$$\alpha_1 = r_{F,GM} = r_{FT_1} r_{T_1 GM}$$

$$\alpha_1^2 = \text{the communality of } F$$

$$\alpha_1^2 + \beta_1^2 = \text{the reliability of } F$$

$$\alpha_1^2 + \beta_1^2 + \gamma_1^2 = 1$$

Since the variables are all in standard score form, the "1" in the last equation is also the variance of F. This equation can then be translated in such a way that it can easily be remembered

$$\alpha_1^2 = \text{the general factor variance of } F$$

$$\beta_1^2 = \text{the specific factor variance of } F$$

$$\gamma_1^2 = \text{the error variance of } F$$

$$1 = \text{the total variance of } F$$

We can then write a mnemonic equation as if the variance of F were being "decomposed" into its parts:

Total Variance of F	=	General Factor Variance of F	+	Specific Factor Variance of F	+	Error Variance

Similar equations can be written for M and indeed for all the aptitudes.

Relationships Between Validity and Coefficients

From the equation developed in the previous section relating the aptitude to its general factor, its specific factor, and its error factor; we can obtain an equation relating the validity coefficients of each factor to the validit of the aptitude itself. That is, we will derive an equation relating the correlation between the aptitude and job performance to the correlations between each factor and job performance. This equation then determines the conclusions drawn in the main test.

For each aptitude X_i, let G_i be its general factor, S_i be its specific factor, and let e_i be its error factor. We then have

$$X_i = \alpha_i G_i + \beta_i S_i + \gamma_i e_i$$

If γ is job performance on some given job, then we have the following covariance formula:

$$\sigma_{X_i\gamma} = \alpha_i \sigma_{G_i\gamma} + \beta_i \sigma_{S_i\gamma} + \gamma_i \sigma_{e_i\gamma}$$

Since all variables are in standard score form, then if we had population correlations they would satisfy

$$r_{X_i\gamma} = \alpha_i r_{G_i\gamma} + \beta_i r_{S_i\gamma} + \gamma_i r_{e_i\gamma}$$

For population correlations, $r_{e_i\gamma} = 0$ for all jobs since e_i is an error factor. Thus for population correlations, we have

$$r_{X_i\gamma} = \alpha_i r_{G_i\gamma} + \beta_i r_{S_i\gamma}$$

In this equation, the three correlations are each validity coefficients: $r_{X_i\gamma}$ is the validity of the aptitude, $r_{G_i\gamma}$ is the validity of its general factor, and $r_{S_i\gamma}$ is the validity of its specific factor. Thus for population correlations, the validity of an aptitude is approximately a weighted average (i.e., the weights are α and β) of the validities of its general factor and its specific factor. If the unit weights hypothesis is also true, then for population correlations:

$$r_{X_i Y} = \alpha_i \, r_{G_i Y}$$

That is, if the unit weights hypothesis is true, then across jobs the validity of an aptitude is directly proportional to the validity of its general factor (where the constant α_i is the constant of proportionality). The validity of the aptitude would then be perfectly correlated across jobs with the validity of its general factor. Since all the aptitudes which measure the same general factor have validities which are perfectly correlated across jobs with the validity of that general factor, the validities of any two aptitudes which measure the same general factor will be perfectly correlated across jobs with each other. Thus, if all validity coefficients could be obtained without sampling error, i.e., if all validation studies could be run with infinite sample size, then the unit weights hypothesis predicts that all aptitudes which measure the same general factor would have validities that are perfectly correlated across jobs.

Alas, most validation studies must be done with small samples (i.e., $N < 500$). Therefore, $r_{e_i Y}$ will not be 0, but will depart from 0 by some random amount in each study. Even if the unit weights hypothesis is true, the correlation $r_{S_i Y}$ will not be 0 on the validation sample but will depart by some random amount from 0. Thus on a validation sample, the unit weights hypothesis becomes

$$r_{X_i Y} = \alpha_i \, r_{G_i Y} + \text{"0"} + \text{"0"}$$

where "0" indicates a quantity which varies randomly about 0. Thus sampling error in $r_{S_i Y}$ and $r_{e_i Y}$ acts across studies like error of measurement acts across test scores; it produces random departures from true values which results in attenuated correlations. Thus for actual validation studies, the unit weights hypothesis predicts that aptitudes which measure the same general factor will be much more highly correlated with each other than with aptitudes which measure different general factors. However the unit weights hypothesis cannot predict the exact size of the correlations without information as to the size of the validation studies and the size of the variation in general factor population validities.