DOCUMENT RESUME

ED 307 310 TM 013 375

AUTHOR	Mellenbergh, Gideon J.
TITLE	Empirical Specification of Utility Functions.
PUB DATE	Mar 89
NOTE	19p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, March 27-31, 1989).
PUB TYPE	Speeches/Conference Papers (150) -- Reports - Evaluative/Feasibility (142)
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Classification; *Cognitive Processes; *Decision Making; *Education; Placement; *Psychology; Selection
IDENTIFIERS	*Decision Theory; Empirical Methods; Mastery Evaluation; *Utility Functions

ABSTRACT

Decision theory can be applied to four types of decision situations in education and psychology: (1) selection; (2) placement; (3) classification; and (4) mastery. For the application of the theory, a utility function must be specified. Usually the utility function is chosen on a priori grounds. In this paper methods for the empirical assessment of utility functions by decision makers are discussed and methods for the analysis of the data are described. These methods are of two types: methods for the analysis of a subject's utility structure, and methods for investigating the type of utility functions. The procedures are illustrated using an example from selection and one from classification. In the example from selection, data from four psychologists and six judges selecting and evaluating applicants for training were analyzed for 10 subjects. In the example from classification, data from 28 subjects with four recommendations each were used. It is concluded that decision makers can specify their own utility functions and that decision theory can be applied to practical situations in education and psychology. Eight figures and two tables provide the illustrative data. (Author/SLD)

Empirical specification of utility functions

Gideon J. Mellenbergh

Department of Psychology, University
of Amsterdam, The Netherlands

I · · presented at the
Annual Meeting of the
American Eductional
Research Association
San Francisco, March 1989

Gideon J. Mellenbergh
Department of Psychology,
University of Amsterdam,
Weesperplein 8,
1018 XA Amsterdam,
The Netherlands.

ABSTRACT

Decision theory can be applied to four types of decision situations in Education and Psychology: selection, placement, classification, and mastery. For the application of the theory a utility function must be specified. Usually the utility function is chosen on a priori grounds. In this paper methods for the empirical assessment of utility functions by decision makers are discussed. Moreover, methods for the analysis of the data are described. These methods are of two types: First, methods for the analysis of a subject's utility structure and, second, methods for investigating the type of utility functions. The procedures are demonstrated using an example from selection and an example from classification. It is concluded that decision makers can specify their own utility functions and that decision theory can be applied to practical situations in Education and Psychology.

Decision theory is a conceptual framework for making rational decisions. The theory was applied in Statistics (see, for example, Ferguson, 1967) and other sciences (French, 1986). The theory was introduced in psychometrics by Cronbach and Gleser (1957), but was only further developed after the publication of a paper of Hambleton and Novick (1973).

In educational and psychological measurement four types of psychometric decision situations are distinguished (Van der Linden, 1985): (1) selection, (2) placement, (3) classification, and (4) mastery. In each of these situations the following elements are involved: First, one or more tests for gathering information on subjects. Second, several treatments that differ among each other. Third, a decision rule to assign subjects to treatments. Fourth, a criterion to evaluate a subject's performance under a treatment. Finally, for the application of formal decision theory a utility function must be specified.

In selection the subjects are usually assigned to two "treatments": accept or reject for a job or educational program. The decision rule is of the following nature: accept applicants with testscores above one or more cutting scores on the test(s) and reject the others. At the end of a certain period the applicants are evaluated on a criterion, such as supervisors' ratings or grade point average. In placement the subjects are assigned to different treatments, where the performance is evaluated on the same criterion. An example is aptitude treatment interaction: Subjects with a low aptitude are assigned to another curriculum than subjects with a high aptitude; at the end of the curriculum all subjects are evaluated using the same achievement test as a criterrion. In classification not only the treatments are different but also the subject's performance is evaluated on different criterions. An example in The Netherlands is the counseling for secondary education. Using achievement or intelligence tests students at the end of primary school are recommended different types of secondary education, varying from lower vocational to pre-university curriculums. These different types of secondary education are closed using completely different final examinations. The mastery situation has two "treaments": Pass or Fail. It differs from the other three situations in its criterion. In mastery the criterion is internal, i.e. the true mastery of the subject matter, whereas the criterion of the other decision situations is external.

The tools of formal decision theory are applied for making optimal decisions, such as the optimal assignment of subjects to treatments. As remarked before for applying formal decision theory a utility function must be specified. The utility function specifies the utility as a function of the criterion and the decision rule. In, for example, selection the utility is a function of the decision rule to accept or reject the applicant based on the testscore and the criterion score, such as success in the job or grade point average.

The utility function can be specified in three different ways. First, the function is specified on economic grounds. For example, in selection the utility of a rejected applicant has a constant negative value, i.e. the costs of testing an applicant. The utility of an accepted applicant represents the money value of the applicant, which is usually a nondecreasing (constant or increasing) function of the applicant's criterion performance. Second, a mathematically convenient function is assumed. Examples are the threshold (Hambleton & Novick, 1973), linear (Van der Linden & Mellenbergh, 1977), normal ogive (Novick & Lindley, 1978), the truncated normal and the extended beta (Chen & Novick, 1982) utility functions. Examples of the threshold, linear, and normal ogive utility functions are given in Figure 1. The third way of specifying utility functions is

Insert Figure 1 about here

empirical: The assessment of utility functions by decisionmakers themselves. The feasibility of this approach for mastery decisions in education was studied by Vrijhof, Mellenbergh, and Van den Brink (1983) and Van der Gaag, Mellenbergh, and Van den

Brink (1988). Using this empirical approach it is possible to study whether the decisionmakers' utility functions resemble specific mathematical forms such as the threshold or linear function. Vrijhof et al. (1983) and Van der Gaag et al. (1988) studied empirical utility functions for Pass/Fail decisions in educational testing. They found that many of the empirical functions were adequately approximated by a linear function.

