Institute of Design
ILLINOIS INSTITUTE OF TECHNOLOGY

Evaluation of Complex Systems

Charles L. Owen

January, 2007

A paper published by:

Design Studies
The international journal for design research
in engineering, architecture, products and systems
Volume 28, No. 1 (2007): 73-101

Keywords: evaluation, planning, modeling, structured planning, systems design

Institute of Design, Illinois Institute of Technology
Evaluation of Complex Systems
/ Charles L. Owen / January, 2007

Abstract

A new process is introduced for evaluating systems. Criteria are represented as a hierarchy of functions or, at a policy level, a hierarchy of policy positions. The system to be evaluated is represented as a hierarchy of its components. Elements of the system are scored against criteria using a bidirectional scale with provision for degrees of both support and obstruction. Assessment values for lowest-level elements are calculated using a series algorithm that relates composite scores uniquely to the scoring scale. Values for categories within both hierarchies are calculated from element values below them, with results presented graphically, in color.

Introduction

Simple things are difficult enough to evaluate, even those for which hard data can be found. For complex entities, especially when subjective evaluations are involved, the problem can be dauntingly difficult—enough so that comprehensive assessments often are just not attempted.

The purpose of this paper is to present a means that combines analysis, modeling and evaluation to assess overall as well as detailed system performance against the needs of system users. The process is comprehensive and is a natural extension of the analytical techniques of Structured Planning, a process developed for advanced planning (Owen, 1992, 1993, 2001a, 2001b, 2005).

Conventional System Evaluation Models

Since system evaluation is admittedly complex, it would be good to have an example system to work with that is familiar enough that the discussion can focus on the process without additional explanation of the content.

An example that meets this criterion is the restaurant. Restaurants are complex, are frequent subjects for evaluation, and are within the personal experience of nearly everyone. Evaluations of them are ubiquitous in travel guides and every city's newspapers and magazines. A restaurant's workings are reasonably familiar to all.

The star model. Travel guides employ the ultimate summary evaluation model: a number of stars or similar symbols. Michelin's system is probably the most prestigious and best known of these. A single Michelin star means a restaurant is "a very good restaurant in its category"; two stars means it has "excellent cooking, worth a detour"; and three stars (very difficult to attain) indicate "exceptional cuisine, worth a special trip". Michelin limits its stars to a maximum of three; other evaluators use four or even five, presumably because they can judge with greater resolution.

How the ranking is done is seldom made explicit. Pat Bruno, restaurant critic of the **Chicago Sun-Times** newspaper comments, "Restaurant owners all too often think it's all about the food. Not. It's a lot more than that. It's the whole package—food, quality, price, service, enjoyment—that provides the gift that keeps on giving, which in the restaurant business translates to steady business" (Bruno, 2004). Critics, usually traveling incognito, bring their experience to bear in a way that integrates many factors in a subjective, but well informed judgment. The process requires highly refined expertise developed over time and, of necessity, is confined to a narrow range of content.

The Multiple Category Model. Restaurant critics who report on television shows or in articles on food frequently divide their evaluations into two or three major categories. The categories usually cover quality of food, excellence of service and, possibly, the ambiance of the environment (which may include the niceties of interior design). James Ward, food critic of Chicago's local ABC TV affiliate, organizes his personal appraisals under the terms

Institute of Design, Illinois Institute of Technology http://www.id.iit.edu

Criteria	Weight	Scale	Scale Factor	Minimum Values for an **Exceptional** Restaurant Min. Values	Calculations	Minimum Values for an **Excellent** Restaurant Min. Values	Calculations	Minimum Values for a **Very Good** Restaurant Min. Values	Calculations	**Values for Restaurant under Evaluation** Actual Values	Calculations
1. Food Quality	**10**	**1 - 10**	**1**	**9**	(10)(1)(9) = 90	**8**	(10)(1)(8) = 80	**7**	(10)(1)(7) = 70	**8**	(10)(1)(8) = 80
2. Presentation	**5**	**1 - 5**	**2**	**5**	(5)(2)(5) = 50	**4**	(5)(2)(4) = 40	**4**	(5)(2)(4) = 40	**3**	(5)(2)(3) = 30
3. Attentiveness	**7**	**1 - 3**	**3.3**	**3**	(7)(3.3)(3) = 70	**3**	(7)(3.3)(3) = 70	**2**	(7)(3.3)(2) = 46.7	**3**	(7)(3.3)(3) = 70
4. Knowledgeability	**6**	**1 - 5**	**2**	**5**	(6)(2)(5) = 60	**4**	(6)(2)(4) = 48	**4**	(6)(2)(4) = 48	**3**	(6)(2)(3) = 36
5. Interior Design	**4**	**1 - 5**	**2**	**5**	(4)(2)(5) = 40	**4**	(4)(2)(4) = 32	**3**	(4)(2)(3) = 24	**3**	(4)(2)(3) = 24
6. Lighting	**6**	**1 - 4**	**2.5**	**4**	(6)(2.5)(4) = 60	**4**	(6)(2.5)(4) = 60	**3**	(6)(2.5)(3) = 45	**4**	(6)(2.5)(4) = 60
	Sum of Weights 38				Sum = 370		Sum = 330		Sum = 273.7		Sum = 300
					Minimum for **Exceptional** = 370/38 = 9.7		Minimum for **Excellent** = 330/38 = 8.7		Minimum for **Very Good** = 273.7/38 = 7.2		**Restaurant Score** = 300/38 = 7.9
					Range for Exceptional = 9.7 to 10		**Range** for Excellent = 8.7 to 9.7		**Range** for Very Good = 7.2 to 8.7		**Merit Index** = 7.9
					☆☆☆		☆☆		☆		☆½

Figure 1 **Criterion Function Method for Calculating a Merit Index**

"breads" and "circuses". The **Zagat Survey** accumulates votes from volunteer members of the dining public using a 1–30 scale and three categories (food, decor and service). Results are averaged for each category and published annually in books specialized for individual cities. For more detail, the **In-Room City Guide** of San Francisco rates restaurants using five categories: (1) decor, ambiance and service; (2) scope of the menu, quality of ingredients, execution of dishes, and how dishes please the palate; (3) cuisine—classic or cutting edge, traditional or fusion; (4) wine list—geographic scope, variety, and prices; and (5), cost of dining—splurge or "casual" affordable.

Ratings in this assessment model are made subjectively, much as is done with the star model. The advantage is that judgments can be narrowed to specific topics, and difficult integration problems can be avoided—such as can occur when the quality of food is excellent, but service unacceptably bad. The separate judgments are passed on to the reader, along with the integration task.

A variant on this model is the "filled circle" version used by institutions such as the magazine **Consumer Reports**. In this version, criteria are represented by circles that can be graphically treated in different ways to indicate degrees of quality. Ratings can be directly qualitative or can translate quantitative information to qualitative (translating, for example, price to high, medium, low). A typical, 5-value scale might have a full filled red circle for highest quality, a half-filled red circle for excellent quality, an open circle for good quality, a half-filled black circle for poor quality, and a full-filled black circle for unacceptably bad quality.

