A PUBLIC WEIGHTING OF FOUR SOCIETAL GOALS IN ARIZONA AND OREGON

D. B. Kimball, R. L. Gum, T. G. Roefs
Department of Hydrology and Water Resources, University of Arizona

In response to the increasing demand for multi-objective planning in the development of water resources, the Technical Committee of the Water Resources Centers of the Thirteen Western States has proposed a new planning methodology which relates water resource use to "social goals" (Peterson, et al., 1971). The planning system, referred to henceforth as the "Strawman," is based on a comprehensive set of nine societal goals which are delineated by an hierarchical array of sub-goals which are connected to water resource policy through social indicators, measures of socially significant phenomena. Provided that the form of connectives between social indicators, sub-goals, and goals have been defined (Gum, et al., 1973) and that the coefficients, or weights, of the connective algorithms are known, the Strawman is capable of providing aggregated information concerning the impacts of alternative plans on an array of goals and/or sub-goals to the policy maker. To incorporate weights into the Strawman planning methodology, two conceptual problems must be resolved. First, "who," the decision-maker, the general public, or interest groups, should be involved in weighting the improvement or attainment of specified goals? Second, what is the appropriate weighting methodology given the requirements of the Strawman planning system and the expertise of the designated individual(s) involved in the weighting process?

Mr. Weston Wilson, a Graduate Associate in Research at the University of Arizona, has addressed himself to the first of these two problems in his paper, "The Cognitive Strawman Planning Methodology: Public Input" (Wilson, et al., 1973). His conclusion is that general public input into the weighting of goals and sub-goals within the Strawman hierarchy is well justified and possible, provided that sub-goals are defined in such a way that they are easily perceived by the public. Mr. Wilson also states that three prime goals, Aesthetic Opportunity, Recreation Opportunity, and Economic Opportunity, and also a sub-goal of Collective Security, Health Security, are most amenable to public perception and also have special relevance to water resource development. Given these recommendations concerning the "who" and the specified goal areas as defined by the "Cognitive Strawman," the following discourse shall concern itself with the development of an appropriate weighting methodology and a test application in two states, Arizona and Oregon.

INTRODUCTION

Undoubtedly, the initial response to the development of a weighted preference methodology must be the basic question, "Can preferences be measured?" Measurement, as defined by Stevens (1946, p. 677), is the assignment of numerals to objects or events according to rules. The ease of application of such a definition to physical objects or processes is readily apparent, but measurement becomes more problematical when subjective entities are involved. Many psychologists have adopted Thorndike's dictum, "Whatever exists at all, exists in some amount," (Thorndike, 1918, p. 16) and have also adopted the corollary that whatever exists in amount, for example the quality of handwriting or the appreciation of a sunset, is measurable. Thus, the idea of measuring preferences in terms of the desirability of sub-goal improvements _per se_ is not overly problematical, but the measurement of such non-physical attributes does imply the necessity of a rigorous consideration of two measurement enigmas. First, in the operation of scaling an attribute on a psychological continuum, what is the functional relationship between the method of scaling and the concept being measured? Second, given that a psychological measurement technique has evaluated a number of stimuli on a linear scale, what type of measurement scale has actually been achieved? Both of these questions have special relevance to the development of an acceptable weighted preference methodology and therefore must receive special attention.

A survey of the psychological literature yielded a plethora of methods, including ranking, rating, paired comparison, fractionation, magnitude estimation, and ratio estimation, all of which are capable of evaluating a stimulus on a linear preference scale. However, the nature of the Strawman planning methodology itself and also the requirement of obtaining weighted preferences from the general public produced constraints on the type of weighting methodology appropriate for inclusion into the Strawman's planning framework.

