10 A Causal Analysis of Cognitive Ability, Job Knowledge, Job Performance, and Supervisor Ratings

John E. Hunter
Michigan State University

Through a meta-analysis of 14 studies, Dr. Hunter investigates the relationships among three variables: ability, job knowledge and performance. Performance is measured in two ways, work sample tests and supervisory ratings. Two causal models are presented which depict the possible relationships among the three variables. The first model suggests a direct impact of ability on performance, on job knowledge, and on supervisor ratings. This implies that there is an indirect relationship impact of job knowledge on supervisory ratings. The alternative model suggests that job knowledge is directly related to supervisory ratings. The results of the path analysis provide support for the latter model.

Dr. Guion states that Hunter's findings provide evidence for the validity of ratings. However, the model is incomplete. He suggests that the model should be enlarged to include such exogenous variables as rater characteristics, ratee characteristics, and context factors.

This paper presents a metaanalysis of 14 empirical studies that assessed the correlations between three variables relevant to performance appraisal: ability, job knowledge, and performance itself. Performance is represented by two kinds of measurement: job-sample tests and supervisor ratings.

Ability should be related to performance in two ways. First, to the extent that the job calls for reasoning, planning, or memory, speed and smoothness of performance will depend on cognitive ability. Second, ability determines the extent to which the person masters the knowledge required for efficient and excellent performance. Ability is especially important if the job requires adjustment to novel circumstances or change in behavior due to changing job requirements. The greater the extent to which the job can be laid out in manuals or training programs, the greater the extent to which ability will be relevant only indirectly through its relation to job knowledge.

Knowledge can be divided into two forms: knowledge of the technical information about objects and concepts required to do the job and knowledge of processes and judgmental criteria required for efficient or correct action on the job. Lack of knowledge reduces job performance in various ways: People must either look up the information or take time away from supervisors or coworkers when asking them for help, or they make errors stemming from incomplete or erroneous information.

Job performance itself has been measured in two ways. First, some authors have constructed simulations of the job in which performance can be directly observed and measured. That is, some researchers measure performance using a job-sample test. Other researchers assess performance by having workers rated by those supervisors who know them well enough. In most cases, the only supervisor who can rate the worker is the immediate supervisor. This was true of all the studies located for this report.

There is a causal model implicit in the previous discussion. According to this model, the work-sample test should be a measure of performance. Thus there should be a causal arrow from ability to the work-sample test representing a direct impact of ability on performance, and arrows from ability to job knowledge and from job knowledge to the work-sample test representing the indirect impact of ability on knowledge and knowledge on performance. If supervisor ratings are determined entirely by job performance, then the model should have an arrow running from the work-sample test to supervisor ratings. If this model is tested using path analysis, then there are two correlations against which the model can be tested: the correlation between ability and supervisor ratings and the correlation between job knowledge and supervisor ratings. If these correlations are correctly predicted by the path model, then the partial correlation between ability and supervisor ratings is zero if either job knowledge or work-sample score is controlled; and the partial correlation between job knowledge and supervisor rating is zero if the work-sample test is controlled.

There is an alternate model for supervisor ratings. Suppose that supervisors rarely observe routine correct performance. A supervisor is called in under two circumstances: The worker has made an error or there is to be a change in job duties. If the worker has done something wrong, then the supervisor asks the worker to explain what he or she did and then tells him or her what he or she did wrong. The supervisor thus gets a much better picture of the worker's job knowledge than of the actual job performance. On the other hand, if the worker is to change procedures, it will be the supervisor who carries out the training. During this training the supervisor will get a good picture of the ease with which the worker flexibly handles two kinds of jobs or jobs under two different sets of instructions. But it is well known that interference from new instructions tends to be greatest if the old knowledge is only partially mastered. Again the supervisor is likely to form an image of the strength of the worker on the basis of job knowledge rather than performance per se. If this alternate model is correct, then there will also be a causal arrow from job knowledge to supervisor ratings. This model can be tested against only one correlation: the correlation between ability and supervisor ratings. If the model fits, then the partial correlation between ability and supervisor ratings is zero if job knowledge is controlled.

ERROR OF MEASUREMENT

If error of measurement is not eliminated from independent variables, then partial correlations and beta weights are biased. Thus a path model can only be correctly tested if error of measurement is controlled. In this study, this is done by correcting correlations for attenuation.

Most studies report reliability coefficients if a test is used in the study. Thus the study usually reports the reliability of ability, job knowledge, or the work-sample test. These reliabilities were recorded or sometimes computed from other tables when possible.