The Criterion Function Model. Extending the multiple category model, a technique developed in engineering design and decision science can be used both to consider more categories and to integrate results (see Cross, 2000; Saaty and Vargas, 2000; Pahl and Beitz, 1996; Roozenburg and Eckels, 1995; Ostrofsky, 1977; and Woodson, 1966 among others). In this model, scales can be individually tailored to each criterion (or objective) and weights can be applied to enable the composite score to reflect the relative importance placed on each criterion by the evaluator or user of the data. Differences in scales can be accommodated by scale factors that convert ratings to a common scale. The final integrated score is often referred to as a *merit index.*

An example will show the advantages of this model in basic form (Figure 1). For this example, the three classic categories of restaurant criteria—food, service and ambiance—are extended to six: food quality, presentation, attentiveness, knowledgeability, interior design and lighting. Any number of criteria, of course, could be accommodated.

In the first column on the left, weights are applied to the six criteria based on the judgment of the evaluator as to how important each is to the overall rating. Scales for rating can be individually specified, and this is done for each criterion in the second column. In the third column, scale factors are given to transform the scales to a common range, 1 to 10 in this example. The scale factors allow criteria to be judged on scales most appropriate to their qualities while preventing unfair calculation advantages due only to differences in the scales of evaluation.

In the next three paired columns, values are given and calculations made for minimum criteria thresholds for each of three generally recognized quality standards: exceptional, excellent and very good. Each calculation multiplies weight by scale factor by value. The evaluator must decide the minimum value that the criterion could achieve and still be awarded the quality distinction under consideration.

In the last column—or columns if more than one restaurant is being compared—the same evaluations and calculations are made for an actual restaurant or restaurants. From the totals, merit ranges and merit indexes are established by dividing the totals by the sum of the criterion weights. Because in this case the scales were converted to a 1 to 10 range, ranges and merit indexes are also in the range of 1 to 10. To bring the final values back into the star system, appropriate stars may be matched to ranges; the sample restaurant would be awarded 1½ stars.

The basic criterion function model may be further enhanced by introducing fuzzy mathematics for uncertainty, range arithmetic, confidence values and other means for making and combining judgments. Objective data (such as costs) can be included with subjective data, and reversed scales where larger numbers represent lower value can be accommodated.

Criterion function modeling represents a significant improvement over the earlier presented methods because it allows criteria to be objectified and weighted, and multiple competing subjects to be more fairly evaluated. The technique is powerful for making choices and even more powerful as a means for understanding the impact of decisions. By playing "what if" games with the data in spreadsheet software, it is possible to gain insight about the choice-making process as well as the choices.

A Hierarchical, Process-oriented Model

In spite of its great flexibility and ability to incorporate a broad range of criteria, the criterion function model has some built-in limitations that diminish its usefulness for evaluating systems, services and other complex man-made entities. This is particularly a problem for planners, designers and managers interested in improving performance or making comparisons among systems for effectiveness of operation.

First, the conventional approach does not very well suit the purposes of planning and design. The focus of evaluation in the criterion function model and similar models is the quality of the delivered experience, rather than the quality of the process that delivers it. Criteria in these models are almost always derived from the characteristics desired in the end result: in the restaurant example, good food, good service, pleasant ambiance, etc. From these, component categories may be extrapolated (food quality, presentation, attentiveness, knowledgeability, interior design, lighting), but all still address the end result, only tenuously connecting quality with the process that creates it. When *food quality* is a criterion for the evaluation of a restaurant, judgment is rendered through a complex assessment of tastes and preparations, all concerned with the product, with negligible consideration—if any—for how the food quality was produced. For the system planner and designer, the process and the system that implements it are the primary subjects of interest.

Second, criteria are not associated with specific, differentiated functions of the system, making it difficult to establish exactly where the design of the system is failing or doing well. In the food quality example, the food that reaches the table for evaluation is the result of the interaction of many human and system functions (directed actions that the system or its users perform) that execute the restaurant's particular approach to the creation of cuisine. Failure of any one or combination of these may be the cause of poor quality; and the special success of one or more may be the source of unusually high quality. Either way, failure to involve system operations explicitly as criteria leaves quality unassociated with the means of production.

Third, the system as the object of evaluation is not expressed in sufficient detail to identify the individual failures and successes of actual system components. Evaluating a "restaurant" as a total concept against a series of criteria requires mentally calling up a different set of the restaurant's properties for each criterion. Those properties are presumed to be known to the evaluator even though they have not been explicitly described! This probably accounts for much of the disagreement that arises among multiple evaluators, who inevitably conjure up different property composites, viewing the subject as the proverbial blind men "see" the elephant.

The Evaluation Tool

To build a tool for evaluation well suited to the needs of the planner and designer as creators of a quality-producing system—as well as the manager as maintainer of quality production—both the nature of the criteria and the characterization of the system being evaluated need to be changed, and the evaluation process itself must be modified to take advantage of the new structures.

Replacing the Criteria. Replacing the criteria, is a "structure" that abstracts the system as an organization of requirements that the system must meet to perform well. This structure acts like a set of criteria in that components of the system being evaluated are scored against each of its elemental requirements for how well they perform. It can be constructed in two different forms that evaluate the system at two different levels of abstraction.

At the higher degree of abstraction, the structure created is a *Policy Structure* grounded in a series of Defining Statements at its lowest level that are the positions on issues of policy that the system should exemplify. Higher levels in the hierarchy are successive layers of policy categories decreasing progressively in number to a *Policy Merit* assessment at the peak. For a thorough discussion of Defining Statements and project definition at the policy level generally, see Owen, 2001a and 2005. For the purposes of this paper, only one criteria structure need be presented and discussed and, because it is more frequently used, that one will be the Function Structure.

A *Function Structure* is at a level of abstraction closely matching that of the system to be evaluated. Rather than high level policy requirements prescribed in a Policy Structure, it designates actions to be taken to fulfill the objectives of the system's operation. Functions express what the system or its users must do. Because it is hierarchical, like the Policy Structure, the Function Structure also allows functionality to be considered at different levels of categorization and detail. It is an operational model of the system as a process—in particular, the process that should exist.

Construction of a Function Structure proceeds top down through three levels of description, each different in kind and appropriate in character to the level of action involved. The three levels, from the top, are: Modes of Behavior, Activities and Functions (Figure 2).