THEORETICAL CONSTRAINTS

The dominant theoretical constraint on the acceptability of a weighting methdoology is the level of measurement which must characterize the achieved preference weights and continua The proposed Strawman planning methodology is a quantified planning structure utilizing numerical weights, mapped-in social indicator values, and a function to aggregate information. The dependence of such a system on quantitative relationships and manipulations requires that the achieved preference weights be analogous in nature to cardinal numbers. That is, the weights must exhibit the properties of cardinal numbers so that all foreseeable mathematical manipulations and comparisons within the hierarchical structure of the Strawman are permissible. The imposition of such a restriction requires that any acceptable weighting methodology must attain an order of measurement equal to that of the cardinal number system.

REVIEW OF MEASUREMENT SCALES

Based on the fact that numerals can be assigned to objects or events under different rules and operations, S. S. Stevens

(1946) recognized that different kinds of measurement and different kinds of scales could be produced. As a result, he advanced a theory concerning scales of measurement which has as its basis the concept of invariance. That is, after a set of numbers has been assigned to reflect the outcome of a series of empirical operations, such as orderings, comparisons, or balancings, the type of measurement achieved can be ascertained by determining how the scale values can be transformed without the loss of empirical information. Using this criterion and the idea that the type of scale achieved depends upon the character of the basic empirical operations performed, Stevens developed a "Classification of Scales of Measurement" (Stevens, 1959, p. 25) which includes four scales of measurement, nominal, ordinal, interval, and ratio.

The lowest scale of measurement, nominal, requires only the determination of equality and thus represents the most unrestricted assignment of numerals. Although nominal scaling is not always thought of as a form of measurement because of its use of names or letters to designate categories and classes, it does perform the important function of classifying and identifying.

Steven's next higher order of measurement, ordinal, not only requires a determination of equality but also a determination of greater or less. This second operation usually takes the form of rank-ordering, or ranking objects in terms of the magnitude of some characteristic without measuring the amount of the characteristic possessed by each object. For example, on Mohs' Hardness Scale, diamond is harder than corundum, but from this scale nothing can be said about the relative difference in hardness or how hard each is in an absolute sense.

An interval scale of measurement, such as the Fahrenheit or Celsius temperature scales, requires the determinations of equality, greater or less, and the equality of intervals or of differences. From this scale of measurement it cannot only be said that, for example, 70^{o}F is warmer than 60^{o}F, but also that 70^{o}F is 10^{o}F warmer than 60^{o}F. The equal intervals of temperature, or degrees, are scaled by noting equal volumes of expansion of mercury in a column. However, the zero point on such a scale is purely a matter of convention or convenience. It cannot be said that 70^{o}F is twice as warm as 35^{o}F. The zero point is arbitrary as can be shown by the fact that the scale remains invariant under multiplication by a constant and addition of a constant (i.e., a linear monotonic transformation). Almost all common statistical measures are applicable with interval scales, provided that such statistics do not require the knowledge of a true zero point. Problems arise, however, in the use of such a scale in higher-order mathematical operations such as logarithmic transformations.

The highest level of quantitative description, the ratio scale of measurement, is possible only when operations have been made for the determination of equality, greater or less, the equality of intervals, and the equality of ratios. Examples of this scale of measurement include the Kelvin or absolute temperature scale, length, time intervals, and the cardinal number system. In all ratio scales, an absolute zero is always implied though the zero point on some scales (e.g., the Kelvin scale) may never be produced. Once a ratio scale is erected, it can be said, for example, that 6 inches is twice as long as 3 inches or $100^{o}K$ is twice as warm as $50^{o}K$. Length or the amount of heat can be thought of in an absolute sense or an amount more than a rational or true zero point. The ratio scale, whose form is invariant only under multiplication by a constant (e.g., converting meters to centimeters) is the most restrictive class of measurement and exhibits properties which allow the application of all types of statistical measures and mathematical equations and manipulations.