Supervisor ratings are subject to two kinds of error. First, there is error in the assessment of the impression of the single rater: for example, a coefficient alpha on the correlations between separate ratings made by the same supervisor. Second, there is halo: the tendency for each rater to form an idiosyncratic impression of the worker that is present in all the ratings made by that rater. If the single impressions are measured perfectly, then this source of error is separately measured by the interrater correlation. If the interrater correlation is given for imperfect ratings, then it simultaneously measures both sources of error.

In the studies reported here, there was no second rater. Thus interrater correlations were not given. However, a metaanalysis of supervisor rating studies was conducted by King, Hunter, and Schmidt (1980), who found that the interrater correlation for perfect measures is virtually constant across studies. That correlation is .60. This, then, represents the upper bound on the reliability of supervisor ratings in the present study. If impression-reliability information was given, then the supervisor rating reliability was correspondingly reduced (four studies); otherwise, the overestimate of .60 was used. However, in most of the studies in which impression reliability was not given, the rating was the sum of independent ratings and was thus not likely to be less than .30 to .90.

RESULTS

Table 10.1 presents the basic data extracted or computed from all studies located that presented correlations for at least three of the four variables in the design. The studies are presented in three sets: (1) the nonmilitary studies with data on all four variables; (2) the military studies, which all had correlations for all four variables; and (3) the nonmilitary studies with data on only three variables. The

TABLE 10.1
Raw Findings from 14 Empirical Studies[a]

Authors	*Occupation*	*Sample Size*	*Reliability*				*Uncorrected Correlations*					
			r_{AA}	r_{KK}	r_{WW}	r_{SS}	r_{AK}	r_{AW}	r_{AS}	r_{KW}	r_{KS}	r_{WS}
Campbell et al. (1973)	Cartographer	443	85	88	49	60	65	48	33	52	40	23
Corts et al. (1977)	Customs Inspector	186	93	67	80	34	53	50	3	49	4	18
O'Leary and Trattner (1977)	Tax Investigator	292	93	64	78	60	54	49	23	41	19	30
Trattner et al. (1977)	Claims Examiner	233	93	81	72	60	55	36	29	34	44	29
Vineberg and Taylor (1972)	Armor Crewman	368		81		60	54	36	6	50	21	13
Vineberg and Taylor (1972)	Armor Repairman	360		76		60	44	32	15	49	20	19
Vineberg and Taylor (1972)	Supply Specialist	380		92		60	36	39	10	64	30	25
Vineberg and Taylor (1972)	Cook	366		84		60	47	35	15	49	27	19
Campbell et al. (1973)	Medical Technican	456	77	85		60	39		28		39	
Drauden (1978)	Social Worker	78	46	56		55	33		29		25	
Campbell et al. (1973)	Investment Management	384	88		95	60		51	29			27
Boyles et al. (19)	Clerical Worker	59	79		68	40		69	40			38
Schoon (1979)	Medical Laboratory Worker	160		91	95	34				72	5	32
van Rijn and Payne (1980)	Firefighter	210	93	78	76		62	50		62		

[a]N = sample size, A = ability, K = knowledge, W = work-sample test, S = supervisor ratings.

military studies did not report reliability estimates for either the ability test or the work-sample measure. The values used were average reliabilities from other research: .80 for ability and .84 for work-sample performance.

Table 10.2 presents the basic correlational data free of the effects of error of measurement. Because performance appraisal is done on incumbent populations, no correction for restriction in range was made. The results for each correlation tend to be very similar across studies. This reflects the fact that these studies have a much higher average sample size than most validation sutides: a total of 3975 across 14 studies, or an average of 284 in comparison to the average of 68 for validation studies in general (Lent, Aurbach, & Levin, 1971).

Table 10.3 presents the data for nonmilitary studies. Correlations were averaged separately for the complete and incomplete studies. If there were large variations in path model parameters in individual studies, then the average correlations across incomplete studies could be inconsistent. Table 10.3 presents the average correlations for each set of studies and the multiple regression results needed for path analyses. The path coefficients are shown in Fig. 10.1. The path model shown is that for the alternative model in which supervisor ratings depend on job knowledge as well as on job performance. The model assuming that supervisors use only performance fails the data dramatically; the average correlation between ratings and job knowledge is higher than the correlation between

TABLE 10.2
Basic Correlations from 14 Empirical Studies[a]

