

# Individual Decision Processes in an Assessment Center

Craig J. Russell

Graduate School of Business, University of Pittsburgh

Past research has shown that assessment centers have accurately predicted managerial success in a number of organizations. The argument is made that understanding assessor judgments is necessary before it can be determined why assessment centers consistently demonstrate predictive validity. Data on the individual ratings made by 46 assessors working in an assessment center were analyzed. Ten assessors' ratings (each of whom had rated more the 200 assesseees) were individually subjected to confirmatory factor analyses (using LISREL) and examined within a lens model framework. Support was found for both a formal and informal method of arriving at an overall assessment rating (OAR). Subgroup analyses indicate tentatively that there is little effect of assessee sex on the way assessors arrive at an OAR. Practical and theoretical implications are discussed.

Assessment centers have gained wide acceptance as tools in the selection of managers. Indeed, their use has spread into many non-managerial measurement applications such as performance in college level courses (Alverno College, 1974) and performance of nurses in examinations of competency for the New York Regents External Degree Program (Lenburg, 1983). Behind the general acceptance enjoyed by assessment centers is an extensive research base indicating a positive relation between assessment center ratings and subsequent success as a manager (Thornton & Byham, 1982).

Klimoski and Strickland (1977) reviewed over 90 studies examining the external validity of assessment centers. They found a median predictive validity of .40. However, they also noted a "curious homogeneity" in the criteria employed in these studies (p. 354). They invoked a warning made by Wallace (1974) about the difference between prediction of an individual's job performance and prediction

of what people say about an individual's job performance. Klimoski and Strickland specifically hypothesized that one possible explanation of the validity of assessment centers for predicting subsequent decisions regarding promotion, salary increases, and performance ratings is that assessors are simply policy capturing, that is, accurately mirroring the future decisions of performance appraisors (which constitute the majority of criteria used in assessment center validity studies).

This alternative explanation reflects a disturbing characteristic of assessment center research: we don't know why assessment center ratings are consistently related to subsequent indices of managerial success. To address the hypothesis raised by Klimoski and Strickland, the assessor rating process must be compared to the performance appraisal rating process (cf. Feldman, 1981). Evaluation of assessment center judgments would have other practical and theoretical implications as well. Information concerning an assessee's EEO protected subgroup classification may have an inappropriate impact on assessor decisions. Examination of the impact of assessee sex on assessor judgments would yield policy implications for the training and selection of assessors and provide a control mechanism to ensure compliance with various EEO regulations.

Sackett has performed the only examination of individual assessors' decision processes to date (Sackett, 1977; Sackett & Hakel, 1979). Sackett and Hakel (1979) suggested

---

Portions of this article were presented at the annual meetings of the American Psychological Association, Anaheim, California, 1983. This research was supported in part by a Faculty Research Grant from the Graduate School of Business, University of Pittsburgh. I would like to thank two anonymous reviewers for their helpful comments on an earlier draft of this manuscript. However, responsibility for the contents herein lie with the author.

Requests for reprints should be sent to Craig J. Russell, Mervis Hall, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

that "the formation of an overall assessment rating appears to be an anomaly" (p. 121) within the assessment center. They argued that the extreme cognitive requirements of differentiating and integrating 15 to 25 pieces of information into an overall assessment rating (OAR) stands in stark contrast to the care taken in selection of exercises, definition of dimensions, and assessor training in the observation and recording of behaviors. Their results indicated that assessors do not differentiate and use observation data to a great extent. They found assessors' dimensional ratings to be dominated by a single common underlying factor. Only five dimensions were required to predict most of the variance in the OAR for each of the 19 assessors examined. Furthermore, there was a weak relation between the assessors' subjective opinions as to which dimensions were most important and which dimensions indeed entered the regression equations predicting the OAR. Sackett and Hakel (1979) concluded that it is the "instruction to assessors to utilize all the dimensions in making the overall rating that is the anomaly" (p. 136).