IMPLICATIONS TO A WEIGHTING METHODOLOGY

Given the previously cited need for weights which are analogous in nature to the cardinal number system and the previous review of measurement scales, it becomes apparent that a weighting methodology appropriate for inclusion into the Strawman's planning framework must exhibit measurement of the highest order, a ratio scale. However, it must be noted that although ratio scales have been utilized in psychophysical studies, the use of ratio scales in general psychological studies of attitudes and preferences has been the exception rather than the rule. Most psychological data are legitimately expressed only as interval scales. It makes little sense to speak, for example, of zero intelligence or to be able to say that one person is 1½ times as anxious as another. However, the idea of a zero desire for an improvement in a sub-goal does make sense, and the ratio comparisons of preferences are essential to the Strawman in terms of the mathematical requirements of the system. Thus, an acceptable weighting methodology, in terms of the theoretical constraints, must be capable of achieving a ratio measurement of sub-goals on a preference dimension.

PRACTICAL CONSTRAINTS

The inclusion of a weighting methodology into the Strawman's planning framework also requires that certain practical constraints be met. Given the rationale for general public involvement in the weighting process, it was decided that any acceptable weighting methodology must be conducive to a mail questionnaire survey. If the target groups had been decision-makers or interest groups, an interview type survey would have perhaps been more suitable, but the desire to query a random sample of the general public over a large area required, due to limited resources, the acceptance of a mail survey. Thus, an acceptable methodology, in terms of the practical constraints, must be short, easily understood, and answerable in an uncontrolled atmosphere.

PROBLEMATICAL METHODOLOGIES

Based on the theoretical constraint of the achievement of a ratio scale of preference, it was realized that the most common psychological measuring techniques were unacceptable. Ranking, the most efficient method of psychological measurement (Eckenrode, 1965, p. 183), requires only that judges arrange stimuli in order of increasing excellence, quantity, or preference with respect to the psychological variable under consideration. Such an operation, however, yields strictly ordinal information and says nothing about the distance or ratio between the ranks. Rating, the most common procedure for measuring human judgments (Schimpeler, 1967, p. 99), involves the location of stimuli on a presumably unique psychological continuum which has been supplied with descriptive phrases to indicate position. However, because judgments are not made in terms of an absolute numerical scale, the resulting evaluations of the stumuli exhibit only interval characteristics. An interesting off-shoot of the rating technique is the Theory of Signal Detection (TSD) rating method (Daniel, et al., 1971), a technique which purports to separate an individual's perception of a stimulus from his "judgmental criteria." However, like the results of simple rating methods, TSD also yields strictly interval information. Another common psychological measuring technique is the successive paired comparison test (Thurstone, 1927). In its typical application, every possible combination of two stimuli is presented to the subject. The number of times each stimulus is preferred to every other one is calculated, and through an information processing algorithm, an interval scale may be derived from the paired comparison judgments. From this brief review of the scales of measurement attained by the most common psychological techniques, it can be discerned that adherence to such a theoretical constraint, the achievement of a ratio scale, is rare in general psychological testing, and thus the investigation of more obscure psychological measurement techniques was required.

A review of the psychophysical measurement literature did yield three ratio-scaling techniques, fractionation (Stevens, 1959), ratio estimation (Stevens, 1959), and magnitude estimation (Stevens, 1966), which were seemingly appropriate for our use on theoretical grounds. However, definite difficulties were quickly discovered when the practical nature (e.g., the complexity of instructions and the difficulty of usage in a mail questionnaire) of each method were considered.

INVESTIGATED WEIGHTING METHODOLOGIES

Although none of the more common attitude and preference measuring techniques nor the better known ratio-attaining psychophysical techniques could be adopted to produce the desired weighted preference system, the literature did yield three techniques, the Thurstone Paired Comparison Test with

Rational Origin Assumption (Thurstone and Jones, 1959), the Comrey Paired Allocation Test (Comrey, 1950), and the Metfessel General Allocation Test (Metfessel, 1947), which were seemingly appropriate and therefore required further investigation.

