MAGNES[SM]

A RATIO-SCALING TOOL FOR DECISION MAKING

Prepared by
Robert L. Kaplan
President
The Rumson Corporation

Prepared for
Third International Symposium
on the
Analytic Hierarchy Process

11-13 July 1994

The George Washington University
Department of Management Science

MAGNES[sm]

A RATIO-SCALING TOOL FOR DECISION-MAKING

I. ABSTRACT

Because the computer is such a powerful tool, analysts are bombarded, continuously, with the new sophisticated, often exotic, tools (algorithms, models, etc.) ostensibly designed to improve the complex problems that characterize our modern society. Amost all incorporate some form of quantification as input.

Careful examination of the tools, however, reveals that often the type of quantification used receives only cursory attention, if at all. The tool developers seem to take the attitude that "numbers are numbers" (input data, that is) and that importance of data quality is secondary to the algorithm/model development.

For example, data gathered by the Likert or category scaling survey technique is often used as input data when trying to draw comparative conclusions among a series of independent issues. Technically speaking, this is mathematically incorrect in most cases. the Likert Scale is ordinal, whereas comparisons can be made only if a logical common denominator, a mathematical ratio is provided.

If then, the foundation (the data) is weak, how can the structure (the model) be sound?

MAGNES[sm] has evolved as a methodology to provide a more substantive basis of quantification for such tools. MAGNES is a ratio-like , psychometric scaling technique based on the same methods used to develop the decibel scale for measuring the intensity of sound. MAGNES has several outstanding characteristics:

- The ability to quantify in mathematically robust form, qualitative, subjective value judgments
- Merge qualitative and quantitative issues on the same scale
- Compare with mathematical precision, highly dissimilar issues (the classic "apples and oranges") on the same scale

The paper will describe the philosophy of MAGNES, how it functions, and how it can be applied.

[sm] MAGNES is the registered Servicemark of The Rumson Corporation

I. PURPOSE AND SCOPE

The purpose of this paper is to describe a mathematically robust tool for assisting decision makers in coming to grips with seemingly insoluable problems so prevalent in our society. The tool or technique is called MAGNES^{sm}, an acronym for magnitude-estimation scaling.

The paper will discuss the problems faced by the decision makers in coping with modern day issues, the characteristics and capabilities of MAGNES as a solution, how MAGNES works, and how MAGNES can be applied.

Space does not permit the presentation of a "how to" text. It is the author's contention that facility with MAGNES-like techniques can only be gained by reading and repeated experimentation.

II. DECISION-MAKING PROBLEMS

A. General Difficulties

Today's social and political climate tends to be fraught with complex issues that defy resolution by traditional analysis. One of the primary impediments normally is the qualitative nature of the many critical factors.

Yet, it is a commonly accepted fact that humans prefer to make decisions based on quantitative guidance. Numbers seem to placate one's moral conscience and ease the dilemma of dealing with almost dimensionless, qualitative issues.

Another problem plaguing the decision process is the usual lack unanimity among those purporting knowledge of the problem at hand. As one wag was quoted, facetiously, "If you have 12 experts in a room, you are liable to get 13, if not 26, solutions".

If by chance, the problem(s) under examination can be described as a series of relevant factors, i.e., a "shopping list", intuition tells the decision maker that some issues are far more important than others. The questions then arise: "Which ones (issues)?", "How much more (are some than others)?" These questions are not trivial since there seldom are sufficient resources available to address all problems for all people. A sense of relative priority, obviously, is needed by the decision maker.

Still another problem assails the decision maker. Examinations of the "shopping lists" invariably indicate that the issues are characteristically or generically dissimilar in nature, the proverbial "apples and oranges" if you will. How can the issues be compared on a rational basis without some form of common denominator?

B. Analytical Inadequacies

The advent of the computer, particularly the high performance PCs, have advanced the formerly limited computational power of man by many orders of magnitude. These capabilities have given rise to the development of sophisticated, if not exotic, mathematical algorithms and models, ostensively designed to solve the complex, modern-day problems. Almost all rely on some form of quantification as input data.

Perhaps the most critical deficiency of the modeling approach is the general inability to incorporate qualitative issues in the equations. The qualitative factors seem to defy quantification, yet some social, political aspects are critical elements of the problem. The qualitative issues tend not to be amenable to quantification; physical laws do not apply, and seldom does robust empirical data exist.

