

Assessing Important and Observable Personal Qualities in the General Selection Interview

Christopher John Jackson¹

Guildhall University, London, UK

Keywords: Job analysis, interview

Typical job analysis methods derive personal qualities (PQs) such as traits, constructs, skills and attributes that are *important* to the job. Such methods are unsuitable to derive measurement criteria for use in general selection interviews because important PQs are not necessarily easy for the interviewer to observe and thus to assess. In this study, a job analysis method was used to derive *important and observable* PQs which were used to assess candidates within a structured, life-history, general selection interview. After correcting for range restriction and adjusting for number of variates, the multiple correlation of the PQs against success at the next stage of training was: 0.41 for non-graduates; 0.28 for staff; and 0.18 for graduates. Two possible explanations, both to do with observability of PQs, are proposed to explain these differences in predictive validity. It is argued that the proposed method can have similar validity to the situational interview for some groups of candidates without the problems and limitations of the situational interview.

Introduction

Observability and Performance Appraisal

Typical methods of job analysis used to derive measurement criteria to appraise existing employees are Kelly's Repertory Grid (Kelly, 1955) and Flanagan's critical incident method (Flanagan, 1954; Smith and Kendall, 1963). Such techniques are used to elicit information about *important* personal qualities (PQs) such as traits, constructs, attributes or skills. For example, the critical incident method has been used in the design and construction of behavioural anchored rating scales (etc.) to measure on-going performance (Bernardin, LaShells, Smith and Alvares, 1976; Burnaska and Hollmann, 1974; Cascio and Valenzi, 1977; Silverman and Wexley, 1984). In these job analysis methods, it is the importance to the job of the PQ that is stressed and not whether it is actually observable to the assessor who is going to perform the assessment. In performance appraisal of employees, observability is usually assumed since the assessor will generally have had multiple opportunities to view the candidate prior to making a performance appraisal.

Observability and the Situational Interview

Contrast the chances to observe PQs in performance appraisal with those in the ubiquitous selection interview (Arvey and Campion, 1982; Robertson and Makin, 1986). Since observability of important PQs can not be assumed in the selection interview because there is a much shorter span of time to assess the potential of a candidate, the standard solution has been to tailor the interview so that important PQs can be observed. For example, the principal aim of the situational interview (Latham, Saari, Pursell and Campion, 1980) and its variants (Janz, 1982; Orpen, 1985) is to design an interview structure which improves the chances of observing important PQs. The situational interview is recently reported to have an average general validity of 0.39 but is limited in application as it does require the candidate to be knowledgeable about the job or for the interview to be considered as a somewhat dubious 'orally administered' intelligence test (Wright, Lichtenfels and Pursell, 1989, p. 197). Such a validity coefficient is however much higher than that com-

1 This study was funded by Science 3 (Air) of the Ministry of Defence.

monly reported for the traditional interview in predicting training success (average validity is 0.1 as reported in the meta-analysis conducted by Hunter and Hunter, 1984).

Observability in the General Selection Interview

The situational interview method is therefore unsuitable when: (a) Some candidates apply for a job in which they have little or no prior experience (e. g., graduate and school leaver selection or career changing); (b) the organisation has not clearly defined the job; or (c) when the job definition is broad (e. g., joining the military as an Officer with choice of specialisation unknown). Anderson and Shackleton (1990) believed the general selection interview has lacked predictive validity in the past – this may well be because researchers have concentrated on deriving measurement criteria that are *important* rather than *important* and *observable*.

Without being able to observe a PQ it is clearly difficult to assess no matter how important it is to the job. Thus, for example, Managerial ability may well be more important than Appearance for a white collar office job, but it is the latter PQ which is sampled at the general selection interview and the former that must be inferred from statements made by the candidate. For an interviewer to assess the truthfulness of a candidate's claims, as opposed to what can be actually observed (or easily surmised), is likely to be a poor interview technique because there is only a small correlation between what employees claim they have done in the past and their current performance (Latham and Saari, 1984).

In other selection type procedures, when observability has been regarded as a problem by researchers, solutions have included tailoring an assessment centre exercise (e. g., Gaugler and Thornton, 1989) or using multiple exercises (e. g., Pynes and Bernardin, 1989).

