

SWP 2/04

**THE CONTRAST EFFECT IN A
COMPETENCY BASED SITUATIONAL
INTERVIEW**

© Mills, T.

Dr Timothy Mills
Research Fellow
Human Resource Research Centre
Cranfield School of Management
Cranfield University
Cranfield
BEDFORD
MK43 0AL
United Kingdom

Tel: +44 (0)1234 751122

Correspondence regarding this paper should be addressed to Timothy Mills
E-mail: t.i.mills@cranfield.ac.uk

ISBN 1 85905 167 7

*The Cranfield School of Management Working Paper Series has been running since 1987 with nearly 500 papers published so far. The aim of the Series is to bring together research in progress from across the School for publication to a wider audience and to facilitate discussion. Papers are reviewed by senior faculty before acceptance into the Series. The agreement of the authors should be sought before referring to its contents in other publications. This paper is available electronically on the School of Management website:
http://www.cranfield.ac.uk/som/research/working_papers*

Editorial Board

Editor: Dr Richard Croucher

Author Profile

Timothy Mills

Dr Tim Mills is Research Fellow in the HR Research Centre at Cranfield School of Management. After leaving school in 1975 Tim joined the Ministry of Defence as an apprentice in electronics. On completion of the apprenticeship he spent a number of years working as a technician in the field of defence research and development. During his time with the Ministry of Defence he studied with the Open University, graduating in 1992 with a first class honours degree in Psychology. In 1995 he completed his MSc. in Applied Psychology with Cranfield University and in that year took up an offer to study for a PhD examining aspects of recruitment in the airline industry with the college of Aeronautics. During the period of the PhD course Tim worked as a consultant on a number of organisational surveys as well as assessment centre design and validation studies. Tim joined the HRRC at Cranfield in September 1998, he was awarded his PhD in 2000 and is a Chartered Psychologist.

During his time at the HRRC Tim has worked on and managed numerous projects relating to Human Resource Management in both public and private sector organisations. These have included staff attitude surveys, organisational surveys and web based HR surveys. He has also worked on the Recruitment Confidence Index as well as the CRANET survey. In recent years Tim has worked increasingly with public sector organisations providing research to inform policy. These projects have included comprehensive research into assessment for work for people with disabilities on behalf of the Employment Service Disability Service and a series of projects since 2001 to assist workforce planning strategy development for the Greater London Authority, the National Health Service and the Home Office in the area of specialist and non-specialist substance misuse workers.

Contrast Effect in a Competency Based Situational Interview

The recruitment interview is used ubiquitously by organisations in the UK as part of the recruitment and selection process. Despite improvements over the years, the method is still prone to error and it is important for organisations to take what steps they can to reduce error in selection decisions. One source of error identified and accepted as a cause of bias within the interview is that of contrast effect. This effect causes assessors to base their judgements of candidates partly on a comparison to earlier performances of other candidates. This has the effect of giving inflated scores to interviewees when others are poor and lower scores when others are good. The presence of this effect is assumed due to various studies carried out within the experimental paradigm without any quantitative evidence collected from real world settings. This study collected data on 694 interviews carried out to recruit cabin crew for a major UK based airline. The data set provided 230 interview pairs for analysis. Correlational analyses showed that the prior performance of one candidate could significantly affect the selection outcome decision of a subsequent candidate. Binary Logistic Regression revealed the scores given to the subsequent candidate mediated this relationship. The implications for practice are discussed. These include improved rater training, the implications of interview timetabling, and rotation of assessor teams.

Introduction and Literature.

Evaluations of Interviewing as a Selection Method

The use of the employment interview for recruitment in the UK is ubiquitous. Keenan (1995) reported that in a survey of 536 UK organisations all reported the interview as part of the selection process. More recently the Recruitment Confidence Index survey (RCI 2003) reported that 97% of 1236 UK organisations surveyed use some form of interview (29% unstructured, 65% CV based, and 63% competency based). These findings show the interview to be the most commonly utilised selection method and presumably implies confidence in the process.