The purpose of this study is to partially replicate and extend the prior work of Sackett and Hakel (1979). Facets of the decision processes used by 46 assessors in an assessment center for the selection of entry-level managers are examined. The prior work of Sackett and Hakel (1979) is extended by examining (a) an explicit organizational heuristic used by assessors to cope with 18 dimensional ratings, (b) how assessors use that organizational heuristic in arriving at an OAR, and (c) whether assessors' use of information in the dimensional ratings is related to the sex of the assessee. The analyses to be reported here examine in detail the decision processes of 10 assessors who assessed at least 200 assessee each. Analyses for another 36 assessors (each of whom individually assessed at least 98 assessee) are summarized for brevity of presentation.

## Method

### *Assesseees and Assessors*

The assessment center was run by a large midwestern organization. Data were obtained on assessment ratings made by 10 assessors who had each assessed at least 200

individuals. Ratings made by an additional 36 assessors who had evaluated at least 98 individuals each were also obtained. Fifty-two percent of the assesseees were female.

Assesseees were either self-nominated or nominated by their supervisors for assessment. All assesseees were current employees of the firm in blue collar or secretarial positions. Assessors were second-level supervisors (one level above the entry-level position being evaluated for) who had gone through a 3-week training program in the observation and rating of behaviors. The assessors worked in teams of four. Three of the team members were on a 3- to 6-month rotation from permanent second-level positions (the firm used a matrix management structure). The fourth assessor was the team leader who was on a longer rotation. An individual usually went through a 3- to 6-month tour as a regular team member before choosing to continue as a team leader.

### *Center Format*

Each assessee was observed in four situational exercises. Sixteen dimensions of managerial skill and ability were evaluated on a 1 to 5-point scale by each assessor. All ratings were made after all exercises had been completed. No dimensional ratings were made within an exercise. If there was a disagreement in the ratings, discussion took place until the difference in ratings was, at most, only 1 point on the 5-point scale. Consequently, for each assessor there were both pre- and postdiscussion ratings on each of the 16 dimensions.

If consensus could not be reached in the discussion, a vote was taken and recorded as consensus rating for that dimension (the most senior assessor received two votes in case of a tie). Consensus dimensional ratings were only obtained to provide (a) a basis for feedback to the assessee, (b) a basis for a needs analysis indicating which training or development steps would be appropriate, and (c) an opportunity for assessor discussion to ensure accuracy in dimensional ratings.

Two additional dimensions of Scholastic Aptitude and Writing Skill were rated by paper-and-pencil instruments, scored by a clerk (on a 1- to 5-point scale) using a standard scoring key, and given to the assessors. Each assessor then individually arrived at an overall assessment rating (OAR) of assessee performance (on a scale of 1 to 4). The assessors were instructed to use their own post-discussion dimensional ratings to arrive at the OAR.

Assessors were also instructed to view the dimensional ratings as forming four categories (personal qualities, interpersonal skills, problem solving skills, and communication skills, or C1 to C4). These categories were to be used as an organizational heuristic to aid the assessor in arriving at the OAR; assessors were trained to use all 18 dimensions, weighing the four categories equally, to arrive at an OAR. The sequence of rating steps can be summarized as follows:

1. *Time 1:* Individual assessors rate 16 dimensions.
2. *Time 2:* If there is lack of consensus, the assessor team discusses each rating. Postdiscussion ratings are recorded for those who change their rating as a result of discussion. The consensus rating is recorded, either as a result of true consensus or a vote.
3. *Time 3:* Individual assessors decide on an OAR using their respective individual postdiscussion ratings on

the 16 dimensions and the two paper-and-pencil scored dimensions. The 18 dimensions are viewed as forming four categories, which are weighted equally by each assessor in arriving at an OAR.

4. *Time 4:* If there is lack of consensus, the assessor team discusses the OAR. Postdiscussion OAR ratings are recorded for those who change their OAR's as a result of discussion. The consensus rating is recorded, either as a result of a true consensus or a vote.

Informal discussion between the researcher and various assessors within the system suggested that an interaction term between the interpersonal skills and problem solving skills categories may also have been used, independent of any formal training, to arrive at the OAR. The rationale verbalized in these discussions was that the weight given to problem solving skills depends on the level of interpersonal skills exhibited by the assessee, because without the interpersonal skills to implement solutions to problems, it does not matter what level of problem solving skills the individual has. Operationally, the assessors described their use of this interaction term as giving more weight to problem solving skills when interpersonal skills are high than when interpersonal skills are low.