It must be noted that these tests represent the two basic lines of development in scaling, "indirect" and "direct" (Ekman and Sjöberg, 1965). In general terms, indirect scaling, as represented in the Thurstone Paired Comparison Test, requires only a minimal amount of information from the subject and then determines the scale from the experimental data using a set of assumptions concerning variability. On the other hand, direct scaling methods such as the Comrey Paired Allocation Test and Metfessel's General Allocation Test require subjects to directly report quantitative judgments. The use of such direct methods requires the acceptance of the assumptions that subjects are capable of estimating quantitative relations between subjective experiences, and, in the case of ratio-attaining direct methods, observers can directly estimate the ratio of two or more psychological entities. A brief discussion of the three weighting methods and the test results of each follows.[1]

Thurstone Paired Comparison Test with Rational Origin Assumption. The Thurstone Paired Comparison Test with Rational Origin Assumption is a technique which utilizes the Law of Comparative Judgment (Thurstone, 1959, p. 34-49) to produce an additive measurement scale of subjective values and which also experimentally determines a subjective, or rational, origin. In such a test, subjects are asked to express their preferences in paired comparison-type questions between each of all single stimuli (e.g., A or B), all combinations of paired stimuli and all single stimuli (e.g., AB or C), and each of all combinations of paired stimuli (e.g., AB or CD). Then, using the Law of Comparative Judgment, all single stimuli and also all combinations of paired stimuli are assigned scale values. The rational origin, or zero point, is then simply determined by calculating the mean difference between the sum of the scale values of all pairs of stimuli (e.g., A and B) and the respective scale values of the combinations of the pairs of stimuli (e.g., AB). Thus, Thurstone's basic assumption is that the utility derived from both A and B (i.e., AB) is equal to the sum of the utility derived from A and from B separately. Providing one accepts this assumption, a zero point exists, a ratio scale of preference is achieved, and thus all mathematical manipulations are permissible.

However, the Thurstone Paired Comparison Test with Rational Origin Assumption has definite practical and theoretical shortcomings. First and foremost is the required length and redundancy of the test. The need to consider single-single, double-single, and double-double stimuli comparisons requires 21 paired

[1]For a more complete discussion, see Gum, et al., 1973.

comparisons for 4 stimuli and 55 paired comparisons for 5 stimuli and thus commonly evokes irreverent comments from subjects such as "redundant, boring, and idiotic." Another shortcoming of the Thurstone technique lies in its basic premise, the additive assumption. Although Thurstone does restrict such linearity to small composites of stimuli (i.e., AB or ABC), the assumption in certain cases must be regarded as questionable, and therefore the basing of a weighted preference system on such an assumption must be considered extremely problematical. On these grounds, the requirement of considering an unwieldy number of paired comparisons and the adherence to the questionable additive assumption, the Thurstone Paired Comparison Test with Rational Origin Assumption was rejected from further consideration.

Comrey Paired Allocation Test. The Comrey Paired Allocation Test utilizes the traditional paired-comparison technique, but requires more than a mere preference judgment between a pair of stimuli. Subjects are also asked to allocate 100 points, votes, dollars, etc., between stimuli pairs, and therefore, not only is simple preference information conveyed, but also the amount and ratio that one stimuli is preferred to another. The Comrey technique requires that all possible pairs of stimuli (e.g., 6 stimuli yields 15 pairs) be considered in the previous manner. Through an information processing algorithm developed by Comrey, such paired allocation data is transformed to a ratio scale of measurement.

The validity of this transformation was tested in a ratio estimation experiment of physical line length. In this experiment, 15 students and faculty at the University of Arizona were presented with pairs of lines of different length and asked to allocate 100 points between the lines in each pair to convey the ratio of the perceived line lengths. For example, if a 3 to 1 ratio was perceived, the appropriate allocation would be 75-25. The responses were processed using Comrey's scaling algorithm, and the resulting measurement scale of perceived ratios of line length correlated highly (correlation coefficient (r) = .975) with the scale of the ratios of true line lengths.

The implicit nature of ratio judgments in the Comrey test itself and the validation of the technique by achieving a very high correlation between judged physical ratios and actual physical ratios confirmed the capability of the Comrey Paired Allocation Test to achieve ratio scales of perceived, physical entities. However, it was realized that the adoption of such a method to obtain ratio measurement of a more subjective entity, preference, required the acceptance of the assumptions previously mentioned in regard to direct scaling methods.