Many early studies suggested that any confidence in the selection interview was misplaced. Reilly and Chao (1982) suggested that the poor predictive validities reported in earlier reviews of interview studies (e.g. Arvey, 1979; Wright, 1969; Ulrich and Trumbo, 1965; Mayfield, 1964; and Wagner, 1949) had led to a dearth of studies around the early 1980s. Reilly and Chao report a mean validity coefficient of .19 (total n=987) derived from 12 validity studies all of which used supervisory ratings as a criterion measure. These authors concluded that the interview offered little practical utility for personnel selection.

Despite these rather disturbing findings the interview has been, and continues to be, widely used. Some rational support for this widespread use can now be given from more recent research findings. To highlight this shift it is worth considering the conflicting findings of two meta-analyses published fourteen years apart. Hunter and Hunter (1984) reported a mean validity coefficient of .14 (similar to the figure reported by Reilly and Chao two years earlier). By 1998 Schmidt and Hunter report a meta-analysis finding of .38 for unstructured interviews and .51 for structured (obtained from meta-analysis calculations by Huffcutt, Roth, and McDaniel 1996). These findings may imply two things. First, that organisations seem to be employing methods within the traditional interview format that is improving its utility. Second, that organisations are also employing a structured approach to the interview process, thus reaping the benefits of a more systematic and scientific approach to the selection process.

The potential benefits of the structured interview were highlighted as early as the 1940's (Wright, 1969). Since then many studies have consistently underlined these benefits

(Weisner and Cronshaw,1988). Campion, Palmer, and Campion (1997) report a range of validities for the unstructured interview of .14 to .33 and for the structured interview of .35 to .62.

Campion et. al. (1997) go on to describe in detail those elements of the structured interview that previous research have shown to be important for improving interview validity. These factors include amongst others basing questions on a job analysis, asking the same questions of each candidate, rating answers using multiple scales with detailed anchors, and using multiple interviewers.

These improvements to the interview process mean that it is possible to be considerably more optimistic about the employment interview than was justifiable two decades ago. Not only are practitioners able to employ the methods recommended by the research findings of many years, but the recent meta-analyses of validity studies carried out suggest they are willing to employ them too. This is positive given its position of popularity.

At the heart of the selection interview process is the assessment and rating of one individual by another. Adherence to guidelines such as those recommended by Campion et. al. (1997) serve to reduce the level of subjectivity within the process. However, rather alarmingly, 22% of respondents to the RCI survey (RCI 2003) reported that they used 'gut reactions' to make the final selection decision. Such decision making practice is likely to contribute to inaccurate selection decisions.

Sources of Error

Error in rating implies that the value assigned to an individual performance is inappropriate in some way. In laboratory based studies such notions as 'rater accuracy' can be described because of the control aspect of the performance. Typically, a target performance that is to be rated by participants of the study is objectively assigned some measure of 'true value'. This may be achieved by the multiple agreement of subject matter experts or the objective scoring of particular behavioural markers. Error in ratings can then be easily quantified by measuring the difference between the ratings assigned and the objectively scored value of the performance.

In field studies such objective classification of performances is more problematic. Therefore, a number of other statistical indicators can be assessed to infer the level of rater accuracy. For example, Halo effect (Thorndike, 1920; Lance, LaPointe and Stewart,1994) is a persistent error that has been widely researched. It manifests itself in the unduly high levels of inter-dimensional correlations found in the ratings of individuals. This should be of particular concern for any assessment process that holds as important the ability to differentiate particular qualities in the assessed person.

Leniency is the tendency for a rater to consistently rate candidates more or less favourably than other raters. Saal, Downey, and Lahey (1980) suggest the following as one conceptual definition of the effect: 'a tendency to assign a higher or lower rating to an individual than is warranted by the ratee's behaviour'. (p.417).

Central tendency is the rater's unwillingness to move ratings away from the central part of the scale. Restriction of range is the rater's unwillingness to give ratings that stray far from a mean rating. Both effects can cause problems in the overall effectiveness of an assessment process.