### Research Questions

The following specific questions are addressed in the analyses that follow:

1. Do the dimensional ratings reflect the four underlying concepts of personal qualities, interpersonal skills, problem solving skills, and communication skills? If not, what subset or alternative set of factors might be reflected in these ratings?

2. How do the assessors arrive at an OAR? Do they use the four *a priori* categories in the way they were trained to? Do they use an interaction term between two of the *a priori* categories? Does assessee sex enter into the way the OAR is decided on?

### Analyses

Two sets of analyses were performed. For the initial research question, common factor analyses were performed to examine the underlying structure of the dimensional ratings. Based on Nunnally's (1978) recommendation that the ratio of observations to variables be at least 10 to 1, factor analyses of individual dimension ratings were performed for only those 10 assessors who had each assessed more than 180 assessees (all 10 assessed more than 200 each). The sample of dimensional ratings for the other 36 assessors were pooled for a between-assessor factor analysis. All factor analyses were performed on the prediscussion 16 ratings and the key-scored ratings for the two paper-and-pencil instruments because it was felt that the prediscussion rating (as opposed to the postdiscussion ratings) more accurately reflected any underlying structure of skills and abilities to which the individual assessor might be sensitive.<sup>1</sup> Only those factors with eigenvalues greater than 1 were examined. Both orthogonal and oblique rotations were performed in search of simple structure.

Confirmatory factor analyses using LISREL (Jöreskog & Sörbom, 1978) to generate large-sample chi-square

tests of goodness of fit were performed to test the appropriateness of the four-factor *a priori* category structure for (a) each of the 10 individual assessors and (b) the pooled ratings of the remaining 36 assessors. LISREL was also used to test whether the *a priori* category structure or factor structure derived above reproduced the assessors' original correlation matrix more accurately (again using a large sample chi-square test).

To address the second research question, regression analyses were performed in an attempt to model the decision processes of the individual assessors. An assessor's prediscussion OAR was used as a dependent variable in applications of the judgment side of the lens model. The independent variables, or decision cues, used in each of these equations were, respectively:

1. Four "scores" for the *a priori* categories.
2. Four scores for the *a priori* categories and an interaction term between the interpersonal skills and problem solving skills scores.
3. Four scores for the *a priori* categories, the interaction term between interpersonal skills and problem solving skills, a dummy coded assessee sex variable, and interaction terms between the assessee sex variable and all previous variables in the equation.
4. Scores for the factors derived from the common factor analyses. For 10 assessors, the number of factors was determined from their individual factor analyses. For the 36 remaining assessors, the number of factor scores was determined from the factor analysis performed on their pooled ratings.
5. Scores for the derived factor analytic solution and an interaction term between the two derived factors that most resembled the interpersonal skills and problem solving skills categories.

6. Scores for the derived factor analytic solution, the interaction term, a dummy coded assessee sex variable, and interaction terms between the assessee sex variable and all previous variables in the equation. Hence, in all, six equations were generated for each of the 46 assessors for a total of 276 regression equations.

The assessors' training did not include the use of any arithmetic computations to arrive at a category score. Informal discussions with the assessors indicated that they used a subjective impression of central tendency in viewing a category's dimensional ratings. The scores for the *a priori* and derived factors used in the regression analyses were obtained by taking arithmetic averages of the postdiscussion dimensional ratings within each category or factor.<sup>2</sup> It was assumed that an arithmetic average of dimensional ratings within categories approximated

<sup>1</sup> It was felt that, as a consequence of discussion with other assessors, any underlying structure used by an individual assessors to organize their observations would be blurred in the postdiscussion dimensional ratings. Nonetheless, parallel factor analyses were performed on the postdiscussion dimensional ratings. Comparison of factor structures generated from pre- and postdiscussion ratings using LISREL yielded nonsignificant large-sample chi-square statistics, regardless of whether the comparison was made on pre- or postdiscussion covariance matrices.

<sup>2</sup> Postdiscussion ratings were used to coincide with the rating step at Time 3.

the assessors' subjective usage of the "category" organizational heuristic.

One limitation to the findings reported below is that the regression equations are assumed to be representative of the way in which assessors integrate information in the dimensional ratings to arrive at an overall assessment rating. In following the lens model requirement of a representative design the decision cues are necessarily correlated. Hence, the choice of studying "real world" assessment decisions prevents examination of the relative weights of the cues.