Incorporating amorphous, qualitative issues into an equation in some meaningful, defensible manner long has been a major problem for most analysts. As an alternative some go so far as to ignore the qualitative factors entirely or assume very narrow scenarios simply because the issues cannot be handled in a convincing fashion.

When quantification of some form is avilable, it is curious to note that despite the mathematical beauty and ingenuity expressed in the design of the model, the algorithm developers seem oblivious or tend to neglect the quality of the input data or "grist". Many developers tend to become overly enamored of the elegance of the equation while forgetting that the model is only as good as the data upon which it operates. Their attitude seems to be "Numbers are numbers" and that seems to be sufficient qualification for use in the model.

This attitude is exemplified by the concept of "quasi quantification". "Quasi quantification" may be defined as the distinction between arithmetic and mathematics. Arithmetic is the mechanical manipulation of numbers, e.g., one (1) plus one (1) equals two (2); whereas, mathematics implies logic. If one (1) represents an apple and another one (1) represents an orange, the two cannot be combined nor compared unless a relationship, i.e., a logic, between the apple and orange is established.

To illustrate the quasi-quantification issue, perhaps the most popular source of numbers is examined, the survey technique. Surveys are used to:

- Provide quantification
- Transpose, ostensively, qualitative issues into quantification.

Survey respondents tend to be individuals who have some knowledge of the subject matter or have formed opinions, so-called "experts". The respondent may be anyone - scientist or average citizen - who might have developed a position regarding the problem.

The two most popular survey techniques employed are the "Question and Answer" (Q&A) and the Category or Likert Scaling method. While powerful tools when applied correctly, each has significant, yet little understood, limitations that can seriously distort the intent and output of the model. The limitations can be classified under the title of quasi quantification.

For example, Question No. 1 may have a response rate of 70 per cent positive, while question No. 2, only 35 per cent. An untuitored interpretation of the data is that No. 1 is twice as significant as No. 2. This is an improper, inaccurate assessment. Question No. 1 may be addressing a relatively trivial issue, whereas Question No. 2 may, literally, have earth-shaking consequences. The two questions are discrete and cannot be compared.

All the percentages have indicated is relative popularity, not intensity or significance of the issues being addressed. No. 1 is an apple and No. 2, an orange. The two responses cannot be compared unless a logic is supplied, i.e., No.1 is "x" times greater or less in significance than No. 2.

The Likert Scale suffers similarly. Several dynamics are in effect. First, the Likert Scale is ordinal and like the Q&A, the responses between any two (2) items cannot be compared, mathematically, unless a common denominator is provided. (Ref. 1) Without a common denominator, the logic for combining or comparing simply does not exist.

Secondly, the Likert Scale places the respondent in an artificial and perhaps insensitive framework (1 to 5, 1 to 10, etc.). The scaling range implies that nothing can be more or less significant than the arbitrarily established limits. In effect, the limits dampen the significance of the most important issues and inflate the least. The net result is a distortion of meaning.

When such quantification is used as input data in any algorithm or model, regardless of mathematical elegance in design, the results are bound to be less than substantial. If the foundation is weak, how can the structure be sound?

III. MAGNES^{sm}, A SOLUTION

A. Characteristics

MAGNES is a psychometric survey technique, psychometrics being science of measuring human response to stimuli. Historically, most of the research has concentrated on stabile, physical stimuli such as sound, color, taste, smell, etc. The decibel scale for measuring the intensity of sound evolved from psychophysical experimentation.

MAGNES, on the other hand, focuses primarily on non-stabile, qualitative stimuli such as social and political issues.

MAGNES is a ratio-scaling technique, the highest order of scaling. (Ref. 2) In other words, inherent in the technique is the capability for comparing highly dissimilar issues or factors (the classic "apples and oranges"), quantitatively, on the same scale. MAGNES is mathematically based, therefore, a logic for comparison is present.

In application, a MAGNES survey is used to:

- Transpose qualitative issues into quantification
- Merge qualitative and quantified issues on the same scale for comparison.
- Measure the relative intensity or significance of issues on the same scale, regardless of dissimilarities.
- Synthesize the disparate opinions/attitudes of respondents without bias or unproductive debate.

The primary output of a MAGNES survey is a series of weights or weighted priorities for each issue attendant to the overall problem.

Figure No. 1 is a typical example. The vertical axis or ordinate is an open-ended scale that represents, quantitatively, the relative significance of a series of issues under examination. The abscissa or horizontal axis lists by code number the issues in descending order of significance as perceived by (in this example) 462 respondents. Issue No. 6 has been rated, collectively, 14.4 times more significant than Issue No. 46, about two (2) times more significant than those rated 7.2. A issue rated 4 is one-third one rated 12 and one-half one rated 8. Similar ratios can be established between any two issue-pairs.

