Hybrid and non-hybrid evaluation models for predicting outdoor recreation behavior : a test of predictive ability

Citation for published version (APA):

Timmermans, H. J. P. (1987). Hybrid and non-hybrid evaluation models for predicting outdoor recreation behavior : a test of predictive ability. *Leisure Sciences*, 9(2), 67-76.

Document status and date:

Published: 01/01/1987

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

Hybrid and Non-Hybrid Evaluation Models for Predicting Outdoor Recreation Behavior: A Test of Predictive Ability

HARRY TIMMERMANS

University of Technology
Eindhoven, The Netherlands

Abstract *This paper compares the internal validity, external validity and ease of implementation of different types of compositional, decompositional and hybrid evaluation models in the context of recreational trip behavior. The results indicate that the full-profile conjoint measurement model outperforms the self-explicated models, the hybrid evaluation models and the trade-off conjoint measurement model in terms of both internal and external validity, although the models differ less in terms of external validity. However, in terms of ease of implementation, the decompositional models may be less attractive.*

Keywords evaluation models, recreational trip behavior, conjoint analysis

Introduction

Over the last decades, the issue of modeling outdoor recreation behavior has been of continuing interest to scholars working in the field of recreation and leisure research. Several modeling approaches have been used for describing recreational behavior patterns and predicting the likely effects of recreation plans on subsequent behavior. Aggregate interaction models were very popular in the 60s and 70s (Ellis and Van Doren 1966; Thompson 1967; Wennergren and Nielsen 1970; Cheung 1972; McAllister and Klett 1976; Baxter and Ewing 1981). More recently, however, the interest has shifted from such aggregate models to various types of individual decision models such as compositional attitude models (Cooksey et al. 1982), discrete choice models (Peterson et al. 1983) and decompositional multiattribute preference models (Lieber and Fesenmaier 1984). A detailed state-of-the art paper in recreation modeling and recreational choice theory was recently written by Cordell et al. (1985). This shift in interest reflects the underlying argument that aggregate spatial interaction models basically describe correlational patterns between behavior and a set of environmental factors and hence do not reveal the rules by which alternative locations are evaluated and choices are made.

Although individual decision models offer a potentially powerful way of uncovering the nature of decision-making and choice processes, there are still some issues which deserve further attention. One of these issues concerns the measurement of individual evaluations and preferences which constitute the basis of individual models. If a model is to yield good predictions, a necessary although not sufficient condition is that such evaluations and preferences be accurately measured. If unreliable or biased measurements are obtained, it is difficult to see how a model can predict the likely effects of recreation plans with any degree of accuracy.

Received February 1, 1985; accepted October 21, 1985

Several measurement procedures can be used for measuring evaluations and preferences. To the best of the author's knowledge, no comparative analysis has been conducted so far in the field of recreation and leisure research to gain more insight into the advantages and disadvantages associated with the different procedures. There exist a few studies in other fields, especially in marketing (Wiley, MacLachlan, and Moinpour 1977; Hoepfl and Huber 1970; Akaah and Korgaonkar 1983), but the results of these studies do not necessarily hold in the field of recreation and leisure research. The aim of the present paper is therefore to compare the usefulness and predictive performance of various measurement procedures associated with individual decision models. As some of these procedures are uniquely related to particular modeling approaches, this paper also presents the results of a test for comparing the predictive performance of different types of individual decision models.

This paper is organized as follows. The next section gives a summary of the measurement procedures that were compared in this study. This is followed by a discussion of the study design; then, the main findings of the empirical analysis are presented. The paper concludes with a discussion of the results of the analysis.

Measurement procedures and model structures'

Individual decision models have in common the assumption that choice alternatives consist of bundles of attributes. An individual's overall evaluation of a choice alternative is assumed to be the result of a cognitive integration process in which an individual combines his or her evaluations of the alternative's various attribute levels according to some algebraic function (or combination rule). The various model approaches can be differentiated in terms of 1) the way in which the nature of this integration process is uncovered and 2) the measurement of its components. At least three approaches can be identified.