When the utility function is specified decision theory can be applied for making optimal decisions. The emphasis in the literature was on the optimal assignment of students and applicants to treatments. Examples are: Optimal cutting scores for Pass/Fail decisions (Hambleton & Novick, 1973; Huyhn, 1976; Mellenbergh, Koppelaar & Van der Linden, 1977; Van der Linden & Mellenbergh, 1977; Veldhuyzen, 1982) and classification decisions (Van der Linden, 1987); optimal cutting scores for culture-fair selection (Gross & Su, 1975; Mellenbergh & Van der Linden, 1981; Petersen, 1976); optimal assignment of students to different treatments in Aptitude Treatment Interaction Research (Van der Linden, 1981). Moreover, the construction of coefficients for tests (Van der Linden & Mellenbergh, 1978), a conceptual framework for the description and evaluation of the qualities of tests (Mellenbergh & Van der Linden, 1979), test length (Van der Linden, 1982), and item selection (Mellenbergh & Van der Linden, 1982) were studied from a decisiontheoretic point of view.

In this paper recent results on the empirical assessment of utility functions are reported. First, the methods for assessing and analyzing these functions are described. Subsequently, two examples are given. One on the assessment of utility in a selection situation and one in a classification situation.

METHODS

Several methods were used for the collection and analysis of data; they are described in this section.

Empirical Assessment of Utility Functions

The utility is a function of the decision rule based on the testscore and the criterion performance. For a given decision the utility is a function of only the criterion performance. For example, in selection the utility is a function of the decision rule to accept or reject the applicants and the applicant's criterion performance; for accepted subjects the utility is a function of criterion performance and for rejected applicants the utility is another function of criterion performance. When these functions are scaled separately for the different decisions a problem arises: the functions might be assessed in different scales. Therefore, also methods are needed to represent the different utility functions on identical scales.

Three methods were used for the empirical assessment of utility functions: Comrey's constant sum method (Torgerson, 1958, p. 105), Bechtel's (1976) method, and a Direct Method (Lamers, Van der Gaag & Mellenbergh, 1988).

For the application of Comrey's and Bechtel's methods a set of criterion performances must be prepared. For example, Lamers et al. (1988) studied a selection situation where the criterion performance was judged on thirteen aspects using five-point scales. The criterion score was the sum of the thirteen aspect scores, which ranged from 13 through 65. They used the following seven, equally spaced, criterion performance points: 19.5, 26, 32.5, 39, 45.5. 52, and 58.5. All possible pairs of the selected criterion performance points were formed. The pairs were collected in a booklet; the order of the pairs was determined using Ross' (1934) table. Using Bechtel's method the pairs were administered to the subjects for the acceptance situation.

Using Comrey's method the subject is asked to indicate for each member of a pair whether the utility is positive or negative. Then, the subject must split 100 points according to the ratio of the utilities of the two members of the pairs. An example of the instruction is given in Figure 2.

Insert Figure 2 through 5 about here

The subject has indicated that the utility for both 39 and 52 is positive and (s)he has split the 100 points into 20 and 80, which means that the ratio of the two utilities is 80/20 = 4. Using Bechtel's method a rating scale is used. An example is given in Figure 3. The subject has indicated to have a moderate preference for the criterion performance of 52.

Using the Direct Method it is not needed to collect paired comparison judgments of the selected criterion performance points. Each of the points is rated; an example is given in Figure 4. The advantage is of course that much less judgements are required. The number of ratings of seven criterion peformance points is seven, whereas the number of paired comparison is 21. The disadvantage of the Direct Method is that internal checks on the judgments are not possible, which will be discussed in the next section.

As remarked before the utility functions of different decisions must be represented on the same scale. For the selection situation this problem was attacked by assessing the point of intersection of the two utility functions. The subjects were instructed to indicate the point of intersection of the utility functions of accepted and rejected applicants. The instruction is reported in Figure 5.

Finally, it is remarked that in our most recent experiments a personal computer was used for collecting the data. One of the above mentioned methods was used for assessing the utility functions and the functions were graphically represented on the screen. The subjects were offered the opportunity to correct their functions, which means that both uncorrected and corrected functions can be analyzed.

Data Analysis

The analysis of the data has two different aspects: First, a model for the preference judgments is specified and the utility functions are estimated under this model. Second, the form of the utility functions is investigated; for example, it is studied whether the functions are of the threshold, linear, or normal ogive type.

Bechtel (1976, chap. 2) proposed a model for the judgments of his method; see Figure 3. A linear model for the ith subject's (i = 1, 2, ..., N) preference strength for the jth criterion performance compared to the kth criterion performance (j, k = 1, 2, ..., n; j ≠ k) is:

$$d_{ijk} = u_{ij} - u_{ik} + y_{jk} + e_{ijk} \quad , \tag{1}$$

with constraints

$$\sum_{j=1}^{n} u_{ij} = 0 \quad (i = 1, 2, \ldots, N) \quad , \tag{2}$$

$$\sum_{k=1}^{n} y_{jk} = 0 \quad (j = 1, 2, \ldots, n) \quad , \tag{3}$$

$$y_{kj} = -y_{jk} \quad (j, k = 1, 2, \ldots, n; j \neq k) \quad , \tag{4}$$

$$y_{jj} = 0 \quad (j = 1, 2, ..., n) \quad . \tag{5}$$

In this model u_{ij} and u_{ik} are subject ith utilities for the jth, respectively, kth criterion performance, and e_{ijk} is the residual term. The interaction term y_{jk} is called the unscalability; it indicates that the preference strength is not additive in the utilities of the two criterion performances. If the interaction term is negligible a more simple model can be used:

$$d_{ijk} = u_{ij} - u_{ik} + e_{ijk} \tag{6}$$

with constraint Formula 2. The parameters in the models are estimated using the least squares method.

