DOCUMENT RESUME

ED 348 732 EA 024 179

AUTHOR Jaeger, Richard M.
TITLE A Comparison of Compensatory, Conjunctive, and Disjunctive Models for Weighing Attributes of School Quality.
PUB DATE Apr 91
NOTE 27p.; Paper presented at the Annual Meeting of the American Educational Research Association (Chicago, IL, April 3-7, 1991).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Assessment; Educational Quality; Elementary Secondary Education; *Evaluation Methods; Evaluation Problems; Institutional Characteristics; *Measurement Techniques; Models; *Multiple Regression Analysis; *Public Schools

ABSTRACT

Findings of a paper that compared three analytic models for classifying public schools are presented in this paper. The first model is the traditional additive multiple linear regression model that assumes the use of a compensatory policy in assigning weights to school attributes. The second is an adaptation of Coomb's (1964) conjunctive model proposed by Einhorn (1970), and the third is an adaptation of Coomb's disjunctive model, also formulated by Einhorn (1970). Data were collected through a "policy capturing" procedure, in which 28 educational leaders in a southeastern state reviewed a series of hypothetical school profiles of 80 elementary and 80 secondary schools and classified them as nonstandard (deficient), standard (acceptable), or exemplary (normatively outstanding). An index of intrajudge consistency was developed to compare the three models of school classification. Multiple regression analysis was used to examine statistical differences. Findings indicate that the distributions of intrajudge consistency indices were largely overlapping. However, the conjunctive model for aggregating school scores was superior in that it clearly differentiated individual school characteristics and provided a more accurate overall picture of school quality. Five tables and three figures are included. (Contains 62 references.) (LMI)

A Comparison of Compensatory, Conjunctive, and Disjunctive Models for Weighting Attributes of School Quality

Richard M. Jaeger[1]
Center for Educational Research and Evaluation
University of North Carolina at Greensboro

Presented at the 1991 Annual Meeting of the American Educational Research Association, Chicago, April 3-7.

The research described in this paper arose in a decision context that required classification of public elementary and secondary schools as "nonstandard," (meaning deficient), "standard," (meaning acceptable) or "exemplary," (meaning normatively outstanding) on the basis of information concerning the quality of certain constituent attributes of those schools. The decision context was legislatively mandated (Section 20-2-282 of the Official Code of Georgia, Annotated, 1985 — see State of Georgia, 1989).

Several methods for classifying entities based on the judged quality of their attributes were considered and evaluated (Jaeger & Usher, 1990). A procedure termed "policy capturing" (Hobson & Gibson, 1983; Barron & John, 1980; Yates & Jagacinski, 1979; Hoffman, 1960; Slovic & Lichtenstein, 1971) was judged to be the method of choice, since it was found to produce significantly more consistent classification judgments than did a competing method, termed "multiattribute utility technology" (Edwards & Newman, 1982; Einhorn & McCoach, 1977; & Jagacinski, 1979; Gardiner & Edwards, 1975).

This paper reports the results of a comparison of three analytic models for classifying schools using data collected through a policy-capturing procedure. The first was the traditional additive multiple linear regression model that assumes the use of a compensatory policy in assigning weights to the attributes of schools. The second was an adaptation of Coombs's (1964) conjunctive model proposed by Einhorn (1970). The third was an adaptation of Coombs's disjunctive model, also formulated by Einhorn (1970). These models were compared in terms of an index of intra-judge consistency in classifying schools. Formal inferential methods were used to examine the statistical significance of differences between mean index values across pairs of analytic models.

Background

The problem of combining judgments of the quality of attributes (components) for the purpose of forming judgments of the quality of an entity has received substantial attention in social psychology, industrial and organizational psychology, and operations research (Dyer & Sarin, 1979; Edwards & Newman, 1982; Eils & John, 1980; Einhorn & McCoach, 1977; Gardiner & Edwards, 1975; Keeney, 1977; 1982; Pitz, 1980; Wiener, & Libros, 1988). Entities that have been used as objects of judgment include social service programs (e.g., drug rehabilitation programs), persons (e.g., persons applying for admission to an education program or for a job), and institutions (e.g., municipal government agencies). Although multiattribute utility technology has been the most popular judgmental approach to this problem, Jaeger and Usher (1990) found that policy capturing resulted in more consistent classification of schools in the application described here.

Hobson and Gibson (1983) have characterized policy capturing as "a general procedure designed to describe statistically the unique information processing strategies of individual raters" (p. 640). In a policy-capturing procedure, judges are presented with hypothetical profiles that portray the results of evaluating the attributes of each of a set of entities. Each profile represents hypothetical multidimensional evaluation results for a single entity. The judges' task is to provide a summative judgment of the quality of each entity on the basis of its hypothetical profile. Judges work independently, and consider a large number of entities. Judges' summative evaluations are analyzed in conjunction with the profiles of evaluations of component attributes, so as to "capture" the policies they used in weighting the evaluations of each attribute in the profiles. A policy is captured for each judge. These policies are then summarized using any of several statistical procedures.