The *compositional* approach typically assumes that the overall evaluation of a multiattribute choice alternative can be arrived at by combining evaluations associated with the attribute levels (or part-worth utilities) of that choice alternative and the subjective importance weights which an individual attaches to the various attributes according to some combination rule. Consequently, this approach involves measuring separately and explicitly an individual's evaluation of perceived attribute levels and subjective importance weights. Several combination rules may be used to describe the way in which these part-worth utilities and importance weights are combined to arrive at some overall evaluation. For example, a linear additive rule assumes that the overall evaluation for a choice alternative i (E_i) can be computed by summing the self-explicated attribute evaluations ($e_{jk}^{(i)}$) weighted with the subjective importances (w_j). In formula,

$$E_i = \sum_{j=1}^J w_j e_{jk}^{(i)} \quad (1)$$

Other specifications are possible (see Timmermans 1984 for a review). In addition, one could use unweighted or normalized versions. Unweighted versions assume that importance weights are implicitly included in the evaluations of the attribute levels. Normalized versions normalize the attribute evaluations and importance weights by dividing them by their respective sums.

The second modeling approach is the so-called *decompositional* measurement approach. Unlike the compositional approach, it typically involves *measuring* rather than

calculating overall evaluations associated with a set of choice alternatives. These measures of overall evaluation are then decomposed into the contributions (or part-worth utilities) associated with each of the attribute levels, given some type of combination rule (e.g., a linear additive one). The functional measurement and conjoint measurement approaches are typical examples of the decompositional approach.

The measurement itself involves the following steps. First, a set of hypothetical choice alternatives is constructed according to the principles of experimental design. That is, each attribute is defined in terms of a series of categories or attribute levels and then these attribute levels are combined, using full factorial designs, in every possible way to yield a set of hypothetical choice alternatives. Next, the individual is requested to express his degree of preference, for example by rating or ranking, for these hypothetical choice alternatives. The estimation problem is then to derive the scale values for the attribute levels (so-called part-worth utilities) which, when combined according to some prespecified combination rule (e.g., a linear additive rule), are as closely related to an individual's expressed overall preference scores as possible. Various estimation methods, ranging from monotonic regression analysis to linear programming techniques and ordinary least squares, may be used to derive the part-worth utilities (see Timmermans 1984 for details).

Mathematically, the conjoint main-effects model can be expressed as:

$$U_i = \sum_{j=1}^J \sum_{k=1}^{K_j} u_{jk} x_{jk}^{(i)} \quad (2)$$

where U_i denotes an individual's overall evaluation of choice alternative i ;
 u_{jk} is the derived part-worth utility associated with level k of the j -th attribute;
 $x_{jk}^{(i)}$ is a dummy variable representing level k of the j -th attribute that corresponds to choice alternative i .

When the number of choice alternatives and/or the number of attribute levels is large, full factorial designs result in a large number of choice alternatives. Under such circumstances the reliability of the measurements may be seriously questioned, since the experimental task might be too demanding both in terms of the number of hypothetical choice alternatives to be evaluated and the complexity of descriptions. Because of this information overload, individuals might simply ignore the less important attributes in providing their evaluations. An alternative method which is less demanding is the so-called trade-off approach, in which the attributes are presented in pairs and individuals are requested to rank the hypothetical choice alternatives which result by combining the levels of the pair of attributes in every possible way. The estimation of the part-worth utilities follows paths similar to those associated with full-profile descriptions, although the estimation differs in terms of the actual goodness-of-fit measures, the design matrix and/or the algorithms involved.

The *third* approach is called the *hybrid* modeling approach. Again, several versions may be identified, but in the present paper only the models suggested by Huber (Huber 1974; Huber Daneshgar and Ford 1971; Huber, Sahney and Ford 1969) will be considered. This approach is an amalgamation of the measurement procedures associated with the compositional (self-explicated) and decompositional multiattribute preference models. The approach requires an individual to rate both an overall evaluation for each choice alternative and an evaluation of the attribute levels characterizing the choice alter-

native. Unlike the self-explicated model, the importance weights are not separately and explicitly measured, but are estimated by multiple regression, where the overall evaluation of the choice alternatives constitutes the dependent variable while the independent variables consist of the self-explicated evaluations of the various attributes of the choice alternatives.