The shape of the distribution of weights, not a curve in the traditional sense, is typical of MAGNES results. It indicates that a few issues are far more significant than others. The lowest rated issues do not imply that they are of no significance - just of those examined, they are the lowest.

Similar figures can be developed for any viable, demographic subgroup of respondents, e.g., by gender, age, location, education. "Viable" refers to the fact that at least 20-25 respondents are required for valid stratification.

As mentioned above, earlier research in magnitude-estimation scaling, traditionally, has focused on stabile stimuli such as sound. By varying the intensity, a characteristic curve for each stimulus, "x", could be described mathematically as:

$$y=ax^b \text{ (Ref. 3)}$$

This equation, described as the "power law", represents the impressions of respondents when exposed to stimuli.

FIGURE 1

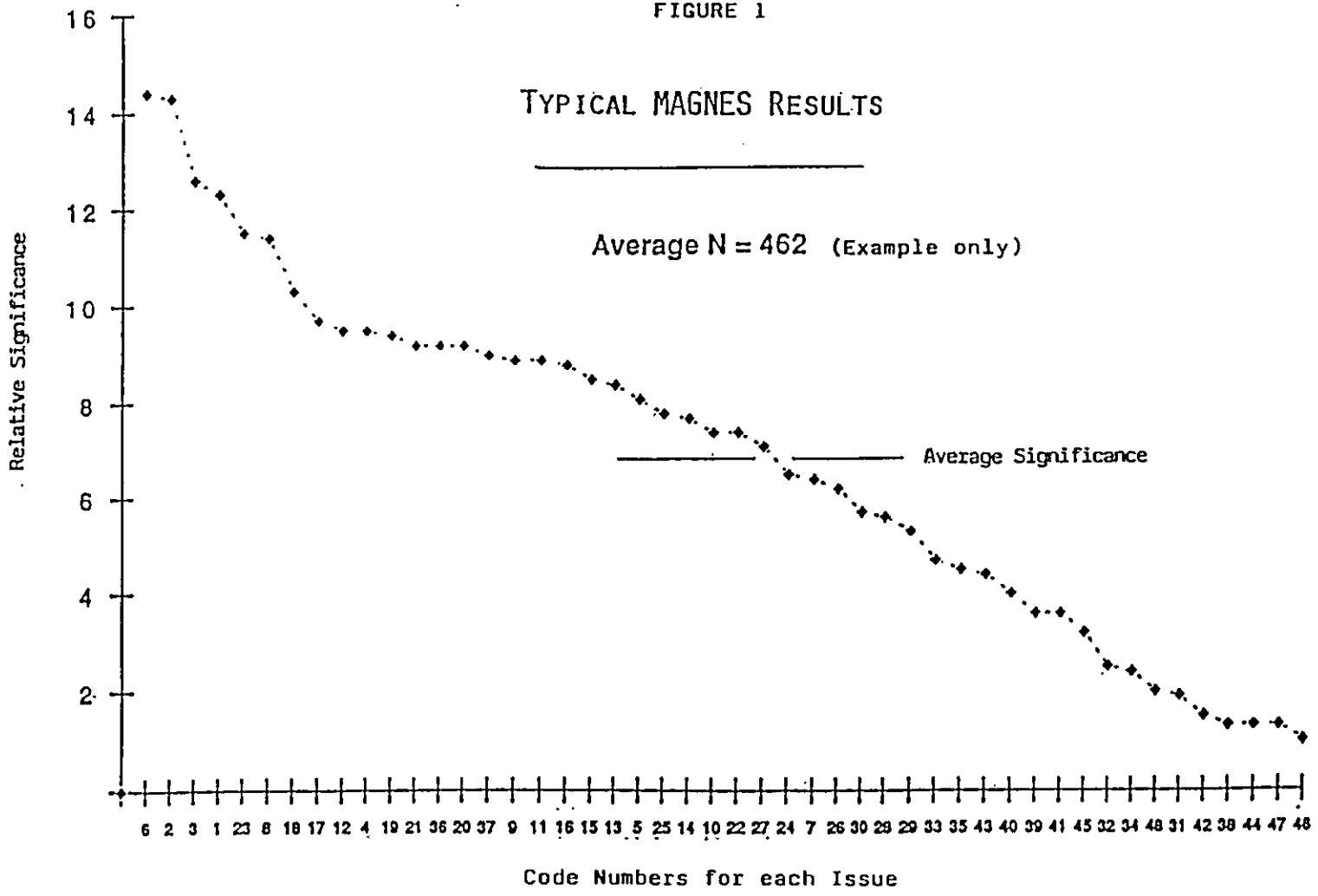

TYPICAL MAGNES RESULTS
Average N = 462 (Example only)
Relative Significance
16
14
12
10
8
6
4
2
Average Significance
6 2 3 1 23 8 18 17 12 4 19 21 36 20 37 9 11 16 15 13 5 25 14 10 22 27 24 7 26 30 28 29 33 35 43 40 39 41 45 32 34 48 31 42 38 44 47 46
Code Numbers for each Issue

Figure 2 depicts families of power-law curves actually defined by MAGNES surveys. The ordinate depicts the relative "desirability" of a series of educational assistance (G.I. Bill type) and cash bonus packages as a function of years of required enlistment,the abscissa. The respondents were high school students with basically very high I.Q.s.

B. Procedural Steps