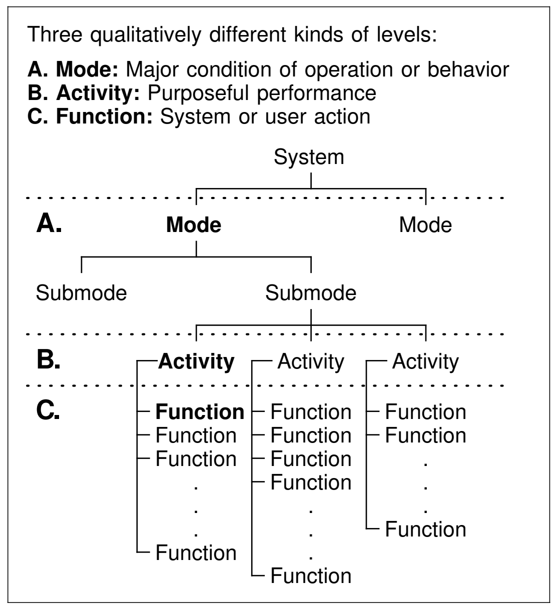


Figure 2 **The Function Structure**

Modes. *Modes* are the major forms of behavior that the system exhibits or works within. The analyst preparing a Function Structure begins at this level, considering the qualitatively different ways in which the system operates. There are seldom more than 10 to 15 modes. To distinguish them from other action entities in the Function Structure, they are named to suggest "classes" of operation. The suffix *-tion* frequently is used to specify a Mode (e.g., *Operation* as opposed to *Operating*), but other noun forms work equally well as long as they carry with them a sense of action. Figure 3 shows a simplified

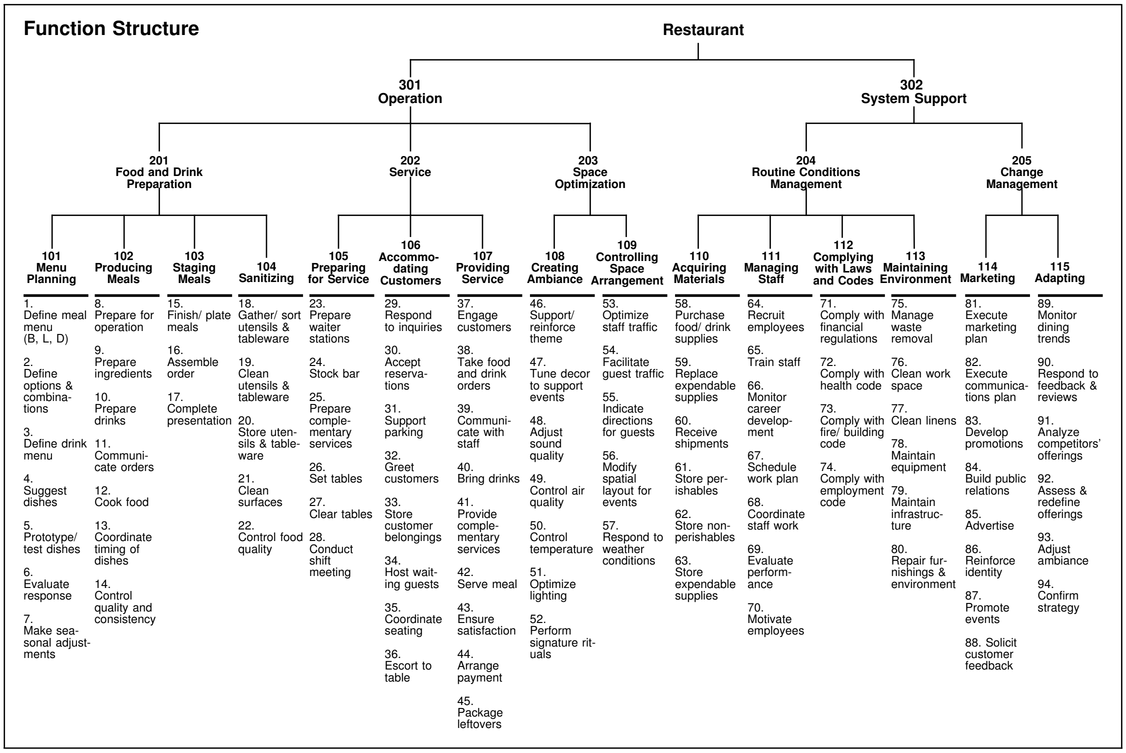


Figure 3 **Sample Function Structure for a Restaurant**

Function Structure for a restaurant. In it, the Modes are just Operation and System Support. Under them are *Submodes*, still major forms of behavior, but at a more specific level of description. For the restaurant example, these are Food and Drink Preparation, Service, Space Optimization, Routine Conditions Management and Change Management.

Activities. Below the high-level Modes and Submodes, are *Activities*. At the Activity level, the naming convention changes to fit a different conception of action, one that brings the analysis to a more dynamic level and closer to specific action descriptions. Activities are "purposeful performances". Much like the scenes of a play, they identify discrete collections of actions that work toward a specific goal.

Activities are named using the gerund form of verbs—verbs with *-ing* endings that turn verb forms into noun forms expressing action with duration. *Operating* is an Activity; *Operation* is a Mode. For the Food and Drink Preparation submode, the Activities are: *Menu Planning, Producing Meals, Staging Meals* and *Sanitizing.*

Functions. *Functions* are the fundamental elements of the Function Structure. At this level, actions are described specifically enough that they can be used as criteria for evaluating the system. A good system should be able to perform all of the prescribed Functions well.

The issue of "completeness" surfaces normally here. Does a hierarchically developed set of Functions completely and uniquely describe the system? The answer most certainly is "no". For any set of Functions, it is always possible to find one more, if not by simple addition, by subdividing a more general Function into two more specific ones. A better question is, "Does a hierarchically developed set of Functions do a more thorough job of covering the description than an otherwise developed list? I think that the answer to that is "yes". The measures of design are relativistic. If a design or plan (or in this case, the basis for a plan) is "better" than what we have been able to do previously, it is a success.

As in the construction of other levels, the selection and specification of Functions should seek to maintain balance and coverage within each Activity. For form, the Function takes a verb phrase, the form best able to capture necessary detail while conveying the sense of action inherent in a system. For the Activity *Producing Meals* in the example, the Functions are: *Prepare for operation, Prepare ingredients, Prepare drinks, Communicate orders, Cook food, Coordi-*

