Journal of Vocational Behavior 29, 340–362 (1986)

Cognitive Ability, Cognitive Aptitudes, Job Knowledge, and Job Performance

JOHN E. HUNTER

Michigan State University

This paper reviews the hundreds of studies showing that general cognitive ability predicts job performance in all jobs. The first section shows that general cognitive ability predicts supervisor ratings and training success. The second section shows that general cognitive ability predicts objective, rigorously content valid work sample performance with even higher validity. Path analysis shows that much of this predictive power stems from the fact that general cognitive ability predicts job knowledge (r = .80 for civilian jobs) and job knowledge predicts job performance (r = .80). However, cognitive ability predicts performance beyond this value (r = .75 versus r = [.80][.80] = .64) verifying job analyses showing that most major cognitive skills are used in everyday work. The third section of the paper briefly reviews evidence showing that it is general cognitive ability and not specific cognitive aptitudes that predict performance.

This paper has three objectives. First, it will review the massive evidence from hundreds of studies showing that general cognitive ability predicts performance on all jobs, including the so-called "manual" jobs as well as "mental" jobs.

Traditional studies measured performance in one of two ways: supervisor ratings and training success. Validity predicting performance is uniformly high for training success but varies for supervisor ratings. Validity predicting ratings is high for jobs of high complexity but lower for jobs of low complexity. The uniformly high validity predicting training success was once thought to be suspect because training might be too "academic." However, supervisor ratings were also suspect because of the subjective nature of ratings. For over 50 years industrial psychologists have pleaded with each other to use objective work sample measures of performance which eliminate both problems. However, few industrial psychologists had the resources to use work sample measures.

Address requests for reprints to John E. Hunter, Michigan State University, East Lansing, MI 48824.

0001-8791/86 $3.00

The second section of this paper will show a meta-analysis of a dozen studies with good (i.e., rigorously content valid) objective measures of job performance. The results are striking. The validity of general cognitive ability is much higher for objective work sample measures of job performance than for subjective supervisor ratings. In fact, the validity of general cognitive ability predicting work sample performance is slightly higher than the validity predicting training success. Most of these studies also had content valid measures of job knowledge. Path analysis shows that much of the predictive power of cognitive ability is explained by the relationship between cognitive ability, job knowledge, and job performance. For civilian jobs, job knowledge is correlated .80 with job performance (measured by objective work sample performance) and general cognitive ability is correlated .80 with job knowledge. Thus a correlation of (.80)(.80) = .64 would be explained by the fact that general cognitive ability predicts who will master the job knowledge and who will not. The correlation of .75 is higher yet, thus supporting the many job analyses in the synthetic validity tradition that have consistently shown that most major cognitive skills are used in everyday performance.

The third section of the paper will very briefly review evidence from 515 U.S. Employment Service studies and more than 500 U.S. Military studies showing that it is general cognitive ability and not specific cognitive aptitudes which predicts job performance. Illustrative data will be presented showing that tailoring aptitude composites to match the job does not improve on general cognitive ability in prediction except in a very small number of special cases.

The phrase "general cognitive ability" is used here in place of the term "intelligence." Most articles on job performance are now read by many lay people both in personnel administration and in law. To most lay people, "intelligence" denotes a person's genetic potential, a potential which is not measured by current intelligence tests. General cognitive ability in the industrial psychology literature means cognitive ability as it has been developed in adult workers or job applicants. General cognitive ability is usually measured by summing across tests of several specific aptitudes, usually verbal aptitude, quantitative aptitude, and sometimes technical aptitude. A typical measure would sum across a vocabulary test, an arithmetic reasoning test, and a test of 3-dimensional spatial patterns. In this work, only power tests were used to measure general cognitive ability. Military data shows that speeded tests contribute little to prediction (Hunter, 1983c, 1984b, 1985c). Speeded tests appear to measure something in addition to general cognitive ability. Speeded tests form their own factor in exploratory factor analysis. Speeded tests have noticeably lower validity. For all jobs in the military except clerical work, the speeded tests have a multiple regression weight of 0 once the power tests are entered in the equation. When multiple regression is done with

a general cognitive ability composite and a composite of the speeded tests, the speed composite adds to prediction only for clerical work.

This report synthesizes the results of hundreds of studies. The key methodology used is called "Meta-analysis" (Hunter, Schmidt, & Jackson, 1982) or "validity generalization" (Schmidt & Hunter, 1977). Meta-analysis combines results across studies. By averaging correlations across studies, the effect of sampling error is eliminated. Meta-analysis also corrects correlations for two other artifactual problems: attenuation due to error of measurement and attenuation due to range restriction on ability.

THE VALIDITY OF GENERAL COGNITIVE ABILITY PREDICTING TRAINING SUCCESS AND PERFORMANCE RATINGS

The fact that general cognitive ability predicts job performance on all jobs need not be theoretically proved. It can be demonstrated by brute force empirical studies showing positive correlations for a large representative sample of jobs from the *Dictionary of Occupational Titles*. This section will summarize the data showing that general cognitive ability predicts performance ratings and training success on all jobs. This correlational data is sufficient to show the pervasive predictive validity for general cognitive ability, but many readers find raw correlations to be too abstract to be convincing. The theoretical basis for validity is not contained in the validation studies reported in this section. The theoretical basis for validity is shown in the data that relates ability, knowledge, and performance. The data on job knowledge shows that cognitive ability determines how much and how quickly a person learns. The data also shows that cognitive ability predicts the ability to react in innovative ways to situations where knowledge does not specify exactly what to do. That data will be presented in the next section.

The main summary data showing the validity of cognitive ability comes from three large data bases: Ghiselli's life work, the 515 validation studies carried out by the U.S. Employment Service, and 30 years work on validation carried out by the U.S. military. These studies looked at performance in terms of either training success or performance ratings. The data for work sample measures of job performance will be presented in the next section on job knowledge.