The Aim of this Study

Having used an original job analysis method to determine PQs which emphasise both *importance* to the job and *observability* within the general selection interview, the aim of the study was to report the predictive validity of this proposed method using different types of candidate. Is the predictive valid-

ity of all types of candidate comparable to that reported for the situational interview or is it similar to that reported for the traditional interview for success at training?

Procedure

Job Analysis to Derive Important and Observable PQs

Subjects

The job analysis technique was implemented on 17 Subject Matter Experts (SMEs) who were knowledgeable of the job and interview. Numbers of SMEs were deliberately kept small: firstly, to ensure that only the most experienced of the available staff within the organisation were used and, secondly, because the proposed job analysis technique involved consensus discussion and this should not be done with large numbers. The use of less experienced SMEs to increase the size of N was not considered.

Method

Firstly, the psychological literature and reports internal to the organisation were scanned for lists of specific and global PQs, describing either Officer potential or Officer ability. The aim was to obtain a complete set of PQs used to describe Officer behaviour that have previously been considered *important*. A total of 104 PQs were chosen. The final list of material was obtained from Borman and Dunnette (1975); Edmends, Stoker and Tapsfield (1982); Helme, Willemin and Grafton (1971); Jones and Harrison (1982); Jones (1984); Nystedt and Smari (1983); Turnage and Muchinsky (1984); Tziner and Dolan (1982); Yukl and Van Fleet (1982); Zedeck, Tziner and Middlestadt (1983) and also from documents internal to the organisation. The items on the list were randomised in terms of order.

Secondly, this checklist² of important PQs was presented to the SMEs who noted what was generally observable, questionably observable or never observable at the general selection interview in question. The aim of this checklist was to familiarise the SMEs with the items and also to record their first impressions.

Thirdly, by consensus discussion, the SMEs decided which items were observable at interview. If there was 100% or near 100% agreement that an

² The checklist is available from the author upon request.

item from the checklist was observable at interview, it was noted for further analysis. The SMEs were only asked if they were observable at interview – they were not assessing the item's relevance or importance since this had been derived from Stage 1. Fourthly, using a black-board, iterative procedure, the SMEs grouped the items trying to avoid excessive overlap.

Fifthly, each group of items was then named. Thus a new list of PQs and definitions was produced which reflected what the literature considered to be important and what the SMEs considered to be observable.

Further information about how this job analysis method derived important PQs from the literature and observable PQs at interview, as well as how the usefulness of the method was assessed, is recorded in Jackson³. The derived list of PQs is shown in Table 1.

Table 1. Important and observable personal qualities derived by job analysis.

Appearance and bearing
Manner and impact
Powers of expression
General knowledge and awareness
Breadth and depth of sports activities
Breadth and depth of other activities
Educational background and academic potential
Motivation
Maturity of character

This is the list of personal qualities that was derived from the job analysis method as described in the text.

Discussion with the SMEs suggested that they expected the PQs to be observable within the interview in a variety of ways. Appearance and bearing, Manner and impact and Powers of expression could be directly observed. General knowledge and awareness, Breadth and depth of sports activities, Breadth and depth of other activities and Educational background and academic potential could be observed by judging candidate responses to direct questions on the appropriate subject. Maturity of character could be observed by judging how well the candidate responded to the stress of the interview and the maturity of views expressed. Motivation was observed by judging the amount of information that the candidate knew about the organisation. High motivation was judged to have been observed if the candidate displayed high knowledge and low moti-

vation was judged to have been observed if the candidate displayed low knowledge.

Predictive Validity of the Different Candidate Groups

Subjects

The Candidates were grouped into three categories according to their background: (a) Civilian direct entry undergraduates and graduates (GRAD, N = 71); (b) Civilian direct entry non-graduates (NGRAD, N = 116); (c) Serving non-commissioned staff seeking a commission on direct entry terms (STAFF, N = 82).

