A New Approach in Requirements Elicitation Analysis

William J. Tastle
Dept of Management, School of Business, Ithaca College
Ithaca, New York, USA
tastle@ithaca.edu

Amjad Abdullat
Dept of Computer Information Systems, West Texas A&M University
Amarillo, Texas, USA
aabdullat@wtamu.edu

Mark J. Wierman
Dept of Computer Science, Creighton University
Omaha, Nebraska, USA
wierman@creighton.edu

***Abstract* — In requirements analysis the task of elicitation of stakeholder need has been a continuing source of error and frustration in systems development. To aid in the acquisition of a set of proper needs that are critical to the design of an effective system, the systems analyst is provided with a new tool to assist in determining when group consensus has been met with respect to the identification of one or more needs. A recently developed measurement tool for measuring subjective concepts like consensus, agreement, and dissent is described. Categorical data are frequently collected using an ordinal scale such as the Likert scale and a new method is available that gives the analyst a different perspective of group-think. The agreement measure is also extended to an agreement distribution and used to calculate a mathematical distance between two separate agreement distributions. With these measures it is easy to calculate the proximity of agreement between two or more groups of stakeholders. This measure is then applied to requirements analysis.**

Keywords — requirements analysis, requirements elicitation, problem identification, agreement, consensus

I. INTRODUCTION

This paper builds on and extends a recently given paper at the International Conference on Information and Communication Systems in Jordan on December 2009. In the domains of software engineering, systems engineering systems analysis, and systems design, it is commonly acknowledged that difficulties in understanding challenging and complex problems, sometimes even perceived as being rather intractable, usually presage the more interesting efforts of building information systems in organizations. Determining the needs for new systems, improvements to existing systems or for new or altered products, can require the systems analyst to address conflicting requirements from the various stakeholders, each viewing the new (or revised) system from a particular perspective.

Being able to sift through the myriad of differing views to identify the set of necessary conditions, rather than a set of symptoms that actual problems may manifest, is critical to the success of the project. To accomplish this end, analysts are charged with the task of producing a set of requirements that are measurable, testable, actionable, and in sufficient detail such that system design can occur with the delivery of a system that truly does solve business needs. Mistakes made at this investigatory level and propagated through to the project conclusion result in very expensive corrections. Textbooks on systems analysis and design (and software engineering) usually identify requirements analysis as consisting of three types of activities: the eliciting of the requirements, the analysis of those requirements, and methods by which these requirements are recorded and documented.

Since the analysis process can be long, arduous and tempting of the patience of management, it is important to identify all the stakeholders (as well as those who move into, and out of, the organization), and take into account their needs while not detracting from the needs of others. There are many techniques used to elicit information from stakeholders and, in fact, this area of requirements analysis, that of *requirements elicitation*, is an area that transcends the domains of business, psychology, sociology, human relations, organizational behavior, statistics, and other disciplines as well as emerging disciplines. The list of domains from which the tools used in requirements elicitation are derived can be quite long, and can vary from project to project. This area of elicitation is fundamental to the entire systems development area, for an error made at this level of

doi:10.4304/jetwi.2.3.221-231

analysis is the most costly (in terms of money and time) to resolve later in the development process.

Requirements elicitation, sometimes referred to as requirements gathering, is the practice of obtaining requirements from the stakeholders (users, customers, clients, suppliers, other companies, etc). The process of elicitation is non-trivial because it is far more than merely the asking of a few questions. One can never really be sure one has all the requirements in spite of using interviews, questionnaires, personal observation, brainstorming techniques, use cases, prototyping and so forth. One frequently used method is the brainstorming technique in which all ideas, regardless of their accuracy or relevance, are encouraged. Unfortunately, those individuals higher in the pecking order sometimes monopolize the discussions until their view becomes the accepted view of all the stakeholders though there are methods by which this kind of behavior can be controlled (see [29] for examples). Discussions at the brainstorming table frequently lack any kind of measurable rigor for the intent is to encourage all ideas without placing anyone in an adverse position. Thus, ideas that might be shared by a large number of people may not be immediately recognized as belonging to a large constituency of stakeholders but rather in varying degrees of acceptance. We offer a method by which such discussions, as well as the results of questionnaires and surveys, can be quickly analyzed.

II. THEORY OF ELICITATION

The motivation for this study is to identify a function that can be easily computed and containing enough meaning such that all individuals engaged in a brainstorming session can easily express their feelings about any situation, idea, suggestion, plan, diagram, discussion, etc. without embarrassing anyone or causing a rift or conflict in the meeting or with one's colleagues. If an idea is being put forth by a stakeholder, and another stakeholder does not believe that the idea being put forth is relevant, or perhaps only practically relevant, an indication as to the degree of agreement can be determined by means of the recently developed agreement measure [14, 26].

Individuals naturally compartmentalize their thinking into categories or fuzzy numbers [30]. A fuzzy number is a convex, normalized fuzzy set whose membership function is at least segmentally continuous. An entire discipline has evolved around fuzzy sets, numbers and systems as reflected in numerous journals (e.g., Fuzzy Sets and Systems, J of Approximate Reasoning). A physician might ask a patient about the degree of pain associated with an injury by giving a number between 1 and 10. The physician has no expectation that the response from the patient is anything more than a categorical estimate, essentially a fuzzy number.

