

Predicting job performance: A comparison of expert opinion and research findings*

Stephen DAKIN
University of Canterbury, Christchurch, New Zealand

J. Scott ARMSTRONG
The Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA

Abstract

A survey was conducted of New Zealand personnel consultants. Their beliefs about the validity of various selection tools and their claimed usage of these tools was then compared with the validities in a previously published meta-analysis. The experts claimed to use the predictors they believed to be most valid. However, their beliefs about validity were unrelated to empirically demonstrated validities (Spearman's $\rho = -0.06$). Suggestions were made on the types of research that are needed to improve predictive ability in selection and on the ways in which practitioners can use existing research.

Keywords: Employee selection, Forecasting, Job performance, Predictor validity, Research vs. expert opinion.

1. Introduction

Research on personnel selection dates from the turn of the century. Meta-analysis has been used to summarize this research for more than 50 years. Recent meta-analyses have examined the results of extensive research with a variety of predictors of job performance ranging from tests of cognitive ability to handwriting analysis. This paper describes the results from a meta-analysis and contrasts these with results from a survey of experts.

2. Meta-analysis on selection

2.1. Research findings

Hunter and Hunter (1984) reviewed the validities of a range of different selection predictors. The authors examined the ability of such predictors as cognitive tests, job tryouts, interviews, biographical data forms, interest tests, age, and education in predicting job performance in a wide variety of occupations. The main criterion of job performance used in this research was ratings of performance made by supervisors.

Overall, for entry-level jobs, Hunter and Hunter found that ability tests were the most accurate single predictors. Ability tests are a combination of cognitive tests (tests of verbal and numerical reasoning) and psycho-motor tests. They claimed that "... ability tests are valid across all jobs in predicting job proficiency" (Hunter and Hunter, 1984, p. 80). For 425 studies the mean correlation of ability with supervisors' ratings of performance was 0.53. Tests of cognitive ability were the most accurate for 'thinking' jobs (e.g., manager, salesperson), while psycho-motor skills were most accurate for jobs requiring manual skills. Job tryouts were also valid predictors, as were biographical inventories. These results reinforce those from previous reviews by Ghiselli (1973), Dunnette (1972), Reilly and Chao (1982), and Vineberg and Joyner (1982). The Hunters also examined ten other types of predictors for entry-level jobs. Interviewing, a highly popular method, fared poorly with an average $r = 0.14$. Age, another popular variable, was the least valid predictor with $r = -0.01$. (See column 1 of Exhibit 1.)

* The authors thank Elmore Alexander, Georgia Chao, Martin Gimpl, John Hunter, Martha Lightwood, Ming Singer, and Myron L. Weber for helpful suggestions on earlier versions of this paper.

Exhibit 1

Estimated validity and use of predictors of managerial success (simple rankings are in parentheses).

Predictor	Research Findings Correlation of predictor to performance (validity)* 1	Expert Opinion Rankings	
		Estimates of validity** 2	Frequency of use** 3
		Cognitive ability tests	0.53 (1)
Job tryout	0.44 (2)	8.0 (9)	9.9 (11)
Biographical inventory	0.37 (3)	6.2 (7)	4.1 (4)
Reference check	0.26 (4)	4.1 (3)	3.8 (3)
Experience	0.18 (5)	2.2 (1)	3.1 (2)
Interview	0.14 (6)	3.6 (2)	2.2 (1)
Training and experience ratings	0.13 (7)	5.6 (4)	7.5 (8)
Academic achievement	0.11 (8)	6.0 (6)	6.3 (7)
Education	0.10 (9.5)	5.9 (5)	5.9 (5)
Interest	0.10 (9.5)	6.7 (8)	6.2 (6)
Age	-0.01 (11)	9.2 (11)	8.0 (9)

* These correlations are from Hunter and Hunter (1984), Table 9, p. 90

** Average ranking given by 21 New Zealand personnel consultants.

A result that surprised us was the low correlation between education and performance ($r = 0.10$). Inasmuch as cognitive ability was the best predictor of supervisors' ratings, and given that those with stronger cognitive abilities go to school longer and get better grades, this low correlation is puzzling. A low correlation also existed between grades in school and length of job tenure ($r = 0.05$).

The Hunters then translated improved personnel predictions into a cost-benefit framework, examining the costs associated with failure to use ability tests, the most accurate predictors. For example, they estimated that for US federal entry-level jobs, the use of ability tests in selection produces a benefit of \$15 billion per year.