In describing the application of policy capturing to the appraisal of personnel, Hobson and Gibson listed the following steps (p. 640):

> (a) the presentation to raters of a series of performance profiles consisting of scores on the major dimensions of performance; (b) instructions to raters to review each profile and then assign an overall rating that best represents or summarizes the available information; and (c) the use of multiple regression analysis to calculate the extent to which an individual rater's overall ratings are predictable, given scores on the separate performance dimensions, and to compute the relative importance of each single dimension in determining overall ratings. The statistical equation resulting from the regression analysis defines the "captured policy" for each individual rater. This captured policy is taken to represent an explicit objective description of the way in which the rater combines and weights dimensional information in arriving at overall ratings.

Although other researchers have introduced variations on the method described by Hobson and Gibson (cf., Barron & John, 1980; Yates & Jagacinski, 1979; Hoffman, 1960; Slovic & Lichtenstein, 1971), their description is essentially accurate.

Policy capturing strategies have been used to describe judges' relative weightings of attributes in such areas as clinical judgment (Hoffman, 1960; Hammond, Hursch & Todd, 1964), personnel selection (Anderson, 1977; Naylor & Wherry, 1965; Taylor & Wilsted, 1974; Zedeck & Kafry, 1977), foreign policy beliefs (Summers & Stewart, 1968; Summers, Taliaferro & Fletcher, 1970), factors influencing choice of an organization or position (Feldman & Arnold, 1978; Sheard, 1970; Zedeck, 1977), human learning (Peterson, Hammond & Summers, 1965), factors related to individual motivation (Dachler & Mobley, 1973; Mitchell & Biglan, 1971), and interpersonal conflict (Summers, Stewart & Oncken, 1968). However, applications in education appear to be rare.

Methodology

Sample of Judges

Judgment data were collected from a group of 28 key educational leaders in a southeastern state. Members of the group included the presidents of the state's associations of elementary school principals, middle-school principals, and high school principals, the state's "Teacher of the Year," the president of the state's teachers' association, and the president of the state's association of educational leaders. Other principals and teachers, curriculum specialists, school media specialists, school counselors, and school superintendents also participated as judges.

Presentation of School Profiles

The 28 judges were presented with a series of hypothetical profiles that portrayed each of eight school programs and achievement test averages as "non-standard," "standard," or "exemplary." Fifteen judges reviewed profiles for 80 secondary schools and classified these schools as "non-standard," "standard," or "exemplary." Thirteen judges classified 80 elementary schools in the same manner. The eight attributes used in these profiles were 1) the quality of the school's reading/language arts program, 2) the quality of the school's mathematics program, 3) the quality of the school's guidance and counseling programs, 4) the quality of the school's special education programs, 5) the quality of the school's programs for "at-risk" students, 6) the quality of the school's personnel evaluation and development programs, 7) the average reading achievement test score of all students in the school, and 8) the average mathematics achievement test score of all students in the school.

On the basis of observations made during a pretest of instruments, each judge was instructed to begin with profiles for 10 "Preliminary Schools" prior to responding to profiles for the 80 schools used in data analyses here. When judges' response latencies were observed during the pretest, it appeared that the judges used an initial set of approximately 10 profiles to establish their policies, and then applied these policies to the remaining profiles. Since research interest focused on policies that were considered and thoughtful, it was decided to omit from analyses, data collected during initial periods of transient policy establishment.

Attribute profiles were presented in a graphical format that contained a set of clearly marked, verbally anchored, numerical scales (1=nonstandard; 2=standard; 3=exemplary). This choice of format was based on review of the literature summarized in Table 1. Profiles were as shown in Figure 1.

Table 1. Summary of policy-capturing literature in terms of types of judges used, objects of assessment, types of stimulus profiles used, and resulting coefficients of determination (R^2).

Reference	**Type of Judge**	**Object of Assessment**	**Type of Profile**	**Coefficient of Determination (R^2)**
Zedeck & Kafry (1977)	35 public health nurses	Quality of public health nurses (40 stimuli)	Narrative descriptions	Minimum 0.40 Median 0.61 Maximum 0.71
Summers, Taliaferro, & Fletcher (1970)	131 university under-graduates	175 hypothetical nations -- level of social development	Numerical list	Median 0.56
Stumpf & London (1981)	43 managers	48 hypothetical candidates for promotion to a management position	Narrative description containing numerical ratings	Mean 0.70; zero, after correction for shrinkage
Hobson, Mendel & Gibson (1981)	20 university faculty members	100 hypothetical faculty members rated for performance	Line graph with numerical ratings	Minimum 0.61 Median 0.77; 0.69 when corrected for shrinkage maximum 0.94
Grizzle & Witte (1984)	33 graduate students	80 simulated social service agencies	Visual profiles with numerical ratings	Minimum 0.53 Maximum 0.95
Yates, Jagacinski & Faber (1978)	35 volunteer students	150 hypothetical college courses	Numerical profiles in 5-point Likert scales	Mean 0.81
Anderson (1977)	164 high school teachers	Unspecified	Verbal and numeric (comparison study)	Mean 0.72 (for numeric profiles)