To better understand MAGNES, the procedures for developing and conducting a survey are discussed briefly.

1. Task A: Issue Identification

Perhaps the most critical step in conducting a MAGNES survey is the identification and definition of the issues attendant to the problem of interest. The task is accomplished in a variety of ways, but the following is more or less standard operating procedure.

Indepth discussions are held with the client to define with precision, the requirements and current understanding of the issues. This effort may include a document search and review, but more often than not, there is an absence of available information.

The most useful means of defining the problem has proven to be the conduct of personal interviews and/or focus-group interviews with individuals who have interest or are affected by the problem in question. This is purely a qualitative exercise. The researcher is interested only in generating opinions, pros and cons, and ideas, not numbers.

It is essential that representatives from all perspectives of problem be interviewed. The researcher should try to generate contravening ideas and points of view. In this regard, he/she must exercise extreme objectivity by assuming there is no right nor wrong in what the interviewees say. The researcher must be sensitive to polarity and all interim perspectives.

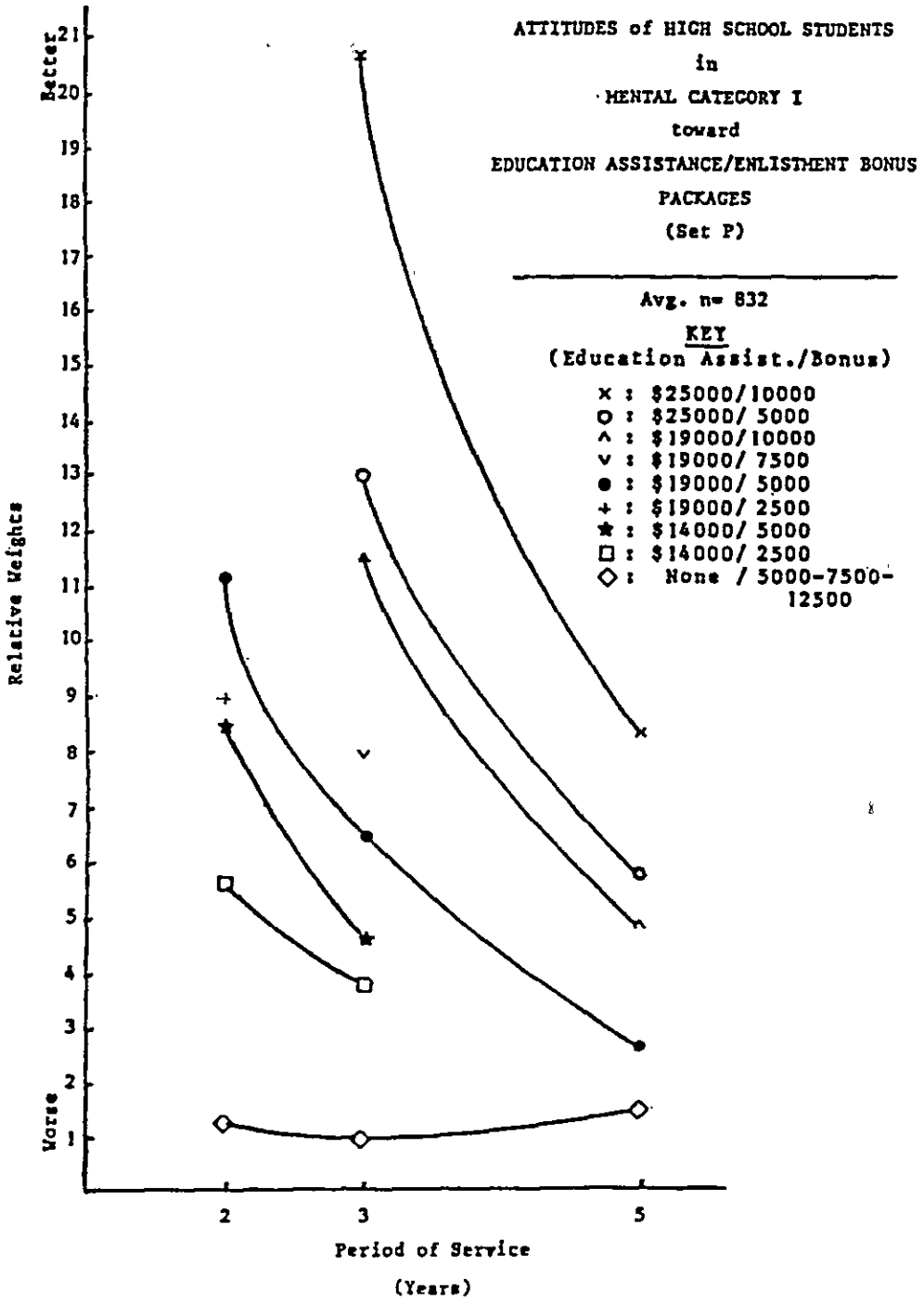
The interview process should be structured to include representation from all factions and locales. Sampling is not a factor but qualitative coverage is important to insure that all facets of the problem have been identified.