It is important to note that the main criterion of job performance used in the Hunters' study was supervisory ratings of performance. In a metaanalysis of the relationship between supervisory ratings and actual performance, Heneman (1986) has shown that the relationship is low, especially when absolute ratings are used. As noted above, Hunter and Hunter provided evidence of some reasonably strong relationships between various predictors and other criteria including training success, tenure, job proficiency, and promotion; however, their cost-benefit analyses were based mainly on the absolute supervisory rating criterion. The limited validity of supervisory ratings of performance weakens the value of these studies.

This criticism aside, however, there is now a substantial body of research dealing with the validity and utility of different predictors for a wide range of occupations. Our central concern is with the extent to which this body of knowledge is disseminated and used by personnel practitioners who make selection decisions. Given the results of these meta-analyses, this paper addresses two questions:

- (1) Do selection practitioners hold opinions about predictor validities that agree with the research findings?
- (2) Do selection practitioners use the most valid predictors?

2.2. Expert opinions

We distributed a one-page questionnaire to a group of experts. These were personnel consultants, mainly from New Zealand, who were attending a conference in 1985. All the consultants were actively involved in middle- to senior-level managerial recruitment and selection, mainly on behalf of private sector employers. The questionnaire listed and defined 11 predictors (derived from Table 9, p. 90 of the Hunters' review). A copy of the questionnaire is provided in the appendix. For a more detailed description of these techniques, see Glueck (1982, Chapter 10). The experts were assumed to be familiar with these techniques.

We have taken two liberties in contrasting the opinions of our experts with meta-analytic research. First, the Hunters' primary data were based on *entry-level* jobs. However, the experts to whom we had access were mainly involved in managerial recruitment. We expected that the experts' backgrounds would lead them to think about their own, experience in selection for middle- to senior-level positions. Therefore, while we asked them to rank selection techniques in general, the questionnaire alluded to managerial selection. However, from the evidence cited in the Hunters' meta-analysis, it appears that the validities of entry-level predictors are similar to the validities of the same predictors for other more senior positions. In almost all the work cited by the Hunters, tests of cognitive ability, job tryout, and biographical information have higher validities than interviews, education, age, and interest regardless of the occupational position concerned. That is, the ranking of the predictors according to the validity seems to be preserved across occupational levels.

Second, because we are dealing with management recruiters, we simply listed "Cognitive ability tests" rather than Hunters' "Ability composite" which includes both cognitive and psycho-motor predictors. The experts who completed our questionnaire were judged unlikely to employ psychomotor tests in selection. Further, the Hunters' re-analysis of Ghiselli's work shows the mean validity of general cognitive ability tests for managerial positions to be identical to the mean ability-composite validity for entry-level jobs ($r = 0.53$).

In short, these two 'liberties', taken because of the nature of the sample, should have no substantive effect on the conclusions presented here.

The consultants showed moderate agreement among one another in their rankings. Kendall's coefficient of concordance (W) for their rankings of validity was $W = 0.41$ (chi-squared = 85.7; $df = 20$; $p < 0.001$). For rankings of use, $W = 0.57$ (chi-squared = 119.5; $df = 20$; $p < 0.001$). The results for the 21 experts who replied are presented in columns 2 and 3 in Exhibit 1.

3. Research vs. practice

Exhibit 1 shows a high correlation between validity estimates and frequency of use by practitioners in selection (Spearman's rho for the two sets of rankings is 0.87). This indicates that selection practice is related to the consultants' beliefs about validity. However, their beliefs have no relationship with the research evidence (rho = 0.06), nor does their practice agree with the research (rho = - 0.06).

Those in the business of personnel selection in New Zealand seem to have an inaccurate view of the validities of various predictors. In particular, it is interesting that tests of cognitive ability were ranked lowest in validity when the reverse is true. Also of interest is the high ranking given to the use of interviews, which have been the subject of extensive research (and criticism) throughout the world (Zedeck and Cascio, 1984). Indeed, a critique of the value of interviews for selection was made in a leading New Zealand business newspaper a year prior to our administration of the questionnaire to the experts (Penney, 1984). The conclusion that the interview is the most widely used method is consistent with the evidence summarized in Robertson and Makin (1986) and in Weekley and Gier (1987).