Hypothetical profiles were constructed by first generating eight sets of uniformly-distributed random numbers on the closed interval [0,1]. Since the attributes of schools used in this simulation were mutually correlated, it was necessary to construct a set of profiles that reflected this correlational structure. An appropriate correlational structure was supported only in part by relevant research. Walberg (1985) reported correlations around 0.35 between average measures of student achievement and various measures of school quality. Jaeger (1984) reported a correlation of 0.84 between median mathematics achievement test scores and median reading achievement test scores for sixth-graders. Shanahan and Walberg (1985) found that correlations between ratings of school quality (defined as ratings of school-level instruction, the schools' reputations in their communities, and ratings of teachers' interest in students) and average scores on various standardized achievement tests ranged between 0.16 and 0.20 for high school sophomores. Other studies (Walberg & Tsai, 1985; Stahl & Fairbanks, 1986) produced correlations between individual characteristics of school programs (e.g., "Does the school provide reading enrichment classes?") and averaged student achievement test scores that were barely above zero (in the range 0.02 to 0.06). These latter correlations were ignored because individual questions about school programs were so narrowly defined that they did not adequately characterize the quality of an instructional program. In addition, the reliabilities of responses to individual questions are typically low and, as a consequence, attenuate the values of correlations involving these variables.

To construct a matrix of correlations between each pair of attributes, a "thought experiment" that incorporated the limited findings of the empirical literature was conducted. Based on the partial correlational structure derived from the literature cited above, the remaining correlations were postulated so as to produce a matrix that was positive definite and substantively reasonable.

In the second step of the procedure used to generate hypothetical profiles, a principal components analysis of the postulated matrix of correlations among attributes was completed. The factor structure matrix resulting from this analysis was then postmultiplied by the eight-dimensional matrix of uniform random variates generated in the first step to form an eight-dimensional matrix of random variates with the desired correlational structure (cf., Kaiser & Dickman, 1962).

In the final step of the procedure, the eight-dimensional matrix of random variates resulting from the second step was transformed, so that elements of each vector assumed integer values on the closed interval [1,3].

The selection of 80 school profiles, rather than some larger or smaller number, was purposeful. Determination of the number of profiles that should be presented in a policy-capturing study involves the same considerations as the determination of a desirable sample size in multiple regression analysis. It is well known (Pedhazur, 1982; Nunnally, 1978) that the stability of estimates of a coefficient of determination and the stability of estimates of regression coefficients are inversely proportional to a function of sample size. When sample size is small in comparison to the number of independent variables, over-fitting of the regression model occurs, with consequent instability of estimates and shrinkage of the coefficient of determination in cross-validation studies. Nunnally (1978) recommended that the sample size in multiple regression studies be not less than 10 times the number of independent variables used. Pedhazur (1982, p. 148) suggested a ratio of 30 observations per independent variable and noted that some researchers recommend a minimum sample size of 400.

In policy-capturing studies, the risk of judges becoming fatigued or bored with the judgment task is an important consideration. Although having judges consider 400 profiles is clearly out of the question, use of 25, 36, or even 40 profiles, as was the case in studies by Taylor and Wilsted (1974), Anderson (1977) and Zedeck and Kafry (1977), appears to ignore the risk of obtaining unstable estimates. A sample size of 80 satisfied the criterion proposed by Nunnally and did not unduly tax judges' stamina or interest.

Data Analysis

Multiple linear regression analysis is the analytic procedure of choice throughout the policy-capturing literature. This model assumes that judges regard each rating dimension as contributing linearly and additively to the overall quality of the entities under consideration. The typical justification for the use of multiple linear regression models without interaction terms is the reasonably high coefficients of determination observed in most policy-capturing studies (cf., Zedeck & Kafry, 1977). In addition, investigations of models that incorporate interaction terms have shown extremely small increments in proportions of predicted variation of overall criterion scores, compared to models that are strictly additive (Slovic & Lichenstein, 1971; Slovic, Fischhoff, & Lichtenstein, 1977; Stahl & Zimmerer, 1984). Stahl and Zimmerer justified the elimination of interaction terms from their models by referencing the robustness of additive linear models in the behavioral decision theory literature (Beach, 1967; Darlington, 1968; Dawes, 1979; Dawes & Corrigan, 1974; Hoffman, 1960; Laughlin, 1978). They might have supported their argument further by citing the more than four decades of theoretical work on integration of information in human judgment or decision making (Asch, 1946; Hammond, 1955; Johnson, 1955; Osgood & Tannenbaum, 1955; Peak, 1955; Rimoldi, 1956; Hoffman, 1960; Meehl, 1960; Willis, 1960; Yntema & Torgerson, 1961; Young & Christensen, 1962; Helson, 1964; Spence & Guilford, 1933; Anderson, 1968).