The models of Formula 1 and 6 can be compared in two ways. First, the null hypothesis that the unscalability parameter y is zero can be tested using the F-statistic. Using this test it is assumed that the residuals e are independently normally distributed with common variance (Bechtel, 1976, sec. 2.5.2 and 2.6.2). This assumption might be violated for paired comparison data. Therefore, product moment correlations can be used as descriptive measures of fit. Using the least squares estimates the reproduced values under model Formula 6 are:

$$d^*_{ijk} = \hat{u}_{ij} - \hat{u}_{jk} \quad . \tag{7}$$

The correlations R_s between observed d_{ijk} and reproduced d^*_{ijk}, computed over all N subjects and all distinct pairs of criterion performances, is a measure of fit for model Formula 6. Although the model may be an adequate overall description for a group of subjects, it is possible that the model does not fit for particular subjects. The correlation R_{si}, computed between d_{ijk} and d^*_{ijk} per subject, describes the fit of model Formula 6 for each of the subjects. The correlation R_u between the observed d_{ijk} and the under model Formula 1 reproduced values over all subjects and all distinct pairs of criterion scores is a descriptive measure of fit for the model Formula 1. The measure R_u is computed using an ANOVA-like table (Bechtel, 1976, pp. 26 and 27). R_u is equal to or larger than R_s because Formula 1 contains more parameters than Formula 6. The more R_u differes from R_s the larger the influence of the unscalability parameter y.

Using Comrey's method (Figure 2) the ratios of the absolute values of the utilities are scaled instead of the difference of the utilities. To apply models Formula 1 and 6 the logarithmic transformation of the utility ratios is used (Vrijfhof et al., 1983). Using the Direct Method (Figure 4) the models cannot be applied because the data are not of the paired comparison type. The disadvantage of this quick and easy method is that the utility structure cannot be investigated and that the structure of model Formula 6 must be assumed.

For Comrey's and Bechtel's methods the utility structure can be investigated. When model Formula 6 is a reasonable description of the data it makes sense to investigate the form of the utility functions. Otherwise, the interaction term y (Formula 1) is needed which means that the utility judgment also depends on the particular pair of criterion performances that is considered. In that case it does not make much sense to investigate the form of the utility functions.

Suppose model Formula 6 is a reasonable description of the data. The utility functions are estimated under this model and the form of the functions can be studied. First, the functions are graphically represented and the graphs are inspected. Second, the fit of some of the apriori functions is investigated. The fit of the threshold function is assessed using the point biserial correlation. For example, Lamers et al. (1988) used in their study seven criterion performances: 19.5, 26, ..., 58.5. The fit of the threshold function was investigated by computing six point biserial correlations: between the utilities and the criterion performance 19.5 scored as 0 and the other six criterion performances scored as 1; between the utilities and the criterion performances 19.5 and 26 scored as 0 and the other five criterion performances scores as 1; and so on. The highest of these squared point biserial correlations was used as a measure of fit of a threshold function. The fit of a linear function was assessed using the squared product moment correlation between the utilities and the criterion performances. The fit of the normal ogive function is investigated by rescaling the utilities into the open interval between zero and one and applying the logit transformation to the rescaled utilities. The squared product moment correlation between these logit transformed utilities and the criterion performance is a measure of fit of the normal ogive function (Van der Gaag, 1989).

EXAMPLES

In this section some results of two studies are briefly described.

Selection

Lamers et al. (1988) studied the selection situation; some information on this experiment was already given in previous sections.

The situation is a multi-stage selection procedure. In the first stage psychological tests are used in selecting applicants for a training. At the end of the practical part of the training applicants are evaluated on thirteen aspects that are scored on a five-point scale. The criterion score is the sum of the thirteen aspect scores (minimum: 13; maximum: 65).

Seven levels of the criterion performance were used, equally spaced over the score range: 19.5 (1.5 on average per aspect), 26 (2.0 on average), ..., 58.5 (4.5 on average). The 21 distinct pairs of these performance levels were formed. Bechtel's method (Figure 3) was used for assessing the utility function for accepted applicants, the Direct Method (Figure 4) was used for the rejected applicants, and the intersection was directly assessed (Figure 5).

The subjects were persons actually involved in the selection process: psychologists doing the selection of applicants for the training and judges doing the evaluation of the trainees' criterion performance of the practical part of the training. Data of four psychologists and six judges were analyzed.

For Acceptance the model of Formula 1 was fitted separately to the group of judges and the group of psychologists. The null hypothesis that the unscalability parameter is zero was tested using the F-statistic. The null hypothesis was rejected at the 1% significance level in each of the two groups. As remarked before the assumptions of the statistical test might be violated. Therefore, also the descriptive measures of fit were computed (Table 1). Comparing the fit of models Formula 1 (R_u) and Formula 6 (R_s)

Insert Table 1 about here

---

shows that the influence of the unscalability parameter is substantial in each of the two groups. Although the fit of model Formula 1 is better than the fit of model Formula 6 Formula 6 is a reasonable approximation for the utility structure of the psychologists. The correlation for the group of psychologists is .907 and the lowest of the individual correlations is .830. For the judges model Formula 6 is a good approximation for no. 3 and 4, might be sufficient for no. 1 and 6, but is insufficient for no. 2 and 5.