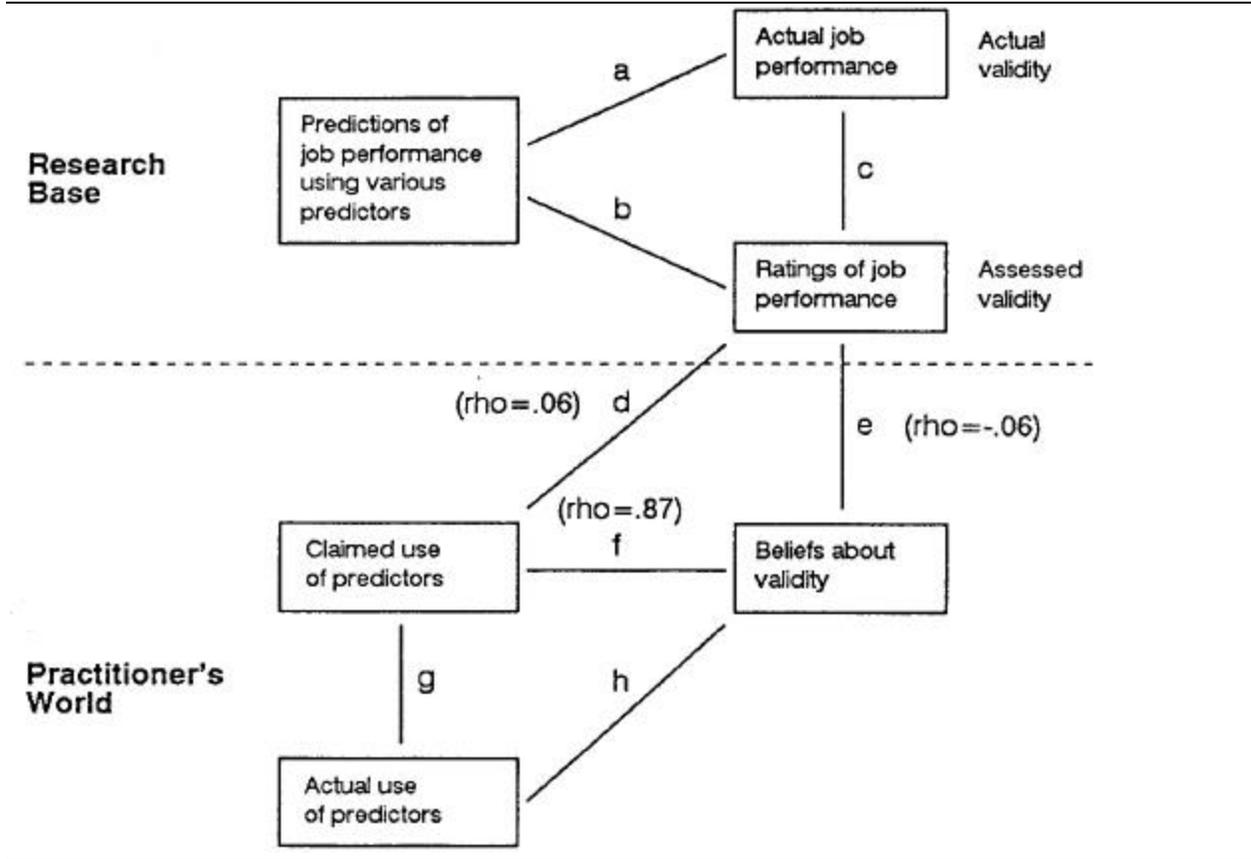
4. Discussion

There is an apparent lack of research knowledge displayed by personnel professionals, and a slow diffusion of research knowledge into practice. For example, major meta-analyses demonstrating the low validities of selection interviews, ranked number two in validity by our experts, have been available for over 30 years. One might expect that professionals should be reasonably current with the research base underlying their profession and would use selection techniques of proven validity. Such is not the case.

Exhibit 2 shows how the study in this paper ties into the prior research. The relationships between various types of predictors and actual performance in different kinds of jobs are denoted by the link 'a' in Exhibit 2. Because actual performance is often difficult to measure accurately, proxies such as supervisory ratings of performance (link 'b') are used. The dotted line represents the barrier between research and practice. In the practitioners' world, the use of the various predictors should be governed by beliefs about validity, although it is likely to be influenced also by convenience and by inertia.

Exhibit 2

Studying the Relationship between Selection Research and Professional Practice



The present study has focused on linkages between claimed use and assessed validity (link 'd'), beliefs and assessed validity (link 'e'), and beliefs and claimed use (link 'f'). Much existing research examines the relationship between predictions and job performance ratings (link 'b'). A smaller, but apparently growing interest is being shown in predictions and actual performance (link 'a') and ratings of performance versus actual performance (link 'c'). Link 'a' is of particular importance; there are too few research studies in which actual performance is used as the criterion. As the Hunters note (personal communication), objective performance measures are not easy to derive, particularly for managerial roles. Nevertheless, in light of work on the accuracy of supervisory ratings (Alexander and Wilkins, 1982; Heneman, 1986), efforts must be made to move away from such proxies. Promising work in this area has been provided by Kritzman (1986), in the case of investment analysts, and by Weekley and Gier (1987), using sales performance. Another useful area for study is athletics, where detailed, accurate performance measures are available over long time periods.