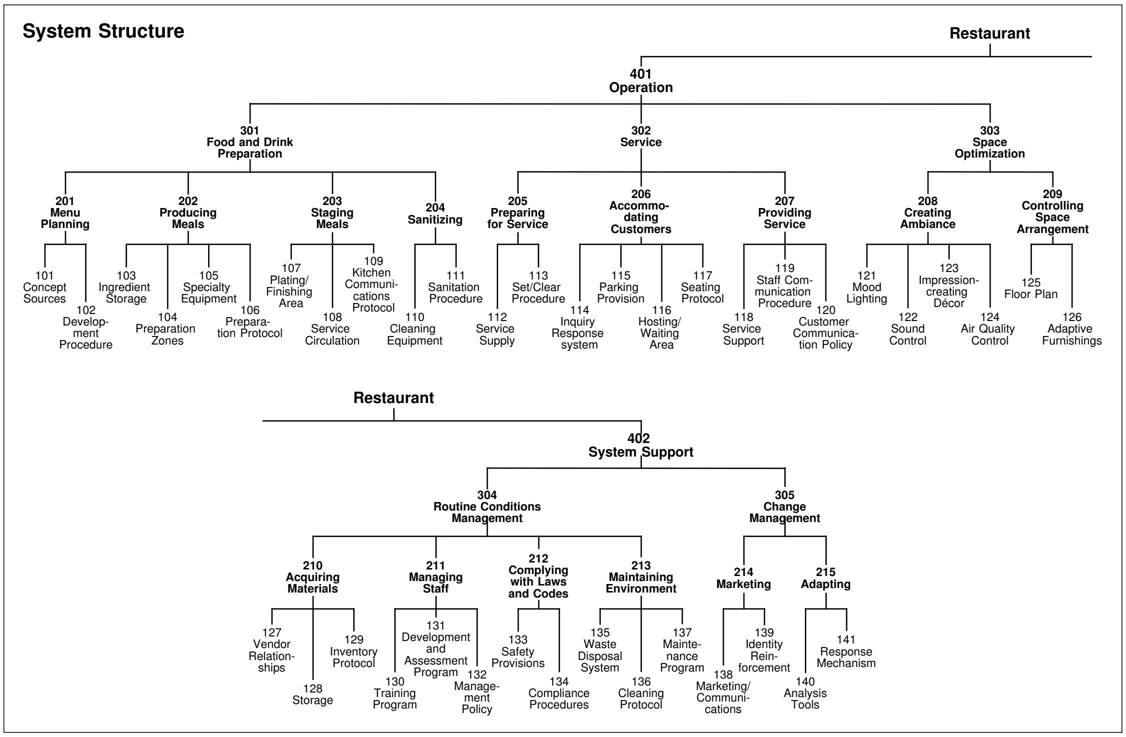


Figure 4 **Sample System Structure Archetype for a Restaurant**

nate timing of dishes, and *Control quality and consistency.*

Replacing the Object of Evaluation. In conventional evaluation processes, the object of evaluation is normally a whole entity. As criteria are addressed, characteristics of importance are summoned and evaluated, criterion by criterion. In the new process, the object of evaluation, like the criteria, is represented hierarchically as a structure—a *System Structure* (Figure 4). This model submits multiple elements of the system to the scrutiny of multiple criteria and enables both system and criteria to be organized and studied at different levels of detail.

The System Structure is constructed from the upper levels of the Function Structure redirected to categories of operational components at lower levels of the hierarchy. These categories collect classes of "actuation" and, ultimately, the actual components that outfit the system. The result is a system *archetype* or template at a class or category level which can be fitted out with actual lowest-level *System Elements* for a specific system (or systems, as they remain within the range defined by the archetype). In the restaurant example, this enables an analyst to compare restaurants within the same approximate competitive range (Figure 5); in this case, covering mid- to upper-mid-range restaurants—not as high as the peak level of fine dining, but well above the level of franchised chain restaurants.

System Structure Levels. A System Structure is essentially a hierarchy of decreasing levels of abstraction that organizes the components of a system. When developed as a product of the Structured Planning process, the structure is a natural byproduct of the synthesis of the system and its System Elements. When developed independently for the purpose of evaluation, it can be created most easily from the upper levels of the Function Structure (which will already have been prepared in earlier steps of the evaluation process).

Mode and Submode levels from the Function Structure can be used directly. Depending on the level of coverage achieved at this point, the Activity level may also be carried over (as was done in the restaurant example of Figure 4), or the style of description may change over at this point toward component depiction. At the highest categorization level, categories answer the question, "What fundamental functional categories are necessary to support this Submode or Activity?" At lower levels of categorization, categories move from general class descriptions to more specific property descriptions until a level is reached at which the next level would require the instantiation of an actual element of