When applied to policy-capturing research, the multiple regression model is inherently compensatory, in the sense that a judge's captured weights are multiplied by "scores" on each profile attribute to compute a predicted evaluation of the overall entity. A low score on some attributes can thus be compensated by a high score on others. As noted by Coombs (1964) and Einhorn (1970), compensatory models fail to capture the judgment policies exhibited by some raters. Although these models are ubiquitous in the policy-capturing literature, they are not universally applicable. Coombs (1964) proposed several alternatives to the additive, linear regression model. His "conjunctive model" assumes that an entity will receive highly positive evaluations only if all of its attributes are positively evaluated. Unlike the compensatory model, the conjunctive model does not permit highly positive evaluations of one or more attributes to compensate for negative evaluations of others. Another model proposed by Coombs (1964) is labeled "disjunctive." The disjunctive model results in an entity being positively evaluated if just one of its valued attributes is positively evaluated. In the case of a school, a disjunctive model might result in a positive evaluation were the school to excel in just one characteristic, such as its reading program. Comparably excellent performance in other programs would not be required. Einhorn* (1970) has developed tractable mathematical models that embody the conjunctive and disjunctive features of models proposed by Coombs (1964). All three models (compensatory, conjunctive, and disjunctive) were applied and compared in this study.

After fitting compensatory, conjunctive, and disjunctive models to the data produced by samples of judges who considered profiles of elementary and secondary schools, distributions of normalized weights for each attribute were computed for both sets of judges. Normalized weights were computed from the standardized regression coefficients and the coefficient of determination that resulted from the responses of each judge to a set of 80 profiles, using a formulation suggested by Hoffman (1960):

$$w_i = \beta_i r_i / R^2 \qquad \text{(Eq. 1)}$$

where w_i is the weight for the ith attribute, β_i is the standardized regression coefficient for the ith attribute, r_i is the correlation between the dependent variable and the ith attribute (in the models used here, the dependent variable was the classification of a school as (1) non-standard, (2) standard, or (3) exemplary), and R^2 is the coefficient of determination for the model. Using data produced by judges who responded to profiles for elementary schools and secondary schools, median values of the distributions of weights for each attribute and median values (as well as other descriptive statistics) of distributions of coefficients of determination, R^2, were then calculated for elementary schools and secondary schools.

Einhorn's (1970) adaptation of Coombs's (1964) conjunctive model was of the form:

$$Y' = \prod_{i=1}^{I} X_i^{b_i} \qquad \text{(Eq. 2)}$$

which, when subjected to a logarithmic transformation, yields a model that is amenable to estimation using the traditional least squares procedure of multiple linear regression:

$$\log Y' = \sum_{i=1}^{I} b_i \log X_i \qquad \text{(Eq. 3)}$$

In this model, Y' is the predicted overall rating of an individual school, which, when rescaled, is interpreted as [0,1.5) = "non-standard," [1.5,2.5) = "standard," and [2.5 and above] = "exemplary;"

This is the formulation used to estimate regression weights under the conjunctive model. The product form of the conjunctive model is such that large values of Y' result only if all of the X_i values are reasonably large. Einhorn noted that this model produces a parabolic response surface with higher values of Y' resulting from values of the component attributes that are large and similar in magnitude.

Einhorn's (1970) adaptation of Coombs's (1964) disjunctive model was of the form:

$$Y' = \prod_{i=1}^{I} [1/(a_i - X_i)]^{b_i} \qquad \text{(Eq. 4)}$$

which, when subjected to a logarithmic transformation, also yields a model that is amenable to estimation using the traditional least squares procedure of multiple linear regression:

$$\log Y' = \sum_{i=1}^{I} -b_i \log(a_i - X_i) \qquad \text{(Eq. 5)}$$

This is the formulation used to estimate regression weights under the disjunctive model. Einhorn noted that this model produces a hyperbolic response surface with relatively large values of Y' when only few component attributes have large values, even if most have relatively small values.

Comparison of Methods

The statistical significance of inter-model differences in mean coefficients of determination was investigated by using a series of paired-difference t-tests. Type I error levels (a) of 0.01 were used in all of these tests to control for experimentwise error rate. The objective was to identify the model that appeared to maximize intra-judge consistency, and to determine whether the largest sample mean value of R^2 exceeded the second-largest value by a statistically significant amount.

Results

The coefficient of determination associated with a judge's captured policy can be interpreted as an indicator of the judge's consistency. The coefficient of determination is also a potent indicator of the "goodness of fit" of a model to a judge's behavior when responding to policy-capturing profiles. One criterion for choosing among alternative models for weighting attributes is therefore the coefficient of determination, R^2; the "best" model is the one with the largest average coefficient of determination.

When analyzed using the compensatory model, the coefficients of determination associated with the policies of the 13 judges who responded to profiles for elementary schools ranged from 0.53 to 0.91 with a mean of 0.73. When a conjunctive model was used, the distribution of R^2 values had a very similar range and a slightly higher mean (0.74). Both the minimum and the mean values of R^2 resulting from use of the disjunctive policy-capturing model were somewhat lower than corresponding values resulting from use of the compensatory or conjunctive models (see Table 2). Statistical comparisons among the mean values of R^2 resulting from use of the three model combinations are summarized in Table 3.

Table 2. Minimum, mean, and maximum coefficients of determination (R^2) resulting from three policy capturing models for weighting eight components of elementary-school quality.