Further testing of the Comrey Paired Allocation Test included an application of the method to groups of sub-goals from two of the Strawman's prime goals and inclusion in a Pre-Test questionnaire which was sent to 200 Arizonans. Although the general response to the questionnaire was relatively good (31% returned), definite practical and theoretical problems were readily apparent. The primary impediment of Comrey's technique is the requirement that every stimulus be compared with every other stimulus. The necessity of considering 15 paired comparisons for 6 stimuli is burdensome even for the most enthusiastic test-taker. Another problem inherent to all tests with a paired comparison format is the possibility of the appearance of intransitivities. The viability of paired comparison scaling methods rests on the premise that the subject's preferences are transitive. That is, if A is preferred to B, and B is preferred to C, then A is preferred to C. However, upon analysis of the scales of preference, as obtained by the Comrey test in the Pre-Test questionnaires, it was realized that highly anomalous results were occurring in many of the individual scores for the six stimuli (15 pair) questions. Further examination of the actual point allocations revealed that these anomalous results were due to the presence of intransitivities. That is, A was preferred to B, B was preferred to C, but C was preferred to A. A review of previous studies on intransitivity (Davis, 1958) revealed that there is no conclusive evidence for the existence of stable intransitivities, intransitivities that occur repeatedly with the same stimuli and observer. Therefore, based on these seemingly uncorrectable enigmas, the required length of the Comrey Paired Allocation Test and its apparent tendency to produce intransitivities, this test was rejected from further consideration as an acceptable weighting methodology.

Metfessel General Allocation Test. Like the preceding method, the Metfessel General Allocation Test utilizes as its basis the assignment of 100 points, and therefore, as Metfessel states, the subject "either actually or symbolically manipulates units of the ratio scale of cardinal numbers, so that his manipulation of the cardinal numbers expresses his judgments of quantitative relations among the items on a given dimension." (Metfessel, 1947, p. 230). More specifically, subjects are asked to distribute or allocate 100 points not just between two stimuli, as in the Comrey test, but simultaneously between all stimuli in question. The resulting assigned numbers serve as immediate sources of the ratio values between the psychological magnitudes corresponding to all stimuli.

Like our experiment with the Comrey test, a similar experiment using line length was conducted with Metfessel's test. The subjects were asked to allocate 100 points among five lines to express the apparent ratios between them. The results were of better quality than the ratios achieved by Comrey's test and yielded a correlation coefficient of .997 to the true line ratios.

Given the apparent ability of Metfessel's test to accurately portray perceptual ratios of physical phenomena and the acceptance of the assumptions of direct scaling methods, Metfessel's General Allocation Test was further tested by applying it to the dendritic structure of a Strawman prime goal in a section of the previously discussed Pre-Test questionnaire. The results of this portion of the questionnaire were encouraging. Respondents could apparently allocate 100 points simultaneously between a number of stimuli to express their desire for an improvement in each stimulus. It appeared that this method, while immediately yielding ratio relationships, was also very economical in terms of a subject's time and effort and thus well suited to the concise style required of mailed questionnaires. The scaling algorithm required with Metfessel's test is very direct for achieving individual ratio-scaled preferences and simply requires calculating the mean for group preference scales. However, this technique is not without minor shortcomings. First, as Metfessel himself notes, the method does require a fair degree of arithmetical sophistication of the subjects. The problem of respondent's allocations not summing to 100 was readily apparent in Pre-Test questionnaires. Other problems, which could be discerned in the Pre-Test and other experiments with Metfessel's technique, were the apparent duplication of an individual's pattern of dividing points among the same number of stimuli and also the apparent tendency of respondents to be more concerned about having their point allocations add to 100 than allocating points to accurately portray their preferences. However, it appears that if some of Metfessel's instructional aids, such as ranking stimuli before allocating points, are added, the occurrence of these problems can be minimized. Therefore, although there are shortcomings and assumptions that must be considered in the use of this technique, the Metfessel General Allocation Test appears to be, given the theoretical and practical constraints previously mentioned, the most acceptable weighting methodology for inclusion into the Strawman's planning framework.