For Rejection the Direct Method was used. Therefore, it is impossible to study the stucture of an utility model.

Using the least squares method the utilities were estimated and the utility functions were plotted. The fit of a linear utility function was assessed by using the squared product moment correlation and the fit of a threshold utility function was assessed by using the highest of the six squared point biserial correlations. The squared correlations are reported in Table 2. For Acceptance in nine of the ten cases the linear

---

Insert Table 2 about here

---

function has a better fit than the threshold function. For all of the ten subjects the linear function is a good description of the utility function: the lowest squared product moment correlation is .790. For Rejection in six of the ten cases the linear function has a better fit than the threshold function. For all of the psychologists and four of the judges (no. 1, 3, 4, and 5) the linear function is an adequate approximation of their utility function, but for the judges no. 2 and 6 the linear function is not adequate. Inspection of the linear approximations of the utility function showed that all functions for Acceptance tend to increase and all functions for Rejection tend to decrease with the criterion performance. Examples of utility functions are given in Figures 6 (Acceptance) and 7 (Rejection). For the psychologist the linear approximation is adequate, whereas the approximation for the judge is bad.

Insert Figures 6 and 7 about here

The point of intersection of the two utility functions was directly assessed. The criterion performances where the two utility functions intersect vary from 30 to 40.

Classification

The classification situation was studied by Van der Heiden (1988). He investigated the recommendations based on achievement or intelligence testscores at the end of primary school. The students are recommended different types of secondary education. Van der Heiden used the most common types of secondary education: Lower Vocational Training (LBO), Lower Administrative Training (MAVO), Higher Administrative Training (HAVO), and Pre-University Training (VWO).

The criterion performance was the Achieved Level after Six Years of Secondary Education. The following seven levels were used: LBO graduated within five years (LBO5), LBO graduated within four years (LBO4), MAVO graduated within five years (MAVO5), MAVO graduated within four years (MAVO4), HAVO graduated within six years (HAVO6), HAVO graduated within five years (HAVO5), and VWO graduated within six years (VWO). The curriculum of LBO and MAVO is four years, of HAVO five years, and of VWO six years. LBO5, MAVO5, and HAVO6 imply that a class is repeated. The points of these levels on a scale Achievement Level after Six Years of Secondary Education were known from another study (Cremers, 1980).

The 21 distinct pairs of the seven criterion performances were formed and Bechtel's method was applied. For each of the four recommendations (LBO, MAVO, HAVO, and VWO) the pairs were administered to three groups of subjects: nine primary school teachers, nine secondary school teachers, and ten parents of students at the end of primary school. The data were collected using a personal computer. The subjects were shown the graphs of their utility functions and they got the opportunity to correct their functions.

In total 4 (recommendations) x 28 (subjects) = 112 utility functions were assessed. For each of the functions seven utility values were estimated. The mean number of corrections per function was 2.77, which means that about 40 percent of the utility values were corrected. Moreover, per function the correlation coefficient between uncorrected and corrected utility values was computed. 99 out of the 112 correlations were above .80.

For each of the three groups of subjects and each of the four recommendations the null hypothesis that the unscalability parameter is zero was tested. In each of the twelve cases the null hypothesis was rejected at the 5 percent significance level. But, for most of the 112 utility functions model Formula 6 is a reasonable description of the utility structure: 75 of these correlation coefficients were above .80.

The corrected and uncorrected utility functions of the first subject are shown in Figure 8. For the highest recommendation (VWO) the utility is an increasing function of

---

Insert Figure 8 about here

---

the criterion Achievement Level after Six Years of Secondary Education (ALSYSE). For the other three recommendations the function increases until a maximum value and then decreases. The maximum corresponds with the recommendation: For the LBO recommendation the utility is maximal for LBO graduated within four years; for the MAVO recommendation the utility is maximal for MAVO graduated within four years; for the HAVO recommendation the utility is maximal for HAVO graduated within five years.

CONCLUSIONS

The results of the statistical tests show that the additive structure of the utilities (Formula 6) does not fit the paired comparison data of our experiments. But, the goodness of fit statistics usually indicate that the additive stucture is a reasonable approximation of the utility structure.

In our studies (Lamers et al., 1988; Van der Gaag, 1988; Vrijhof et al., 1983) it was found that the linear function is mostly an adequate description of the type of the utility function; there are, however, subjects that specify different types of functions.

Finally, it is remarked that insight in the structure of the utilities and the type of utility function might be of practical importance. When the utility function is of a given type, e.g. threshold, linear, or normal ogive, then decision theory can be applied to educational and psychological testing.

REFERENCES

Bechtel, G.G. (1976). Multidimensional preference scaling. 's-Gravenhage, The Netherlands: Mouton.

Chen, J.J., & Novick, M.R. (1982). On the use of a cumulative distribution as a utility function in educational or employment selection. Journal of Educational Statistics, 7, 19-35.

Cremers, P.J.G. (1980). Konstruktie van een schaal voor Bereikt Niveau van Voortgezet Onderwijs (B.N.V.O. schaal). Tijdschrift voor Onderwijsresearch, 5, 80-91.