			Correlations					
Authors	*Occupation*	*Sample Size*	r_{AK}	r_{AW}	r_{AS}	r_{KW}	r_{KS}	r_{WS}
Campbell et al. (1973)	Cartographer	443	75	74	46	79	55	42
Corts et al. (1977)	Customs Inspector	186	67	58	5	67	8	35
O'Leary and Trattner (1977)	Tax Investigator	292	70	58	31	58	31	44
Trattner et al. (1977)	Claims Examiner	233	63	44	39	45	63	44
Vineberg and Taylor (1972)	Armor Crewman	368	66	44	9	60	30	18
Vineberg and Taylor (1972)	Armor Repairman	360	54	39	22	59	28	27
Vineberg and Taylor (1972)	Supply Specialist	380	44	48	14	77	43	35
Vineberg and Taylor (1972)	Cook	366	58	43	22	59	38	27
Campbell et al. (1973)	Medical Technician	456	50		41		55	
Drauden (1978)	Social Worker	78	65		58		45	
Campbell et al. (1973)	Investment Management	384		56	32			36
Boyles et al. (19)	Clerical Worker	59		86	71			66
Schoon (1979)	Medical Laboratory Worker	160				77	9	56
van Rijn and Payne (1980)	Firefighter	210	73	59		81		

[a]N = sample size, A = ability, K = job knowledge, W = work-sample test, S = supervisor rating.

TABLE 10.3
Correlations for the Nonmilitary Studies[a]

		Complete Studies (N = 1154)				Incomplete Studies (N = 636)			
		A	*K*	*W*	*S*	*A*	*K*	*W*	*S*
Cognitive ability	*A*	100	70	61	34	100	58	60	41
Job knowledge	*K*	70	100	65	43	58	100	79	50
Work sample	*W*	61	65	100	42	60	79	100	42
Supervisor rating	*S*	34	43	42	100	41	50	42	100

[a]N = sample size, A = ability, K = job knowledge, W = work-sample test, S = supervisor rating.

Multiple regression of work-sample performance on job knowledge and cognitive ability:

$R = .68$ $R = .81$

$\beta_K = .44\ \beta_A = .30$ $\beta_K = .67\ \beta_A = .21$

Multiple regression of supervisor rating on job knowledge and work-sample performance:

$R = .47$ $R = .50$

$\beta_K = .27\ \beta_W = .24$ $\beta_K = .45\ \beta_W = .07$

ratings and the work-sample measure in both data sets (i.e., $r_{KS} = .43$ versus $r_{WS} = .42$ in the complete studies; $r_{KS} = .50$ versus $r_{WS} = .42$ in the incomplete studies). The model predicts a correlation of .33 between ability and supervisor ratings in the complete data, where the actual correlation is .34: an error of .01. The predicted correlation in the incomplete studies is .30, where the actual correlation is .41: an error of .11. For a sample size of 635, this difference

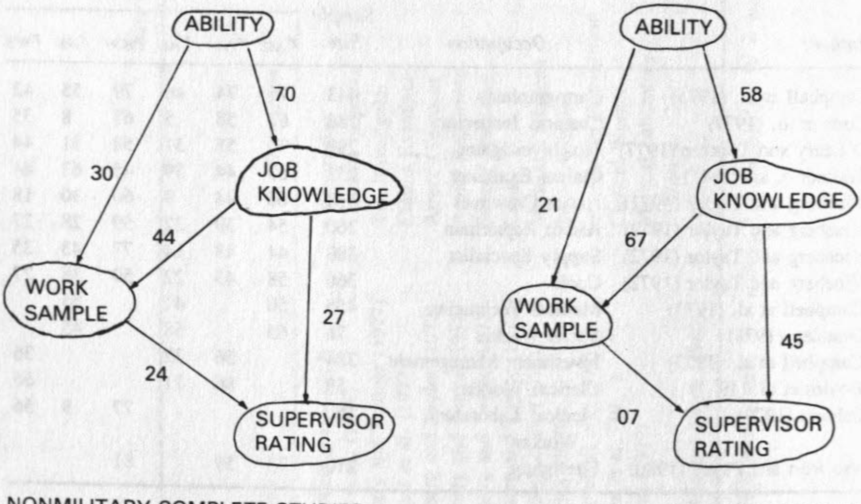

FIG. 10.1. The path models that fit the nonmilitary data with complete and incomplete studies analyzed separately.

TABLE 10.4
A Comparison of Military and Nonmilitary Correlations[a]

		Nonmilitary Studies N = 1790				Military Studies N = 1474			
		A	*K*	*W*	*S*	*A*	*K*	*W*	*S*
Cognitive ability	*A*	100	66	61	36	100	55	44	17
Job knowledge	*K*	66	100	70	45	55	100	63	35
Work sample	*W*	61	70	100	42	44	63	100	27
Supervisor rating	*S*	36	45	42	100	17	35	27	100

[a]N = sample size, A = ability, K = job knowledge, W = work-sample test, S = supervisor ratings.