APPLICATION OF THE WEIGHTING METHODOLOGY: ARIZONA AND OREGON

Based on the acceptance of the Metfessel General Allocation Test as an appropriate weighting methodology and the desire to query the general public via mail questionnaire in regard to improvements in sub-goals and goals as specified in the "Cognitive Strawman," a demonstration of the proposed weighting methodology was undertaken in two states, Arizona and Oregon. These states were deemed suitable sites for such a "test" because, in addition to providing different physical settings, it was thought that an analysis of the resulting weights could lead to some valuable insights into the apparent environmental policy divergence between Arizona and Oregon and could perhaps answer the following hypothetical question: "Are the desires of the people in two states different or are the governmental officials' desires different in regard to state policy concerning protection of the environment?"

Therefore, a questionnaire was developed which included the Metfessel General Allocation Test applied to groups of sub-goals in the dendritic structures of three prime goals, Recreation Opportunity, Aesthetic Opportunity, and Economic Opportunity, and a sub-goal of Collective Security, Health Security (see example below). The test was also applied to all three prime goals and Health Security simultaneously.

Example: Distribute 100 points to indicate your desire for improvements in the following aspects of air aesthetics: Visibility (the distance you can see), Odor, and Eye Irritants (eye discomfort caused by airborne substances).

Visibility	_____
Odor	_____
Eye Irritants	_____
Sum	100

A biographical section was included in the questionnaire to obtain socio-demographic data which was thought to bear relevance in regard to later analysis of the weights. More specifically, the questionnaire contained questions relating to age, sex, residence, race, years of education, occupation, employment sector, family income, political affiliation, and self-rating questions regarding environmental knowledge and environmental activity. A rating of ten state problems, which were thought to be appropriate to the areas under study, was also included.

To achieve a representative public weighting of the elements in the specified goal hierarchies, a random sample of individuals in the two test states was desired. Therefore, a random smaple (N=2500) of Arizonans was obtained from a private marketing firm, while the Oregon Department of Motor Vehicles provided co-workers at Oregon State University with a random sample (N=2000) of individuals in Oregon. Questionnaires, preceded by a "warm-up" letter and followed by a "reminder" letter, were then sent to individuals in these random samples.

PRELIMINARY RESULTS

After a 3½ week questionnaire-return period, the return rate of usable Arizona and Oregon questionnaires was 18% and 13% respectively. Analysis of the socio-demographic data revealed that individuals responding to the questionnaire were on the average much alike in both states and could be characterized as being more educated and having higher incomes than individuals drawn randomly from the two states. In regard to the interest concerning similarities or differences in the desires of Arizonans and Oregonians with respect to the protection of their environment, the following three indices were considered: the rating of state problems, the results of the environmental activity and knowledge questions in the biographical section, and the mean sub-goal and goal weights.

As can be discerned in Table 1, the individuals in the responding samples of the two states rated the seriousness of these ten problems in a very similar manner. More precisely, a comparison of the average rating scores for the two states yielded a correlation coefficient of .907 . Table 1 also includes the resulting rank of each state problem by state. An analysis of the ranking differences between the two states produced a Spearman Rank Correlation Coefficient of .952 (significant at .001).

TABLE 1

Average Rating and Ranking of State Problems by State

	Normalized Rating Scale (Σ=100)		Rank Order	
	Arizona	Oregon	Arizona	Oregon
Uncontrolled Growth	10.52	10.62	5	4
Water and Air Pollution	12.63	13.53	1	1
Taxes	12.23	13.47	2	2
Flood Control	8.06	5.83	8	10
Crime	11.99	10.88	3	3
Employment and Wages	9.50	9.68	7	7
Water Conservation	9.77	9.97	6	6
Drugs	10.54	10.13	4	5
Transportation	7.34	7.65	10	9
Welfare System	7.42	8.25	9	8

To determine if individuals in the Oregon responding sample rated themselves significantly higher on the environmental activity and environmental knowledge scales than responding Arizonans, a Chi-Squared Test at the 5% level of significance was applied to the results of these two questions in the biographical section of the questionnaire. The results of the test revealed that the state in which an individual resides and the environmental activity or knowledge levels are independent. That is, the responding individuals in Arizona and Oregon are not significantly different in the way they rate themselves concerning their environmental activity or knowledge.