Cronbach, L.J., & Gleser, G.C. (1957). Psychological tests and personnel decisions. Urbana: University of Illinois Press.

Ferguson, T.S. (1967). Mathematical Statistics: A decision theoretic approach. New York: Academic Press.

French, S. (1986). Decision theory: An introduction to the mathematics of rationality. Chichester: Ellis Horwood limited.

Gross, A.L., & Su, W.H. (1975). Defining a "fair" or "unbiased" selection model: A question of utilities. Journal of Applied Psychology, 60, 345-351.

Hambleton, R.K., & Novick, M.R. (1973). Toward an integration of theory and method for criterion referenced tests. Journal of Educational Measurement, 10, 159-170.

Huyhn, H. (1976). Statistical considerations of mastery scores. Psychometrika, 41, 65-79.

Lamers, L.M., van der Gaag, N.L., & Mellenbergh, G.J. (1988, submitted for publication). Empirical utility functions for selection.

Mellenbergh, G.J., Koppelaar, H., & van der Linden, W.J. (1977). Dichotomous decisions based on dichotomously scored items: A case study. Statistica Neerlandica, 31, 161-169.

Mellenbergh, G.J . & van der Linden, W.J. (1979). The internal and external optimality of decisions based on tests. Applied Psychological Measurement, 3, 257-273.

Mellenbergh, G.J., & van der Linden, W.J. (1981). The linear utility model for optimal selection. Psychometrika, 46, 283-293.

Mellenbergh, G.J., & van der Linden, W.J. (1982). Selecting items for criterion-referenced tests. Evaluation in Education: An International Review Series, 5, 177-190.

Novick, M.R., & Lindley, D.V. (1978). The use of more realistic utility functions in educational applications. Journal of Educational Measurement, 15, 181-191.

Petersen. N.S. (1976). An expected utility model for "optimal" selection. Journal of Educational Statistics, 1, 333-358.

Ross, R.T. (1934). Optimum orders for the presentation of pairs in the method of paired comparisons. Journal of Educational Psychology, 25, 375-382.

Torgerson, W.S. (1958). Theory and methods of scaling. New York: Wiley.

Van der Gaag, N.L. (1989). Mastery decisions at the end of Dutch secondary education: linear and normal ogive empirical utilities. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco.

Van der Gaag, N.L., Mellenbergh, G.J., & van den Brink, W.P. (1988). Empirical utility functions for Pass/Fail situations. Methodika, (accepted).

Van der Heiden, E. (1988). Empirische utiliteitsfuncties. Master Thesis. Amsterdam, The Netherlands: Department of Psychology, University of Amsterdam.

Van der Linden, W.J. (1980) Decision models for use with criterion-referenced tests. Applied Psychological Measurement, 4, 469-492.

Van der Linden, W.J. (1981). Using aptitude measurements for the optimal assignment of subjects to treatments with and without mastery scores. Psychometrika, 46, 257-274.

Van der Linden, W.J. (1982). Passing score and length of a mastery test. Evaluation in Education: An International Review Series, 5, 149-164.

Van der Linden, W.J. (1985). Decision theory in educational research and testing. In T. Husen & T.N. Postlewaite (Eds.), International encyclopedia of education: Research and studies (pp. 1328-1333). Oxford: Pergamon Press.

Van der Linden, W.J. (1987). The use of test scores for classification decisions with threshold utility. Journal of Educational Statistics, 12, 62-75.

Van der Linden, W.J., & Mellenbergh, G.J. (1977). Optimal cutting scores using a linear loss function. Applied Psychological Measurement, 1, 593-599.

Van der Linden, W.J., & Mellenbergh, G.J. (1978). Coefficients for tests from a decision theoretic point of view. Applied Psychological Measurement, 2, 119-134.

Veldhuyzen, N.H. (1982). Setting cutting scores: A minimum information approach. Evaluation in Education: An International Review Series, 5 141-148.

Vrijhof, B.J., Mellenbergh, G.J., & van den Brink, W.P. (1983). Assessing and studying utility functions in psychometric decision theory. Applied Psychological Measurement 7, 341-357.

Table 1

Measures of Fit (Correlation) Utility Models for Acceptance in Selection

Subjects		Correlation		
		R_u	R_s	R_{si}
Psychologists		.950	.907	
	no. 1			.968
	no. 2			.830
	no. 3			.909
	no. 4			.926
Judges		.841	.756	
	no. 1			.770
	no. 2			.451
	no. 3			.875
	no. 4			.934
	no. 5			.620
	no. 6			.748

Table 2

Squared Product Moment Correlations Linear Utility Function

and Optimal Squared Point Biserial Correlations Threshold Function

Subjects		Situation			
		Acceptance		Rejection	
		Linear	Threshold	Linear	Threshold
Psychologist	no. 1	.984	.741 (3)	.905	.823 (2)
	no. 2	.983	.775 (3)	.988	.828 (3)
	no. 3	.910	.920 (3)	.815	.952 (3)
	no. 4	.966	.859 (3)	.897	.865 (2)
Judge	no. 1	.869	.832 (2)	.892	.845 (4)
	no. 2	.790	.770 (3)	.526	.768 (2)
	no. 3	.985	.794 (3)	.973	.816 (4)
	no. 4	.968	.845 (4)	.785	.787 (1)
	no. 5	.921	.792 (2)	.974	.832 (2)
	no. 6	.888	.738 (3)	.625	.765 (4)

Note. Between parentheses the optimal threshold, e.g. the number 2 indicates that the criterion scores 19.5 and 26 are scored 0 and the other criterion scores are scored 1.