Humans simply do not have the ability of ranking or evaluating anything along such the continuous number line unless they use some piece of equipment by which their perceptions can be augmented. It is the use of categories that people are comfortable utilizing as in their propensity to agree or disagree, with an issue. Human vocabulary is replete with categorical rankings: cold, warm, tepid, hot; low vs. high; innocent vs. guilty. Thus, stakeholders around a table engaged in dialogue or brainstorming react to ideas in the form of categories: they will strongly agree, moderately agree, moderately disagree, or strongly disagree with what is on the table at the moment. If all the stakeholders are in complete agreement with the idea being presented, then the overall consensus around the table should be 100%. If, on the other hand, the stakeholders are completely split with half being in strong agreement and the other half in strong disagreement, then the overall consensus around the table should be 0%. Every other possible category assignment by the stakeholders would be somewhere between these extremes. Thus an interval exists between 0 and 1, or 0% and 100%. This is simply to understand and offers a quick guide as to the degree to which the group is in consensus or, stated another way, the degree to which the stakeholders are in agreement as to a need the new system must address.

It is not required that the stakeholders agree completely on an idea in order to accept it for preliminary analysis shows that a consensus of 80% approximates 95% significance [28]. What is important is that stakeholders can discuss their perceptions of system needs in an open and non-threatening environment but can also "quietly" indicate their level of support to the current idea.

It is difficult enough to get stakeholders to a table to discuss their perceptions of system inadequacies or new system requirements, let along control the discussions along some particular direction. The systems analyst must engage in this kind of activity and do so in an efficient manner. By using this measurement tool it may be possible to direct, in real time, discussions on ideas that appear to have stronger merit based on a consensus calculation. It is to this concept that the remaining portion of the paper is directed.

III. THE CONNECTION BETWEEN CONSENSUS AND ORDINAL SCALES

The basis for the study of uncertainty, in the sense of imprecision, was first established by Zadeh [24] when he characterized grades of set membership by a function that assigned a value between zero and one to each member. The succeeding papers (many hundreds of them that have since spawned the discipline of fuzzy sets and fuzzy measures (now sometimes referred to as general measures)) have further developed the new discipline, but the overwhelming majority or papers have dealt with the collection of data based on interval and/or ratio scales although laudable efforts were made to connect, for example, category theory and systems theory [2] and fuzzy clustering to ordinal and nominal scales [3].

The original purpose in studying fuzzy sets as they relate to ordinal scales was to establish a connection between group consensus making and a calculable measure. It was intuitively felt that the various forms of fuzzy set and fuzzy measure theory could be the basis for such a measure. The result of the initial meeting was the consensus measure (see below). In the original formulation an “average” was calculated (the mean), but after further study it was decided that the median was more conceptually correct when dealing with ordinal measures than the mean, for an mean calculation (addition and division) is predicated on the values being summed, and then divided by the number of values, properties that properly belonged to the interval (or ratio) scale and was thus inappropriate for the consensus measure.

The difficulty in using interval and ratio scale measures on ordinal scales lies in the use of permissible statistics (see figure 9 in the Appendix). While the median and percentile are permitted (and the mode and chi square, permissible on nominal scales and thus permitted on higher-level scales), the mean, standard deviation, and other statistical measures are limited to at least interval scales and hence not permitted on nominal or ordinal scales. It is unfortunate that means are regularly applied to ordinal scale data, typically in the form of evaluations of Likert measures, no evidence exists to support an existence of a regular interval between ordinal categories, though totally ordered and monotonically increasing; the literature, however, is replete with examples (none are specifically cited for the purpose of propriety) of studies that assume an interval is the absence of evidence [8].

There are several examples below to illustrate the value of using consensus theory measures to evaluate ordinal data.

IV. Typical Methods of Ordinal Scale Analysis

It is typical for researchers to apply the mean to ordinal data, for the available statistical tools are quite limited. There are nonparametric tests (see any introductory statistics text) that can be used to make sense of ordinal data (e. g., Wilcoxon signed ranks test, Spearman test, Gamma coefficient, and the like), but all these methods were devised before the onset of fuzzy set theory, fuzzy measure theory and information-theoretic measures. These are the basis for the mathematics of vagueness and impreciseness, two qualities that are present in most systems investigations undertaken by the systems analyst as one pursues the requirements identification phase. Papers abound (see example below) in information systems (as well as other disciplines) in which Likert data (based on the selection of a category from an ordered sequence such as "strongly agree," "agree," "neutral,", "disagree," and "strongly disagree" which is represented in this paper as SA, A, N, D, and SD, respectively) is presented in terms of a category mean, standard deviation, confidence intervals, t-test, and other such statistics. The authors argue this is equivalent to saying that the average of warm and hot is warm-and-a-half. Ordinal scales possess no inherit interval between categories. Sometimes researchers place category labels on a number scale, i.e., SA = 1, A = 2, etc. to give an impression that they are interval. These are called *Likert-like* measures but are actually ordinal in disguise. To paraphrase an old phrase, painting stripes on a horse does not make it a zebra! We offer a different method that is conceptually sound and mathematically proven [12-19].

V. Measures of Consensus, Dissent and Agreement

A. Concept

The underlying concept behind the measures of consensus and agreement, and their complementary measures of dissent and disagreement, respectively, (disagreement is not discussed in this paper) is centered on the existence of a perceived relative distance between ordered categories (called the intra-categorical distance) that may or may not be equal, and may or may not be similar to the distances in the minds of others, but the distance from one extreme category to the other extreme category is always 100% of whatever the mindset. Hence, from SA to SD the overall intra-categorical distance is 100%; from cold to hot is 100%. Given a stakeholder team or an even number of people, if half select SA and the other half select SD, then the group consensus should be zero, for the group is equally partitioned at their extremes (SA and SD). Similarly, because the group is at maximum opposition the dissent should also be maximized at 100%. Consensus and dissent are measures that characterize the entire set of stakeholders and are thus measures of the collective and are directly related.