Central tendency and restriction of range can lead to candidates being awarded very similar ratings. In selection this causes difficulties in differentiating the good candidates from the less good. Furthermore, it makes it difficult to discriminate between candidates who are more closely matched. These problems in a selection interview can lead directly to inappropriate selection decisions.

Error in assessment can come from many sources. Examination of the literature reveals the wide spectrum of potential causes of inaccurate rating of candidates. For example, the inter-relationships between assessors and assessees including Sex effects (Walsh, Weinberg, and Fairfield; 1987) Race effects (Prewett-Livingstone, Field, Veres, and Lewis; 1996) and age differences within the candidate group (Finkelstein, Burke, and Raju; 1995). The type of rating scale used has also been shown to impact rating accuracy, for example, Campbell, Dunette, Arvey, Hellervick and Lowell (1973) and Hartel (1993). The assessment process can also affect the amount of rater error, including, length of interview (Tullar, Mullins and Caldwell; 1997) amount of rater training given (Woehr and Huffcutt, 1994) and exercise order in multiple exercise assessment (Bycio and Zoogah, 2002).

Other Ratee Performance as a Source of Error

A possible source of error that has attracted considerable interest over the last thirty years has been that of *other ratee performance*. Many researchers (Wexley, Yukl, Kovacs, and Sanders, 1972; Murphy, Balzer, Maurer, Lockhart and Eisenman, 1985; Buckley, Villanova, and Benson, 1989; Maurer and Alexander, 1991; Gaugler and Rudolph, 1992; Becker and Villanova, 1995; Alter, 1997) have concerned themselves with the nature of the impact that the quality of the performance of others has on the assessments made of target individuals.

The perceived quality of others could have one of three possible affects upon ratings given to a ratee under assessment. Firstly, it may be that the performance of others has no impact upon ratings given. This would be the ideal in assessment environments where a number of assessees or candidates are present and multiple assessments are being made.

Secondly, the performance of one individual might draw the ratings of others toward it. In practice this would mean that the good performance of one assessee would bias the ratings of others in a positive direction and conversely, poor reference performances would bias the ratings of others in a negative direction. This effect is known as Assimilation bias.

Finally, the ratings given to an individual may bias the ratings of others in a reverse direction. This would mean that poor ratings given to one individual may lead to inflated ratings given to others, or conversely high ratings may bias others in a negative direction. This has come to be termed the Contrast effect.

Wexley et al carried out an important study in 1972, and it is worth considering in detail, as the methodology used became influential in future experimental designs. The Wexley et al. study was constructed to examine how much influence other people's performances could have on the ratings given to a candidate in an interview setting.

Wexley et al. (1972) presented participants with videotaped performances of candidates at interview. Each participant observed two interviews that set the frame of reference and then a third target interview that was rated. Experimental conditions were determined by the quality of the frame of reference performances (context) followed by the target performance. The design was set such that High, Average and Low reference frames were each followed by High, Average and Low target performances (except the Average frame was not followed by an Average target). This design generated eight conditions. Table one illustrates the Wexley et al. experimental framework. It can be seen for example that in condition 2 participants were presented with two positive interview performances and then an average performance, which they were required to rate. In this way, both context and target performances observed by participants were closely controlled.

Table 1. The eight experimental conditions used by Wexley et al. (1972)

Condition	Context Performances	Target Performance
1	High High	High
2	High High	Average
3	High High	Low
4	Low Low	High
5	Low Low	Average
6	Low Low	Low
7	Average Average	High
8	Average Average	Low

Wexley et al. found small but significant effects in the High and Low target performances accounting for 1% or 2% of error variance. In the Average target ratings however 80% of sample error variance was accounted for by the frame of reference. These findings strongly suggest that contrast effect can be potentially very damaging to the accuracy of those selection decisions based on evidence from serial assessments.