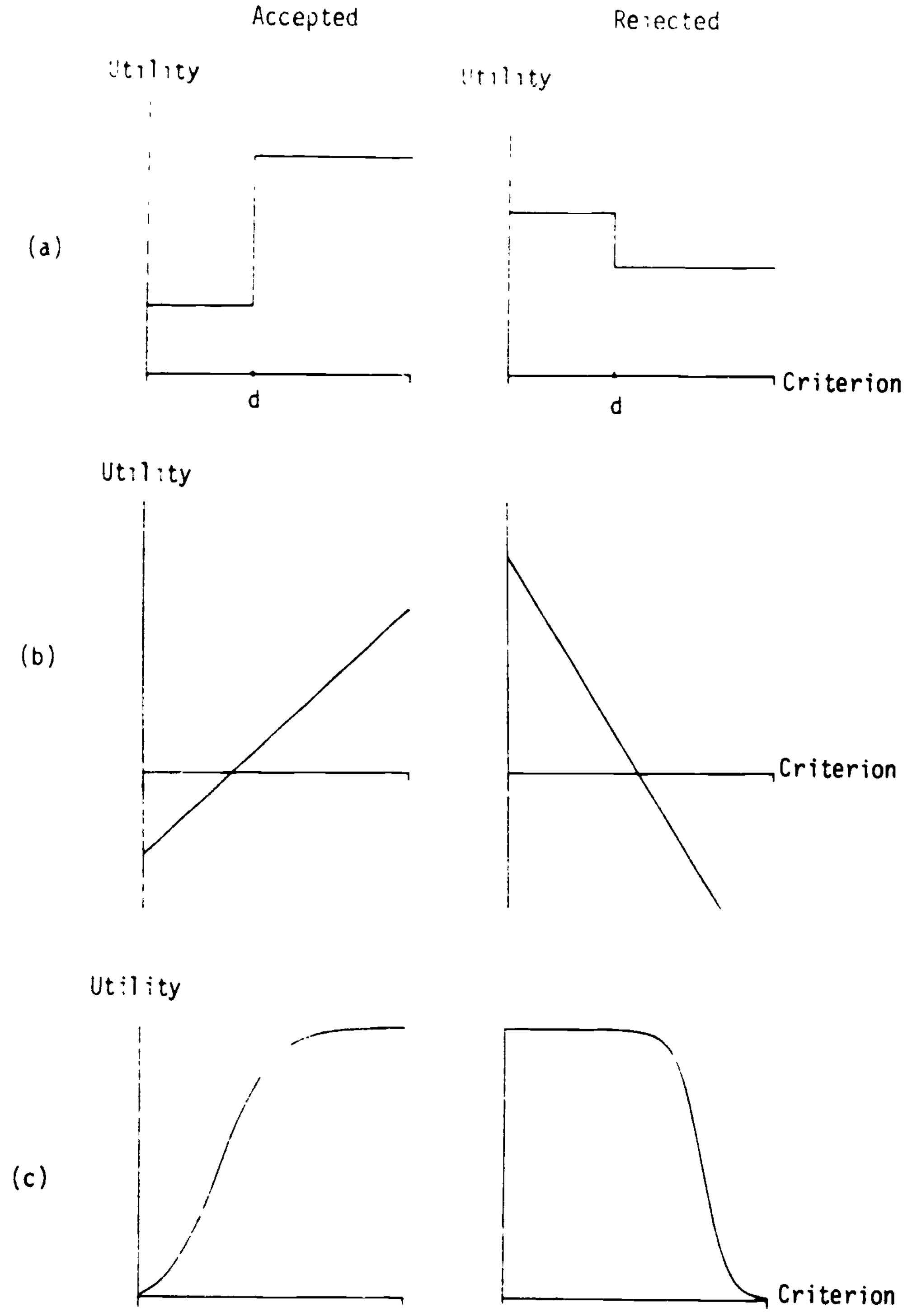


Figure 1. Example of a (a) threshold, (b) linear, and (c) normal ogive utility function, selection situation.

Figure 2

Example of Comrey's Method, Selection, Acceptance

The applicant has been accepted: The decision has for me at the criterion performance level

		39	52	
	an	+	+	utility
The utility ratio is		20	80	

Figure 3

Example of Bechtel's Method, Selection, Acceptance

Indicate the degree in which you prefer one option above the other

An applicant is accepted and has a criterion score

39 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 52

Figure 4

Example of the Direct Method, Selection, Rejection

Indicate the degree in which you prefer one option above the other

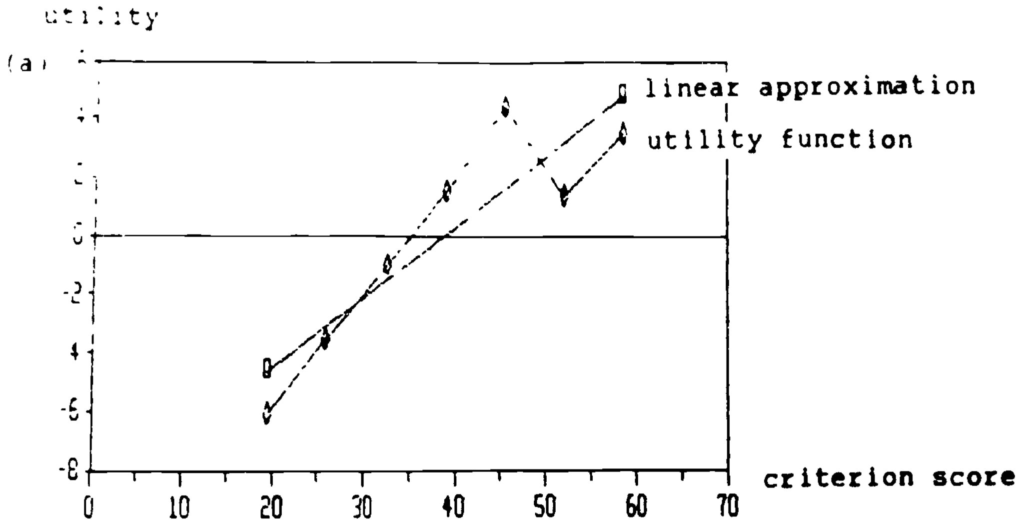
I evaluate the decision to reject an applicant with a criterion score of 39