This is similar to the Congress of the US in which the principal two parties (Democrats and Republicans) each hold half of the membership. A consensus may never be attained. If one person moves from SD to D, or from SA to A, then the consensus should increase to some value above zero for the group is no longer balanced on the extremes. A consensus does not require 100% agreement, and it is usually the committee chair who must recognize when a consensus has been met in order to move the group on, but how does the committee chair (or the systems analyst) know that point has been reached? This is a matter for the psychologists and sociologists to research, but we can establish a criterion *a priori*. A percentage value that determines the threshold for consensus, from 0 to 100%, should be agreed upon before this analysis is applied.

If is reasonable to assume that a consensus is represented by a super majority, hence 51% probably does not represent a consensus. Clearly a consensus is met when 100% of the participants agree on a single Likert category, be it to agree or disagree with the statement under review. A group could even form a consensus around neutral in the sense that they have come to an agreement that they are all unsure. It is the selection of a number in the gray area between 50% and 100% that is the challenge. The US Senate requires a 60% super majority to pass legislation, and that value could be used to indicate a consensus. A recent study [11] on the establishment of a curriculum for dermatology students has used 80% as an indicator of consensus among the dermatology medical community as to the importance of certain items being in a basic curriculum for dermatology students. Whatever the threshold, the following procedure can be used to determine the degree of consensus (and/or dissent) of the group towards a Likert statement.

B. Measure of Consensus

A consensus measure has been introduced, desirable properties have been proven, and applications have been demonstrated [12-21]. It is starting to receive outside attention.

The measure of consensus, whose form was inspired by the Shannon entropy, requires an ordinal scale, but it can be used with interval and ratio scales, though the standard statistical measures are probably a better choice in those particular cases. The equation for consensus is:

$$Cns(\mathbf{X}) = 1 + \sum_{i=1}^{n} p_i \log_2 \left(1 - \frac{|X_i - \mu_x|}{d_x}\right) \quad (1)$$

where X represents the list of categories (Strongly Agree (SA), Agree (A), Neutral (N), Disagree (D), and Strongly Disagree (SD)), X_i is an element of X, μ_X is the mean of X and d_X is the width of X, $d_X = X_{\max} - X_{\min}$. Let us assume that we have a five-attribute Likert scale: SA, A, N, D, and SD. Let us further assign an arbitrary numerical scale of SA = 1, A = 2, N = 3, D = 4, and SD = 5. Then X = {1, 2, 3, 4, 5} and X_1 = 1, X_2 = 2, etc. The width of X, d_X, is $X_{\max} - X_{\min}$. In this case d_X = 5 - 1 = 4. The mean, or expected value, of X is given by the usual formula $E(X) = \sum_{i=1}^{n} p_i X_i = \mu_X$.

C. Measure of Dissent

Let X be a discrete random variable of size $n > 2$ with probability distribution $p(x)$. As usual $E(X) = \sum_{i=1}^{n} p_i X_i$ is the mean μ_X of X. Let $d_X = X_{\max} - X_{\min}$ be the width of X and set $d_i = |X_i - \mu_x|$ as the absolute deviation of X from the mean. The Dissention, $\mathtt{Dnt}(\mathbf{X})$ is then defined to be

$$\mathtt{Dnt}(\mathbf{X}) = -\sum_{i=1}^{n} p_i \log_2 \left(\frac{d_X - d_i}{d_X}\right)$$
$$= -\sum_{i=1}^{n} p_i \log_2 \left(1 - \frac{|X_i - \mu_X|}{d_X}\right)$$

If there is no chance of confusion then we will drop the subscripts and write:

$$\mathtt{Dnt}(\mathbf{X}) = -\sum_{i=1}^{n} p_i \log_2 \left(1 - \frac{|X_i - \mu|}{d}\right) \quad (2)$$

If $n = 1$ then there is no dissention and we will set $\mathtt{Dnt}(\mathbf{X}) = 0$.

$Cns = 1 - Dnt$ or

$$\mathtt{Cns}(\mathbf{Y}) = 1 + \sum_{i=1}^{n} p_i \log_2 \left(1 - \frac{|Y_i - \mu_Y|}{d_Y}\right) \quad (3)$$

Table 1 A simulation of categorical assignments made by a group of 10 stakeholders. Consensus is compared to the mean and standard deviation.

	SA	A	N	D	SD	Cns	Mean	Std Dev
1	7	3	0	0	0	0.838	1.3	0.458
2	0	0	0	3	7	0.838	4.7	1.458
3	6	4	0	0	0	0.815	1.4	1.490
4	5	5	0	0	0	0.807	1.5	0.500
5	8	1	1	0	0	0.802	1.3	0.640
6	0	0	8	1	1	0.802	3.3	0.640
7	7	2	1	0	0	0.773	1.4	0.663
8	5	4	1	0	0	0.760	1.6	0.663
9	6	3	1	0	0	0.759	1.5	0.671
10	8	1	0	1	0	0.703	1.4	0.917
11	5	4	0	1	0	0.693	1.7	0.900
12	7	2	0	1	0	0.685	1.5	0.922
13	6	3	0	1	0	0.682	1.6	0.917
14	7	0	3	0	0	0.649	1.6	0.917
15	8	0	1	1	0	0.637	1.5	1.025
16	5	4	0	0	1	0.577	1.8	1.166
17	7	0	2	1	0	0.569	1.7	1.100
18	6	3	0	0	1	0.548	1.7	1.187
19	6	0	3	1	0	0.533	1.9	1.136
20	7	2	0	0	1	0.532	1.6	1.200
21	5	0	4	1	0	0.528	2.1	1.136
22	8	1	0	0	1	0.527	1.5	1.204
23	7	0	0	3	0	0.420	1.9	1.375
24	8	0	0	1	1	0.403	1.7	1.418
25	8	0	0	0	2	0.278	1.8	1.600
26	6	0	0	3	1	0.258	2.3	1.616
27	5	0	0	4	1	0.251	2.6	1.625
28	7	0	0	1	2	0.210	2.1	1.700
29	7	0	0	0	3	0.119	2.2	1.833
30	6	0	0	1	3	0.101	2.5	1.857