Method	Minimum	Mean	Maximum
Policy Capturing, Compensatory Model (Elementary Schools)	0.53	0.73	0.91
Policy Capturing, Conjunctive Model (Elementary Schools)	0.52	0.74	0.91
Policy Capturing, Disjunctive Model (Elementary Schools)	0.49	0.70	0.90

As noted earlier, paired-difference t-tests of the statistical significance of mean differences between R^2 values were completed for all comparisons of interest: Policy capturing with the conjunctive model vs. policy capturing with the compensatory model; policy-capturing with the compensatory model vs. policy-capturing with the disjunctive model; and policy capturing with the conjunctive model vs. policy-capturing with the disjunctive model. Because coefficients of determination resulting from analyses of judges' captured policies using different models were highly correlated, standard errors of differences between mean values of R^2 tended to be very small. As a result, the differences between mean values of R^2 produced by the various models were often statistically significant when computed for elementary-school profiles, even though sample mean differences were small (see Table 3).

Table 3. Tests of significance of differences between mean values of coefficients of determination, R^2, resulting from three models for weighting attributes of elementary schools.

Methods	Sample Mean Difference	Standard Error of Difference	t-Statistic and Statistical Significance
Policy Capturing, Conjunctive Model Minus Policy Capturing, Compensatory Model	0.013	0.005	t = 2.762 Significant (0.01 < p < 0.05)
Policy Capturing, Compensatory Model Minus Policy Capturing, Disjunctive Model	0.025	0.006	t = 4.292 Significant (p < 0.01)
Policy Capturing, Conjunctive Model Minus Policy Capturing, Disjunctive Model	0.038	0.008	t = 4.725 Significant (p < 0.01)

For elementary-school profiles, application of the conjunctive model resulted in a mean coefficient of determination that was significantly larger ($p < 0.05$) than that resulting from use of the compensatory model. As the results in Table 3 indicate, policy capturing analyzed with the conjunctive model dominated every other procedure investigated in terms of average coefficient of determination.

When analyzed using the compensatory model, the coefficients of determination associated with the policies of the 15 judges who responded to profiles for secondary schools ranged from 0.35 to 0.87 with a mean of 0.69 (see Table 4). The minimum value in this distribution was an outlier; the next lowest value in the distribution was 0.48. When a conjunctive model was used, the distribution of R^2 values had a very similar range and a slightly higher mean (0.71). Both the minimum and the mean values of R^2 resulting from use of the disjunctive policy-capturing model were somewhat lower than corresponding values resulting from use of the compensatory or conjunctive models (see Table 4). Statistical comparisons among the mean values of R^2 resulting from use of the three models are summarized in Table 5.

Table 4. Minimum, mean, and maximum coefficients of determination (R^2) resulting from three policy capturing models for weighting eight components of secondary-school quality.

Method	Minimum	Mean	Maximum
Policy Capturing, Compensatory Model (Secondary Schools)	0.35	0.69	0.87
Policy Capturing, Conjunctive Model (Secondary Schools)	0.38	0.71	0.86
Policy Capturing, Disjunctive Model (Secondary Schools)	0.33	0.67	0.86

Paired-difference t-tests of the statistical significance of mean differences between R^2 values were completed for all comparisons of interest: Policy capturing with the conjunctive model vs. policy capturing with the compensatory model; policy-capturing with the conjunctive model vs. policy-capturing with the disjunctive model; and policy capturing with the compensatory model vs. policy-capturing with the disjunctive model. As noted earlier, because coefficients of determination resulting from analyses of judges' captured policies using different models were highly correlated, standard errors of differences between mean values of R^2 tended to be very small. As a result, the differences between mean values of R^2 produced by the various models were often statistically significant when computed for secondary-school profiles, even though sample mean differences were small (see Table 5).

Table 5. Tests of significance of differences between mean values of coefficients of determination, R^2, resulting from three models for weighting attributes of secondary schools.

Methods	Sample Mean Difference	Standard Error of Difference	t-Statistic and Statistical Significance
Policy Capturing, Conjunctive Model Minus Policy Capturing, Compensatory Model	0.027	0.005	t = 5.229 Significant ($p < 0.01$)
Policy Capturing, Compensatory Model Minus Policy Capturing, Disjunctive Model	0.020	0.005	t = 3.689 Significant ($p < 0.01$)
Policy Capturing, Conjunctive Model Minus Policy Capturing, Disjunctive Model	0.047	0.006	t = 7.281 Significant ($p < 0.01$)

For secondary-school profiles, application of the conjunctive model resulted in a mean coefficient of determination that was significantly larger ($p < 0.01$) than that resulting from use of the compensatory model. As the results in Table 5 indicate, the policy capturing analyzed with the conjunctive model dominated every other procedure investigated in terms of average coefficient of determination. The consistency of this finding for elementary and secondary schools suggests that policy capturing, with data analyzed using the conjunctive model, produces the largest mean value of R^2. However, it should be noted that sample mean differences between the R^2 values resulting from the conjunctive and compensatory models were quite small.