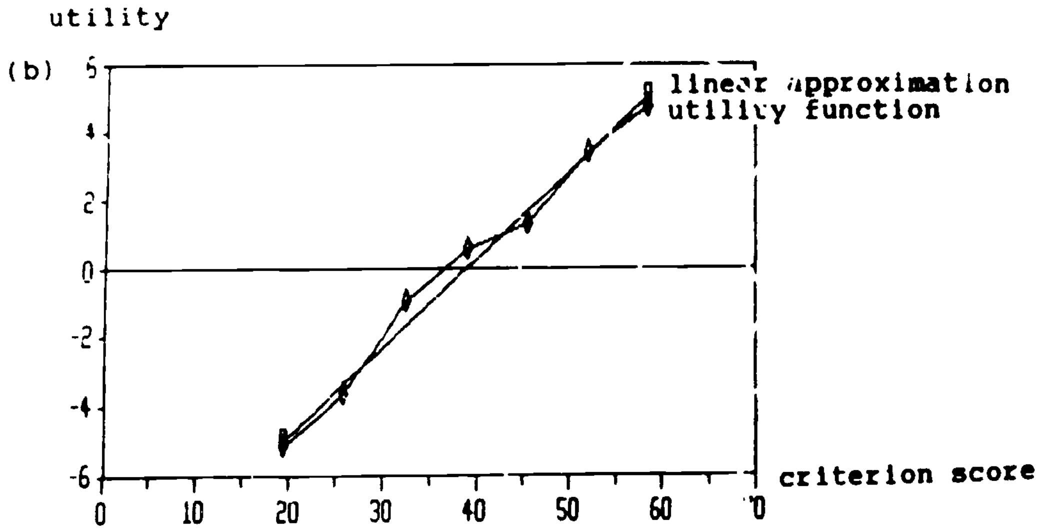
negative 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 positive

Figure 5

Instruction for Determining Point of Intersection, Selection

At which of the criterion scores you equally evaluate the decision to accept or reject? It can be each of the scores between 13 and 65. At a score of ... I am indifferent whether the applicant is accepted or rejected.





<u>Figure 6.</u> Utility function and linear approximation (a) judge no. 2, and (b) psychologist no. 1, acceptance situation.