D. Measure of Agreement

The measure of agreement has been previously introduced [14-15] and shown to be a method by which differing opinions of stakeholders can be justifiably assembled to yield a single value upon which there exists maximal agreement [19]. This does not mean that all stakeholders need to have selected a particular category, for it is possible to agree on a category to which no one has made a selection. For example, row 21 in Table 1 shows a mean of 2.1, or 1/10 of the way between Agree and Neutral. If the Agree category is selected, it is apparent that the category in greatest agreement is that one which no one selected. Having an "average value" (see the *mean* column in table 1) refer to a category that reflects a zero frequency is not uncommon (note rows 23-30 with a mean representing a category with an assignment of zero).

We define agreement as a harmony of opinion or action. To attain a harmony of opinion does not require that all involved individuals express the same view.

The consensus measure was found to be modifiable to determine the degree of agreement associated with every frequency in a given distribution by assigning each category to the position of the *mean* in equation (1). In other words, consensus depends on the calculation of a mean value against which the distance of each category from that mean value is calculated. Hence the consensus is a measure of the degree of attraction to a mean value. This equation is modified slightly to calculate the degree of attraction to each individual category value. It should be noted that in lieu of the mean it is appropriate to use the median. In fact, the median is more conceptually accurate than the mean when working with ordinal scales.

The desire to *target* the consensus led the authors to examine different expressions for the log term. This has led to the development of an Agreement measure:

$$\mathbf{Agr}(\mathrm{X},\tau)=1+\sum_{i=1}^{n}p_i\log_2\left(1-\frac{|X_i-\tau|}{2d_X}\right) \quad (4)$$

τ represents the target category such as SA, A, etc. and the denominator is changed to $2d_X$ to reign in the range of the measure.

VI. EXAMPLE

A. Illustration of Consensus

Assume a group of 10 stakeholders in a meeting conducted by a systems analyst for the purpose of defining the problem that needs to be addressed. Each stakeholder has an interest in having the eventual solution satisfy their particular set of needs, and no individual stakeholder has full knowledge of the extent of the problem under investigation. It is up to the systems analyst to determine the actual nature of the problem. Suppose the analyst begins by raising a series of issues that seem to be plausible indicators of the problem, and each stakeholder is asked to respond to the statement indicating their level of agreement. A Likert scale is chosen and the questions are similar in format to the following example:

The problem centers on the inability of our customers to adequately access our web site to initiate orders.

For purposes of illustration let us select a subset of possible responses that compare the consensus with the traditionally calculated mean and standard deviation.

The mean is expectedly different for each of these frequencies indicating the average value without any dispersion as evidenced by the standard deviation (see Table 2). The consensus yields a value of 1.0 to show that these frequencies are in complete agreement. However, means (μ) and standard deviations (SD) do not convey a sense of agreement in the same way as the consensus measure (Cns).

Table 2 Three frequency distributions centered on strongly agree, neutral, and strongly disagree respectively, with the consensus, mean and standard deviation for each distribution.

SA	A	N	D	SD	Cns	μ	SD
10	0	0	0	0	1.000	1	0
0	0	**10**	0	0	1.000	3	0
0	0	0	0	**10**	1.000	5	0

Table 3 shows another possible distribution of responses from our group of 10 stakeholders. Note that the mean and standard deviation properly and accurately depict the information gathered by these distributions, but the interpretation of these results is in question. Each distribution has the same mean of 3.0 = Neutral, but the standard deviations show different dispersions. Reflecting only the meaning of the standard deviation, the reader visualizes a normal distribution with a slightly wider stance in the left and right legs of the curve. There is nothing special in this statistical interpretation. However, the consensus measure of 0.585. or 58.5%, induces a mental image of a consensus that reflects too much dissention to be acceptable.

Row 2 of table 3 shows an even further dispersion as the consensus is 0.0 or 0%. This is the situation when neither side accepts the argument of the other. This gridlock is serious and reflects a situation that the analyst needs to further investigate. When the stakeholders differ in their perception of the problem at this great a degree, the actual problem is likely to be something other than that being presented.

Table 3. Two possible distributions having the same mean value, with increasing standard deviations, and consensus values that tell strikingly different stories.

	SA	A	N	D	SD	Cns	Mean	Std Dev
1	0	5	0	5	0	0.585	3.0	1.054
2	5	0	0	0	5	0.000	3.0	2.108

Table 4 shows four different distributions that contain identical pairs of means, standard deviation and consensus measures. Note that the presence of a mean does not require the presence of any values in that category (e.g., row 1 has a mean of 2 = Agree, but the frequency for that element is zero. Means can change and still retain the same standard deviation. The consensus measures show about a 43% consensus around the frequencies in Rows 1 and 2, and a 31% degree of consensus in Rows 3 and 4. Since these frequencies are symmetrical to each other (row 1 and row 2 are reversals of each other), it should be expected that the degree of consensus is the same. What matters is whether or not a degree of consensus has been met to move forward in identifying the problems to be addressed in the new system.

Table 4 Pairs of consensus', means, and standard deviations associated with different distributions.