Following the work of Wexley et al. and subsequent researchers into contrast effects in assessment settings (e.g. Murphy et al. 1985, Becker and Villanova, 1995) the presence of contrast error within selection interviews has become an accepted risk that requires addressing through relevant training (e.g. Alter, 1997; Cook, 1998). But how confident should we be that contrast effects exist in interviews *in practice*? The research to date, which has examined this rater error, has been entirely carried out in the laboratory, usually with student participants. The assumption is that professional assessors, working within operational recruitment interviews are prone to the same errors as students taking part in experiments with mock interviews. This may or may not be true, but there is no evidence from field data that this is the case. This is not because data collected in the field fail to reveal contrast error, but rather that the field data does not exist.

Thirty years ago, Landy and Bates (1973) called for more research to be carried out in field settings to examine contrast errors. They acknowledge the practical difficulties in collecting and analysing such data but make clear the value of examining real world rating decisions. This study provides such an examination.

Method

Data were collected from an operational assessment centre utilised by a large, UK based airline to recruit cabin crew. The recruitment cycle is seasonal and runs from September until April. Candidates are invited to take part in an assessment centre based upon CV, application form data, a numeracy test and a bespoke, pre-assessment questionnaire. The assessment centre operates over two days and includes pen and paper tests, group exercises and a final

selection interview. To reach the interview stage, candidates have to perform sufficiently well in the group exercises to be invited to stay.

The interview is structured to ensure that all candidates are given the same questions. Questions are behaviourally and situationally based, for example 'can you tell me of a time at work when you have done something to save your company money?' The interview assesses the company competencies *anticipating problems, communicating, cost awareness, self-control, action orientation* and *service orientation*. Each dimension has two or three questions that relate directly, and these are asked together. In this way the interview moves through sections that are competency specific.

Two assessors interview each candidate. One assessor is given the task of asking the questions whilst the other records the responses on a sectioned note sheet. Following the completion of the interview, the two assessors discuss the evidence recorded and assign a score for each dimension to be rated. Each dimension is rated on a -1 to 5 scale and the final decision to select is made upon the scores achieved. There is no cut score that determines this choice, but a general guide is that successful candidates should obtain a score of at least 2 across the six dimensions rated. Assessors are trained to assign a score of 2 to applicants that offer the minimum acceptable response to questions.

Assessor names, data, candidate details and ratings with comments are recorded on an assessment sheet. For the purpose of this study, assessors were also asked to record the time of the interview to permit a reconstruction of the order of interviews. Copies of assessment sheets were collected and subsequently collated. Assessor teams remained constant for the duration of each assessment day.

A total of 1599 candidates took part in the exercise stage of the assessment centre. 694 were invited to attend the interview stage of which 443 were offered jobs.

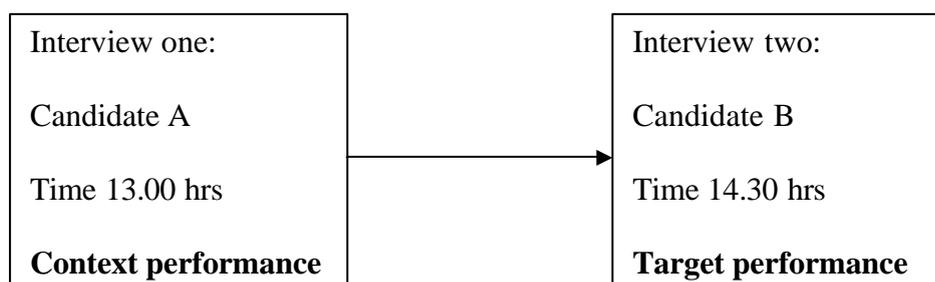
Analysis and Results

Data for each candidate were entered into an SPSS database. Variables recorded included the six competency dimension scores, the time and date of interview, the assessor identity codes and the selection decision.

From these data two new variables were created. First, a composite variable was constructed from the six dimension scores to represent overall performance in the interview. This was achieved by producing a mean value for each candidate derived from the six dimension scores. An unweighted combination of scores was appropriate, given the requirement to achieve minimum scores on all dimensions.

Second, by reordering the file the combined overall assessment score given to the preceding candidate was entered into a new variable for each case where this was possible. This variable was taken to represent the context or *frame of reference* performance. Figure 1 illustrates the relationship between the context and target assessments.