Three model specifications were considered in this study. First, the linear additive hybrid multiattribute preference model was considered. It assumes that the overall evaluation of choice alternatives is the weighted sum of the evaluation of the attributes. It can be expressed as:

$$U_i = \beta_0 + \sum_{j=1}^J \beta_j u_{jk}^{(i)} \quad (3)$$

where β_0 is an intercept term;

β_j is a coefficient that reflects the relative contribution of the j -th attribute to the overall evaluations.

The second model specification that was examined is the multiplicative model, which can be expressed as:

$$U_i = \prod_{j=1}^J \beta_0 u_{jk}^{(i)\beta_j} \quad (4)$$

After transforming the above equation by taking the natural logarithm on both sides, the β_0 and β_j can be estimated by multiple regression analysis. Whereas the linear additive model represents a truly compensatory decision-making process, the multiplicative model is capable of representing noncompensatory decision-making processes. The third model examined was the addilog hybrid model, which can be expressed as:

$$U_i = \beta_0 + \sum_{j=1}^J \beta_j \log u_{jk}^{(i)} \quad (5)$$

The addilog hybrid model was suggested on the basis of empirical evidence which suggested that an individual's perception of differences in physical stimuli is proportional to the logarithm of the difference and not to be the distance itself. The model can be considered as a variation of the additive hybrid model, it takes the log of the utilities, and hence it reflects also compensatory decision-making processes.

Study design

A random sample of 81 respondents was selected through personal interviews during August 1984. All interviews took place at the respondents' homes. The survey was conducted by a specially trained interviewer. All respondents live in Tilburg, a major city in the south of The Netherlands. The sample was diverse in terms of socioeconomic background variables.

The operationalization of the models required several decisions. First, the factors influencing recreation behavior had to be elicited. Several methods may be used for the identification of influential factors (see, e.g., Allton and Lieber 1983). In the present

study, an extensive literature search and a few in-depth interviews yielded six major attributes influencing recreational choice behavior: type of area, quietness, distance, facilities, admission, and number of recreation opportunities. Each of the six attributes was conceptualized at three levels. Table 1 gives the attributes and their levels as used in this study. Second, the decompositional multiattribute preference models required the choice of an experimental design. Since the nine models included the full-profile and the trade-off approaches, both these measurement procedures were adopted. However, since a full factorial experimental design would yield $3^6 = 729$ hypothetical choice alternatives, a fractional factorial design involving only 27 choice alternatives and which allowed estimating all main effects was selected. Six additional hypothetical choice alternatives which served as holdout profiles were constructed to assess the external validity of the models.

The data collection itself involved the following measurements. First, the self-explicated evaluations of the attribute levels were measured by showing the three levels of each of the six attributes to each respondent, who was then asked to evaluate the attribute levels in terms of their desirability on a zero-to-10-point rating scale. Each respondent was asked first to assign a value of 10 to the most preferred attribute level. The respondent was then asked to rate the remaining attribute levels in terms of their desirability in relation to the most preferred attribute level. Ties were permitted. Each attribute was shown in turn; hence these measurements were conducted for each set of attribute levels separately. Second, after all six sets of attribute levels had been evaluated, respondents were requested to express the relative importance of each of the six attributes for recreational choice behavior on a zero-to-100 scale, allocating 100 points over the six attributes. Again, respondents were first requested to identify the most important attribute

Table 1
Attributes and Attribute Levels

1. Type of area	—natural park —amusement park —watersport area
2. Quietness	—very quiet —average —crowded
3. Distance	—0–30 minutes —31–60 minutes —61–90 minutes
4. Facilities	—non-existent —only a few present —many facilities
5. Admission	—free —average fee —high fee
6. Number of recreation opportunities	—one opportunity —a few opportunities —many opportunities

and assign some importance weight, and then to express the importance of the remaining attributes by allocating the remainder of 100 points. Again, ties were permitted. Third, respondents were shown the full-profile descriptions of the 27 hypothetical choice alternatives and asked to evaluate each of these choice alternatives on a zero-to-100-point scale in terms of overall preference. Moreover, respondents were shown a series of six trade-off matrices, a balanced subset of all 15 possible trade-off matrices, and asked to rank the hypothetical choice alternatives, which thus differed on only one or two attributes, in terms of overall preference. Respondents were told to assume that the choice alternatives did not differ on any other attribute. Finally, respondents were invited to evaluate each of the six holdout hypothetical choice alternatives on a zero-to-100-point scale in terms of overall preference; ties were permitted.