The Rumson Corporation utilizes a qualitative, factor analysis technique to filter the information gained from all sources. Irrelevant issues are discarded. The result is a synthesized list of pertinent factors, both pro and con, that outline the problem.

Depending on the problem, it should be possible to generate a list not exceeding 50 issues; the less the better.

Normally, the synthesized list is reviewed by the client. The researcher, however, must not allow client biases to interfere

FIGURE 2


ATTITUDES of HIGH SCHOOL STUDENTS
in
MENTAL CATEGORY I
toward
EDUCATION ASSISTANCE/ENLISTMENT BONUS
PACKAGES
(Set P)
Avg. n= 832
KEY
(Education Assist./Bonus)
× : $25000/10000
○ : $25000/ 5000
^ : $19000/10000
v : $19000/ 7500
● : $19000/ 5000
+ : $19000/ 2500
★ : $14000/ 5000
□ : $14000/ 2500
◇ : None / 5000-7500-12500
Better
Worse
Relative Weights
21
20
19
18
17
16
15
14
13
12
11
10
9
8
7
6
5
4
3
2
1
2
3
5
Period of Service
(Years)

with the finalization of the list. Objectivity is a must!

2. Task B: Survey Instrument Development

Four (4) main elements form the survey instrument:

- Anonymous, demographic-profile questionnaire used for establishing the credentials of the respondent to participate, and to permit stratification of subgroup responses according to such parameters as age, gender, occupation, location.[1]

- Respondent context that describes the objectives and frame of reference with respect to the problem in which to respond.

- Instructions in verbal and graphic form

- Randomized listing of issues on one or two pages, depending on the number identified

With respect to the instructions, the respondent is provided with a reference issue, selected at random, that is assigned a value of 15, arbitrarily. [Note: The value of 15 has been found to avoid a host of methodological problems too detailed for purposes of this paper.] The respondent is then asked to compare each subsequent issue on the list in turn to the reference issue only. The issue in question may be equal, more, or less significant to the respondent than the reference. The respondent has the numerical range, theoretically, from the infinitesimal to the infinite available to express his/her sense of value; although most are fairly conservative in their selection of values.

It is this simple comparison using a fixed common denominator, the reference issue, that establishes the ratio-like scale and permits the mathematically legitimate comparison of the "apples and oranges" on the same scale.

3. Task C: Selection of Respondents

This exercise depends on the nature of the problem under examination. Standard texts on survey development may be used for reference. (example: Ref. 5)

It should be added that in the past, MAGNES surveys have been administered to a broad range of respondents from high school students to geriatrics, Army privates to generals, social service workers to law enforcement officials, etc. The instructions and language used throughout, therefore, must be linguistically tailored to be read and understood by the lowest common denominator of the

[1] Identification of the respondent by name should be avoided to reduce the possibility of inhibiting the response.

respondent target group.