## Results

### *Factor Analyses*

All of the common factor analyses judged to yield the cleanest loadings used oblimin rotation with delta set to either 0 or  $-.5$ .<sup>3</sup> A loading heuristic was used to determine which dimension would be identified with which factor. Loading rules ranged from (a) at least .40 on the major factor and .30 on all other factors to (b) at least .45 on the major factor and .25 on all others for the 10 assessors and pooled 36 assessors. In all cases, the dimensional loadings were extremely clean.

A three-factor structure was most parsimonious for 4 of the 10 assessors. Two of the factors are readily interpretable (in comparison to the *a priori* categories) as Interpersonal Skills and Problem Solving Skills. The third factor contained the two paper-and-pencil instruments. One of the 10 assessor's factor analyses generated a two-factor solution (Interpersonal Skills and Problem Solving Skills, with the paper-and-pencil instruments falling in the Problem Solving Skills factor). A four-factor solution was generated for 5 of the 10 assessors. For each of these five assessors, the factor that had been identified as Interpersonal Skills for the previous six assessors was now split between two factors (the range of correlations among these "split" factors was .42 to .51). Results of the common factor analyses performed on the pooled sample of 36 provide support for three factors identified as Interpersonal Skills, Problem Solving Skills, and the two paper-and-pencil instruments. Table 1 contains the frequencies with which each dimension loaded onto the (a) three-factor solution ( $N = 4$  assessors) and (b) four-factor solution ( $N = 5$  assessors).

Correlations between the factors most congruent with the *a priori* categories of Inter-

personal Skills and Problem Solving Skills ranged from .62 to .39. Correlations with and between other factors tended to be lower, though, in five instances the correlation of the factor containing the two paper-and-pencil instruments achieved correlations of .40 to .43 with the Interpersonal Skills and Problem Solving Skills factors. Estimates of internal reliability ranged from .78 to .93 for all subscales.

For 4 of the 10 assessors, the initial factor could be considered to be dominated by problem solving skill dimensions. Five have the first factor dominated by dimensions that could be considered interpersonal skills, whereas only one assessor has the first factor dominated by the two paper-and-pencil instrument dimensions. The first factor is approximately four times greater in terms of variance explained than any of the other factors in each of the 10 assessors' factor structures. This also true for the results of the factor analysis performed on the pooled sample of 36 assessors. The Interpersonal Skills factor dominates the two remaining factors of Problem Solving Skills and paper-and-pencil instruments for the 36 assessors.

Table 2 contains the results of the large-sample chi-square test of goodness of fit for the four *a priori* categories and the factor structures derived from the initial dimensional ratings. All of the chi-square tests reject the null hypothesis that the respective category or factor structures adequately reproduce the covariance matrix of the 18 pre-discussional and paper - and - pencil dimension ratings. However, the sensitivity of this particular test to minor deviations from underlying assumptions is such that the chi-square statistic is easily inflated. Because the *a priori* "category" structure is independent of the derived factor structure, the two chi-square tests of goodness of fit for the respective models are independent. Hence, the difference between the two chi-square statistics for any single assessor is also distributed as chi-square with degrees of freedom equal to the differences in degrees of freedom for the two respective tests (see Jöreskog

<sup>3</sup> Factor loadings and correlations for the 10 individual assessors and the pooled 36 assessors are available from the author on request.

& Sörbom, 1978, for a detailed explanation of this procedure).

These differences in chi-square values fail to support the four-factor a priori category structure. The a priori category structure and derived factor structure are significantly different in their fit for eight of the assessors, indicating that the derived factor structures tend to be more representative of any underlying structure in the data. This evidence suggests that, at least for 8 of the 10 assessors, the four a priori categories are not representative of the underlying concepts that they

are sensitive to in making their dimensional ratings. The derived factor structure for the pooled ratings of the 36 assessors was also found to provide a significantly better fit than the four a priori categories.