It would be useful to explore the extent to which claimed usage and actual usage are related (link 'g'). The present results suggest that selection professionals place too much emphasis on predictors that have low validities, while ignoring those with higher validities. However, this conclusion is based on the assumption that actual usage and claimed usage are related. Some evidence questions how much insight professionals have into their own decision-making behavior (Webster, 1964; Slovic, Fleissner, and Bauman, 1972; Nisbett and Wilson, 1977). That is, personnel consultants may not be fully aware of the variables they actually use in selection or how they combine them. Their claimed use of different predictors may not be as closely related to actual use as found in this paper.

Failure to use the more effective predictors may be due to:

- (1) Lack of knowledge: The fact that *estimates of validity* were uncorrelated with *assessed validity* is partial evidence of lack of information about the results of recent research.
- (2) Disbelief: As in our case, personnel practitioners may wish to have better evidence of the validities of different predictors. Many personnel specialists continue to believe that *they* are

effective interviewers, even when confronted with the evidence that interviews in general have low validities. This 'persistence of judgmental confidence' results partly from the absence of adequate feedback about interview decisions (Arvey and Campion, 1982, p. 316).

- (3) Lack of motivation to change current practice: There is probably considerable inertia in the marketplace. For example, employers and job applicants believe that job interviews are a necessary part of the hiring process. Also, consumer resistance to the use of objective psychological testing has developed over decades of testing.

Personnel selection research has focused on the utility of improved predictive ability (e.g., Janz and Etherington, 1985). Empirical evidence supports the view that the more objective selection techniques yield substantial economic returns (e.g., Hunter and Hunter, 1984). These economic validation studies should encourage use of the research findings, and encourage practitioners to adopt more valid selection procedures.

5. Conclusions

Despite massive amounts of research in the past, the challenges to researchers are formidable. Can we improve techniques to predict actual performance in specific situations? Meta-analyses should help to address this issue.

Practitioners can draw some conclusions from this paper that should assist them to improve selection decisions. These relate to the communication of research findings, to an emphasis on objective methods, and to improving the interview as described below.

5.1. *Communication of research findings*

Ensure that those making selection decisions are familiar with the central conclusions to be drawn from the selection literature. People involved in selection need to know that the validities of objective selection procedures are higher than for the more subjective procedures such as interviewing. Because of perceptual distortions (Arvey and Campion, 1982), it can be difficult to convince practitioners that unstructured interviews lack validity. One way to shake established beliefs is through interactive seminars in which the concepts of validity and utility are explained. One approach is to have participants complete the questionnaire in the Appendix, and then to feed back results for discussion.

5.2. *Emphasize objective methods*

Selection practitioners should be encouraged to place more emphasis on objective selection procedures (testing, biodata) and less on subjective procedures such as unstructured interviews. In part, greater knowledge of selection validities may change selection practices, but as Arvey and Campion (1982) have pointed out, interviews are often used for reasons that have little to do with improved selection. For example, people may continue to interview, knowing that validities are low, for public relations reasons. Under these circumstances, one way to encourage greater use of objective procedures is to add objective methods to the existing subjective procedures, but to ensure that different methods are used for different purposes. Thus, objective procedures could be used to make selection decisions or to provide short lists of acceptable candidates, while interviews can be undertaken separately for the express purpose of public relations or to facilitate self-selection, or to provide a realistic job preview.

5.3. *Improve interviewing*

Orpen (1985) cites his own unpublished study showing that selection officers were unwilling to give up on interviews even when informed that psychological tests were more valid predictors of success for a particular job.

Inasmuch as the interview will continue to be widely used, why not try to improve the interview? Unstructured interviews are typically ineffective. Advocacy of structured interviews has been strong for many years (Wagner, 1949). Structured interviews can be used to obtain valid information, such as information on prior experience (Arvey, Miller, Gould and Burch, 1987). High validities have been obtained when the structured interviews are based on a formal job analysis (Wiesner and Cronshaw, forthcoming). Impressive results have also been obtained using interviews that evaluate the candidate's behavior in previous similar situations (Orpen, 1985) or the candidate's intentions relevant to subsequent behavior (Weekley and Gier, 1987). In effect, these procedures

incorporate aspects of the job tryout (second in the rankings of actual validities) into the interview. Validation research on these procedures is relatively recent (Campion, Purrell and Brown, 1988). The procedures themselves have been around for a long time (e.g., see Wagner 1949).

Appendix

Selection procedures questionnaire

We use many different techniques to select managers. Some techniques are more useful than others. Read through the following list of selection procedures, and then rank each of them in terms of how useful you feel they are and the frequency with which you use them.