Sampling may or may not be used. In drawing the sample, however, the researcher must base the estimate of sample size on the degree of stratification desired ("viability" issue) and expected response rate.

4. Task D: Polling

When possible, the conduct of the survey or polling should be conducted at gatherings such as conventions or meetings. This permits the researcher to review the instructions and control the retrieval of the completed surveys. This procedure is not always practical nor realistic.

Mailing is a viable alternative, however, a sizable non-response normally can be expected.

5. Task E: Computation and Analysis

Each completed response is subjected to a quality-assurance screening before being entered into the data base. Procedural errors, excessive omissions and repetitions, failure to complete the demographic questionnaire, are typical reasons for rejecting the response in entirety.

Proprietary MAGNES software is used to facilitate the process.

The raw responses are converted to logarithms. The geometric mean or arithmetic mean of the logarithms is used in computing the raw average score or weight for a given issue. Since the geometric mean in effect is taking the "Nth" root of the repsonses, the numerical damping is significant. The occasional aberrant response, i.e., ones abnormally large or small, therefore, have little impact.

It is important at this point, however, to be aware of two (2) methodological limitations in applying MAGNES:

- The first has to do with the aforementioned "viability" issue. As mentioned above, with "N" respondents, the geometric mean calculates the "Nth" root of the product of raw scores. Twenty (20) or fewer respondents renders the mean vulnerable to distortion by virtue of insufficient damping of the aberrant response. The effect is minimized by requiring a desired minimum of 25 respondents in a viable cell or subgroup.

- A second limitation applies when comparing the responses between any subgroup pairs. Only the rank orders for each subgroup can be examined. The Spearman's "rho" correlation coeficient is used. Magnitudes cannot be compared because there is an inverse effect, i.e., the smaller the subgroup population, the larger the the magnitudes. For this reason, distribution-free statistical

methods are used to identify significant differences in perception of value between subgroup pairs. (Ref. 6)

One feature of MAGNES output should be borne in mind. The ratio-like scaling permits the mathematically legitimate addition, subtraction, multiplication, and division of weights, a feature not found in any other popular survey technique.

Another key characteristic should be noted. Although theoretically providing the same mathematically rigorous results, the "paired comparisons" technique requires the respondent to make $q(q-1)/2$ value judgments. A survey of 25 issues thus requires the respondent to make 300 judgments, obviously an unacceptable burden which few will willingly entertain.

A MAGNES survey, on the other hand, requires only 24 judgements.

Products of the computation include:

- Demographic profiles of the entire respondent pool
- Quantitatively weighted priorities for each issue representing the collective attitudes of the entire respondent pool
- Weighted priorities for each issue as perceived by selected, viable subgroups
- Comparisons between subgroup pairs identifying specific issues in sharp disagrrement
- Demographic profiles of selected, viable subgroups.

6. Task F: System Development

Depending on the objectives of the research, it sometimes is possible to develop a system for using the findings of the survey on a recurring basis. For example, assessment/evaluation systems have been developed employing predetermined criteria that have been weighted before the fact by a MAGNES survey.

7. Task G: Reporting (self explanatory)

IV. APPLICATIONS

Past MAGNES applications include:

- Monitoring the intensity of military operations
- Setting R&D priorities

- Articulating user requirements for new products or product improvements
- Developing assessment/evaluation systems for complex issues, e.g, personnel selection and performance rating, proposals for high technology R&D programs
- Measuring financial concerns of banking customers
- Establishing training priorities

Possible applications include:

- Customer satisfaction monitoring
- Forecasting
- Political campaign strategy development
- Input parameters for mathematical modeling

The mind is the only inhibitor in applying MAGNES.

REFERENCES

1. Lodge, M., Magnitude Scaling, Quantitaive Measurement of Opinions, Series/No. 07-25, Sage Publications, London, 1981.
2. Guilford, J.P., Fundamental Statistics in Psychology and Education, McGraw-Hill Book Company, New York, 1956.
3. Stevens, S.S., Psychophysics, J. Wiley & Sons, New York, 1976.
4. Kaplan, R.L. w/ P.T. Harris, The Measurement of High School Students' Attitudes toward Recruiting Incentives, The Rumson Corporation, McLean, VA. 1985.
5. Moser, C.A. & G. Kalton, Survey Methods in Social Investigation Basic Books, Inc., New York, 1972.
6. Maritz, J.P., Distribution-Free Statistical Methods, Chapman and Hall, London, 1984.

[NOTE]

See Reference No. 3 for extensive bibliography.