Table 2, a comparison of the mean weights by state, reveals that only 5 of the 18 sub-goal groups have a different rank order for the two states and that the mean weights of Arizona and Oregon on the whole are not significantly different. More precisely, the results of a Student's T-Test at the 5% level of significance demonstrated that of the 61 sub-goals and goals weighted, only 18 of the mean weights are significantly different between the two states. In regard to some of the specific differences, it appears that responding Oregonians were more concerned with improvements in their leisure time,

TABLE 2

MEAN SUB-GOAL AND GOAL WEIGHTS BY STATE

	Arizona	Oregon
Recreation		
*Leisure Time	45.06	49.74†
Income	54.94	49.50†
Access	31.69	26.67†
Admission Cost	24.17	24.08
Capacity of Recreation Activities	41.14	47.72
Facilities	47.50	44.05
Ability to Recreate	52.50	54.80
*Camping	21.02	26.19†
Fishing	19.55	18.04
Hunting	11.50	12.95
Swimming	15.09	14.77
Boating	12.84	11.32
Picnicking	20.02	15.60†
Quality of Recreation Activity	38.82	35.48
Scenic Aesthetics	61.18	64.14
Supply and Ability to Use Recreation Facilities	29.50	29.98
Quality	25.88	24.29
Variety	18.68	19.25
Equality of Opportunity	25.94	25.72
Aesthetics		
Intermittent Sound	43.68	41.89
Background Sound	56.32	56.59
Visibility	43.31	37.78†
Odor	25.78	30.48†
Eye Irritants	30.91	30.59
*Clarity	38.68	35.09†
Odor	25.63	27.22
Floating Objects	35.69	37.31
*Urban	37.00	39.97
Mountain	10.70	8.13†
Desert	11.43	6.57†
Agricultural	9.51	10.54
Forest	10.88	12.25
Water	20.48	21.77

*Different sub-goal group rank order for the two states.

†Mean Weights significantly different between the two states at 5 percent level.

Table 2. Continued

	Arizona	Oregon
Aesthetics (Continued)		
Population	42.46	42.53
Variety	28.13	28.83
Location	29.40	25.60†
*Air	26.27	22.82†
Water	19.37	23.50†
Landscape	15.04	14.32
Biota	11.80	11.97
Sound	11.78	13.24
Equality of Opportunity	15.75	13.78
Health		
Decrease in the presence of Health Hazards	53.82	53.10
Number of Medical Facilities and Personnel	46.18	46.14
Economics		
Prices	50.35	45.78†
Quality	30.80	32.06
Selection	18.85	21.77†
Income Level	29.96	26.73†
Consumption of Goods and Services	16.51	16.62
Leisure Time	17.11	20.45†
Stability of the Economy	36.41	36.20
Potential for Future Employment	27.20	28.55
Potential for Savings and Investments	36.97	36.28
Potential of Retirement Plans	35.83	34.79
Present Standard of Living	30.58	30.40
Future Standard of Living	37.21	37.50
Equality of Opportunity	32.21	31.72
Recreation	18.13	19.08
Health	28.84	25.68†
Aesthetics	20.52	20.81
Economics	32.51	34.43

*Different sub-goal group rank order for the two states.

†Mean Weights significantly different between the two states at 5 percent level.

while responding Arizonans desired greater improvements in their income level. Weighting differences which are seemingly explained by physical divergences between the two states include the following: the greater Arizona concern for improvements in the air, mountains, and desert, the greater concern of responding Oregonians for improving the condition of their waters, and the significant lack of Oregon concern for improvements in mountain and desert areas.