	Effectiveness rank	Use rank
1. Tests of cognitive ability (e.g., intelligence tests)	_____	_____
2. Job tryout (person tries out the job for trial period)	_____	_____
3. Biographical inventory (e.g., application forms)	_____	_____
4. Reference check (telephone call to referees)	_____	_____
5. Work experience (amount and relevance of experience)	_____	_____
6. Interviews (standard employment interviews)	_____	_____
7. Training and experience ratings (ratings of performance in training and work)	_____	_____
8. Academic achievement (grades in education)	_____	_____
9. Education (type of education; qualifications)	_____	_____
10. Interest tests (level of interest expressed in different areas of work)	_____	_____
11. Age	_____	_____

Now, in the column marked 'Effectiveness rank,' rank each of the selection procedures in terms of how useful you think it would be in selection. Give the most useful procedure a rank of '1', the second most useful procedure a rank of '2', and so on. Think of usefulness in terms of validity - the most useful device is one which would predict future job performance most accurately.

Second, in the column marked 'Use Rank', rank each of the procedures in terms of the frequency with which you use them. Give the procedure you use the most a rank of '1', and so on to rank 11.

References

- Alexander, Elmore R. III and R. D. Wilkins, 1982, "Performance rating validity: The relationship of objective and subjective measures of performance," *Group and Organization Studies*, 7, 485-498.
- Arvey, Richard D. and J. E. Campion, 1982, "The employment interview: A summary and review of recent research," *Personnel Psychology*, 35, 281-322.
- Arvey, Richard D., H. E. Miller, R. Gould, and P. Burch, 1987, "Interview validity for selecting sales clerks," *Personnel Psychology*, 40, 1-12.
- Campion, Michael A., E. D. Pursell and B.K. Brown, 1988, "Structured interviewing: Raising the psychometric properties of the employment interview," *Personnel Psychology*, 41, 25-42.
- Dunnette, Marvin D., 1972, *Validity Study Results for Jobs Relevant to the Petroleum Refining Industry*. American Petroleum Institute, Washington, D.C.
- Ghiselli, Edwin E., 1973, "The validity of aptitude tests in personnel selection," *Personnel Psychology*, 26, 461-477.
- Glueck, William F., 1982, *Personnel* (Plano, Texas: BPI Publications).
- Heneman, Robert L., 1986, "The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis," *Personnel Psychology*, 39, 811-826.
- Hunter, John E. and R.F. Hunter, 1984, "Validity and utility of alternative predictors of job performance," *Psychological Bulletin*, 96, 72-98.
- Janz, Tom and L. Etherington, 1985, "Using forecasted net benefits in designing improved recruitment and selection systems," *International Journal of Forecasting*, 1, 287-296.
- Kritzman, Mark, 1986, "How to detect skill in management performance," *The Journal of Portfolio Management*, 12, no. 2, 16-20.
- Nisbett, Richard E. and T.D. Wilson, 1977, "Telling more than we can know: Verbal reports on mental processes," *Psychological Review*, 84, 231-259.
- Orpen, Christopher, 1985, "Patterned behavior description interviews versus unstructured interviews: A comparative validity study," *Journal of Applied Psychology*, 70, 774-776.
- Penney, John, 1984, "The job interview - a dinosaur in its own time," *National Business Review*, Wellington, NZ: Fourth Estate Publishing, p. 48, June 25.
- Reilly, Richard R. and G. Chao, 1982, "Validity and fairness of some alternative employee selection procedures," *Personnel Psychology*, 35, 1-62.
- Robertson, Ivan T. and P.J. Makin, 1986. "Management selection in Britain: A survey and a critique," *Journal of Occupational Psychology*, 59, 45-57.
- Slovic, Paul, D. Fleissner and W.S. Baumann. 1972, "Analyzing the use of information in investment decision making: A methodological proposal," *The Journal of Business*. 45, 283-301.
- Vineberg, R. and J.H. Joyner, 1982, *Prediction of Job Performance: Review of Military Studies*. Human Resources Research Organization, Alexandria, VA.
- Wagner, R.E., 1949, "The employment interview: A critical summary," *Personnel Psychology*, 2, 17-16.
- Webster, Edward C., 1964, *Decision making in the employment interview* (Eagle, Montreal).

Weekley, Jeff A. and J.A. Gier, 1987, "Reliability and validity of the situational interview for a sales position," *Journal of Applied Psychology*, 72, 484-487.

Wiesner, Willi H. and S.F. Cronshaw, 1988. "A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview," *Journal of Occupational Psychology*, 61, 275-290.