Ghiselli

Table 1 presents a summary of Ghiselli's life work. Ghiselli collected validation studies over the years 1949–1973 and presented several summaries (1966, 1973). Hunter and Hunter (1984) presented a reanalysis of his work on performance ratings. A reanalysis of his findings for training success was done for this report. For this analysis of training success, Ghiselli's means were corrected for error of measurement using the average reliability of .81 found for training success measures in the U.S.

TABLE 1
The Validity of General Cognitive Ability as Found by Ghiselli (See Hunter & Hunter, 1984) for Proficiency Criteria and for Training Success

Job family	Proficiency	Training success
Manager	53	51
Clerk	54	71
Salesperson	61	—
Protective professions	42	87
Service worker	48	66
Trades and crafts	46	65
Elementary industrial	37	61
Vehicle operator	28	37
Sales clerk	27	—

Navy. The correction for restriction in range on ability used the incumbent/applicant standard deviation ratio $u = .60$ found as the average for 90 studies of training success done by the U.S. Employment Service; for performance ratings, the average reliability was .60 and the standard deviation ratio was .67.

The occupations in Table 1 are arranged in order of complexity as I judge it. Ghiselli had no formal job analysis such as was used in the U.S. Employment Service data presented below. The validity of general cognitive ability predicting performance ratings (traditionally labeled "proficiency") drops as complexity drops. That is, the more complex the job, the better cognitive ability predicts performance ratings. However, even for the simplest occupations in Ghiselli's categories, the validity is .27. This is high enough to provide for substantial improvement in work force productivity if cognitive ability is used for selection.

Ghiselli found studies of training success in almost all categories. Sales and sales clerk were the only exceptions. The figures show variability from one category to the next, but there is no simple trend across complexity. It may be that the mean validity for training success is almost constant for these occupations. The sample size for each correlation is not so large as to rule out sampling error as an explanation for the observed differences. Ghiselli did not provide sample sizes, and so this hypothesis cannot be tested directly. However, the other major data bases also show little variation in validity for training success. The validity of cognitive ability predicting training success is very high for all occupations.

U.S. Employment Service Studies

The U.S. Employment Service developed the General Aptitude Test Battery (GATB) in about 1945. This battery is still state of the art. They then did 515 validation studies over the next 30 years. There were 425

TABLE 2
The Validity of General Cognitive Ability as Found in 515 Validation Studies Conducted by the U.S. Employment Service (Hunter & Hunter, 1984); 425 Studies of Job Proficiency and 90 Studies of Training Success

Size of the data base				
	Performance ratings		Training success	
	Number of studies	Number of workers	Number of studies	Number of workers
General job families				
High complexity	36	2,455	24	1863
Medium complexity	151	12,933	54	3823
Low complexity	201	14,403	8	575
Industrial families				
Setup	17	1,114	4	235
Feeding/offbearing	20	1,219	0	0
Total	425	32,124	90	6496

Validity			
	Validity for proficiency	Validity for training	%Workers
General job families			
High complexity	58	50	15
Medium complexity	51	57	63
Low complexity	40	54	18
Industrial families			
Setup work	56	65	3
Feeding/offbearing	23	—	2

studies assessing the prediction of proficiency with a total of 32,124 workers participating. There were 90 studies of training success with a total sample size of 6496. The results were subjected to meta-analysis by Hunter (1980a, 1980b, 1981a, 1981b) and presented in summary form by Hunter and Hunter (1984). The key results are presented in Table 2, and reflect the same corrections for unreliability and restriction in range that were applied to Ghiselli's data.

The U.S. Employment Service validation data base included job analysis data gathered using six different job classification systems. All job classification systems had one dimension that was relevant to the validity of cognitive ability predicting performance ratings. However, it was the same dimension in all systems: a dimension which Hunter labeled "complexity." This dimension is well captured by Fine's (1955) "data" dimension which is available for all jobs in the *Dictionary of Occupational Titles* published by the Department of Labor (1977). The three main complexity families in Table 2 are defined by that dimension. Fine's "things" dimension

TABLE 3
The Average Validity of General Cognitive Ability Predicting Training Success for Military Jobs (Hunter, 1985c)

Job family	Number of studies	Number of workers	Validity
Mechanical	277	156,143	62
Clerical	104	42,832	58
Electronic	160	92,758	67
General technical	287	180,806	62
Total	828	472,539	62

also defined two specialized industrial categories as well: setup work and feeding/offbearing work ("feeding" is putting something into a machine; "offbearing" is taking something off a machine). Industrial setup jobs are between high and medium complexity. Feeding/offbearing jobs define the bottom category in terms of complexity. As complexity decreases, the validity of cognitive ability predicting performance ratings drops from .58 to .51 to .40 to .23. The more complex the job, the better cognitive ability predicts performance ratings. However, even for the least complex jobs, the validity is still .23. This value is high enough to yield considerable utility if cognitive ability is used for selection. However, the same studies showed that psychomotor ability is also relevant to many of the simpler jobs and has even higher validity than cognitive ability for some of these jobs. Note that there are no jobs for which cognitive ability is not valid.

The validity of cognitive ability predicting training success shows no systematic variation across job complexity. The validity is uniformly high. It is true that there are no studies for training success in feeding/offbearing jobs, but the validity is probably equally high there. There are many laboratory studies of tasks even simpler than feeding/offbearing jobs, and general cognitive ability tests predict performance on those tasks [see, for example, Bachelder & Denny (1977a, 1977b) for work on discrimination reaction time and other simple tasks].

U.S. Military Studies

The third and most massive validation data base is the set of studies of training success conducted by the U.S. military. These have been analyzed by Hunter (1983b, 1983c, 1984a, 1985c). The summary data for the four main job families in military work are presented in Table 3. Data was analyzed for 828 schools and nearly half a million military personnel. The correlations were corrected for restriction in range by each military service using either applicants or recruits for the reference population. No corrections for unreliability were made for either predictors or criteria.

Table 3 shows that there is little variation in validity for different job

categories. This is also true of individual schools if the sample size is large. The one school for which validity is lower is combat (not shown here), where the validity is usually about .45 rather than .62. Combat schools place a considerable emphasis on physical ability with scores for marksmanship, obstacle courses, strength and endurance tests, etc. Thus the military data shows that cognitive ability predicts training success with high validity in every line of work.