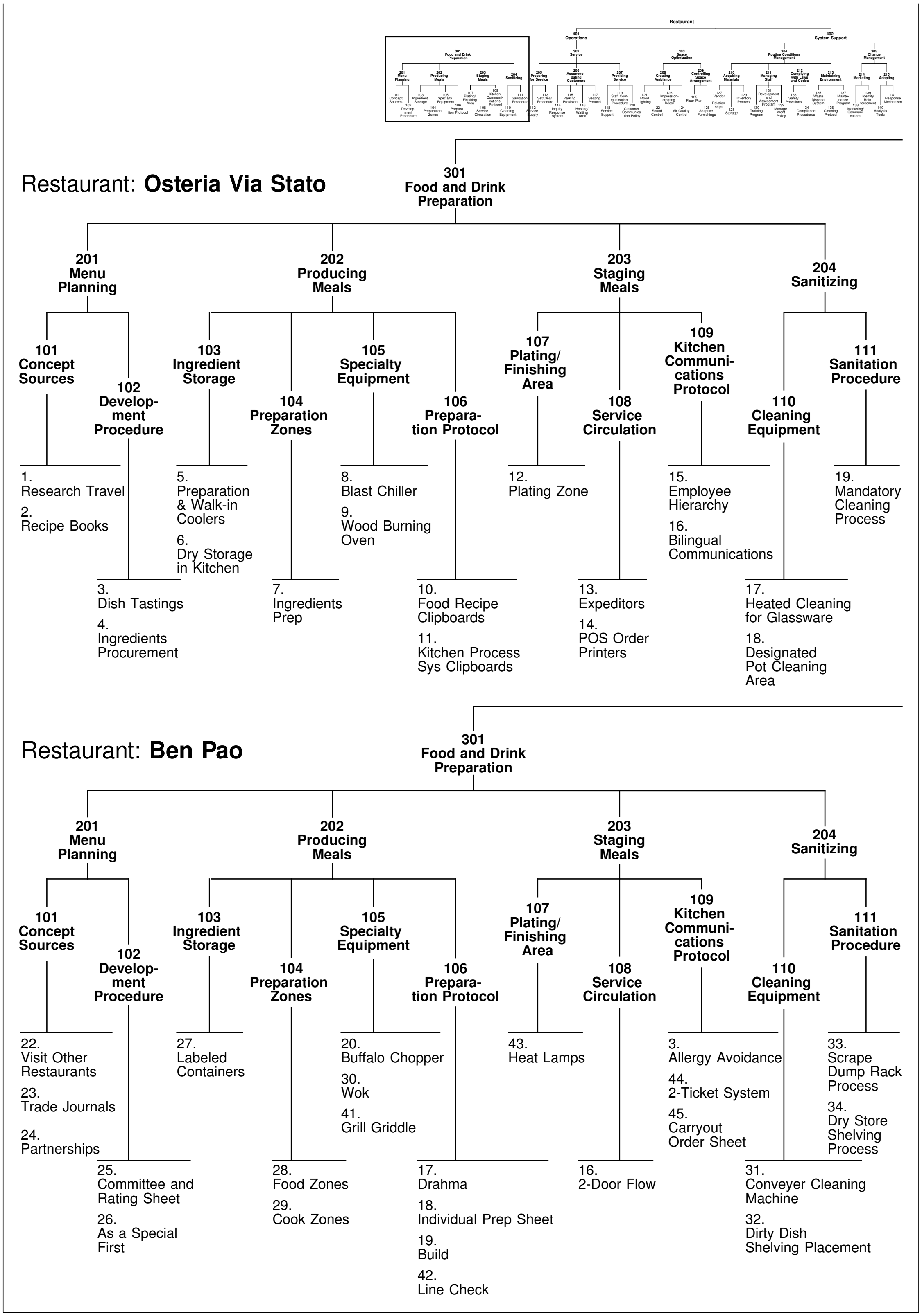
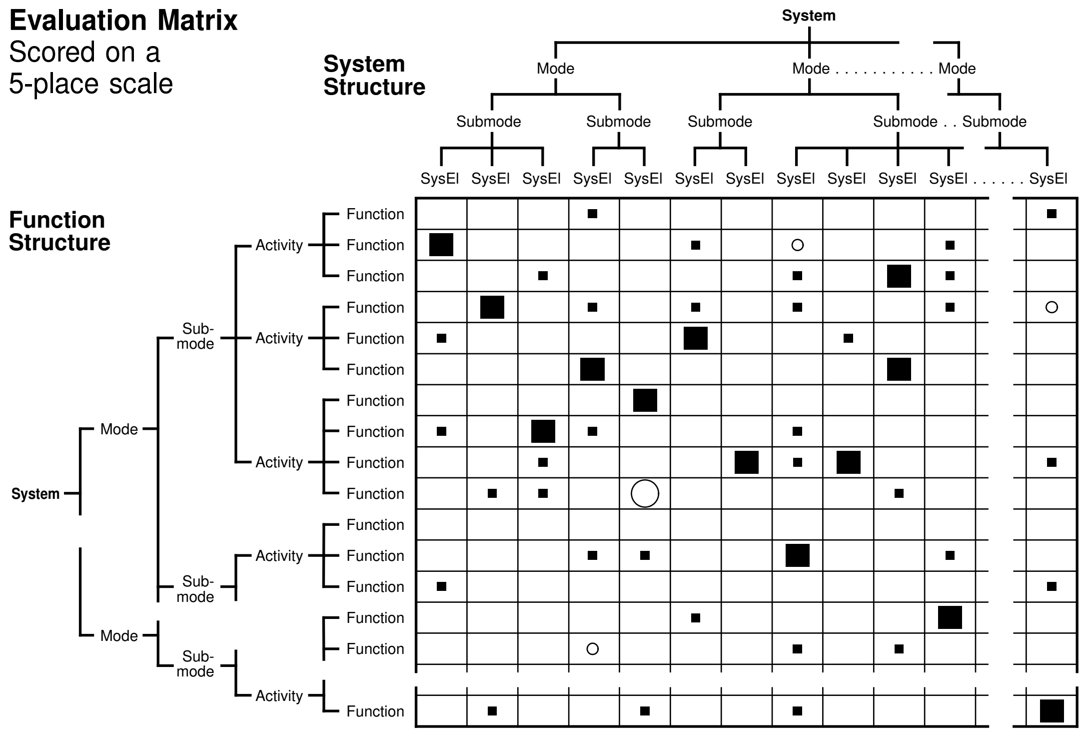


Figure 5 **Comparison of Two Actual Restaurants Using the Same System Structure Archetype**

the system under study. At this point, the System Structure is complete as an archetype usable for any system within the comparison group.

The Evaluation Process

The evaluation process is one of scoring and value accumulation through calculation. System



5-Place Evaluation Scale

- **+2** ***Strongly Supports:*** System Element is an excellent solution or provides additional strong support toward fulfilling the Function
- **+1** ***Supports:*** System Element is a good solution or provides additional support toward fulfilling the Function
- **0** ***No Effect:*** System Element has no bearing on the Function
- **–1** ***Obstructs:*** System Element is not too good a solution or obstructs fulfilling the Function
- **–2** ***Strongly Obstructs:*** System Element is a bad solution or seriously obstructs fulfilling the Function

7-Place Evaluation Scale

- **+3** ***Strongly Supports:*** System Element is an excellent solution or provides additional strong support toward fulfilling the Function
- **+2** ***Firmly Supports:*** System Element is a good solution or provides additional good support toward fulfilling the Function
- **+1** ***Supports:*** System Element is an adequate solution or provides some support toward fulfilling the Function
- **0** ***No Effect:*** System Element has no bearing on the Function
- **–1** ***Obstructs:*** System Element is not too good a solution or mildly interferes with fulfilling the Function
- **–2** ***Firmly Obstructs:*** System Element is a poor solution or actually obstructs fulfilling the Function
- **–3** ***Strongly Obstructs:*** System Element is a bad solution or seriously obstructs fulfilling the Function

Figure 6 **5-place and 7-place Scales for Scoring Relationships**

Elements are scored for how well they support the Functions. Because in some cases they may actually obstruct or impede fulfillment of Functions unintentionally, provision is made for both negative and positive scoring.

The Scoring Scale

Scoring is qualitative, conducted on a bidirectional, five- or seven-place scale (Figure 6). At the central point of this scale, a rating is neutral indicating that the System Element has no effect on the Function, neither supporting its fulfillment nor obstructing it. Positions on the positive side

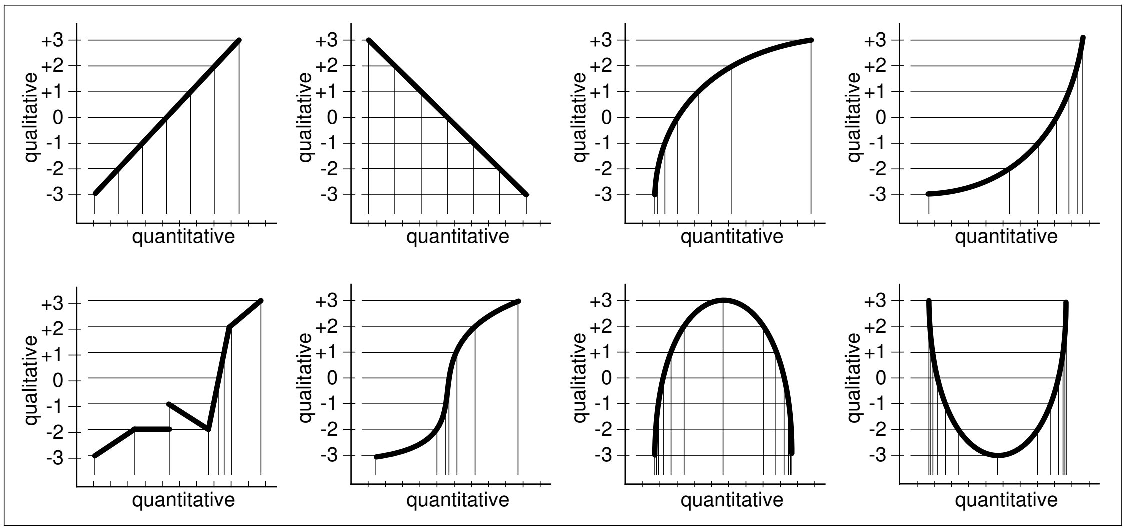


Figure 7 **Mapping Quantitative Data to a 7-place Qualitative Scale**

of the scale allow increasing degrees of support; positions on the negative side register increasing degrees of obstruction.