In summary, the measurement tasks resulted in the following sets of data:

- (1) An 81×27 matrix of each respondent's overall evaluation of the 27 hypothetical recreation objects;
- (2) An 81×6 matrix of each respondent's overall evaluation of the 6 holdout profile descriptions of hypothetical recreation objects;
- (3) An 81×6 matrix of subjectively explicated importance weights, one for each of the six attributes;
- (4) An 81×18 matrix of the self-explicated evaluations of the three levels of each of the six attributes;
- (5) Six 81×9 trade-off matrices representing a respondent's overall preference for each of the nine combinations of pairs of three-level attributes.

Analysis and results

Three principal research questions guided the analysis:

- (1) How well does each of the nine evaluation models represent each respondent's data?
- (2) How well does each of the nine evaluation models predict each respondent's evaluation of the holdout profiles?
- (3) Do the measurement procedures underlying the nine evaluation models differ in terms of their demand of a respondent's ability to express his or her preferences and evaluations?

The examination of the internal and predictive validity of the evaluation models was based on three measures. The test on the internal validity involved computing independently, for each respondent, a Pearson's product moment correlation coefficient and a Spearman's rank correlation coefficient between the respondent's overall evaluation of the 27 hypothetical recreation objects and the overall evaluation as predicted by fitting the model. Next, these measures were averaged across respondents. The test on the external or predictive validity involved using for each respondent separately his or her self-explicated data and/or estimated parameter value, depending on the type of evaluation model, and then computing independently for each respondent the product-moment correlation coefficient and Spearman's rank order correlation coefficient between the respondent's overall evaluation of the six holdout profiles and the overall evaluation as predicted by the respondent's specific model. Again, these measures were averaged across the 81 respondents.

Table 2 provides a summary of the test on the internal validity of the nine evaluation

Table 2
The Internal Validity of the Models

Model	Product-Moment Correlation Coeff.	Rank Correlation Coefficient
1. weighted additive model	0.61	0.61
2. unweighted additive model	0.58	0.57
3. normalized weighted additive model	0.61	0.60
4. normalized unweighted additive model	0.57	0.57
5. additive hybrid model	0.81	0.82
6. multiplicative hybrid model	0.68	0.78
7. logadditive hybrid model	0.82	0.83
8. full profile conjoint measurement model	0.91	0.94
9. trade-off conjoint measurement model	0.64	0.64

models. Several conclusions may be drawn by examining this table. First, it indicates that the conjoint models generally outperform the self-explicated models and the hybrid evaluation models. The rank order correlations and product-moment correlations show that on average the respondents' overall evaluations of the 27 hypothetical recreation objects is best represented by the full-profile conjoint measurement model, followed by Huber's hybrid evaluation models. The self-explicated models and the trade-off conjoint measurement model are the poorest in representing the respondents' overall evaluations of the hypothetical recreation objects. This result is not surprising in that, unlike the full profile conjoint models and the Huber hybrid evaluation models, the fitting by the self-explicated models and the trade-off conjoint measurement model does not explicitly involve the respondent's overall evaluations of the 27 recreation objects. In addition, it should be remembered that the conjoint models have a larger number of parameters than the self-explicated and the hybrid evaluation models.

Second, Table 2 illustrates that the hybrid evaluation models yield comparable results. The average rank correlation coefficient of the addilog model is .83, that of the additional model .82 and that of the multiplicative model .78. The Pearson's product-moment correlation coefficient gives similar results, although the predictive performance of the multiplicative models now drops somewhat more ($r = .68$), which suggests that for the given degree of monotonicity the multiplicative hybrid evaluation model deviates more from linearity than the other two hybrid evaluation models.