Summary of Validity Data on Performance Ratings and Training Success

General cognitive ability predicts performance ratings in all lines of work, though validity is much higher for complex jobs than for simple jobs. General cognitive ability predicts training success at a uniformly high level for all jobs.

GENERAL COGNITIVE ABILITY, JOB KNOWLEDGE, JOB PERFORMANCE, AND PERFORMANCE RATINGS

Work Sample Performance Measurement

Most psychologists believe that the best measure of job performance is a work sample measure. A work sample measure of performance is obtained by setting up work stations where performance can be directly observed and measured. Performance scores are then added across work stations (possibly with more weight given to more important stations). Performance can sometimes be physically measured. Otherwise, performance is compared to preset benchmarks. Work sample performance measures provide a direct objective measure of job performance.

Most psychologists would prefer to do validation studies using work sample measurement of performance. However, this is usually infeasible. First, for jobs like police officer, there would be an immense number of work stations—more than 150 tasks in the typical job analysis. Second, tasks such as crowd control may be difficult to simulate. Finally, even when work sample measurement is feasible, it is usually prohibitively expensive. Thus where there are thousands of validation studies done with supervisor ratings of performance, there are only a handful of studies using work samples. However, these studies are of very high quality and provide a unique theoretical perspective.

Two Theories of Ability and Performance

This section will spell out two theories of the relationship between cognitive ability, learning, and job performance. The classic psychological theory of Edward Thorndike and other pre-behaviorist students of performance asserted that performance of any kind is primarily dependent on learning. Since data shows that all complex learning is predicted by general cognitive ability (which they called intelligence), performance in

all complex tasks will be closely predicted by general cognitive ability. In contrast is a theory which was originally developed by extreme behaviorist psychologists but which has been recently widely promulgated by antitesting lawyers. This theory asserts that a paper-and-pencil test can only measure "academic ability" and cannot predict actual concrete performance. This theory admits that general cognitive ability tests predict scores on job knowledge tests, but behaviorist theory asserts that this is because the job knowledge test is also a paper-and-pencil test. The theory then asserts that there will be no relationship between job knowledge tests and actual job performance. Behaviorists have also asserted that findings of validity for cognitive ability predicting training success are irrelevant for predicting performance because training success also measures mere knowledge. The two theories will be stated in empirically testable form. The two theories make sharply contrasting predictions about certain correlational data. The relevant data will be presented and the two theories will be pitted against one another.

Behaviorist theory. Antitesting lawyers in recent court cases have revived a behaviorist theory of learning and performance that is virtually dead within the mainstream psychology of learning. This theory asserts the complete irrelevance of articulate verbal processes—i.e., thought and articulated knowledge—for concrete performance. This theory takes two forms. The extreme form is the behavioristic theory which says that all cognitive activity is irrelevant to behavior and hence job performance. The less extreme form of the theory admits that cognitive processes might be relevant but asserts that the processes elicited by paper-and-pencil tests are not the processes used on the job. This is the theory of "method variance." High correlations between ability tests and job knowledge tests are assumed to be due to "common method variance", i.e., the fact that both tests are paper-and-pencil tests. Ironically, these apparently hard boiled theories are actually extremely abstract. The actual nature of the paper-and-pencil elicited cognitive processes are never explained. For example, no explicit theory is given as to how there could be two kinds of memory: the memory used when a clerk remembers a stock number and the memory used by an applicant carrying out instructions in a clerical coding test. Both the extreme behaviorist and the methodological behaviorist theories are disconfirmed by the validation research done with work sample performance measures to be reported below.

Behaviorists reject the strongest evidence of validity in the empirical literature: the high correlation between general cognitive ability and training success and the high correlation between general cognitive ability and job knowledge tests. They argue that the job knowledge tests and training success measure mere knowledge which is irrelevant to performance. Their phrase for knowledge is "academic criterion measure." They reach this conclusion on the basis of a theory of performance that

stands completely at odds with the classic theory of performance which has grown up in the psychological study of learning.

Behaviorists have relied on supervisor ratings for their measurement of performance. They assumed that the problem with traditional ratings is that the rater acted as evaluator rather than as observer and recorder. Instead, they asked judges to rate the worker's performance on each important task on one or more behaviorally anchored rating scales (BARS). While they admit that there are problems with supervisor ratings, they claim that ratings are at least a measure of actual performance rather than academic knowledge. However, BARS summated rating scales have proven to be perfectly correlated with traditional trait summated rating scales.

The classic theory of learning and performance. The classic theory from psychology has over 100 years of research behind it. The theory which fits the data is essentially the theory of E. L. Thorndike and his contemporaries evolved in the 1920s. The classic theory relating ability to performance derives its predictions from the learning process. Learning may take place in a formal training environment or it may take place on the job. The parameters of learning are different for the two environments. Learning in a formal training program means absorbing knowledge which is presented directly to the student with the important features of the knowledge already emphasized. Learning on the job requires two steps. First, if a relevant event takes place, the worker must recognize the event as significant. Second, the worker must be able to formulate the lesson inherent in the event in such a way as to learn from it. Cognitive ability is critical to the recognition process because the worker must link current information to the knowledge already in memory. Cognitive ability is necessary to learning from recognition because the information must be restructured to a form relevant to *future* recognition. Thus learning on the job will be more dependent on cognitive ability than learning in a formal program.

According to the classic theory, performance is bounded by learning. If the worker has not learned what to do in a given situation, then the worker cannot respond correctly. Thus there should be a high correlation between learning and performance.

Learning is a necessary but not sufficient condition for performance. Performance may require that the worker go beyond knowledge of the job. Consider a police officer responsible for crowd control. The officer may have mastered the recognized principles of crowd control, but any actual situation will only be approximately like the situation described when the rules were given. Thus the officer must innovate to meet the specific conditions in the actual situation. Crowd control where spectators are enthralled with a fire is different from crowd control when a parade has gotten out of control which is different from crowd control when a

criminal is firing into the street. Thus the classic theory of performance predicts that cognitive ability will correlate with performance above and beyond the correlation determined by the relationship between cognitive ability and learning.