The coefficients of determination resulting from these analyses compare favorably with those found in the seven policy-capturing studies summarized in Table 1. Recall that the range of coefficients reported by Zedeck and Kafry (1977) was 0.40 to 0.71; Hobson, et. al (1981) reported a range of 0.61 to 0.94; and Grizzle and Witte (1984) reported a range of 0.53 to 0.95. The distribution of median coefficients of determination reported in the seven studies ranged from 0.56 to 0.81, and the median of the distribution of medians was 0.72, a value that is just below the smallest of the two medians (0. 77 and 0.74 for elementary school profiles and secondary school profiles, respectively) found in our analyses. Distributions of R^2 values resulting from elementary-school and secondary-school stimuli are summarized in the box-and-whisker charts shown in Figures 2 and 3.

Fig. 2. Coefficients of Determination by Model, Elementary Schools

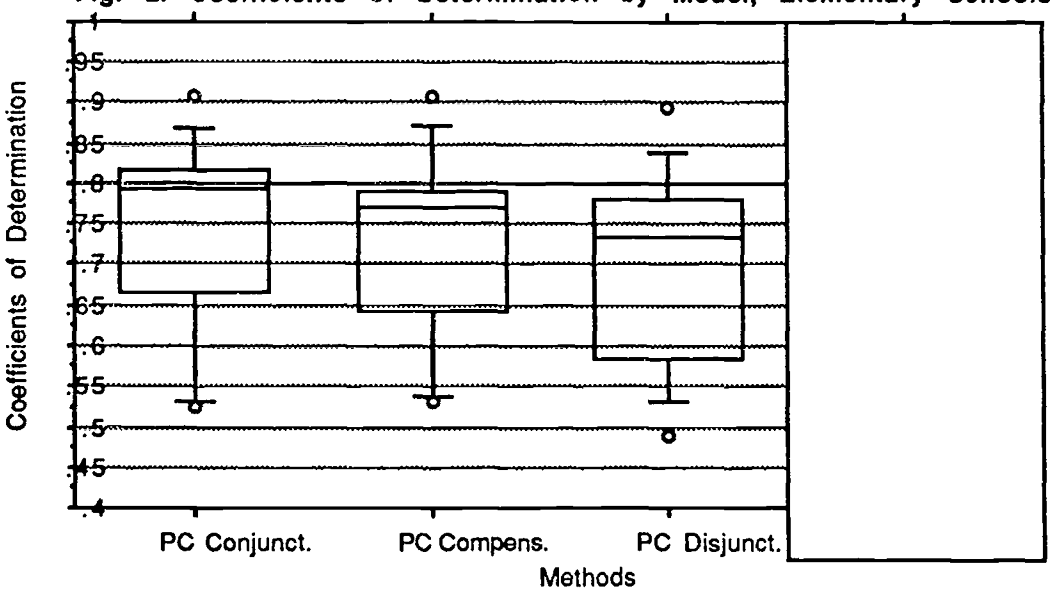
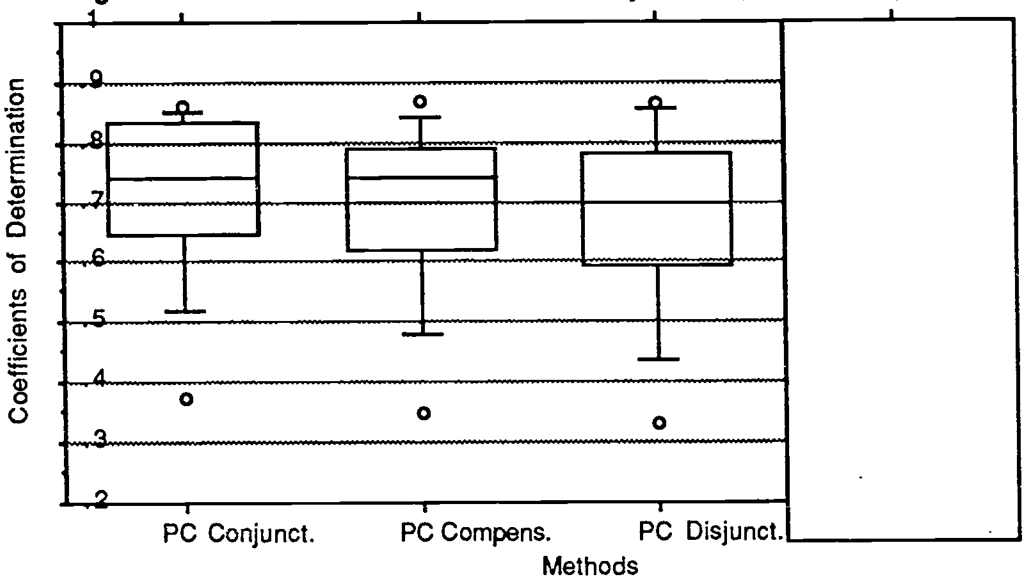


Fig. 3. Coefficients of Determination by Model, Secondary Schools



Conclusions

The three analytic models applied to the policy-capturing data collected in this study produced distributions of intra-judge consistency indices that were largely overlapping (see Figures 2 and 3). However, as shown in Tables 3 and 5, the conjunctive model provided mean coefficients of determination that were, statistically, significantly larger than corresponding means resulting from use of the compensatory or disjunctive models ($p < 0.01$ in three of four comparisons; $0.01 < p < 0.05$ in the fourth comparison).