Although it requires additional processing, quantitative information, where it exists, may be incorporated in evaluations by mapping values to the qualitative scale. Figure 7 illustrates the process diagrammatically for a variety of classic relationships. The range of possible quantitative values for a System Element's performance relative to a Function is plotted against the 5- or 7-place scale being used for qualitative judgments, and the resulting plot is used to translate scores for the given System Element.

Although component elements of any system are designed to meet the needs of specific functions, it is not unusual that they have effects on other functions. The more complex the system, the more likely it is that there will be collateral effects of System Elements on Functions other than those for which they were intended. There may even be surprises, some of them undesirable. In any case, it is important to examine the full range of possible relationships and to recognize active associations, positive or negative, so that they may be directly addressed. The structure-against-structure formulation of the evaluative process guarantees this.

Graphic data collection forms of various kinds can be used to simplify the task of scoring. The form in Figure 8 was created for use in the example restaurant evaluation. The master form showing all Functions was multiply replicated with the title for each System Element added by hand. Focusing on the System Element allowed the analysis team to consider the System Element vs all Functions (rather than Function vs all System Elements) as the easier of the two ways to approach the scoring.

Accumulating Values for Functions and System Elements

$$V_{pos} = v_1^+ + \frac{1}{[(v_{max}^+ + 2) - v_2^+]^2} + \frac{1}{[(v_{max}^+ + 2) - v_3^+]^3} + \ldots + \frac{1}{[(v_{max}^+ + 2) - v_{n_{pos}}^+]^{n_{pos}}} \quad (1)$$

$$V_{neg} = v_1^- - \frac{1}{[(v_{max}^+ + 2) + v_2^-]^2} - \frac{1}{[(v_{max}^+ + 2) + v_3^-]^3} - \ldots - \frac{1}{[(v_{max}^+ + 2) + v_{n_{neg}}^-]^{n_{neg}}} \quad (2)$$

$$V_{final} = V_{pos} + V_{neg} \quad (3)$$

Values for Functions and System Elements are calculated similarly using an algorithm and equations that force accumulating values to be asymptotic to limits readily interpretable in terms of the scoring scale. Steps in the algorithm are:

1. If the rating scale does not have 0 for its neutral value, transform the scale and convert all values. For data entry purposes, it is convenient to use unsigned scales such as 1 - 5 or 1 - 7,

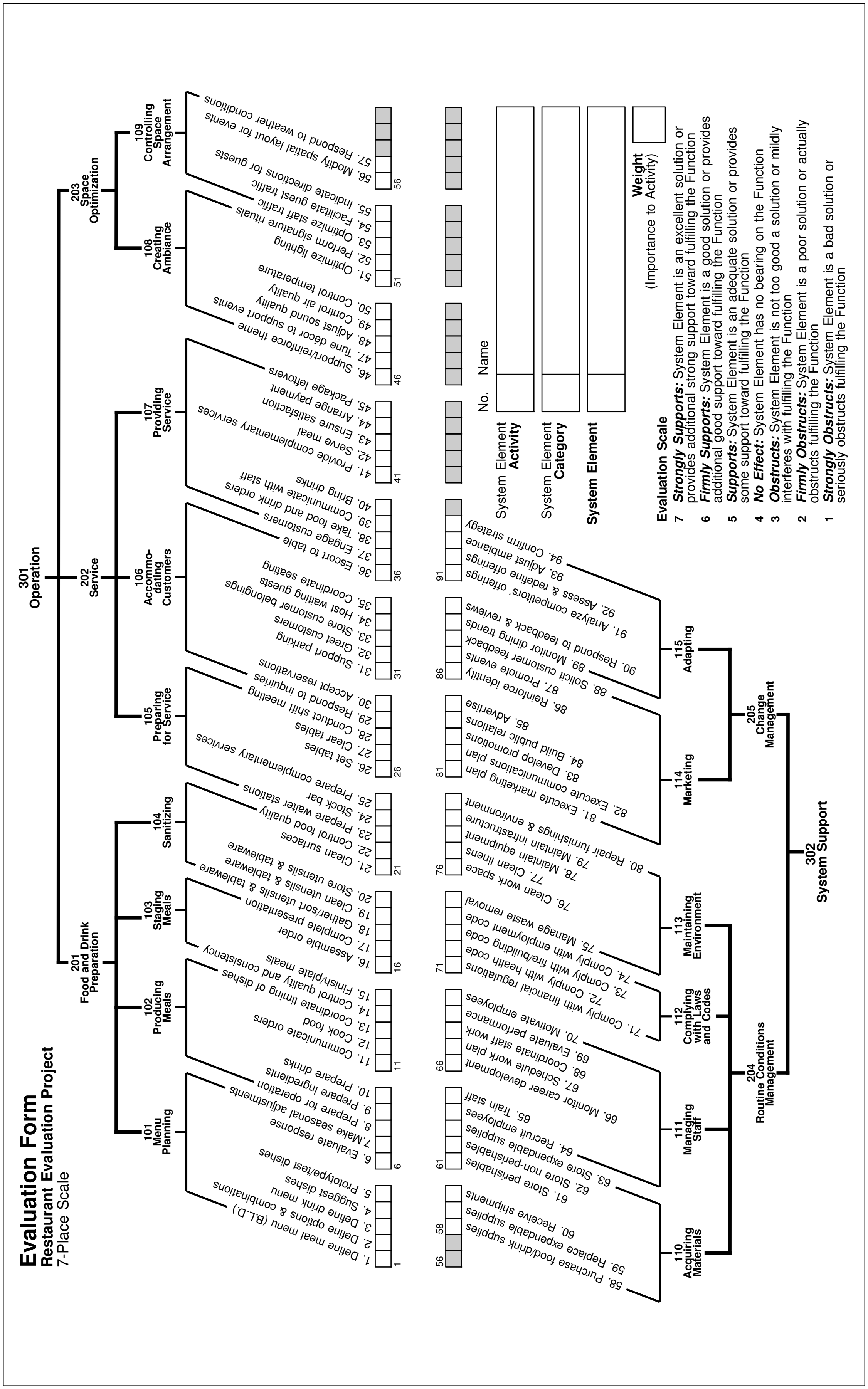
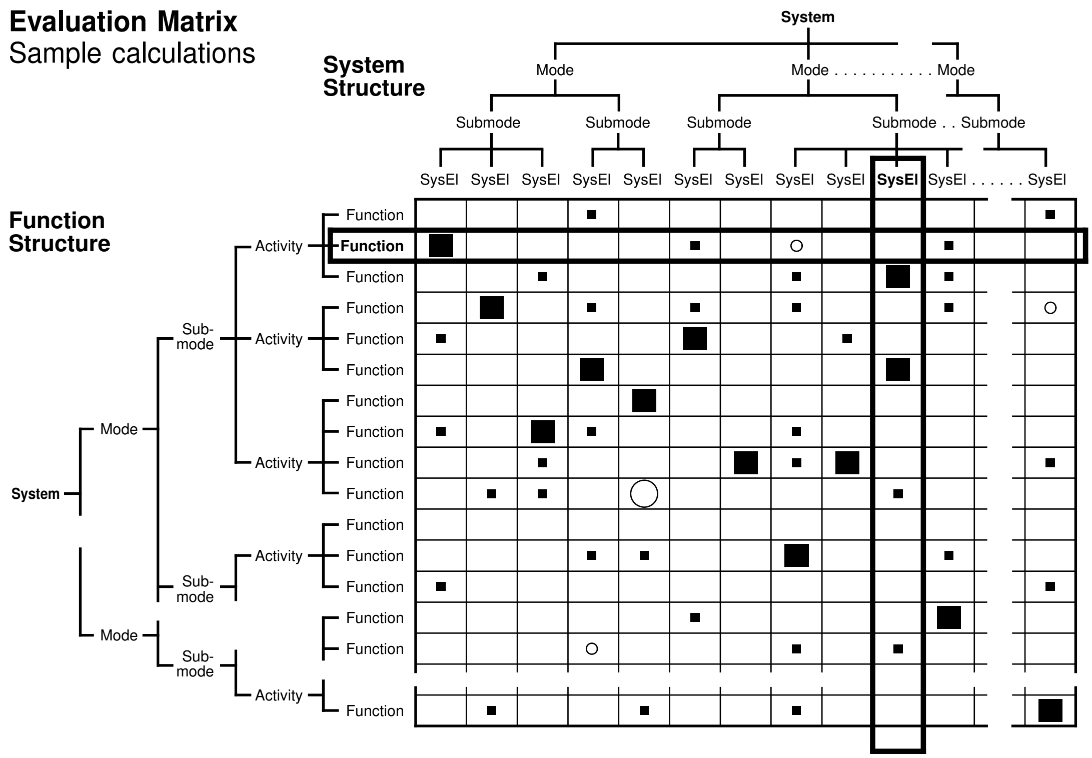