Figure 1. Interview order showing context and target.



Clearly, those interviewed first in a series, or those interviewed outside a series of interviews were unable to have a score entered. Furthermore, missing data contributed to an attrition of numbers for analysis. A total of 237 target candidates were assigned context scores, which represented the quality of the previous candidate's performance. Of these 237 candidates 152 (64%) were selected for employment with 85 (36%) being rejected.

Context and overall interview scores were examined using Kolmogorov-Smirnov tests to check for normality of distribution. Both variables were found to lack a normal distribution and therefore this informed the selection of tests selected for subsequent analyses.

To examine the inter-relationships between the three variables of Previous Candidate's Score, Target Candidate's score and Interview Outcome, Spearman's correlational analyses were carried out. The Interview Outcome variable was coded dichotomously with 1 representing a positive outcome and 0 representing a reject decision. Table 2 presents the results of these correlations.

Table 2. Spearman correlations between Interview outcome, Previous Candidate's score and Target Candidates Score.

N = 237	Previous candidate's score	Target Interview score	Interview outcome
Previous candidate's score (Context)	1	-.105 .107	-.145* .026
Target Interview score		1	.724** .000

It can be seen in table 2 that, not surprisingly, the biggest predictor of interview outcome is the performance of the target candidate. However, it should be noted that the performance of the previous candidate also significantly predicts whether the target candidate is selected. The negative direction of this correlation implies that candidates are more likely to be selected if the previous candidate had not performed well and less likely if the previous candidate had scored highly.

The association between the previous candidate's score and target candidates score approaches significance at $p=.107$. It should be noted that the ratings in both context and target interviews are given by the same assessor teams. Given the potential impact of leniency or harshness, it may be expected that these scores would be positively correlated. After all, if a team is lenient to the first candidate they may be lenient to the second. This effect may actually lessen the contrast effect observed in these scores and lead to an underestimate of the true impact.

This pattern of results suggests that the performance observed prior to a candidate being interviewed can bias the selection outcome. Given the observed direction of the association between the two interview scores it is likely that the relationship between the previous candidate's performance and the present candidate's chances of being selected will be mediated through the score given to the present candidate.

To test this assumption a Binary Logistic Regression was carried out, with the Interview outcome as the criterion and the two interview scores serving as the covariate predictors. Analysis of outliers lead to 7 cases with greater than 2.5 SD residual scores being removed (North, 2002). Bivariate Spearman correlations were again run to ensure the bivariate relationships had not changed given the removal of 7 cases. Table 3 presents these correlations, and shows the pattern of results holds with the relationship between interview scores becoming slightly more negative.

Table 3. Spearman correlations between Interview outcome, Previous Candidate's score and Target Candidates Score with outliers removed.

N = 230	Previous candidate's score	Target Interview score	Interview outcome
Previous candidate's score (Context)	1	-.114 .086	-.147* .026
Target Interview score		1	.782** .000

Table 4. Binary Logistic Regression results.

Statistic	Coefficient
-2 Log Likelihood	77.945
Goodness of Fit	112.111
Chi Square for model	217.98 df = 2 p < .0001
Overall percentage correctly predicted cases	93.9%
Number of cases in analysis	230

Variables in the equation.

Variable	B	S.E.	Wald	df	Sig	R	Exp (B)
Previous candidate score	-.5896	.5155	1.3083	1	.2527	.0000	.5546
Overall interview score	8.1657	1.3708	35.4819	1	.0000	.3364	3518.1
Constant	-14.38	2.7451	27.4476	1	.0000		

This analysis reveals that when both interview scores are entered together into the regression, only the candidate's own score predicts the interview outcome. This indicates that the variance shown to be shared in the bivariate analysis between the previous candidate's score and the interview outcome is shared with the subsequent interview score. This strongly suggests that any effect prior performances may have on another candidate's chances of selection is mediated through the ratings given to the later candidate.