Method

The general selection interview of this study took place within a well-established assessment centre (Dexter, 1984; Jackson, footnote 2) which aimed to select candidates who would be successful at passing initial training within an extremely large and uniformed military organisation. Candidates who pass initial training join the organisation as Junior Officers. The interview was conducted by two uniformed Senior Officers of differing rank. They were highly trained and maintained structure by using an aide-memoire. One interviewer asked the questions for the first half of the interview while the other made subjective assessments with notes. They then exchanged tasks. As such the interview was highly structured so that it rigidly gave about equal chances to observe all aspects of the background of the candidate from birth to present (pre-school, school, further education and previous work). After the interview, the interviewers completed an assessment form in which they rated each of the important and observable PQs.

Criterion

Candidate result at the next stage of training was coded as: 3 (Pass on time); 2 (Pass after recourse or medical recourse); 1 (failed recourse or withdrew, failed medical after recourse, suspended after recourse); and 0 (Voluntary withdrawal, Resigned, failed for medical reasons, suspended).

Multiple regression

PQs were rated by the interviewers on a scale of +4 to -2 where 0 was equal to satisfactory, +4 was ex-

³ Jackson, C. J. (1989). *Improvement to the subjective rating methods used to select Officers for the RAF Science 3 (Air)*, Ministry of Defence, Lacon House, Theobalds Road, London, WC1X 8RY.

tremely satisfactory and -2 was unsatisfactory. A linear multiple regression model was applied to ratings of observable and important PQs (shown in Table 1) against the criterion. Since predictive validity derived from a multiple regression equation will tend to increase as the number of independent variates increase, even when added variates are of no real value, the multiple correlation was adjusted for number of variates according to the algorithm described in MINITAB (1991). A further adjustment to these measures of predictive validity was necessary because of restriction of range. Predictive validity of all candidates (i.e., those who pass and fail the interview) should be higher than just those who pass. As such range restriction was corrected for by weighting the correlation according to the standard deviation of the restricted and unrestricted predictor variates using the method outlined by Ghiselli, Campbell and Zedeck (1981).

Results

The multiple correlation, for each of the three candidate groups (Table 2), demonstrated that the predictive validity was significantly different from zero for NGRAD and STAFF only.

Table 2. Predictive validity of the observable and important personal qualities.

	Candidate groups		
	GRAD	NGRAD	STAFF
R (adj. crr)	0.18	0.41**	0.28**

** = $p < 0.01$

The criterion is result achieved at training after having been selected. The personal qualities are shown in Table 1
GRAD = Civilian direct entry undergraduates and graduates

NGRAD = Civilian direct entry non-graduates

STAFF = Serving non-commissioned member of staff

R(adj.crr) = Predictive validity measured as a multiple correlation, adjusted for number of variates. (R will tend to become larger as more variates are added to the equation – even if they are of no real value. Adjusting for number of variates provides an unbiased estimate of the population R.) It was also corrected for restriction of range.

Discussion

Predictive Validity of Important and Observable Measurement Criteria

Predictive validity of important and observable measurable criteria varied between 0.18 to 0.41 depending upon candidate type. At its best, it compared favourably with situational interviews that have an average validity of 0.39 (Wright, Lichtenfels and Pursell, 1989) and, at its worst, it was a poor method with a validity that was little better than that of the traditional interview (average validity is 0.1 against training success as reported by Hunter and Hunter, 1984). As such, there is evidence that the rating of observable and important PQs of non-graduates and staff within a structured life-history interview can produce validity coefficients that are very similar to those reported when using a situational interview, but without its problems and limitations. It may be that the time and effort used in the design of situational interviews is not necessary if observable and important PQs are rated, and the proposed method is also much more appropriate to general selection than the situational interview because the candidate does not need to be knowledgeable about the job and/or the job definition can be broad or poorly understood.

Why had some interviewers the ability to predict potential performance of non-graduates and staff but poor ability to predict potential performance of graduates? Different leniency or severity bias for different types of candidate was rejected as an explanation since these errors have little effect on correlations unless the rating scale restricts ratings due to a ceiling effect (Jackson⁴) and there was little evidence in favour of a ceiling effect of the predictor variates. Moreover the interviewers were well trained to play their roles correctly and training can reduce types of assessment error such as leniency and the invalid halo rating error (Borman, 1979; Ivancevich, 1979; Latham, Wexley and Pursell, 1975; Pulakos, 1984; Thornton and Zorich, 1980). Rating according to different prototype errors was also unlikely since Anderson and Shackleton (1990) provided recent evidence that raters tend to rate according to the same prototype for different occupational groups and the three groups of this study are certainly not that dissimilar from each other.