Results of the multiple regression analyses are presented in Table 3. The multiple  $R^2$ s for the equations are quite high (.69 to .82), as one would expect in so structured a decision environment. The multiple  $R^2$ s for the equations involving category subscale scores are larger for eight of the assessors than the  $R^2$ s for the equations involving derived factor

Table 1  
*Frequency of Factor Loadings Across Assessors*

Dimensions	Three-factor solution ( $N = 4$ assessors)			Four-factor solution ( $N = 5$ assessors)			
	Inter- personal skills	Problem solving skills	Paper and pencil tests	Inter- personal skills (A)	Inter- personal skills (B)	Problem solving skills	Paper and pencil tests
C1							
D1: Energy	4	—	—	5	—	—	—
D2: Self-objectivity	1	—	—	1	2	—	—
C2							
D3: Impact	4	—	—	1	4	—	—
D4: Leadership	4	—	—	1	4	—	—
D5: Awareness of social environment	3	—	—	1	4	—	—
D6: Behavior flexibility	4	—	—	—	4	—	—
D7: Autonomy	4	—	—	3	1	—	—
C3							
D8: Fact finding—oral	—	1	—	1	—	—	—
D9: Fact-finding—written	—	3	—	—	—	3	—
D10: Organizing	—	4	—	—	—	4	—
D11: Interpreting information	—	4	—	—	—	4	—
D12: Planning	—	4	—	—	—	5	—
D13: Decisiveness	—	3	—	—	—	3	—
D14: Decision making	—	4	—	—	—	5	—
C4							
D15: Oral presentation	1	—	—	1	—	1	—
D16: Oral defense	2	—	—	—	—	1	—
D17: Writing fluency	—	—	4	—	—	—	5
D18: Scholastic aptitude	—	—	4	—	—	—	5

*Note.* Categories 1–4 (C1–C4) are labeled personal qualities, interpersonal skills, problem solving skills, and communication skills, respectively. The derived factor subscales (F1–F3) were labeled Interpersonal Skills, Problem Solving Skills, and Paper-and-Pencil Instruments. Where four factors were derived (F1–F4) there were two factors labeled Interpersonal Skills.

Table 2  
*Confirmatory Factor Analyses: A Priori  
 Categories and Derived Factors*

Assessor no.	Four a priori categories		Derived factor structures	
	$\chi^2$	df	$\chi^2$	df
1	542	133	398	129
2	692	133	554	124
3	511	133	335	120
4	584	133	420	120
5	433	133	307	115
6	547	133	376	117
7	755	133	566	129
8	409	133	305	124
9	423	133	433	133
10	440	133	459	130

subscale scores. This is to be expected because the category subscale scores are calculated using all 18 dimension ratings. The derived factor subscale scores use only the ratings for dimensions found to load on a particular factor. Fifteen to 17 dimensions could be identified as loading onto individual factors in all factor analyses. Furthermore, for five of the assessors, the number of derived factors are less than the number of categories. Slightly larger  $R^2$ s would be expected as a result of more independent variables in equations using category subscale scores. Regardless,  $R^2$  is viewed here as an inappropriate criterion of fit for a cognitive model (Birnbaum, 1973, 1974).

When regressing the OAR onto the a priori categories, the regression coefficient for the interaction term between the Interpersonal Skills and Problem Solving Skills categories was significantly different from zero for 6 of the 10 assessors. For assessors whose derived factor structure included two Interpersonal Skills factors, the problem solving by interpersonal skills interaction term was calculated by multiplying the problem solving skills score by an arithmetic average of the dimensional ratings in both Interpersonal Skills factors. Interaction terms arrived at in this way are labeled "Int" in Table 3. When the OAR was regressed onto the "scores" from the derived factor structures, the interpersonal skills by problem solving skills interaction was significant for 7 of the 10 assessors. The corresponding analyses performed on the ad-

ditional 36 individual assessors indicate that the interpersonal skills by problem solving skills interaction was significant for eight assessors using the category subscale scores and nine assessors using the factor subscale scores.<sup>4</sup> Examination of the eigenvalues from the factor analyses indicates that there is no relationship between the significance of the interaction term and whether the assessors derived factor structure was dominated by an Interpersonal Skills or Problem Solving Skills factor.

The interpersonal skills by problem solving skills interaction was significant for Assessor 1 when both the category and factor subscale scores were used. Assessor 1 was one of approximately a dozen assessors who verbally described using this interaction in discussions with the researcher (other assessors who described this interaction are not included in these analyses because of small sample sizes).