In the view of many evaluators, a conjunctive model for aggregating attribute scores is philosophically superior to the traditional compensatory model or to a disjunctive model. They would hold that a school's superiority in terms of some valued characteristics should not be allowed to compensate for its deficiencies in terms of others, and that a school's superiority in terms of a single characteristic should never be regarded as sufficient (as implied by the disjunctive model). Equity issues provide an obvious case in point, whether applied to gender groups or racial groups. Results of this study that suggest the analytic superiority of the conjunctive model (however slight its analytic advantage) are therefore comforting.

The generalizability of these results beyond the set of school attributes investigated here cannot be assured. The compensatory model might provide a higher mean index of intra-judge consistency when applied to other sets of school attributes. Even so, the substantial degree of overlap between distributions of coefficients of determination resulting from the conjunctive and compensatory models in this study suggests that the analytic advantage of one model, compared to the other, might be substantively trivial (in terms of average stability coefficient) and that these results might well apply to other sets of school attributes.

[1]It is a pleasure to acknowledge the assistance of Ms. Claire H. Usher in conducting the research underlying this paper.

References

Anderson, B. L. (1977). Differences in teachers' judgment policies for varying numbers of verbal and numerical clues. *Organizational Behavior and Human Performance, 29,* 68-88.

Anderson, N. H. (1968). A simple model for information integration. In R.P.
Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258-290.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41,* 258-290.

Barron, F. H., & John, R. (1980). Reference effects: A sheep in wolf's clothing. *Organizational Behavior and Human Performance, 25,* 365-374.

Beach, L. R. (1967). Multiple regression as a mo del for human information utilization.. *Organizational Behavior and Human Performance, 2,* 276-289.

Coombs, C. (1964). *A theory of data.* New York: John Wiley.

Dachler, H. P., & Mobley, W. H. (1973). Construct validation of an instrumentality-expectancy-task-goal model of work motivation: Some theoretical boundary conditions. *Journal of Applied Psychology, 58,* 397-418.

Darlington, R. B. (1968). Multiple regression in psychological research and practice. *Psychological Bulletin, 69,* 161-182.

Dawes, R. M. (1979). The robust beauty of improper linear models in decision making. *American Psychologist, 34,* 571-582.

Dawes, R. M., & Corrigan, B. (1974). Linear models in decision making. *Psychological Bulletin, 81,* 95-106.

Dyer, J. S., & Sarin, R. K. (1979). Measurable multiattribute value functions. *Operations Research, 27,* 810-822.

Edwards, W., & Newman, J. R. (1982). *Multiattribute evaluation.* Beverly Hills, CA: Sage Publications.

Eils, L. C., & John, R. S. (1980). A criterion validation of multiattribute utility analysis and of group communication strategy. *Organizational Behavior and Human Performance, 25,* 268-288.

Einhorn, H. J., & McCoach, W. (1970) The use of nonlinear, noncompensatory models in decision making. Psychological Bulletin, 73, 221-230.

Feldman, D. C., & Arnold, H. J. (1978). Position choice: Comparing the importance of organizational and job factors. *Journal of Applied Psychology, 63,* 706-710.

Gardiner, P. C., & Edwards, W. (1975). Public values: Multiattribute utility measurement for social decision making. In M. F. Kaplan & S. Schwartz (Eds.). *Human judgment and decision processes.* New York: Academic Press.

Grizzle, G. A., & Witte, A. D. (1984). Evaluating multidimensional performance: A social judgment theory approach. *Evaluation Review, 8,* 777-800.

Hammond, K. R. (1955). Probabilistic functioning and the clinical method. *Psychological Review, 62,* 255-262.

Hammond, K. R., Hursch, C., & Todd, F. (1964). Analyzing the components of clinical inference. *Psychological Review, 71,* 438-456.

Helson, H. (1964). *Adaptation-level theory.* New York: Harper and Row.

Hobson, C. J., Mendel, R. M., & Gibson, F. W. (1981). Clarifying performance appraisal criteria. *Organizational Behavior and Human Performance, 28,* 164-188.

Hobson, C. J., & Gibson, F. W. (1983). Policy capturing as an approach to understanding and improving performance appraisal: A review of the literature. *Academy of Management Review, 8,* 640-649.

Hoffman, P. J. (1960). The paramorphic representation of clinical judgment. *Psychological Bulletin, 47,* 116-131.

Jaeger, R. M. (1984). *Sampling in education and the social sciences.* New York: Longman.

Jaeger, R. M. , & Usher, C. H. (1990). *An evaluation of alternative methods for classification in Georgia schools.* Greensboro: University of North Carolina at Greensboro.

Johnson, D. M. (1955). *The psychology of thought and judgment.* New York: Harper and Row.

Kaiser, H. F., & Dickman, K. (1962). Sample and population score matrices and sample correlation matrices from an arbitrary population correlation matrix. *Psychometrika, 27,* 179-182.

Keeney, R. L. (1982). Decision analysis: An overview. *Operations Research, 30,* 803-838.

Keeney, R. L. (1977). The art of assessing multiattribute utility functions. *Organizational Behavior and Human Performance, 19,* 267-310.