	SA	A	N	D	SD	Cns	Mean	Std Dev
1	6	0	3	0	1	0.426	2.0	1.414
2	1	0	3	0	6	0.426	4.0	1.414
3	7	0	0	2	1	0.309	2.0	1.633
4	1	2	0	0	7	0.309	4.0	1.633

	SA	A	N	D	SD	Agr(SA)	Dnt	Cns	Mean	StDev
1	0	0	0	0	15	0.000	0.000	1.000	5.000	0.000
2	0	0	0	1	14	0.021	0.048	0.952	4.933	0.249
3	0	0	0	2	13	0.043	0.089	0.911	4.867	0.340
4	0	0	0	3	12	0.064	0.124	0.876	4.800	0.400
5	0	0	0	4	11	0.086	0.151	0.849	4.733	0.442
6	0	0	1	4	10	0.125	0.213	0.787	4.600	0.596
7	0	0	2	4	9	0.164	0.259	0.741	4.467	0.718
8	0	0	3	4	8	0.203	0.291	0.709	4.333	0.798
9	0	1	3	4	7	0.257	0.347	0.653	4.133	0.957
10	0	2	3	4	6	0.310	0.389	0.611	3.933	1.062
11	1	2	3	4	5	0.377	0.490	0.510	3.667	1.247
12	3	3	3	3	3	0.543	0.566	0.434	3.000	1.414
13	5	4	3	2	1	0.709	0.490	0.510	2.333	1.247
14	6	4	3	2	0	0.775	0.389	0.611	2.067	1.062
15	7	4	3	1	0	0.820	0.347	0.653	1.867	0.957
16	8	4	3	0	0	0.866	0.291	0.709	1.667	0.798
17	9	4	2	0	0	0.893	0.259	0.741	1.533	0.718
18	10	4	1	0	0	0.921	0.213	0.787	1.400	0.596
19	11	4	0	0	0	0.949	0.151	0.849	1.267	0.442
20	12	3	0	0	0	0.961	0.124	0.876	1.200	0.400
21	13	2	0	0	0	0.974	0.089	0.911	1.133	0.340
22	14	1	0	0	0	0.987	0.048	0.952	1.067	0.249
23	15	0	0	0	0	1.000	0.000	1.000	1.000	0.000

Table 6 Twenty-three distributions from a stakeholder group of 15 individuals with agreement (Agr) calculated on the Strongly Agree (SA) category.

As the distributions become more random in the permutations of stakeholder values, making sense of the means and standard deviations become increasingly difficult. Table 1 above shows a subset of randomly selected rows, sorted by consensus, then by mean.

If we assume that Salmoni [11] is correct and we accept 80% as the threshold for consensus, it is easy to justify rows 1 through 6 in Table 1 as being acceptable to the systems analyst. While we can accept that the group of stakeholders has arrived at a consensus, it is still a matter of interpretation as to the winning category. There is also the matter of those means that are midway between categories, like rows 4, 9, 12, 22, and 26. What is the arbiter of the consensus category in these situations? While the consensus measure gives a theoretical justification for having attained consensus, it does not permit the unambiguous selection of a particular category of ownership (see table 5 for the identification of first 10 rows

Row	Category
1	SA
2	SD
3	SA
4	SA or A
5	SA
6	N
7	SA
8	A
9	SA or A
10	SA

Table 5. A subjective determination of category based on the mean.

of table 1 into categories based on the mean value). To solve this problem we must modify the consensus equation to permit us to calculate the amount of agreement the group of stakeholders has for each individual category.

B. Illustration of Dissent

An examination of Table 6 shows another set of 23 distributions that could be elicited from some group of 15 stakeholders. For each distribution the agreement (Agr), dissent (Dnt), consensus (Cns), mean and standard deviation (StDev) are calculated.

For each value associated with Cns the value of Dnt can be seen to equal 1 – Cns (see equation 3 above). Note that the consensus is maximized in rows 1 and 23 and minimized, in this particular set of distributions, at row 12. Since 15 cannot be evenly divided by 2, it is not possible to have a zero consensus. The dissention can be interpreted as a measure of dispersion around the mean. Dissention as a measure of dispersion gives no new information to the statistical measure of the standard deviation except to say that it is far easier to understand. Figure 1 (see Appendix) shows the linear relationship between dissent and the standard deviation. The covariance of Dnt and StDev in Table 6 is 0.0635, and the R^2 is 0.9913 (see Figure 1 after references). Dissention as a measure of dispersion gives no more information than the standard deviation except that it is a far easier statistic to understand.

Since both measures give the same mathematical sense of variance (dispersion), one could use either one. The authors argue that it is easier to visualize and understand dispersion when it is represented in terms of a percentage than by a number that grows with the magnitude of the numbers that comprise the calculation, and since the calculation does not assume an interval value, it is the more conceptually correct measure. Thus, given our group of stakeholders, it is likely that everyone will better understand the dispersion in their ratings when it is represented in row 11 of table 6 as 49% rather than the 1.291 given by the calculated standard deviation; if the numbers in the distribution changed from {1,2,3,4,5} to {100,200,300,400,500} the standard deviation would change from 1.464 to 146.385. Again, the value of the standard deviation is driven by the size of the numbers used in the calculation while the dissent (as dispersion) is a percentage that retains a constant mental reference of a value between 0-100%.

C. Agreement

Table 6 shows the Agreement based only on the strongly agree category, and table 7 shows the Agreement for each category in the grayed row below its frequency distribution. Each gray row is called a *set of agreement measures* (that can also be considered an *agreement 5-tuple*). Since we assume that the survey data represents an entire population (like a group of stakeholders attempting to identify a systems problem) we make no claims with respect to samples and populations and leave that area for future research activities.

Table 7 Agreement calculated for each response.