The results of this analysis, when taken in conjunction with the preceding bivariate correlations show contrast effects to be present within the ratings awarded to candidates within an operation, competency-based selection interview.

Discussion and Conclusions

It has been supposed that serial recruitment interviews are susceptible to contrast effect (e.g. Wexley et al 1972; Cook, 1998). This assumption has been based entirely on data collected within the experimental paradigm (e.g. Wexley, Yukl, Kovacs, and Sanders, 1972; Murphy, Balzer, Maurer, Lockhart and Eisenman, 1985; Buckley, Villanova, and Benson, 1989; Maurer and Alexander, 1991; Becker and Villanova, 1995; Alter, 1997). This study examined data collected from operational interviews carried out to recruit cabin crew for a major UK based airline.

As shown above, this study identified a contrast effect between the previous candidate's performance and the subsequent candidate's performance. The prior candidate's score was directly correlated with the outcome of the subsequent assessor's selection decision and this was shown to be mediated through the scores awarded to the second candidate. The size of this relationship was small and appeared to support previous, lab-based studies, indicating an overall effect of between 1 to 2% error.

It should be noted however that some artefacts of the study might serve to suppress the overall size of this effect. The scores examined in this study were generated by 70 assessors working together in random pairs negotiating the scores to be given. There is likely to be a certain amount of leniency/harshness bias to be found in the scores given by all assessors (De Cottis, 1977) and although moderated by the interaction of the assessors, this may still be present. The effect of such a bias is likely to build in a positive association between scores awarded by any given assessor pair. This will serve to reduce the apparent size of the contrast effect, as it has to overcome the positive bias working in the opposite direction (Kravitz and Balzer, 1992).

The numbers of candidates reaching the final interview stage is restricted necessarily by the earlier selection process. Only 694 candidates of the 1599 who took part in the assessment centre were assessed in the interview. This should also be seen in the light of the 5000 or so applications received by the company. Thus, only some 14% of applicants reach this stage in the selection. This is likely to create a restriction in the range of performances observed at the interview stage, and will therefore also suppress the size of any correlation coefficient derived from their scores.

Another factor that may serve to minimise the effect of the context performance is the familiarity the observers are likely to have with the candidates. By the second day of assessments it is likely that the assessors will have seen candidates performing in other exercises. This observation may well bias their interview ratings and again serve to minimise the impact of the context performance on the targets' ratings.

Furthermore, it should be noted that in laboratory studies (e.g. Wexley et al. 1972; Alter, 1997) strict control over the quality combinations of context and target performances could generate much larger levels of error. The implication of this is that occasionally, when circumstance creates the *ideal* conditions, very bad decisions can be made concerning candidates, particularly of average ability. It may, therefore, be that 1 or 2 percent of error understates the actual size of the effect and the potential harm it can have.

It is apparent that when designing the interview process there are some simple steps that can be taken to alleviate the risk of contrast effect impacting recruitment decisions. For example, rater training, allowing assessors to become aware of this effect, and observing numerous example interview performances, will serve to minimise the effect (Wexley, 1973). Furthermore, the exaggerated risk associated with average performers should be made explicit to assessors.

Another method of reducing the impact of the previous interview performance would be to timetable the interviews with a greater time between candidates. Murphy et al. (1985) showed

that contrast effects are sensitive to the time between reference and target performances. These researchers found that significant levels of contrast error disappeared if a day elapsed between the reference and target performances. It is unlikely that such a long period between interviews is feasible in an organisational setting. Nevertheless, the principle holds that greater time between interviews is likely to reduce contrast errors.

The interview process examined in the present study had a short time scheduled between interviews. This time was spent dealing with the administrative issues brought up by the interview. Thus, assessors were discussing the previous performance just prior to the next candidate. In an assessment convened in this manner, assessors should be encouraged to take a break between interviews. Perhaps a discussion of all the candidate performances that day could be entered into with the other assessors before final scores are awarded.

Wherever possible, assessor teams should be rotated to maximise the range of performances observed. Rotation of assessors should be standard practice in assessment centres (Andres and Kleinman, 1993) and should be carried out wherever practical in serial interview schedules.