CONCLUSIONS

In a general sense, the Arizona and Oregon mean weights are identical. This evidence for the congruity of desires for improvements in these goal areas is further substantiated by the similarities revealed in the rating of state problems and the environmental activity and knowledge levels. Thus, given the limited size of the responding sample and the realization that perhaps the true sources of divergent public desire were not measured or for some reason could not be measured, the tentative conclusion, based on preliminary results, must be that Oregon's image of being more progressive than Arizona with respect to environmental concern and action is apparently the result of divergences in state leadership and/or interest group orientation.

In regard to the viability of the weighted preference methodology, the Metfessel General Allocation Test, the Arizona-Oregon demonstration has shown that this weighting technique can be done by the general public and is appropriate for a mail survey. Although the return rate of the questionnaires was adequate for a demonstration of the methodology, it must be noted that for the weighting methodology to be truly useful in an actual planning situation, more intensive follow-up techniques and other efforts, such as interviews, must be undertaken to insure greater and more balanced input from all segments of the population.

REFERENCES CITED

Comrey, Andrew L., "A Proposed Method for Absolute Ratio Scaling," Psychometrika, Vol. 15, No. 3, September, 1950, pp. 317-325.

Daniel, Terry C., Lawrence Wheeler, Ronald Boster, and Paul R. Best, Jr., "Quantitative Evaluation of Landscape: A Preliminary Application of Signal Detection Analyses to Selected Forest Management Alternatives," unpublished manuscript, 1971.

Davis, J. M., "The Transitivity of Preferences," Behavioral Science, Vol. 3, No. 1, 1958, pp. 26-32.

Eckenrode, Robert T., "Weighting Multiple Criteria," Management Science, Vol. 12, No. 3, November 1965, pp. 180-191.

Ekman, G. and Lennart Sjöberg, "Scaling," Annual Review of Psychology, Vol. 16, 1965, pp. 451-474.

Gum, R. L., T. G. Roefs, M. D. Bradley, R. M. Judge, D. B. Kimball, and W. W. Wilson, "Cognitive Strawman: Public Input to the Water Resources Planning and Evaluation Process," Report to the Technical Committee of the Water Resources Centers of the Thirteen Western States, unpublished, 1973.

Metfessel, Milton, "A Proposal for Quantitative Reporting of Comparative Judgments," The Journal of Psychology, Vol. 24, 1947, pp. 229-235.

Peterson, D. F., et al., "Water Resources Planning and Social Goals: Conceptualization Toward a New Methodology," Utah Water Research Laboratory Publication, PRWG-04-1, September, 1971.

Schimpeler, C. L., "A Decision-Theoretical Approach to Weighting Community Development Criteria and Evaluating Alternative Plans," Ph.D. thesis, Purdue University, 1967.

Stevens, S. S., "Measurement, Psychophysics, and Utility," Measurement: Definitions and Theories, eds. C. W. Churchman,and P. Ratoosh, John Wiley and Sons, Inc., New York, 1959, pp. 18-61.

__________, "On the Operation Known as Judgment," American Scientist, Vol. 54, No. 4, 1966, pp. 385-401.

__________, "On the Theory of Scales of Measurement," Science, Vol. 103, 1946, pp. 677-680.

Thorndike, E. L., Measurement of Educational Products Seventeenth Yearbook of the National Society for the Study of Education, Public School Publishing Company, Bloomington, Illinois, 1918.

Thurstone, L. L., "The Law of Comparative Judgment," The Measurement of Values, ed. L. L. Thurstone, University of Chicago Press, Chicago, Illinois, 1959, pp. 34-49.

__________, "Psychophysical Analysis," American Journal of Psychology, Vol. 38, 1927, pp. 368-389.

Thurstone, L. L. and Lyle V. Jones, "The Rational Origin for Measuring Subjective Values," Measurement of Values, ed. L. L. Thurstone, University of Chicago Press, Chicago, Illinois, 1959, pp. 145-210.

Wilson, W. W., R. L. Gum, and T. G. Roefs, "The Cognitive Strawman Planning Methodology: Public Input," Hydrology and Water Resources in Arizona and the Southwest, Vol. 3, 1973.