	SA	A	N	D	SD
1	0	0	0	0	15
$Agr_1(\tau)$	0.000	0.322	0.585	0.807	1.000
2	0	0	0	1	14
$Agr_2(\tau)$	0.021	0.339	0.600	0.820	0.987
3	0	0	0	2	13
$Agr_3(\tau)$	0.043	0.357	0.615	0.833	0.974
4	0	0	0	3	12
$Agr_4(\tau)$	0.064	0.375	0.629	0.846	0.961
5	0	0	0	4	11
$Agr_5(\tau)$	0.086	0.392	0.644	0.859	0.949
6	0	0	1	4	10
$Agr_6(\tau)$	0.125	0.424	0.672	0.859	0.921
7	0	0	2	4	9
$Agr_7(\tau)$	0.164	0.457	0.700	0.859	0.893
8	0	0	3	4	8
$Agr_8(\tau)$	0.203	0.489	0.727	0.859	0.866
9	0	1	3	4	7
$Agr_9(\tau)$	0.257	0.534	0.742	0.844	0.820
10	0	2	3	4	6
$Agr_{10}(\tau)$	0.310	0.580	0.757	0.829	0.775
11	1	2	3	4	5
$Agr_{11}(\tau)$	0.377	0.612	0.757	0.797	0.709
12	3	3	3	3	3
$Agr_{12}(\tau)$	0.543	0.704	0.757	0.704	0.543

Notice row 12 in table 7. The agreement is not equi-distributed over the categories because each category in this ordered set gives strength to those other categories contiguous to itself. Hence, the three individuals who selected SA would also permit the selection of A. Those who chose N are strengthened by those who selected A and D. Row 1 show the strongest possible agreement with the SD category, and the agreement drops off rather quickly. If we adopt Salmoni's [11] threshold of 80% for acceptance of agreement, then row 10 shows an agreement in support of the Disagree category even though more people selected SD.

This measure is particularly useful when soliciting the response from a set of knowledgeable individuals who are acting separately in making a categorical assignment. It is apparently not atypical to make a solicitation of experts[1] and then to make a claim that the experts assert a particular position or category. With this new agreement measure it is possible to assign a real number to that level of agreement. This was particularly significant in an application of this measure to the assignment of colors to the terrorist threat levels (green, yellow, orange, and red) [19]. One problem with that application was the inability to map changes in the agreement distribution over time. That is to say, as the agreement frequencies changed over time, so would the agreement measures associated with each category, but how could one follow the degree by which the change occurred? It is now possible to measure the distance between these agreement distributions.

D. Measuring Distance between Agreement 5-Tuples

Given two frequency distributions (see table 8), F_1 and F_2, for which the agreement distributions, Agr_1 and Agr_2, are calculated using equation (5), a distance between the distributions can be determined. For each category in each frequency distribution there is a corresponding agreement value (see figure 2 for an illustration). The distance is calculated using

$$c_n \sqrt{\sum_{i}^{n} (Agt_1^i - Agt_2^i)^2} \qquad (5)$$

where n = the number of categories, c_n is a constant for each n (for a five category Likert scale $c_n = 0.63612$), Agr_1 uses probabilities derived from the frequencies F_1, and Agr_2 uses probabilities derived from the frequencies F_2. Agr (X, X_i) is the Agreement of the categories X with the i^{th} category.

The illustration in row 1 of Table 8 shows the maximum possible distance between two distributions in which the survey participants have chosen extreme positions.

		SA	A	N	D	SD	Dist
1	F_1	0	0	0	0	5	1.000
	F_2	5	0	0	0	0	
2	F_1	0	0	0	1	4	0.957
	F_2	5	0	0	0	0	
3	F_1	0	0	1	1	3	0.864
	F_2	5	0	0	0	0	
4	F_1	0	0	1	2	2	0.830
	F_2	5	0	0	0	0	
5	F_1	0	0	1	2	2	0.777
	F_2	4	1	0	0	0	
6	F_1	0	0	1	2	2	0.670
	F_2	3	1	1	0	0	
7	F_1	0	0	1	2	2	0.512
	F_2	2	1	1	1	0	
8	F_1	0	0	1	2	2	0.283
	F_2	0	1	2	2	0	
9	F_1	0	0	1	2	2	0.063
	F_2	0	0	2	1	1	
10	F_1	0	0	1	2	2	0.000
	F_2	0	0	1	2	2	

Table 8. Computed distance between agreement distributions. First column indicates a row number (from 1 to 10), column 2 distinguishes the two rows of distributions, columns 3-7 represent the values in the Likert distribution, and the last column is the calculated distance.

Row 1 of table 8 shows two frequency distributions, each of five stakeholders. The agreement distribution for each row is calculated (not shown in this table) and compared to the agreement distribution for the other row in this couple. Each row in table 8 contains two distributions, a "top" distribution and a "bottom" distribution. It is the agreement distribution for each of these that are compared and a distance calculated. Thus, the agreement distances become smaller as the agreement values become equal (see rows 8-10). It is important to understand that actual frequency values are not compared, rather, the agreement measure calculated on those frequencies. This permits us to calculate a distance without regard to the number of items constituting the frequency distribution.