Third, an examination of Table 2 indicates that normalizing the importance weights and evaluations of attribute levels does not substantially influence the internal validity of the self-explicated evaluation models. Last, this table shows that the weighted self-explicated evaluation models somewhat better represent on average the respondent's overall evaluations of the 27 hypothetical recreation objects than do the unweighted self-explicated evaluation models.

A more stringent test of the validity of hybrid and non-hybrid evaluation models is a test of cross-validity: the ability of a model to predict the respondents' overall evaluation of the six holdout profiles. The results of this test are provided in Table 2. Table 3 indicates that the rank ordering of the nine evaluation models in terms of Pearson's

Table 3
The External Validity of the Models

Model	Product-Moment Correlation Coeff.	Rank Correlation Coefficient
1. weighted additive model	0.62	0.61
2. unweighted additive model	0.57	0.53
3. normalized weighted additive model	0.62	0.60
4. normalized unweighted additive model	0.57	0.54
5. additive hybrid model	0.72	0.71
6. multiplicative hybrid model	0.66	0.72
7. logadditive hybrid model	0.74	0.73
8. full profile conjoint measurement model	0.78	0.77
9. trade-off conjoint measurement model	0.62	0.58

product-moment correlation coefficient is very similar to the rank ordering implied by Table 2, which expresses the results of the test on the internal validity of the models. The full profile conjoint measurement model still gives the best predictive results ($r = .78$), followed by the hybrid evaluation models. Again, the self-explicated models and the trade-off conjoint model are poorest in predictive performance, with an average product-moment correlation coefficient ranging from .58 to .62. Table 3 also illustrates that normalizing the importance weights and evaluations of the attribute evaluations does not have any effect on the predictive ability of the self-explicated models. Probably this result is due to the restrictions imposed upon the measurements, which basically set the origin and the unity of the scales. Table 3 also shows that the weighted versions of the self-explicated models perform somewhat better on this cross-validity test than do their unweighted counterparts. Finally, the addilog hybrid evaluation model on average performs somewhat better than the additive and multiplicative hybrid evaluation models, at least in terms of the product-moment correlation coefficient. Spearman's rank correlation coefficient gives a somewhat different picture in that the three hybrid models now perform almost equally well, while the multiplicative model now ranks above the additive model. Another important observation which may be drawn from Table 3 is that, although on average the full profile model and the hybrid evaluation model still outperform the self-explicated models and the trade-off conjoint model, the differences in predictive ability are less clear-cut than in the case of the calibration data. The average product-moment correlation associated with the full profile conjoint measurement model has dropped from .91 to .78, and the average product moment-correlation coefficient for the three hybrid evaluation models has decreased by respectively .09, .02 and .08 units, whereas the average product-moment correlation coefficients for the self-explicated models have dropped at most .01 units or have even increased .01 unit.

Obviously, one expects the measurement procedures underlying the nine evaluation models to differ in terms of what they are demanding from a respondent's ability to express his or her preferences and evaluations. To gain more insight into this aspect, respondents were asked to evaluate the three measurement procedures after finishing the

questionnaire. The rating procedure underlying the self-explicated models was considered to be the most simple procedure. Many respondents expressed their difficulty in grasping fully the full-profile descriptions. The trade-off task was considered less complex, although it was sometimes difficult to assume independence among the attributes. At the very least, the trade-off task took considerably less time to finish than the task involving the full-profile descriptions of the choice alternatives, notwithstanding the fact that it required more judgments. Apparently, thus, the full-profile method may yield reliable results but the respondents require much more time to grasp and compare the descriptions of the choice alternatives.

Conclusions and discussion

The primary aim of this study was to assess the predictive ability of a set of hybrid and non-hybrid evaluation models in the context of recreational choice behavior. In particular, three general approaches of modeling preferences and evaluations—the compositional (self-explicated) approach, the decompositional (conjoint measurement) approach and the hybrid (Huber) approach, involving nine different models—were compared in terms of their internal and external validity and the ease of implementing these approaches.