It has been observed that experienced assessors may be more resistant to bias due to the immediate performance of observed candidates (Maurer and Alexander, 1991; Wexley et al, 1973). This underlines the importance of rigorous training for assessors. Furthermore, restricting the size of the assessor pool within a company would increase the experience of those assessors used. For example, the centre studied here typically uses around seventy assessors each year. Assessors participate in assessment centres when it is convenient to do so. This practice is adopted to meet the other commitments of the assessors, and perhaps reflects the priority given to recruitment duties. Reducing the assessor pool would have associated pros and cons. Training costs would be reduced and resources could be better focused on those assessors selected to take part. Against this, some inconvenience and even restructuring of duties may be needed to allow assessors more time to be allocated to recruitment.

It should also be noted that motivational aspects of assessment may contribute to apparent context effects. In the present study it was observed that there was a perceived organisational imperative to hit recruitment targets. Over a week of assessments there was an implicit expectation that a certain number of recruits needed to be selected. If the overall quality of the candidate pool was lower than average then it is possible that some assessors may feel pressured into over rating borderline candidates simply to fill quotas. It should be made clear to assessors that correct selection decisions are more important to the organisation than the painless attainment of recruitment targets.

Locke (1986) makes the point that most commentators assume that laboratory findings generalise to real world settings if the essential elements of both environments equate. He goes on to suggest that an equally valid approach is to seek inductive evidence for the generalisability assumption. In this, the importance of actually looking for an effect in both settings is made clear. The findings of this study justify this assertion. Many of the laboratory studies carried out in this research area have been supported by these results. Both in size, and direction of effect, assessors within a real world recruitment setting appear to be affected by the context of other candidate's performances as much as participant assessors in the laboratory. This study supports the view that the general assertions made by the laboratory studies looking at context effects in assessment would appear to legitimately generalise to the typical employment interview and as such is the first, so far as the author is aware, to do so.

References

- Alter, P. (1997), 'Resistance of a past-behaviour interview to contrast error', *Dissertation Abstracts International Section A: Humanities and Social Sciences* Vol 58 (4 a), Oct 1997, p.1195.
- Arvey, R. (1979), 'Unfair discrimination in the employment interview: Legal and psychological aspects', *Psychological Bulletin*, Vol. 36, pp. 736-765.
- Becker, G and Villanova, P. (1995), 'Effects of Rating Procedure and Temporal Delay on the Magnitude of Contrast Effects in Performance Ratings', *The Journal of Psychology*, Vol. 129 (2), pp. 157-166.
- Buckley, R.M. Villanova, P. and Benson, P.G. (1989), 'Contrast Effects in Performance Ratings: Another Look Across Time', *Applied Psychology: An International Review*. Vol. 38. (2), pp. 131-143.
- Bycio, P. and Zoogah B. (2002), 'Exercise order and assessment centre performance', *Journal of Occupational and Organizational Psychology*. Vol. 75 (1), pp. 109-114.
- Campbell, J.P. Dunnette. Marvin, D. Arvey. Richard, D. Hellervik, Lowell, V. (1973), 'The Development and Evaluation of Behaviourally Based Rating Scales', *Journal of Applied Psychology*. Vol. 57 (1), pp. 15-22.
- Campion, M.A Palmer, D.K and Campion, J.E. (1997), 'A Review of Structure in the Selection Interview', *Personnel Psychology*. Vol. 50. (3), pp. 655-702.
- Cook, M. (1988), *Personnel Selection and Productivity*. John Wiley and Sons Ltd. New York.
- Finkelstein, L.M, Burke, M.J, and Raju N.S. (1995), 'Age Discrimination in Simulated Employment Contexts: An Integrative Analysis', *Journal of Applied Psychology*, Vol. 80. (6), pp. 652-663.
- Gaugler, B. and Rudolph, A. (1992), 'The Influence of Assessee Performance Variation on Assessors' Judgements', *Personnel Psychology*, Vol. 45, pp. 77-98.
- Hartel, C. (1993), 'Rating Format Research Revisited: Format Effectiveness and Acceptability Depend on Rater Characteristics', *Journal of Applied Psychology*. Vol. 78 (2), pp. 212-217.
- Huffcutt, A. Roth, P. and McDaniel, M. (1996), 'A Meta-analytic Investigation of Cognitive Ability in Employment Interview Evaluations: Moderating Characteristics and Implications for Incremental Validity', *Journal of Applied Psychology*. Vol. 81, pp. 459-473.
- Hunter, J. and Hunter, R. (1984), 'Validity and Utility of Alternative Predictors of Job Performance', *Psychological Bulletin*, Vol. 96, pp. 72-98.
- Keenan, T. (1995), 'Graduate recruitment in Britain: A survey of selection methods used by organizations', *Journal of Organizational Behavior*, Vol. 16, pp. 303-317.