VII. Conclusion

One of the most challenging areas in requirements determination is the identification and selection of the actual problem to address. The challenge is compounded in that finding solution methods frequently means crossing disciplines from organizational behavior to psychology, from statistics to measure theory, from the structured to the subjective. Currently available statistical tests such as the Wilcox-

[1] Private discussions with corporate representatives at the Risk Symposium 2008, Santa Fe, NM, 11-13 March 2008.

on signed rank test, are commonly used, but the underlying concept for their use is circumspect. The use of the mean and the standard error assumes an interval measure, something for which the evidence is, at best, sketchy using a Likert or similar scale. Scientists have, nonetheless, been using these kinds of tests since they were created in mid-1940 with good results. Now there are other concepts that have been developed such as the Dempster-Shafer evidence theory in 1976 and fuzzy set theory by Lotfi Zadeh in 1965. Building upon the works of others produced the concept of applying a variation from information theory to what is presented here as consensus theory. This method offers a different way of examining ordered category data, not to exclude or otherwise dispose of the currently used measures, but rather as an additional tool that brings a different view of categorical data analysis.

Beginning with the basic concept of deriving a measure of consensus given an ordered set of choices, such as the Likert scale, the concept was extended to provide a degree of focus on each category in a distribution (the agreement and dissent measures), and finally to being able to make a reasonable calculation of distance between two such distributions (the agreement distribution and the agreement distance). Using the agreement distance it is possible to easily track the success (or lack of success) in securing an agreement among stakeholders as to a particular position and when to claim that there is sufficient consensus from the stakeholders to accept the position and move onto another item.

REFERENCES

[1] Antunes, P., Costa, C.J. and Pino, J.A. (2006) "The use of genre analysis in the design of electronic meeting systems" *Information Research*, **11**(3) paper 251 [Available at http://InformationR.net/ir/11-3/paper251.html].

[2] Arbib, M. A. and Manes, R. G. (1975) “A Category-Theoretic Approach to systems in a Fuzzy World,” Synthese, 30(3-4), pp. 381-406.

[3] Benedictis, L D and Maio, M D (2009) “Economists’ View about the Economy. Evidence from a Survey of Italian Economists,” Centre for Economic and International Studies, University of Rome.

[4] Brouwer, R. (2007) “Fuzzy Clustering of Feature Vestors with some Ordinal Valued Attributed using Gradient Descent for Leanring,” Unternational Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 16(2), pp. 195-218.

[5] ComputerWorld.com 2008 "Top 10 Corporate Information Technology Failures," accessed 2008 at http://www.computerworld.com/computerworld/records/images/pdf/44NfailChart.pdf.

[6] Dennis, Alan and Wixom, Barbara H. (2000) Systems Analysis and Design: An Applied Approach. John Wiley & Sons, New York.

[7] Farley, J and Ceroni, M (2007) “Redefining “Productive”: Implications for Sustainable Forest Management,” Northeastern States Research Cooperative, www.uvm.edu/envnr/nsrc.

[8] Jamieson, S. (2004) “Likert scales: How to (ab)use them?.” Medical Education, 38, pp. 1217-1218.

[9] Loughman, T., Fleck, Jr,, R. and Snipes, R. 2000 "A cross-disciplinary model for improved information systems analysis," Journal of Industrial Management & Data systems, 100(8), pp. 359-369.

[10] Saldarini, R., (1989) Analysis and Design of Business Information Systems, Macmillan, New York.

[11] Salmoni, A. J., Coxall, S., Gonzalez, M., Tastle, W., and Finley, A. (under review) " Defining a postgraduate curriculum in dermatology for general practitioners: a needs analysis using a modified Delphi method."

[12] Sidky, A. S., Sud, R. R., Bhatia, S., and Arthur, J. D. 2002 "Problem Identification and Decomposition within the Requirements Generation Process," Technical Report TR-02-31, Computer Science, Virginia Tech.

[13] Stevens, S. S., 1946, "On the theory of scales of measurement. *Science*, 103, pp. 677-680.

[14] Tastle, W. and Wierman, M. (2005b) "Consensus and Dissention: Theory and Properties," North American Fuzzy Information Processing Society (NAFIPS) Conference, Ann Arbor, MI.

[15] Tastle, W. and Wierman, M. (2005c) “A Tool for the Analysis of Ordinal Scale Sate: Measuring Consensus, Agreement, and Dissent,” 5th International Conference on Methods and Techniques in Behavioral Research, Wageningen, The Netherlands.

[16] Tastle, W. and Wierman, M. (2005d) “Measuring Consensus in Team Settings Doing Non-routine Knowledge Work,” 4th AIS Symposium on Research in Systems Analysis and Design, Univ of Cincinnati [2005], April 23-24, 2005, Univ of Cincinnati.

[17] Tastle, W. and Wierman, M. (2006b) “An Information Theoretic Measure for the Evaluation of Ordinal Scale Data,” Journal of Behavior Research Methods, 38(3), pp. 487-494, August 2006.

[18] Tastle, W. and Wierman, M. (2006c) “Consensus and dissention: A measure of ordinal dispersion,” International Journal of Approximate Reasoning, 45(3), pp. 531-545.

[19] Tastle, W. and Wierman, M. (2007) “Determining Risk Assessment Using the Weighted Ordinal Agreement Measure,” Journal of Homeland Security, http://www.homelandsecurity.org/newjournalArticles/displayArticle2.asp?article=157, June 2007.

[20] Tastle, W. and Wierman, W. (2005a) "Consensus and Dissention: A New Measure of Agreement," North American Fuzzy Information Processing Society (NAFIPS) Conference, Ann Arbor, MI.

[21] Tastle, W. and Wierman, W. (2006a) “Adjusting the Consensus Measure to Satisfy Ordinal Scale Arguments,” North American Fuzzy Information Processing Society (NAFIPS 2006) Conference, Montreal, Canada.

[22] Tastle, W., Russell, J. and Wierman, M. (2005) "Ordinal Scales: Using the Consensus Measure to Compare Likert Scale Data," Information Systems Education Conference (ISECON 2005), Columbus, Ohio.