The findings of this study suggest that the full-profile conjoint measurement model (decompositional approach) outperforms the self-explicated models and the hybrid models both in terms of internal and external or cross-validity. The advantage of this approach is that respondents are forced to consider the hypothetical choice alternatives carefully and make comparisons continuously. At the same time, however, this feature of the decompositional approach might be a drawback, since the approach might be too demanding for some respondents in terms of their capability of grasping the profile descriptions and expressing their preferences for these profiles. Hence, the choice between these alternative modeling approaches might ultimately depend on practical circumstances such as available budget, type of respondents, complexity of the problem, etc.

As has been noted in the introductory section, similar analyses have been conducted by marketing scientists, and by and large they have obtained similar findings, although the predictive performance of the self-explicated models was somewhat better in this study. Hence, it seems that the superiority of the decompositional approach is not restricted to a particular field of study but can be generalized to different contexts. Nevertheless, it is evident that these findings on recreational behavior are based on a single sample. There is a clear need for further empirical testing, especially in connection with different measurement procedures, types of recreational behavior, and socio-economic and cultural background of respondents. It is hoped that this paper will stimulate such research efforts.

References

- Akaah, I. and P.K. Korgaonkar. 1983. An empirical comparison of the predictive validity of self-explicated, Huber-hybrid, traditional conjoint and hybrid conjoint models. *Journal of Marketing Research* 20:187–198.
- Allton, D.J., and S.R. Lieber. 1983. Trail area attributes of metropolitan Chicago. *Leisure Sciences* 5:197–220.
- Baxter, M.J., and G. Ewing. 1981. Models of recreational trip distribution. *Regional Studies* 15:327–344.

- Cheung, H.K. 1972. A day-use park visitation model. *Journal of Leisure Research* 4:139–156.
- Cooksey, R.W., T.L. Dickenson, and L.L. Loomis. 1982. Preference for recreational environments: theoretical considerations and a comparison of models. *Leisure Sciences* 5:19–34.
- Cordell, H.K., D.R. Fesenmaier, S.R. Lieber, and L.A. Hartmann. 1985. Advancements in methodology for projecting future recreation participation. Paper presented at the Trends Conference, Myrtle Beach, South Carolina, February.
- Ellis, J.B., and S. van Doren. 1966. A comparative evaluation of gravity and system theory models for statewide recreational traffic flows. *Journal of Regional Science* 6:57–70.
- Hoepfl, R.T., and G.P. Huber. 1970. A study of self-explicated models. *Behavioral Science* 15:408–414.
- Huber, G.P. 1974. Multiattribute utility models: a review of field and field-like studies. *Management Science* 20:1393–1402.
- Huber, G.P., R. Daneshgar, and D.L. Ford. 1971. An empirical comparison of five utility models for predicting job preferences. *Organizational Behavior and Human Performance* 6:267–282.
- Huber, G.P., V. Sahney, and D.L. Ford. 1969. A study of subjective evaluation models. *Behavioral Science* 14:483–489.
- Lieber, S.R., and D.R. Fesenmaier. 1984. Modelling recreation choice: a case study of management alternatives in Chicago. *Regional Studies* 18:31–43.
- McAllister, D.M., and F.R. Klett. 1976. A modified gravity model of regional recreation activity with an application to ski trips. *Journal of Leisure Research* 8:21–34.
- Peterson, G.L., J.F. Dwyer, and A.J. Darragh. 1983. A behavioural urban recreation site choice model. *Leisure Sciences* 6:61–81.
- Thompson, B. 1967. Recreational travel: a review and pilot study. *Traffic Quarterly* 21:527–542.
- Timmermans, H.J.P. 1984. Decompositional multiattribute preference models in spatial choice analysis: a review of some recent developments. *Progress in Human Geography* 8:189–221.
- Wennergren, E.B., and D.B. Nielsen. 1970. Probability estimates of recreation demands. *Journal of Leisure Research* 2:112–122.
- Wiley, J.B., D.L. MacLachlan, and R. Moinpour. 1977. Comparison of stated and inferred parameter values in additive models. *Advances in Consumer Research* 4:98–105.