$R = .73$ $R = .64$

$\beta_K = .53$ $\beta_A = .26$ $\beta_K = .56$ $\beta_A = .13$

Multiple regression of supervisor rating on job knowledge and work-sample performance:

$R = .48$ $R = .36$

$\beta_K = .31$ $\beta_A = .21$ $\beta_K = .30$ $\beta_A = .10$

of .11 is barely significant at the .05 level. The average error of .06 across the two sets of studies is not significant.

Because there is little difference between the results for complete and incomplete studies, all correlations across nonmilitary studies were averaged. These averages are presented in Table 10.4 along with the results for the military studies. Table 10.4 also presents the multiple regression results required for the path analysis. The resulting parameter estimates are shown in Fig. 10.2. Again

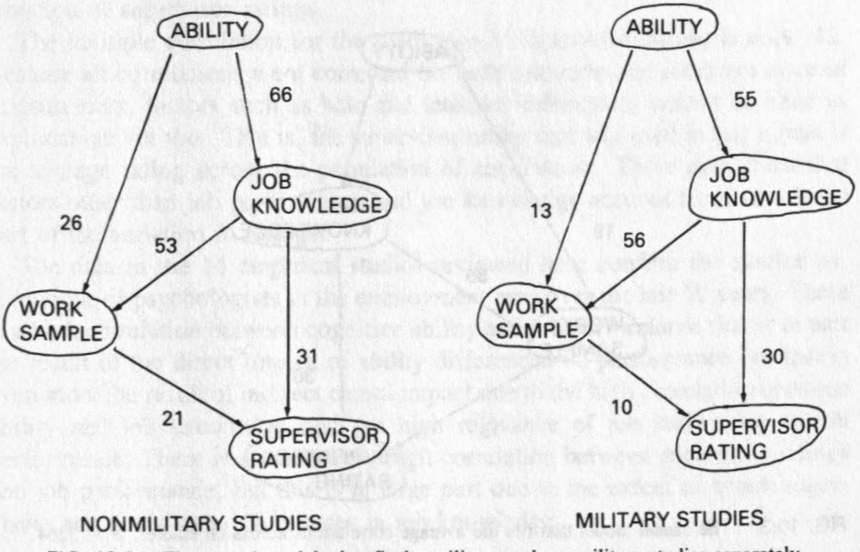

FIG. 10.2. The causal models that fit the military and nonmilitary studies separately.

TABLE 10.5
Correlations Across All Studies[a]

		All Studies N = 3264			
		A	*K*	*W*	*S*
Cognitive ability	*A*	100	61	53	27
Job knowledge	*K*	61	100	67	40
Work sample	*W*	53	67	100	35
Supervisor ratings	*S*	27	40	35	100

[a]N = sample size, A = ability, K = job knowledge, W = work-sample test, S = supervisor rating.

Multiple regression of work-sample performance on job knowledge and cognitive ability:

$$R = .68$$
$$\beta_K = .55 \ \beta_A = .19$$

Multiple regression of supervisor ratings on job knowledge and work-sample performance:

$$R = .42$$
$$\beta_K = .30 \ \beta_A = .15$$

the assumption that supervisors rely only on performance was disconfirmed in both data sets; the correlation for job knowledge was higher in both cases (for nonmilitary studies $r_{KS} = .45$ versus $r_{WS} = .42$ and for military studies $r_{KS} = .35$ versus $r_{WS} = .27$). The model predicts a correlation of .33 between ability and supervisor ratings in the nonmilitary data and the actual correlation is .34: an error of .01. For military data, the predicted correlation is .21 and the

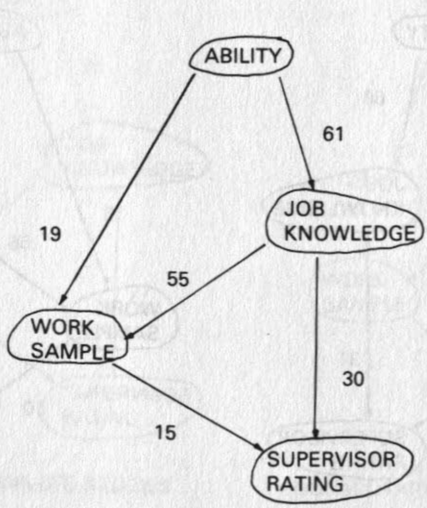

FIG. 10.3. The causal model that fits the average correlations across all studies, $N = 3264$.

actual correlation is .17: an error of −.04. Neither error is statistically significant.