[23] Tastle, W., Wierman M. and Dumdum, U. R. (2005) "Ranking Ordinal Scales Using the Consensus Measure," Int'l Assn of Computer Information Systems Conference (IACIS 2005), Atlanta, GA, pp. 96-102.

[24] Van Horn, R. L., Schwarzkopf, A. and Price, R. (2006) *Information Systems Solutions: A Project Approach*, McGraw-Hill-Irwin, New York City.

[25] Velleman, P. F. and Wilkinson, L. (1993) "Nominal, Ordinal, Interval, and Ratio Typologies are Misleading. The American Statistician, 47(1), pp. 65-72.

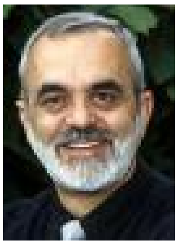
[26] Wierman, M. and Tastle, W. (2006) “Placing the dissonance measure in the context of generalized information theory,” In North American Fuzzy Information Processing Society (NAFIPS 2006) Conference, Montreal, Canada.

[27] Zadeh, L. (1965) “Fuzzy Sets,” Information and Control, 8, pp. 338-353.

[28] Tastle, W. (2009) “Measuring Faculty Instructional Performance.” In The Proceedings of the Information Systems Educaiton Conference 2009, v 26 (Washington DC): §1532. ISSN: 1542-7382.

[29] Toubia, O. (2006) “Idea Generation, Creativity, and Incentives,” Marketing Science, 25, pp. 411-415.

[30] Nasseri, H (2008) "Fuzzy Numbers: Positive and Nonnegative," International Mathematical Forun, 3(36), pp. 1777-1780.



William J. Tastle is an associate professor in the department of management at Ithaca College, New York, a Research Fellow at the Semeion Research Centre of Rome, Italy, a Research Professor at the University of Iceland, and a Fellow of the Association of Information Technology Professionals (educaton). He is also a senior member of both IEEE and ACM. His PhD is from the State University of New York at Binghamton in Advanced Technology with specialization in Systems Science, holds MS and MBA degrees, and his research interests include measure theory, offshoring of IT functions, and information systems curriculum development. He is currently the conference chair for the North American Fuzzy Information Processing Society Conference in 2010.

Amjad Abdullat serves as the head of the Department of Computer Information Systems and associate professor in the College of Business. He has a B.S. and B.A. from California State University at Sacramento, an M.B.A. from National University of San Diego and a Doctorate. from Pepperdine University. Dr. Abdullat's dissertation was "The Need for Paradigm Shift in Information Technology Planning." Dr. Abdullat is one of the founding principles and member of the board of directors of Edmin.com. Edmin is a comprehensive technology enterprise providing learning organizations with the next generation of web-based products, applications and services.

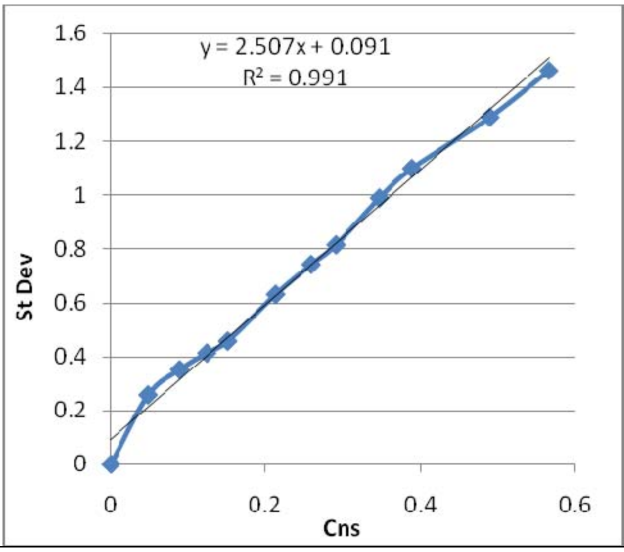
Mark J. Wierman BA mathematics, Purchase College, Purchase NY, 1978. MA mathematics, Binghamton University, Binghamton NY, 1980. PhD systems science, Binghamton University, Binghamton NY, 1994.

He has been an Instructor at SUC Oneonta, done research for Rome Labs, and programmed for IBM. Currently is an Assistant Professor of Computer Science at Creighton University in Omaha, NE. He co-authored two books, *Uncertainty-Based Information* with George Klir and *Applying Fuzzy Mathematics to Formal Models in Comparative Politics* with Terry D. Clark, Jennifer M. Larson, John N. Mordeson, and Joshua D. Potter.

Dr. Wierman major areas of research are Generalized Information Theory and the mathematics of Fuzzy Set Theory. Recently he has been studying measures of consensus as well as the application of fuzzy set theory to Political Science. He is a member of the Society for New Mathematics as well as a NAFIPS board member.

APPENDIX A

Figure 1 Comparison of Dissent (Dns) to Standard Deviation (StDev) as measures of dispersion. Note the correlation coefficient of 0.99.



Scale Type	Permissible Statistics	Admissible Scale Trans	Mathematical structure
Nominal (also denoted as categorical or discrete)	Mode, chi square	One to one (equality (=))	Standard set structure (unordered)
Ordinal	Median, percentile	Monotonic increasing (order (<))	Totally ordered set
Interval	Mean, standard deviation, correlation, regression, analysis of variance	Positive linear (affine)	Affine line
Ratio	All statistics permitted for interval scales plus the following: geometric mean, harmonic mean, coefficient of variation, logarithms	Positive similarities (multiplication)	Field

Table 9 Classification of the four different types of scales, Stevens (1946, 1951), taken from Wikipedia. Note that each scale type includes the permissible statistics of the previous types; hence, ordinal statistics include those in the nominal category (mode, chi square).