According to the classic theory, supervisors are mainly observers of performance. However, a supervisor's perceptions of performance will be colored by a variety of nonwork related factors. That is, a supervisor will be influenced by all the factors known to influence person perception, factors such as personal appearance, moral conventionality, etc. Furthermore, the classic theory would predict that supervisor perceptions will be influenced by idiosyncratic factors such as the match or mismatch between the personality of the worker and the personality of the supervisor. The classic theory predicts that supervisor performance ratings will be only an indirect measure of performance. Ratings of the same worker by different supervisors will disagree to the extent that perceptions are influenced by idiosyncratic factors. An average rating across a population of raters would eliminate the idiosyncratic component to ratings, but it would still leave nonwork factors which are common to all raters. Thus even if idiosyncratic factors are eliminated, the purified ratings will still not correlate perfectly with performance.

Correlational Predictions from the Two Theories

Behaviorist theory is stark. Behaviorists believe that the only reason that ability correlates with job knowledge is because the job knowledge test is a paper-and-pencil test and the ability test is a paper-and-pencil test. Behaviorists believe that "academic" knowledge has nothing to do with performance. Thus behaviorists predict that the correlation between job knowledge and job performance will be 0. Behaviorists believe that supervisor ratings measure performance. Thus behaviorists predict a zero correlation between knowledge and supervisor ratings. Finally, behaviorists believe that since paper-and-pencil tests measure nothing but academic as opposed to practical performance, cognitive ability tests cannot correlate with job performance.

The classic theory relating job performance to cognitive ability makes a number of contrasting correlational predictions. Because the rate and amount of learning is determined by cognitive ability, the classic theory predicts a high correlation between cognitive ability and learning. Since the amount of learning is measured by a content valid job knowledge test, the classic theory predicts a high correlation between ability and knowledge. Because performance is believed to be learned, the classic theory predicts a high correlation between knowledge and performance. Because innovative adaptation is required by most actual work situations, the classic theory predicts that cognitive ability will be even more highly correlated with performance than would be predicted from the high cor-

relation between ability and knowledge. Supervisors are predicted to be imperfect measures of performance in two ways: (1) supervisor perceptions will disagree because supervisors are influenced by idiosyncratic nonwork factors such as the personality match or mismatch with the worker, and (2) supervisors will be influenced by nonwork factors that influence all supervisors.

Testing the Two Theories

The predictions of the classic theory can be tested empirically. In order to do this, each factor must be made observable. Cognitive ability was made observable by the testing research of the first 40 years of this century (Vernon, 1957; Tyler, 1965). The learning process can be measured after the fact by measuring job knowledge. The greater the worker's job knowledge, the greater the learning which has taken place. Job performance can be measured using work sample methods.

For theoretical purposes, there are at least four key variables to be observed in validation: general cognitive ability, job knowledge, job performance, and performance ratings. For simplicity, abbreviated language for these will be used in the following discussion. The word "ability" will mean general cognitive ability. The word "knowledge" will mean job knowledge. The word "performance" will mean work sample performance. The word "ratings" will mean supervisor performance ratings. Each variable will be considered to be perfectly measured. Empirical data will be fully corrected for attenuation and range restriction to provide estimates of the corresponding correlations in an applicant population.

Once the classical theory has been mapped into observed variables, then the theory can be tested by checking the obtained correlations against predictions. The theory predicts a high correlation between ability and knowledge. The theory predicts a high correlation between knowledge and performance. The theory predicts that the correlation between ability and performance will be even higher than the correlation predicted by the high correlation between ability and knowledge. This prediction can be tested using the multiple regression of performance onto ability and knowledge together. The theory predicts that the β weight for ability will be positive and large for jobs which require a high degree of innovation on the job. Since the supervisor is aware only of the worker's performance and job knowledge and is unaware of the workers ability, a path model for the four variables should have no direct link between ability and ratings.

Behaviorist theory predicts that ability and knowledge will be highly correlated because both are paper-and-pencil tests. However, neither will be correlated with actual concrete performance and, hence, neither will be correlated with "hands on" work sample performance. On the

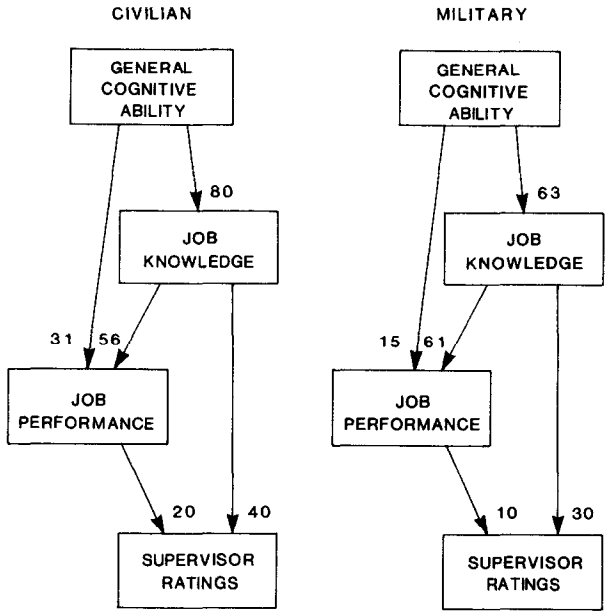

FIG. 1. A path analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. (Numbers refer to standardized path coefficients.)

other hand, corrected for error of measurement, supervisor ratings should be almost perfectly correlated with work sample performance.