4 Jackson, C.J. (1989). *Relations between rater bias, accuracy and perceived task difficulty*. Unpublished PhD Thesis, Coventry University, Coventry.

This leaves two competing 'observability' explanations for the differences in predictive validity between the three types of candidate. Firstly, it is likely that graduates possess greater cognitive complexity than non-graduates and staff and thus need to be assessed on PQs which are cognitive in nature and thus which are unobservable within the interview (e.g., Attitude towards responsibility).

A second explanation results from the moderating effect of the background of candidates. The highly structured interview was designed to assess a candidate from birth to present and although it possessed a certain amount of flexibility, its structure was similar for all candidates, whatever their background and age. It is possible that graduates have undergone rapid and recent changes in their outlook, expectations and perhaps maturity as a result of higher education which the interviewers failed to observe sufficiently as a result of their life-history perspective.

It is possible that this proposed method is therefore a very poor predictor of graduate success. Nevertheless, the emphasis on limiting interview assessment to only observable and important PQs does at least provide strong grounds for assessing candidates with a battery of tests that would allow observation of many more important abilities.

Further Research on the Observability of PQs

The issue of observability is important within the interview and needs further investigation especially since the classification of PQs into those which are observable and unobservable is indistinct. For the purpose of this study, the SMEs used a broad definition of observability which included those PQs which were directly observed (such as Powers of expression) and those which were easy to observe by assessing candidate response (such as Breadth and depth of sports activities). SMEs regarded PQs as non-observable (e.g., Managerial ability) if they required greater judgement skills on the part of the interviewer. Such unobservable assessments are likely to be a source of the attribution error as described by Herriot (1981) and Herriot and Rothwell (1981, 1983) and this would be a worthwhile area of investigation.

Previous recognition that motivation was important to measure but hard to observe led to the idea that only observable facets of motivation should be judged by the interviewers. As such, some interview questions aimed to assess motivation by asking

about the candidate's knowledge of the organisation in question and the role that it played in world affairs. Whilst this meant that this PQ was more easily observed, there is some doubt about whether or not this is the most sensible definition of motivation. The utility of this way of observing motivation needs to be assessed since it would appear to have considerable promise.

Author's Address:

Dr Chris Jackson
Department of Psychology
London Guildhall University
Old Castle Street
London E1 7NT
UK

References

- Anderson, N., & Shackleton, V. (1990). Decision making in the graduate selection interview: A field study. *Journal of Occupational Psychology*, 63, 63–76.
- Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology*, 35, 281–322.
- Bernardin, H. J., LaShells, M. B., Smith, P. C., & Alvares, K. M. (1976). Behavioural expectation scales: Effects of developmental procedures and formats. *Journal of Applied Psychology*, 61, 75–79.
- Borman, W. C., & Dunnette, M. D. (1975). Behaviour-based versus trait orientated performance ratings: An empirical study. *Journal of Applied Psychology*, 60, 561–565.
- Borman, W. C. (1979). Format and training effects on rating accuracy and rater errors. *Journal of Applied Psychology*, 64, 410–421.
- Burnaska, R. F., & Hollmann, T. D. (1974). An empirical comparison of the relative effects of rater response biases on three rating formats. *Journal of Applied Psychology*, 59, 307–312.
- Cascio, W. F. & Valenzi, E. R. (1977). Behaviourally anchored scales: Effects of education and job experience of raters and ratees. *Journal of Applied Psychology*, 62, 278–282.
- Dexter, R. A. (1984). Officer selection in the eighties. *Air Clues*, September, 348–350.
- Edmunds, K. J., Stoker, P., & Tapsfield, P. (1982). *Instructor judgements of student potential in flying training – A study using repertory grid technique and consensus methodology*. Chief Scientist (Royal Air Force), CS (RAF) Record No. 21/82.
- Flanagan, J. C. (1954). The critical incident technique. *Psychological Bulletin*, 51, 327–358.
- Gaugler, B. B., & Thornton, G. C. (1989). Number of assessment centre dimensions as a determinant of assessor accuracy. *Journal of Applied Psychology*, 74, 611–618.
- Ghiselli, E. E., Campbell, J. P., & Zedeck, S. (1981). *Mea-*