Figure 8 **Master Form for Scoring Function Data for a System Element**



Function:

$$V_{pos} = 2 + \frac{1}{[(2+2)-1]^2} + \frac{1}{[(2+2)-1]^3} = \boxed{2.148}$$

$$V_{neg} = \boxed{-1.000}$$

$$V_{final} = 2.148 + (-1.000) = \boxed{\mathbf{1.148}}$$

System Element:

$$V_{pos} = 2 + \frac{1}{[(2+2)-2]^2} + \frac{1}{[(2+2)-1]^3} + \frac{1}{[(2+2)-1]^4} = \boxed{2.299}$$

$$V_{neg} = \boxed{0.000}$$

$$V_{final} = 2.299 + 0.000 = \boxed{\mathbf{2.299}}$$

Figure 9 **Sample Calculations for the Value of a Function and a System Element**

where the neutral point is 3 or 4. Conversion is handled easily within the calculation section of the computer program. The maximum scale value is V^{+}_{max}.

2. Sort positive ratings, $\mathbf{v}^{+}$, highest first; sort negative ratings, $\mathbf{v}^{-}$, lowest first.

3. Calculate composite values for the positive series and negative series using equations (1) and (2).

4. Combine the composite values for the final value using equation (3).

For the positive calculation, the diminishing series limits the value to a value .5 above the maximum value given during scoring (+1 or +2 for a 5-place scale; +1, +2 or +3 for a 7-place scale). For the negative calculation, the series calculation limits the value to .5 below the minimum value given during scoring (–2 or –1 for a 5-place scale; –3, –2 or –1 for a 7-place scale). Values are calculated for each Function using the scores given it for each System Element. Values are also calculated for all System Elements using the scores given them for each Function. Because of the asymptotic behavior of the series calculation, a Function will have for its overall positive score a value no lower than its single highest score, and no higher than that value plus .5. Similarly, the negative portion of its score

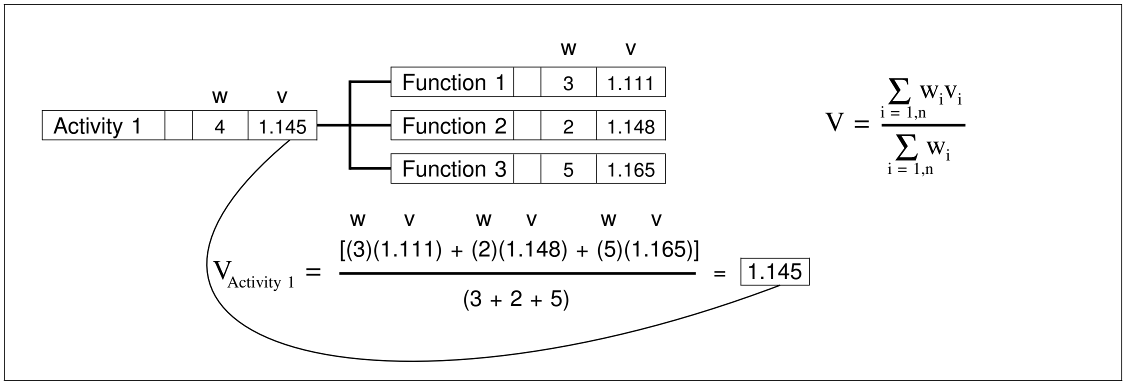


Figure 10 **Calculation of Nodal Values Using Weights**

will always be at least as bad as its single lowest score, and no worse than that value minus .5. This supports a policy that holds that a Function or System Element is served as well as the best and worst evaluations given for it. Subsequent positive or negative evaluations can add to or subtract from that in the series calculation, but not substantially enough to raise or lower the ultimate value to the next scale level. Then, when rated on a 7-place scale, a Function that receives a single –3 score can never achieve an overall rating of +1 no matter how high or how many positive scores it receives: similarly, a System Element that receives a single +3 score can never drop overall to the –1 level no matter how low or how many negative scores it receives.