Laughlin, J. E. (1978). Comment on "Estimating coefficients in linear models: It don't make no nevermind." *Psychological Bulletin, 85,* 247-253.

Meehl, P. E. (1960). The cognitive activity of the clinician. *American Psychologist, 15,* 19-27.

Mitchell, T. R., & Biglan, A. H. (1971). Instrumentality theories: Current uses in psychology. *Psychological Bulletin, 76,* 432-454.

Naylor, J. C., & Wherry, R. J. (1965). The use of simulated stimuli and the "JAN" technique to capture and cluster the policies of raters. *Educational and Psychological Measurement, 25,* 969-986.

Nunnally, J. C. (1978). *Psychometric theory* (2nd. ed.). New York: McGraw-Hill.

Osgood, C. E., & Tannenbaum, P. H. (1955). The principle of congruity in the prediction of attitude change. *Psychological Review, 62,* 42-55.

Peak, H. (1955). Attitude and motivation. In M. Jones (Ed.), *Nebraska symposium of motivation, 1955.* Lincoln, Nebraska: University of Nebraska Press, 149-188.

Pedhazur, E. J. (1982). *Multiple regression in behavioral research* (2nd. ed.). New York: Holt, Rinehart and Winston.

Peterson, C. R., Hammond, K. R., & Summers, D. A. (1965). Optimal cue weighting in multiple probability learning. *Journal of Experimental Psychology, 70,* 270-276.

Pitz, G. F. (1980). Assessing the utility of multiattribute utility assessments. *Organizational Behavior and Human Performance, 26,* 65-80.

Rimoldi, H. J. A. (1956). Prediction of scale values for combined stimuli. *British Journal of Statistical Psychology, 9,* 29-40.

Shanahan, T., & Walberg, H. J. (1985). Productive influences on high school student achievement. *Journal of Educational Research, 78,* 357-363.

Sheard, J. L. (1970). Intrasubject prediction of preferences for organizational types. *Journal of Applied Psychology, 54,* 248-252.

Slovic, P., Fischhoff, B., & Lichtenstein, S. (1977). Behavioral decision theory. *Annual Review of Psychology, 28,* 1-39.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. *Organizational Behavior and Human Performance, 6,* 649-744.

Spence, W., & Guilford, J. P. (1933). The affective value of combinations of odors. *American Journal of Psychology, 45,* 495-501.

Stahl, S. A., & Fairbanks, M. M. (1986). The effects of vocabulary instruction: A model-based meta-analysis. *Review of Educational Research, 56,* 72-110.

Stahl, M. J., & Zimmerer, T. W. (1984). Modeling strategic acquisition policies: A simulation of executives' acquisition decisions. *Academy of Management Journal, 27,* 369-383.

State of Georgia (1989). *Georgia School Laws.* Charlottesville, VA: The Michie Company, 1989.

Stumpf, S. A., & London, M. (1981). Capturing rater policies in evaluating candidates for promotion. *Academy of Management Journal, 24,* 752-766.

Summers, D. A., & Stewart, T. R. (1968). Regression models for foreign policy beliefs. *Proceedings of the 76th Annual Convention of the American Psychological Association,* 195-196.

Summers, D. A., Stewart, T. R., & Oncken, G. R. (1968). Interpersonal conflict in heterocultural dyads. *International Journal of Psychology, 3,* 191-196.

Summers, D. A., Taliaferro, J. D., & Fletcher, D. J. (1970). Subjective vs. objective description of judgment policy. *Psychonomic Science, 18,* 249-250.

Taylor, R. L., & Wilsted, W. D. (1974). Capturing judgment policies: A field study of performance appraisal. *Academy of Management Journal, 17,* 440-449.

Walberg, H. J., & Tsai, S. (1985). Correlates of reading achievement and attitude: A National Assessment study. *Journal of Educational Research, 78,* 159-167.

Wiener, R. L., Frank, M. A., & Libros, T. (1988). Termination guidelines for a magnet school review committee: Application of decision theory. *Professional Psychology: Research and Practice, 19,* 316-322.

Willis, R. H. (1960). Stimulus pooling and social perception. *Journal of Abnormal and Social Psychology, 60,* 365-373.

Yates, J. F., & Jagacinski, C. M. (1979). Reference effects in multiattribute evaluations. *Organizational Behavior and Human Performance, 24,* 400-410.

Yates, J. F., Jagacinski, C. M., & Faber, M. D. (1978). Evaluation of partially described multiattribute options. *Organizational Behavior and Human Performance, 21,* 240-251.

Yntema, D. B., & Torgerson, W. S. (1961). Man-computer cooperation in decisions requiring common sense. *IRE Transactions of the Professional Group on Human Factors in Electronics, 2,* 20-26.

Young, P. T., & Christensen, K. R. (1962). Algebraic summation of hedonic processes. *Journal of Comparative and Physiological Psychology, 55,* 332-336.

Zedeck, S. (1977). An information processing model and approach in the study of motivation. *Organizational Behavior and Human Performance, 18,* 47-77.

Zedeck, S., & Kafry, D. (1977). Capturing rater policies for processing evaluation data. *Organizational Behavior and Human Performance, 18,* 269-294.