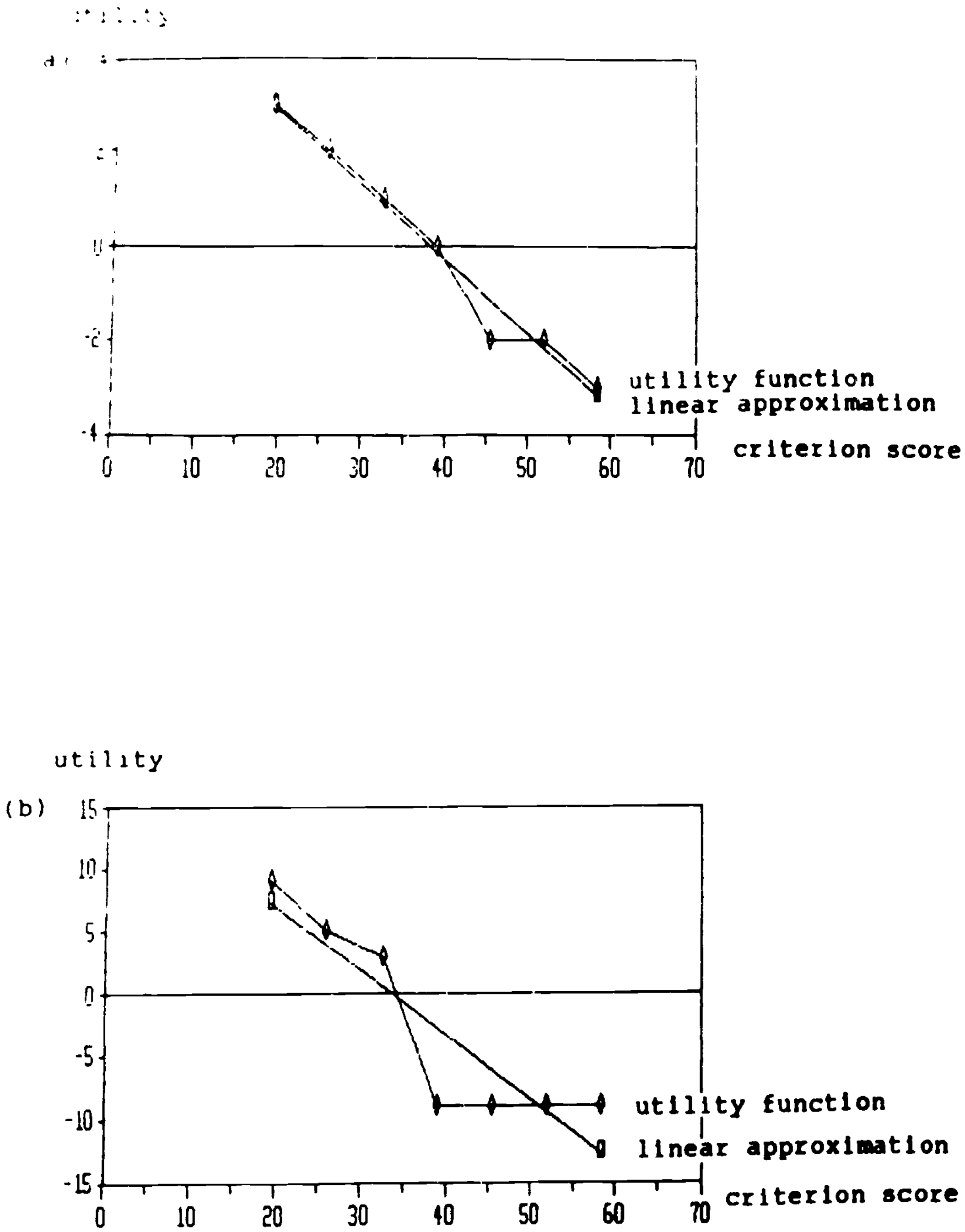


Figure 7. Utility function and linear approximation (a) judge no. 3, and (b) psychologist no. 3, rejection situation.

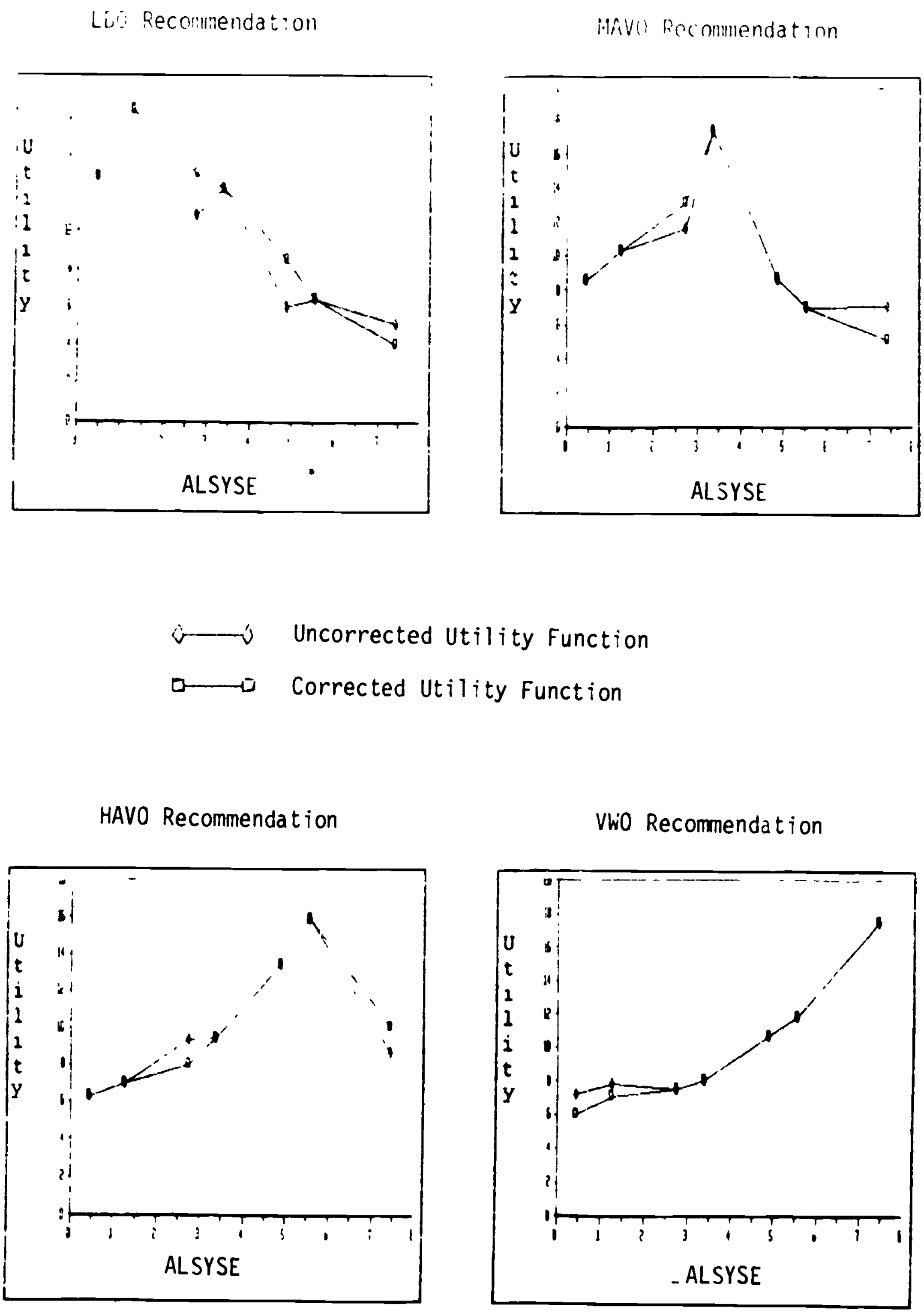


Figure 8. Uncorrected and corrected utility functions first subject, classification