Table 3 also shows the results of adding a dummy coded assessee sex variable and its interaction with all previous independent variables in the equations. Assessee sex interacted with other variables in only one of the 10 assessors' equations using the a priori category subscales. For Assessor 7, the interaction of sex and the problem solving skills category (C3) was significant, indicating that for women, the problem solving skills category received less weight in arriving at the OAR than for male assesseees. Sex of the assessee also entered as a main effect for Assessor 7, indicated that females overall were receiving a lower OAR.

Assessee sex entered the regression equations containing the factor subscale scores as a main effect for Assessors 4, 5, and 8. In each case female assesseees tended to receive a lower OAR. Assessee sex did not enter into any of the equations in interaction with other variables.

### Discussion

The results presented above are interesting for a number of reasons. First, they represent an extension of the prior efforts of Sackett

<sup>4</sup> Regression analyses for the additional 36 assessors are available from the author on request.

and Hakel (1979) in examining the decision process of individual assessors. The results of this examination indicate that (a) assessors' perceptions of assessee behaviors as reflected in their initial dimensional ratings are dominated by a single factor (generally either interpersonal skills or problem solving skills), (b) the a priori category groupings do not represent the concepts underlying the 18 dimensional ratings as well as a derived factor structure does, and (c) for approximately 33% of all the assessors (70% of those subjected to within-subject analyses), the interpersonal skills by problem solving skills interaction was found to explain a significant proportion of variance in the OAR. Examination of the tables indicated that there is no relation between the significance of the interaction term and whether the assessors derived factor structure was dominated by the Interpersonal Skills or Problem Solving Skills factor.

These findings are disturbing in that they indicate assessor judgment processes do not coincide with assessment center architects' assumptions of how (a) the dimensional ratings group together or (b) how assessors use dimensional ratings to arrive at an OAR. Although disturbing, the findings should not be that surprising. The literature on individual differences does not lend itself to a theory of managerial skills and abilities. In the absence of such theory, the (a) grouping of observations under dimensional headings and (b) decision hueristics are at best educated guesses. The best guesses of assessment center architects and job analysts should not be expected necessarily to exhibit rigorous evidence of construct validity. What is surprising is that assessment centers exhibit any predictive validity at all (cf Howard, 1974). Although assessors are apparently not doing what assessment center architects thought they were doing, the reasons behind assessment center predictive validity remain unknown.

Another important contribution of this study is the examination of the impact of non-job-related information on the assessors' decision making process. If it is assumed that the factor analytically derived subscales are more representative of the underlying cognitive organization assessors apply to the dimensional ratings, it can be concluded that

sex of the assessee has almost no impact on the way in which the OAR is arrived at for the 10 assessors. Whereas assessee sex has a significant main effect for Assessors 4, 5, and 8, because this is a field study with no random selection of assessees one cannot be sure that systematic sampling bias did not cause women of lower ability to be nominated for assessment in the current sample and assigned to these assessors. The regression analyses for the remaining 36 assessors indicated that assessee sex interacted with the way factor subscales were related to the OAR for only five additional assessors (again, with female assessee receiving less weight for components of their subscale profiles).

This finding is not surprising in light of control processes operating within this and many other assessment centers. Assessors' OARs are monitored periodically to check for continually high or low OARs as a function of assessee race or sex. However, an assessor could conceivably be arriving at an OAR in different fashion for male and female assessee while still giving comparable OARs. An example of this situation would be where, because of nonrandom selection, women nominated to attend an assessment center have substantially higher skills and abilities than their male counterparts. An assessor who systematically rated female assessee lower would then appear to be treating men and women comparably in terms of the OARs.

These results tend to support those reported by Sackett (1977) and Sackett and Hakel (1979). The factor structures derived for individual assessors were dominated by a single factor. Sackett (1977) reported that the initial factors for individual assessors explained four to eight times the variance of the second factor. Sackett and Hakel (1979) report substantial individual differences in both the number of factors and factor loadings. The results reported above indicate that, although there are differences in the number of factors derived, individual assessors are very similar in the qualitative profiles of loadings that result. The results reported above provide support for an informal integration hueristic used by assessors to arrive at an OAR. This hueristic does not use all of the information or subscales available to the assessor (indicated by the nonsignificant coefficients in Table 3).