The Data

Hunter (1983a) located 14 studies which measured at least three of four key theoretical variables. He analyzed the correlations between them for an incumbent population because he was interested in performance appraisal where the focal population is incumbents. For purposes of discussing personnel selection, the relevant population is the applicant population. Therefore the correlations considered here will be corrected for range restriction using the average incumbent/applicant standard deviation ratio of .67 found for 425 proficiency studies done by the U.S. Employment Service (Hunter, 1980b). These correlations were given in Hunter (1984a, 1985c). The basic results are presented in Table 4 and Fig. 1. Table 4 presents the obtained correlations between ability, knowledge, performance, and ratings in civilian and military work. Figure 1 presents the path analysis which fits this data.

The data is broken down separately for military and civilian studies. For the purpose of pitting behaviorist's theory of performance against the classic learning theory, the civilian and military results turn out to be equivalent. The differences in the results are quantitative rather than

TABLE 4
The Correlations between General Cognitive Ability, Job Knowledge, Job Performance, and Supervisor Ratings

Definition of variables

A = Ability : General cognitive ability—that factor which explains the high correlations between different cognitive aptitudes (primary factors) and between achievement in diverse areas—estimated by a composite across various aptitudes such as quantitative and verbal aptitude

JK = Job Knowledge—a content valid (based on job analysis) measure of job knowledge

WS = Job Performance—a content valid (based on job analysis) work sample measure of job performance—performance at work stations where performance can be objectively measured

SR = Supervisor Ratings—supervisor ratings of job performance—correlations are corrected for interrater reliability so these ratings are free of both random response error and halo

Correlations

	Civilian (N = 1790)					Military (N = 1474)			
	A	JK	WS	SR		A	JK	WS	SR
A	100				A	100			
JK	80	100			JK	63	100		
WS	75	80	100		WS	53	70	100	
SR	47	56	52	100	SR	24	37	31	100

Multiple correlations (see path analysis)

R for WS : 82	R for WS : 71
R for SR : 57	R for SR : 38

qualitative. Discussion of these quantitative differences is beyond the scope of this paper (see Hunter, 1983a, 1985e). In this paper, behaviorist theory and the classic learning theory will be pitted against each other.

Behaviorist Theory versus Learning Theory

The predictions of the classic learning theory are borne out in the data. The theory predicts a high correlation between ability and knowledge. The correlation between ability and knowledge is .80 in the civilian data and .63 in the military data. The classic theory predicts a high correlation between knowledge and performance. The correlation between knowledge and performance is .80 in civilian data and .70 in military data. The classic theory predicts a high correlation between ability and performance. The correlation between ability and performance is .75 in civilian data and .53 in military data. The classic theory predicts that ability will be more highly correlated with performance than can be explained solely

on the basis of the correlation between ability and knowledge. The β weight for ability (with knowledge implicitly held constant) is +.31 in the civilian data and +.15 in the military data. The classic theory predicts that ability will be correlated with ratings. The correlation between ability and ratings is .47 in the civilian data and .24 in the military data. The classic theory predicts that a path model will fit the data without a direct link between ability and ratings. This too is true for both the civilian and the military data. Thus the classic theory created by learning psychologists and supported by so much other data also fits the validation data. Every major prediction made by the classic theory is verified by the data.

Behaviorist theory predicts that a paper-and-pencil job knowledge test cannot be correlated with practical job performance. That is, in contrast to the classic theory, behaviorists predict a correlation of zero between knowledge and performance. Instead, the correlation between knowledge and performance is .80 in civilian data and .70 in military data. Behaviorists predict that a paper-and-pencil ability test cannot be correlated with practical performance. Thus in contrast to the classic theory, behaviorists predict a correlation of zero between ability and performance. Instead, the correlation between ability and performance is .75 in civilian data and .53 in military data. Behaviorists believe that ratings measure actual performance. Since behaviorists predict that a paper-and-pencil job knowledge test will correlate zero with practical performance, behaviorists predict that knowledge will be uncorrelated with performance ratings. Instead the correlation between knowledge and ratings is .56 in the civilian data and .37 in the military data. Behaviorists assert that ratings are a measure of actual job performance. Thus behaviorists predict a correlation of 1.00 between ratings and performance. Yet the correlation between ratings and performance is only .52 in the civilian data and only .31 in the military data. Thus every major prediction made by behaviorist theory is falsified by the data.

In particular, behaviorists claim that ratings measure actual performance while paper-and-pencil knowledge tests are totally irrelevant. In actuality for the civilian data, knowledge correlates .80 with performance while ratings correlate only .56. For the military data, knowledge correlates .70 with performance while ratings correlate only .37. Thus in both data sets, knowledge is a much better measure of performance than ratings.

The contrast between learning theory and behaviorist theory is stark. Every prediction made by classical learning theory is verified while every major prediction made by behaviorist theory is falsified. The classic theory fits the data perfectly. Behaviorist theory appears to be totally without merit.

Summary

For purposes of practical validation, it does not matter why cognitive ability correlates with performance. The only practical question for selection is just how high is the validity? This question is answered by the data on work sample performance: an average validity of .75 for civilian work. However, it is very important to explain why general cognitive ability predicts job performance. The data on job knowledge strongly supports the classical theory of Edward Thorndike. Articulate knowledge is critical to "complex" performance (i.e., performance that goes beyond single, short laboratory tasks). Thus Thorndike was right to expect high validity for general cognitive ability for all jobs. The data sharply disconfirm behaviorist theories which deny the importance of knowledge and other cognitive processes in the determination of concrete performance.

Cognitive ability predicts job performance in large part because it predicts learning and job mastery. Ability is highly correlated with job knowledge and job knowledge is highly correlated with job performance. The path analysis shows that this indirect causal path accounts for a majority of the effect of ability on performance.

However, the β weight for ability net of knowledge is large for civilian jobs and not negligible for military jobs. Thus ability is related to performance itself, not just to job knowledge. This may be because high ability workers are faster at cognitive operations on the job, are better able to prioritize between conflicting rules, are better able to adapt old procedures to altered situations, are better able to innovate to meet unexpected problems, and are better able to learn new procedures quickly as the job changes over time. In a variety of applied settings, content validity is being argued for general cognitive ability on complex jobs with a high degree of judgement, reasoning, and planning. The positive β weight for ability supports those linkage analyses. However, the jobs in Hunter's (1983a) data base include cooks and other workers at a much lower level of complexity than those where content validity has been claimed. This data actually suggests that all jobs involve considerable judgement and that general cognitive ability may well be content valid for all jobs.