Table 10.5 presents the average correlations across all 14 studies with a total sample size of 3264. Table 10.5 also contains the multiple regression results needed for the path analysis. The parameters for the path model are shown in Fig. 10.3. Again the only model presented is that that assumes that supervisors are sensitive to job knowledge as well as to job performance. Both the beta weight and the zero order correlation are larger than predicted by the performance-only model. In fact, the beta weight for job knowledge is not only greater than 0, but it is also greater than the beta weight for performance. The correlation between ability and supervisor ratings predicted by the model is .26 and the actual correlation is .27: an error of .01, which is far from large enough to be statistically significant even with a sample size of 3264.

DISCUSSION

The model shown in Fig. 10.3 shows virtually perfect fit to the average correlations across 14 studies with a total sample size of 3264. The high correlation of .53 between ability and performance derives in part from a direct causal impact of .19 and from an indirect effect of (.61) (.55) = .34 representing the effect of ability on job knowledge. The correlation of .67 between job knowledge and job performance is largely a direct causal impact of job knowledge on job performance—that is, though there is a small "spurious" influence due to the common impact of ability on both (.61) (.19) = .07. The causal analysis shows job knowledge to be twice as important as job performance in the determination of supervisor ratings.

The multiple correlation for the prediction of supervisor ratings is only .42. Because all correlations were corrected for both intrarater and interrater error of measurement, factors such as halo and leniency differences cannot be cited as explanations for this. That is, the supervisor rating that was used in this model is the average rating across the population of supervisors. These data show that factors other than job performance and job knowledge account for a very large part of the variation in ratings.

The data in the 14 empirical studies reviewed here confirm the routine assumptions of psychologists in the employment area over the last 50 years. There is a high correlation between cognitive ability and job performance that is in part the result of the direct impact of ability differences on performance but that is even more the result of indirect causal impact due to the high correlation between ability and job knowledge and the high relevance of job knowledge to job performance. There is a moderately high correlation between supervisor ratings and job performance, but this is in large part due to the extent to which supervisors are sensitive to differences in job knowledge.

REFERENCES

Boyles, W. R., Palmer, C. I., & Veres, J. G. *Empirical validation of a content validated test.* Montgomery: Center for government and public affairs, Auburn University.

Campbell, J. T., Crooks, L. A., Mahoney, M. H., & Rock, D. A. *An investigation of sources of bias in the prediction of job performance: A six year study.* Princeton, N.J.: Educational Testing Service, 1973.

Corts, D. B., Muldrow, T. W., & Outerbridge, A. M. *Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for customs inspectors.* Washington, D.C.: Personnel Research and Development Center, U.S. Office of Personnel Management, 1977.

Drauden, G. *Multijurisdictional empirical validation of an entry level social worker test battery in Minnesota.* Minneapolis: Hennepin County Personnel Department, 1978.

King, L. M., Hunter, J. E., & Schmidt, F. L. Halo in a multidimensional forced choice performance evaluation scale. *Journal of Applied Psychology,* 1980, *65,* 507–516.

Lent, R. H., Aurbach, H. A., & Levin, L. S. Predictors, criteria, and significant results. *Personnel Psychology,* 1971, *24,* 519–533.

O'Leary, B. S. & Trattner, M. H. *Research base for the written portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for internal revenue officers.* Washington, D.C.: Personnel Research and Development Center, U.S. Office of Personnel Management, 1977.

Schoon, C. G. *Correlation of performance on clinical laboratory proficiency examinations with performance in clinical laboratory practice.* New York: *Professional Examination Service,* 1979.

Trattner, M. H., Corts, D. B., van Rijn, P. P., & Outerbridge, A. M. *Research base for the written test portion of the Professional and Administrative Career Examination (PACE): Prediction of job performance for claims authorizers in the social insurance claims examining occupation.* Washington, D.C.: Personnel Research and Development Center, U.S. Office of Personnel Management, 1977.

van Rijn, P., & Payne, S. S. *Criterion related validity research base for the D.C. firefighter selection test.* Washington, D.C.: Personnel Research and Development Center, U.S. Office of Personnel Management, 1980.

Vineberg, R., & Taylor, E. N. Performance in four army jobs by men at different aptitude (AFQT) levels: 3. The relationship of AFQT and job experience to job performance. Alexandria, Va.: Human Resources Research Organization (HUMRRO), 1972.