- surement theory for the behavioural sciences. U.S.A.: W. H. Freeman and Company.
- Helme, W. H., Willemin, L. P., & Grafton, F. C. (1971). *Dimensions of leadership in a simulated combat situation*. Technical Research Report 1172, July, Behaviour and Systems Research Laboratory, Arlington, Virginia.
- Herriot, P., & Rothwell, C. (1981). Organizational choice and decision theory: Effects of employers' literature and selection interview. *Journal of Occupational Psychology*, 54, 17–31.
- Herriot, P., & Rothwell, C. (1983). Expectations and impressions in the graduate selection interview. *Journal of Occupational Psychology*, 56, 303–314.
- Herriot, P. (1981). Towards an attributional theory of the selection interview. *Journal of Occupational Psychology*, 54, 165–173.
- Hunter, J. E., & Hunter, R. F. (1984) Validity and utility of alternative predictors of job performance. *Psychological Bulletin*, 96, 72–98.
- Ivancevich, J. M. (1979). Longitudinal study of the effects of rater training on psychometric error in ratings. *Journal of Applied Psychology*, 64, 502–508.
- Janz, T. (1982). Initial comparisons of patterned behaviour descriptions interviews versus unstructured interviews. *Journal of Applied Psychology*, 67, 577–580.
- Jones, A., & Harrison, E. (1982). Prediction of performance in initial officer training using reference reports. *Journal of Occupational Psychology*, 55, 35–42.
- Jones, A. (1984). A study of pre-assessment centre candidate short-listing. *Journal of Occupational Psychology*, 57, 67–76.
- Kelly, G. A. (1955). *The psychology of personal constructs*. New York, Norton, Vols 1 and 2.
- Latham, G. P., & Saari, L. M. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology*, 69, 569–573.
- Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology*, 65, 422–427.
- Latham, G. P., Wexley, K. N., & Pursell E. D. (1975). Training managers to minimize rating errors in the observation of behaviour. *Journal of Applied Psychology*, 60, 550–555.
- MINITAB Reference Manual, Release 8. (1991). Minitab Inc., 3081 Enterprise Drive, State College, PA 16801–3008, USA.
- Nystedt, L., & Smari, J. (1983). Mixed standard scales and behaviourally anchored scales: A multimethod-multidimension-multirater approach. *Reports from the Department of Psychology, Number 607*, August, The University of Stockholm, Stockholm, Sweden.
- Orpen, C. (1985). Patterned behaviour description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology*, 70, 774–776.
- Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69, 581–588.
- Pynes, J. E., & Bernardin, H. J. (1989). Predictive validity of an entry-level police officer assessment centre. *Journal of Applied Psychology*, 74, 831–833.
- Robertson, I. J., & Makin, P. J. (1986). Management selection in Britain: A survey and critique. *Journal of Occupational Psychology*, 59, 45–57.
- Silverman, S. B., & Wexley, K. N. (1984). Reaction of employees to performance appraisal interviews as a function of their participation in rating scale development. *Personnel Psychology*, 37, 703–709.
- Smith, P. C., & Kendall, L. M. (1963). Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. *Journal of Applied Psychology*, 47, 149–155.
- Thornton, G. C., & Zorich, S. (1980). Training to improve observer accuracy. *Journal of Applied Psychology*, 65, 351–354.
- Turnage, J. L., & Muchinsky, P. M. (1984). A comparison of the predictive validity of assessment centre evaluations versus traditional measures in forecasting supervisory job performance: Interpretive implications of criterion distortion for the assessment paradigm. *Journal of Applied Psychology*, 69, 595–602.
- Tziner, A., & Dolan, S. (1982). Validity of an assessment centre for identifying future female officers in the military. *Journal of Applied Psychology*, 67, 728–736.
- Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology*, 62, 191–199.
- Yukl, G. A., & Van Fleet, D. D. (1982). Cross-situational, multimethod research on military leader effectiveness. *Organizational Behaviour and Human Performance*, 30, 87–108.
- Zedeck, S., Tziner, A., & Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology*, 36, 355–370.