GENERAL COGNITIVE ABILITY VERSUS SPECIFIC APTITUDES

From 1917 to 1928, psychologists did many studies showing that general intelligence predicted job performance in a wide variety of jobs. However, when Hull (1928) published his book on aptitude testing, he set off a quest for improved prediction which is still in motion. Hull started from two facts: (1) The most common method of measuring intelligence was (and is) to add across scores on tests of several different specific aptitudes such as verbal aptitude, quantitative aptitude, and technical aptitude. (2)

Multiple regression offered an optimal procedure for combining test scores. Hull hypothesized that instead of scoring tests to create one score for general cognitive ability, we might combine test scores with different weights for different jobs. By using multiple regression to tailor the weights to the job, we might greatly increase predictive validity. This section will summarize the empirical data showing that Hull's hypothesis is false. Only very meager gains are made by considering cognitive aptitudes above and beyond the measurement of general cognitive ability. Furthermore, there is only a handful of jobs where there is any gain at all.

Over the next 40 years, the differential aptitude hypothesis came to be widely accepted even though it was never empirically verified. In part this was because the differential aptitude theory was interpreted in a manner acceptable to methodological behaviorists. For example, Wernimont and Campbell (1968) argued for the differential aptitude hypothesis by claiming that the test that will best predict performance will be the test which is most similar to the job in terms of the behaviors sampled. A math test should predict work with numbers while a verbal test should predict work with words. Note, however, that their examples do not consider the actual work behaviors which occur inside the head, but rather they refer to the external stimulus materials that workers use in doing their work. This is important because the processes of learning and thinking are not directly visible and hence were overlooked by the behaviorists.

Throughout this period, there were always a few individuals who argued that the data did not fit the differential aptitude hypothesis. They argued that the data suggested that predictive power lay in general cognitive ability rather than the specific aptitudes. Early arguments were made by Vernon (1957), McNemar (1964), and Humphreys (1962, 1979). More recent arguments have come from Hunter (1980a), Jensen (1984), and Thorndike (1985).

Hull's differential aptitude hypothesis asserts that a tailored composite can *improve* over general cognitive ability. Note that it does not assert that general cognitive ability might not have a high validity in its own right. It does not assert that some aptitudes are irrelevant (i.e., invalid) for some jobs. It merely asserts that tailored weights will yield higher validity than scoring the battery for general cognitive ability. I will review evidence from contemporary military data disconfirming this hypothesis. Composites tailored to the job predict no better than composites created for other jobs and predict with slightly lower validity than a composite measuring general cognitive ability. I will focus on the military data since it is methodologically more straightforward. [see Hunter (1983b, 1983c, 1984a, 1984b, 1985c) and Hunter, Crossen, & Friedman (1985) for the original reports]. However, there is also an equally disconfirming large

data base in the 515 validation studies done in the private sector by the U.S. Employment Service (see Hunter, 1980a, 1980b).

Classification and Counseling

One reason that Hull's hypothesis was so attractive is that many benefits would follow from that hypothesis. Hull's hypothesis would be true if different jobs used different cognitive aptitudes. If different aptitudes were used on different jobs, then it would be possible to achieve a very high level of average job performance if people were matched to jobs. There would be few people who could not do some job well. Since people want to do well at their job, they would welcome counseling that would assure them that there is some job that they would do well. Brogden (1959) quantified the gains that would arise from optimal classification and showed that gain depends strongly on the size of the correlation between the aptitude composites tailored to different jobs. Hull's hypothesis can be restated as asserting that the correlation between composites is low.

Military Research

Military researchers have long been aware of Brogden's (1959) mathematical work on classification. He showed that large gains in classification depend on having low correlations between the composite scores for different jobs. Every report since his time has bemoaned the high correlations between actual composites. The only way to keep these correlations in the .80's or low .90's is to restrict the number of tests in each composite and to artificially make the composites as close to nonoverlapping as possible. Confirmatory factor analysis shows that these "reduced" correlations are only artifactually lower than .95 because of error of measurement. If the correlations were corrected for attenuation, only the clerical composites would differ from the others. The clerical composites contain the speeded tests and thus differ slightly from the others.

Whenever large samples were available, the military researchers would find positive β weights for most of the tests on each job. Actually, careful checking shows that the exceptions are the speeded tests, which are poor measures of cognitive ability. A meta-analysis across hundreds of studies shows that the speeded tests make no contribution to the prediction of success in any occupational area except clerical, and even there the contribution is minor (Hunter, 1985c). Ironically, multiple regression on large samples leads to composites that differ only trivially from the composite that best estimates general cognitive ability [for an early statement of this fact see Humphreys (1962, 1979); for a recent meta-analysis, see Thorndike (1985)]. Meta-analysis has shown that nearly all of the increase in multiple correlation due to using tailored composites has been

TABLE 5

The Average Validity for Four Tailored Composites and for a General Cognitive Ability Composite Predicting Training Success in Four Different Occupational Areas (Compiled from Hunter, 1985c)[a]

Occupational area		NT	NS	ME	EL	SS	CL	G
Mechanical	ME	50	23,269	60	60	60	54	61
Electronic	EL	41	13,537	61	66	64	60	66
Skilled services	SS	45	33,652	57	61	60	58	62
Clerical	CL	21	11,979	55	64	63	66	64
Average		157	82,437	58	63	62	60	63

[a] NT = Number of training programs; NS = Number of trainees; ME = Aptitude composite tailored to mechanical jobs; EL = Aptitude composite tailored to electronic jobs; SS = Aptitude composite tailored to skilled services jobs; CL = Aptitude composite tailored to clerical jobs; G = General cognitive ability composite.

due to sampling error. That is, had it not been for capitalization on chance and small samples, simple mechanical statistical regression would have led to the discovery of the dominance of general cognitive ability over specific aptitudes in predicting job performance. Instead, this discovery had to wait for meta-analysis. Hunter (1985b) listed the few cases in which meta-analysis has shown that there is any profit in adding any specific cognitive aptitude to general cognitive ability for predicting performance.