- Landy, F. and Bates, F. (1973), 'Another look at Contrast Effects in the Employment Interview', *Journal of Applied Psychology*, Vol.58 (1), pp. 141-144.
- Maurer, T.J. & Alexander, R. A. (1991), 'Contrast effects in behavioral measurement : An investigation of alternative process explanations', *Journal of Applied Psychology*, Vol.76 (1), pp. 3-10.
- Mayfield, E. (1964), 'The Selection Interview: A Re-evaluation of Published Research', *Personnel Psychology*, Vol. 17, pp. 239-260.
- Murphy, K.R. Balzer, W.K. Maura, C. Lockhart, M.C. & Eisenman, E.J. (1985), 'Effects of previous performance on evaluations of present performance', *Journal of Applied Psychology*, Vol. 70, (1), pp. 72-84.
- Prewett-Livinstone, A. Field, S. Veres, J. and Lewis, P. (1996), 'Effects of Race on Interview Ratings in a Situational Panel Interview', *Journal of Applied Psychology*. Vol. 81, pp. 178-186.
- Reilly,R.R and Chao, G.T. (1982), 'Validity and Fairness of Some Alternative Employee Selection Procedures', *Personnel Psychology*, Vol. 35., pp. 1-62.
- RCI (2003) *The Recruitment Confidence Index: Executive report*. Cranfield University. Cranfield UK.
- Saal, F.E. Downey, R.G. & Lahey, M.A. (1980), 'Rating the Ratings : Assessing the Psychometric quality of rating data', *Psychological Bulletin*, Vol. 88, (2), pp. 413-428.
- Schmidt, F. and Hunter, J. (1998), 'The Validity and utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings', *Psychological Bulletin*. Vol. 124 (2), pp. 262-274.
- Tullar, W.L, Mullins, T.W. & Caldwell, S.A. (1979), 'Effects of interview length and applicant quality on interview decision time', *Journal of Applied Psychology*, Vol.64, pp. 669-674.
- Ulrich, L. and Trumbo, D. (1965), 'The Selection Interview Since 1949', *Psychological Bulletin* Vol. 63, pp. 100-116.
- Wagner, R. (1949), 'The Employment Interview: A Critical Review', *Personnel Psychology*. Vol. 2, pp. 17-46.
- Walsh, J.P. Weinberg R.M. and Fairfield M.L. (1987), 'The Effects of Gender on Assessment Centre Evaluations', *Journal of Occupational Psychology*. Vol. 60 (4), pp. 305-309.
- Weisner, W.H and Cronshaw, S.F. (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*, Vol. 61, pp. 275-290.

Wexley, K.N. Yukl, G.A. Kovacs, S.Z. & Sanders, R.E. (1972), 'Importance of contrast effects in employment interviews', *Journal of Applied Psychology*, Vol.56, (1), pp. 45-48.

Woehr, D. and Huffcutt, A. (1994), 'Rater Training for performance appraisal: A quantitative review', *Journal of Occupational and Organisational Psychology*, Vol.67(3), pp. 189-205.

Wright, C.R. (1969), 'Summary of Research on the Selection Interview Since 1964', *Personnel Psychology*. Vol 22, pp. 391-413.