Sackett and Hakel reported that only 5 of 17 dimensional ratings were required to predict the large majority of variance in assessors' OARs. Finally, Sackett and Hakel compared (a) the assessors' subjective impressions as to which dimensions were most important to (b) the significance of the dimensions' regression coefficients in predicting the OAR. Whereas they found no relation, the dimensions consistently rated most important were the dimensions that explained the major portion of variance in the OAR. The results presented above indicate that the assessors' subjective verbalizations of how they arrive at an OAR are also supported for a number of individual assessors. See Russell (1983) for

an analysis of how various characteristics of the assessment center may influence the use of subjective decision hueristics.

These findings have at least two practical implications. First, the regression analyses reported provide a more accurate method of monitoring potential bias in ratings. Differential regression weights for subscale scores in predicting the OAR as a function of EEO protected subgroup could be detected quickly and corrective action taken. Second, analyses such as these would allow for the systematic evaluation of the impact of changes made in assessor training and/or the format of assessment center exercises. For example, because of the time and resources required, many

Table 3  
*Regression Analyses*

Subscale score independent variables	Assessor									
	1 (N = 269, NF = 3)		2 (N = 312, NF = 3)		3 (N = 200, NF = 4)		4 (N = 257, NF = 4)		5 (N = 256, NF = 4)	
	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b
1. Categories 1-4 (C1-C4)	.78	—	.80	C1	.80	—	.77	C1	.73	—
2. C1-C4 & C2 × C3 Interaction (Int)	.79	—	.80	C1, Int	.80	Int	.77	C1, Int	.74	—
3. C1-C4, Int, sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.79	C2, Sex, Int, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.80	C1, C4, Int, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.81	C1, Int, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.78	C1, C4, Int, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.75	C1, C4, Int, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int
4. Derived factors (F1-F3)	.75	F3	.77	—	.81	—	.73	—	.71	—
5. Derived factors and interaction (Int)	.76	F3	.77	Int	.81	Int	.74	—	.72	—
6. Derived factors, Int, sex, Sex × Int, and Sex × All Derived Factors	.76	F3, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × Int	.77	F2, Int, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × Int	.82	F2, F4, Int, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × F4, Sex × Int	.74	F1, F2, F4, Int, Sex × F1, Sex × F2, Sex × F3, Sex × F4, Sex × Int	.73	F3, F4, Sex × F1, Sex × F2, Sex × F3, Sex × F4, Sex × Int

Note. Categories 1-4 (C1-C4) are labeled personal qualities, interpersonal skills, problem solving skills, and communication skills, respectively. The derived factor subscales (F1-F3) were labeled Interpersonal Skills, Problem Solving Skills, and Paper-Pencil Instruments. Where four factors were derived (F1-F4) there were two factors labeled Interpersonal Skills. NF = no. of factors; ns/b = nonsignificant regression coefficient at  $\alpha \leq .05$ .



firms consider changes in the structure of the assessment center. Pretesting candidates with components of the assessment center and/or modification of certain exercises to decrease time required are common considerations in the evolution of ongoing assessment centers.

Whereas this study does not tell us why assessment centers have exhibited such consistent evidence of predictive validity, it does provide a more complete description of (a) the organizational hueristics used by assessors to cope with numerous assessment ratings and (b) how assessors use their information contained in the organization hueristics to arrive at an OAR. Future research needs to examine the assessors' and performance ap-

praisors' cognitive processes for purposes of comparison.

Finally, Schmidt, Hunter, and Pearlman (1981) have speculated that situational specificity may occur at the level of regression coefficients. The validity generalization results presented by Schmidt et al. (1981) indicated that test validities are not moderated by small differences in task or behavior requirements and are moderated to a only a small degree by large task differences. Schmidt et al. went on to critique the influence of logical positivism on personnel psychology for causing the field to accept the tenet of situation specificity for so long. They argued that applications of cognitive psychology would better enable an