The severe limits of tailored composites have been made clear in the meta-analyses of military data done by Hunter cited above. One table based on recent Armed Services Vocational Aptitude Battery (ASVAB) data will be presented here to give the flavor of typical findings. Hunter's reports contain several dozen tables which list results by job family and by military branch of service going back to the mid 1950s. Table 5 is based on results from 157 training programs with performance data from 82,437 trainees. All correlations were computed without correction for attenuation in either the predictor or the criterion. Thus the correlations in Table 5 are understated by about 10%. The correlations were corrected by the military for multivariate range restriction.

Table 5 presents the average validity for four tailored composites and for a general cognitive ability composite predicting training success in four different occupational areas (compiled from Hunter, 1984b). That is, the military researchers have put together four different composite scores to be used for prediction, one for each of the four major occupational areas that have emerged from studying validity data across the years. An extreme form of the differential aptitude hypothesis would predict that "irrelevant" tests will not predict job performance in other occupational areas. For example, the composite tailored to clerical work should not

predict mechanical work and vice versa. The first row of Table 5 contains the correlations between mechanical performance and the various composites, the column for ME being the correlation for the mechanical composite. According to the extreme hypothesis, the first row of correlations in Table 5 would have a high correlation in the ME column but would be .00 in the other three columns. Instead, we see that performance in mechanical work is as well predicted by the electronic and skilled services composites as by the mechanical composite. The composite for general cognitive ability has a validity of .61 which is slightly higher than the validity for the tailored composite which is .60. The same pattern holds for electronic work and for skilled services work; performance is predicted as well by composites constructed to predict performance in other occupational areas as by the composite in its own area.

The only break in the uniformity is that the composite for clerical work predicts clerical performance slightly better than the other composites. Hunter (1985c) has shown that this is because of the presence of the speeded tests in the clerical composite. However, the β weight for speeded tests is not large and the correlation between the optimal clerical composite and the composite for general cognitive ability is .94.

The detailed analyses showing that general cognitive ability dominates specific aptitudes in the prediction of performance is presented in Hunter's (1983b, 1983c, 1984a, 1984b, 1985c; Hunter et al., 1985) military technical reports and in the detailed analysis of the U.S. Employment Service data found in Hunter (1980a).

Summary of Work on the Specific Aptitude Hypothesis

The specific aptitude hypothesis asserted that if tests were combined using weights tailored to the job, there could be a substantial increase in predictive validity over predicting performance in all jobs using general cognitive ability. A massive data base gathered by the U.S. Employment Service and even more data gathered by the U.S. military have shown the specific aptitude hypothesis to be false. The military data considered only training success but the U.S. Employment Service data was largely performance ratings.

Table 5 is illustrative of military data showing that general cognitive ability predicts training success better than tailored aptitude composites. This suggests that it is general cognitive ability on which job learning is based rather than specific cognitive aptitudes. This is consistent with the data on job knowledge which shows that the main reason for the validity of general cognitive ability is that it predicts learning. It is also consistent with the hypothesis that ability predicts performance above and beyond its prediction of job knowledge because it measures the ability to innovate and prioritize in dealing with situations that deviate from those encountered in prior training.

SUMMARY AND DISCUSSION

This paper has reviewed the evidence of hundreds of studies showing that general cognitive ability has high validity predicting performance ratings and training success in all jobs. Evidence on other predictors reviewed by Hunter and Hunter (1984) shows that if people are to be trained for their job after hiring, then there is no other predictor with validity nearly as high. This paper has reviewed the results of a small number of studies which show that general cognitive ability predicts objective work sample measures of performance with even higher validity, $r = .75$ for civilian work. No other predictor is currently known to have similar validity. This means that general cognitive ability is the best basis for job selection for all jobs where training follows hiring, which includes nearly all entry level positions.

The research using work sample measures of performance also included measures of job knowledge. This opened up a theoretical analysis of the predictive power of general cognitive ability. The correlation matrix for general cognitive ability, job knowledge, work sample performance, and performance ratings could be subjected to path analysis. This made possible a test of two theories of job performance: behaviorism and classic learning theory. Classic learning theory saw no boundaries between cognitive learning and learning on the job whereas behaviorists believed that cognitive processes are irrelevant to "real" behavior and hence job performance. The data showed that cognitive ability has a very high correlation with objective hands-on work sample measures of job performance. Classic learning theorists believed that job knowledge was the basis for job performance and often used job knowledge tests in preference to ratings as an indicator of performance. Behaviorists believed that no paper-and-pencil test could be related to behavioral performance and declared job knowledge tests to measure only "academic" learning. Behaviorists dismissed the high correlation between cognitive ability tests and job knowledge tests as mere "method variance," claiming that any two paper-and-pencil tests would be highly correlated. The data show a correlation of .80 between paper-and-pencil job knowledge tests and hands-on work sample performance. Thus the empirical data consistently support classic learning theory and sharply disconfirm behaviorist theories.

The path analysis shows that the main reason that general cognitive ability predicts job performance is because it predicts job knowledge. That is, general cognitive ability predicts the learning of the job. Job analyses have also consistently shown that all major cognitive processes such as planning and judgement and memory are used in day to day performance. General cognitive ability is correlated with all such processes and is virtually perfectly correlated with a composite measure that assesses all processes. These job analyses predict that since general cognitive

ability is used in daily performance, it will be more highly correlated with job performance than is predicted by its impact on job knowledge alone. This was supported by the path analysis. For civilian work, the correlation between ability and performance explained by the impact of general cognitive ability on job knowledge is (.80) (.80) = .64 whereas the actual correlation is .75, which is higher as predicted. The multiple regression of performance onto knowledge and ability showed a very high weight for knowledge (.56) but also showed a substantial weight (.31) for ability. Thus general cognitive ability is used in daily performance as well as in learning the job.

Hunter and Hunter (1984) noted that for purposes of certification or promotion to a similar job, general cognitive ability predicted future performance ratings about as well as current job performance. This is consistent with the path analysis since current job performance is a measure of the job knowledge that is carried over to the new job. If the jobs are similar, then that knowledge is highly relevant. High correlations found in the military between training success and later job performance are also explained by the high relevance of job knowledge for job performance. Indeed, if behaviorists had been right that knowledge gained in training is irrelevant to actual performance, the military would have been foolish for having such extensive training programs. The fact that general cognitive ability predicts day to day performance as well as learning means that cognitive ability is also a valuable predictor in promotion settings. That is, general cognitive ability should be used along with job knowledge (and objective performance measures, if available).

Since learning the job is the key to job performance, and general cognitive ability predicts learning, it is to be expected that general cognitive ability will be the key predictor of job performance. This makes the failure of the specific aptitude hypothesis more understandable. It would appear that most jobs do not build on specific aptitudes but rather build on general knowledge. It may also be that the knowledge in nonprofessional jobs is largely new to workers so that new learning swamps old learning in the determination of long term performance. Certainly recent work on job experience shows that learning on the job goes on at a high rate for at least 5 years and continues at a slower rate out to 20 years, which is as far as the data goes (McDaniel, 1986). Thus even simple jobs require far more learning than is evident to outsiders.

REFERENCES

Bachelder, B. L., & Denny, M. R. (1977a). A theory of intelligence. I. Span and the complexity of stimulus control. *Intelligence*, **1**, 127–150.

Bachelder, B. L., & Denny, M. R. (1977b). A theory of intelligence. II. The role of span in a variety of intellectual tasks. *Intelligence*, **1**, 237–256.

Brogden, H. E. (1959). Efficiency of classification as a function of number of jobs, percent rejected, and the validity and intercorrelation of job performance estimates. *Educational and Psychological Measurement*, **19**, 181–190.

Fine, S. A. (1955). A structure of worker function. *Personnel and Guidance Journal*, **34**, 66–73.

Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.

Ghiselli, E. E. (1973). The validity of aptitude tests in personnel selection. *Personnel Psychology*, **26**, 461–477.

Hull, C. L. (1928). *Aptitude testing*. Yonkers-on-Hudson, NY: World Book.

Humphreys, L. G. (1962). The organization of human abilities. *American Psychologist*, **17**, 475–483.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence*, **3**, 105–120.

Hunter, J. E. (1980a). *The dimensionality of the General Aptitude Test Battery (GATB) and the dominance of general factors over specific factors in the prediction of job performance*. Washington, DC: U.S. Employment Service.

Hunter, J. E. (1980b). *Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB)*. Washington, DC: U.S. Employment Service.

Hunter, J. E. (1981a). *The economic benefits of personnel selection using ability tests: A state-of-the art review including a detailed analysis of the dollar benefit of U.S. Employment Service placements and a critique of the low-cutoff method of test use*. Washington, DC: U.S. Employment Service.

Hunter, J. E. (1981b). *Fairness of the General Aptitude Test Battery (GATB): Ability differences and their impact on minority hiring rates*. Washington, DC: U.S. Employment Service.

Hunter, J. E. (1983a). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 257–266). Hillsdale, NJ: Erlbaum.

Hunter, J. E. (1983b). *Validity generalization of the ASVAB: Preliminary report*. Rockville, MD: Research Applications.

Hunter, J. E. (1983c). *Validity generalization of the ASVAB: Higher validity for factor analytic composites*. Rockville, MD: Research Applications.

Hunter, J. E. (1984a). *The prediction of job performance in the civilian sector using the ASVAB*. Rockville, MD; Research Applications.

Hunter, J. E. (1984b). *The validity of the Armed Services Vocational Aptitude Battery (ASVAB) high school composites*. Rockville, MD; Research Applications.

Hunter, J. E. (1985a). *The validity of the NCPOST as determined by the cumulative research findings on general cognitive ability*. Report prepared for United States vs. Nassau County.

Hunter, J. E. (1985b, April). *Are validities generalizable? An empirical assessment*. Paper presented at American Educational Research Association, Chicago.

Hunter, J. E. (1985c). *Differential validity across jobs in the military*. Rockville, MD: Research Applications.

Hunter, J. E., (1985d). *Attrition artifacts from promotion: The inapplicability of partial correlation and an alternative way to eliminate the reduction of validity caused by experience differences*. Report prepared for United States vs. Nassau County.

Hunter, J. E. (1985e). *Sample validity of the NCPOST*. Report prepared for United States vs. Nassau County.

Hunter, J. E., Crossen, J. J., & Friedman, D. H. (1985). *The validity of the Armed Services Vocational Aptitude Battery (ASVAB)*. Rockville, MD: Research Applications.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, **96**, 72–98.

Hunter, J. E., Schmidt, F. L., & Jackson, G. B. (1982). *Meta-analysis*. Beverly Hills, CA: Sage.

Jensen, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, **7**, 93–118.

McDaniel, M. A., (1986). *The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance.* Dissertation, George Washington University.

McNemar, Q. (1964). Lost: Our intelligence? Why? *American Psychologist,* **19,** 871–882.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology,* **62,** 529–540.

Thorndike, R. L. (1985). The central role of general ability in prediction. *Multivariate Behavioral Research,* **20,** 241–254.

Thorndike, R. L. (in press). *Is there any future for intelligence?* Wesleyan University Sesquicentennial.

Tyler, L. E. (1965). *The psychology of human differences.* New York: Appleton–Century–Crofts.

U.S. Department of Labor. (1977). *Dictionary of occupational titles* (4th ed.). Washington, DC: U.S. Government Printing Office.

Vernon, P. E. (1957). *The structure of human abilities* (Rev. Ed.). London: Methuen.

Wernimont, P. F., & Campbell, J. (1968). Signs, samples, and criteria. *Journal of Applied Psychology,* **52,** 372–376.

Received: August 15, 1986.