Assessor									
6 (N = 291, NF = 4)		7 (N = 348, NF = 3)		8 (N = 210, NF = 4)		9 (N = 213, NF = 2)		10 (N = 222, NF = 3)	
R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b	R <sup>2</sup>	ns/b
.75	C1	.73	—	.79	—	.78	—	.70	C1
.75	Int	.74	—	.80	—	.79	—	.71	C1
.75	C1, C4, Int, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.74	C1, C4, Int, Sex × C1, Sex × C2, Sex × C4, Sex × Int	.81	C1, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int	.80	C1, C4, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int,	.71	C1, C4, Sex, Sex × C1, Sex × C2, Sex × C3, Sex × C4, Sex × Int
.74	—	.71	—	.79	—	.77	—	.69	—
.74	Int	.72	—	.80	—	.79	—	.71	—
.74	F1, F2, F4, Int, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × F4, Sex × Int	.72	F3, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × Int	.81	F1, F2 Int, Sex × F1, Sex × F2, Sex × F3, Sex × Int	.79	Sex, Sex × F1, Sex × F2, Sex × Int	.71	F2, Sex, Sex × F1, Sex × F2, Sex × F3, Sex × Int

understanding of the groupings of test validities around task-heterogeneous jobs that are similar in problem solving and information processing requirements (Ghiselli, 1966; Maier & Fuchs, 1969, 1972). The study reported here indicates that one direction for future research may lie in an alternate application of cognitive psychology. Specifically, in an assessment center the information processing and problem solving requirements of a managerial position are defined in the perceptions and judgments of an assessor. Examination of assessor cognitive processes across jobs and situations would provide evidence of any moderator effects. This study indicates that one moderator, assessee sex, has very little impact on assessor cognitive processes.

### References

- Alverno College. (1974). *Competence based learning at Alverno College*. Milwaukee, WI: Alverno College.
- Birnbaum, M. H. (1973). The devil rides again: Correlation as an index of fit. *Psychological Bulletin*, 79, 239-242.
- Birnbaum, M. H. (1974). Reply to the devil's advocates: Don't confound model testing and measurement. *Psychological Bulletin*, 81, 854-859.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive process in performance appraisals. *Journal of Applied Psychology*, 66, 127-148.
- Ghiselli, E. E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- Howard, A. (1974). An assessment of assessment centers. *Academy of Management Journal*, 17, 115-134.
- Jöreskog, K. G., & Sörbom, D. (1978). *LISREL: Analysis of linear structural relationships by the method of maximum likelihood*. Chicago: National Educational Resources.
- Klimoski, R. S., & Strickland, W. (1977). Assessment centers: Valid or merely prescient. *Personnel Psychology*, 30, 353-363.
- Lenburg, C. (1983). *Nursing competence*. Paper presented at the 11th International Congress on the Assessment Center Method, Williamsburg, VA.
- Maier, M. H., & Fuchs, E. F. (1969). *Development of improved aptitude area composites for enlisted classification* (Tech. Research Rep. 1159). Arlington, VA: U.S. Army Behavioral Science Research Laboratory. (NTIS No. AD-701 134).
- Maier, M. H., & Fuchs, E. F. (1972). *Development and evaluation of a new ACB and aptitude area system* (Tech. Research Note 239). Arlington, VA: U.S. Army Behavior and Systems Research Laboratory. (NTIS No. AD-751761).
- Nunnally, J. (1978). *Psychometric theory*. New York: McGraw Hill.
- Russell, C. J. (1983, July). *An examination of assessor's clinical judgments*. Paper presented at the 11th International Congress on the Assessment Center Method, Williamsburg, VA.
- Sackett, P. R. (1977). *Information utilization in an assessment center*. Unpublished master's thesis, The Ohio State University.
- Sackett, P. R., & Hakel, M. D. (1979). Temporal stability and individual differences in using assessment information to form overall ratings. *Organizational Behavior and Human Performance*, 23, 120-137.
- Schmidt, F. L., Hunter, J. E., & Pearlman, K. (1981). Task differences as moderators of aptitude test validity in selection: A red herring. *Journal of Applied Psychology*, 66, 166-185.
- Thornton, G. C., & Byham, W. C. (1982). *Assessment centers and managerial performance*. New York: Academic Press.
- Wallace, S. R. (1974). How high the validity? *Personnel Psychology*, 27, 397-407.

Received September 4, 1984

Revision received February 6, 1985 ■