If the values were simply averaged instead of using the series calculation, the contribution of a single strong system Element to fulfilling a Function could be quickly lost. For example, the set of scores {+3, +1, +1, +1} produces an average value of 1.5, where the series calculation produces 3.08. In the averaged calculation, the Function appears not to be well fulfilled, whereas in the series calculation, the single +3 score guarantees an evaluation of "strong fulfillment". Similarly, averaging {–3, –1, –1, –1} results in –1.5, quite different from the –3.08 calculated with the series equation. The series calculation is based on the concept that a function of a system needs only to be fulfilled once; multiple contributions to fulfillment are welcome supplements for redundancy and reliability, but proper evaluation should recognize and communicate the strength of the highest score.

When positive and negative results are merged, however, cancellation unavoidably obscures some of the transparency built into the series calculation. Fortunately, in practice, negative scores are usually scarce, far fewer than positive in almost any evaluation; but the situation does arise. To prevent inadvertent misinterpretation of evaluation scores, a "marking" system is built into the graphic presentation of final results. It will be introduced and discussed in the Communication section of this paper.

Using the data from the simplified example of Figure 6, the values for a sample Function and System Element are calculated in Figure 9: 1.148 and 2.299 respectively. The calculations are shown in Figure 9.

Hierarchical Calculations

Because of the hierarchical formulation of both evaluative criteria and the system under evaluation, it is possible to measure performance at nodal positions throughout both structures. This enables the analyst to view the system at varying levels of detail from the elemental to the all inclusive. At the apex, a single value integrates all values below into a Functional Merit Index, a Policy Merit Index or a System Merit Index.

Weights. Once constructed, a Function Structure and a System Structure archetype can be used to compare systems. In most cases, the only adjustments necessary are the instantiations in the archetype of System Elements specific to the systems being compared. If there are more fundamental variations between systems—usually caused by differences in the contexts in which the systems operate—weights can be employed to emphasize those aspects of functionality most relevant to the individual contexts.

Weights may be used at all structural levels to reflect relative importance. Various strategies may be employed for determining weights, but usually these will reflect the relative importance

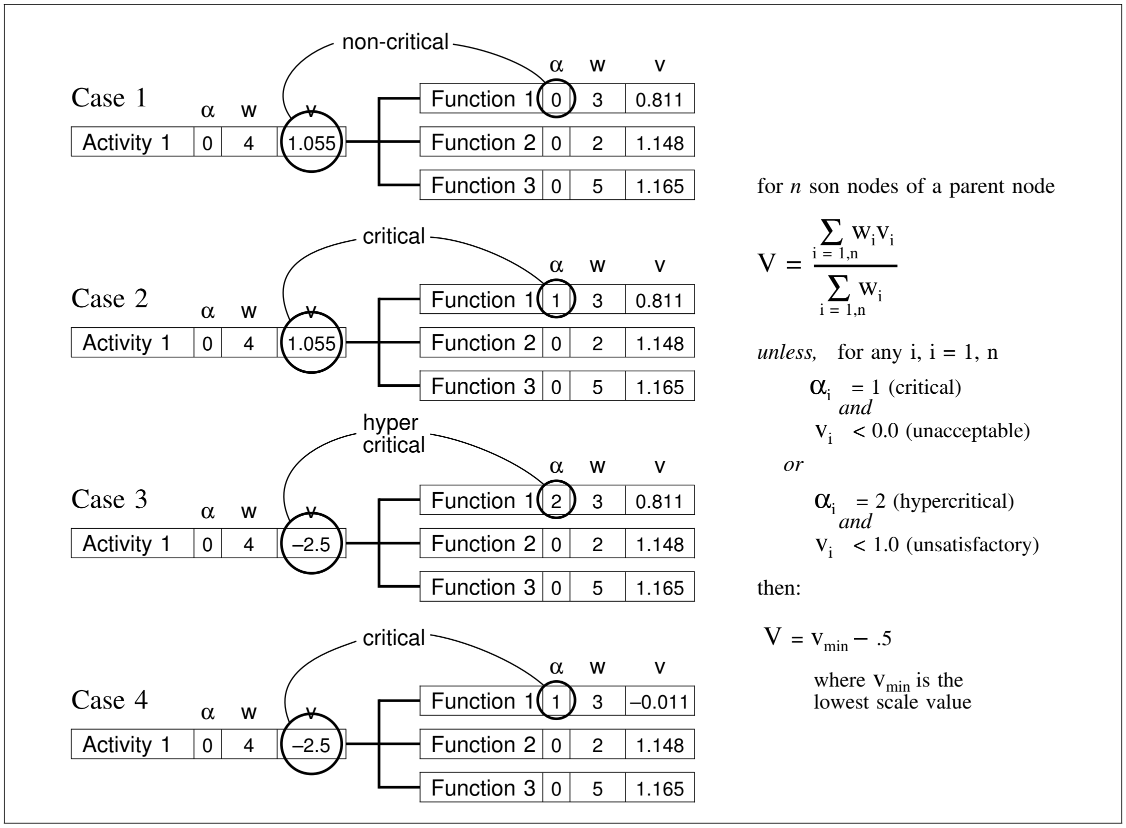


Figure **11 The Effect of Criticality on Values**

to each other of elements within a category, given the contextual circumstances.

Figure 10 diagrams the inclusion relationship between an Activity and its three member Functions. Each of the Functions has a value, **v**, calculated from the scores given it for System Elements. Each also has a weight, **w**, establishing its importance relative to the others in performing the Activity. The value for the Activity is calculated as a simple weighted average of the values for the Functions. The actual calculation for this example is also shown in Figure 10.

Critical Values. In many systems, certain component Functions or elements are critical. Their failure would be more serious than single-component failure and might cause failure of their parent processes, or even the entire system. To accommodate that possibility, the evaluation process has criticality factors, α, that can be assigned to any element of either the criteria or system hierarchies.

Criticality is assigned at three levels. At level 0 (the normal case) performance of the designated element is not critical, and its value will be averaged with other values no matter what its score. At level 1, performance is considered *critical*, and a value below 0.0 will cause the value of the parent process or category to go to minimum calculable value (–2.5 or –3.5) without regard for sibling component values. At level 2, performance is considered *hypercritical*, and performance below 1.0 is cause for the value of the parent process or category to be driven to minimum calculable value. In all cases, satisfactory performance results in the value being normally averaged with other sibling values no matter what the criticality factor.

Figure 11 diagrams the calculation process for an Activity with three Functions scored on a 5-place scale. In Case 1, α for Function 1 is set to 0 (off), and the three values for Functions 1, 2 and 3 are averaged with their weights normally using the equation shown to establish the value for Activity 1 (1.055). In Case 2, α is set to *critical* (1) for Function 1, but the function's performance is sufficiently high (> 0.0) that the activity value is still calculated normally. In Case 3, α is set to *hypercritical* (2), and the same value (0.811) in the zone between unsatisfactory (< 1.0) and unacceptable (< 0.0) causes the Activity value to go to –2.5, the lowest value possible on the 5-place scale. In Case 4, α set to *critical* (1) is sufficient to drive the activity value to –2.5 when Function 1's value drops to an unacceptable level (< 0.0).

Communicating the Evaluation

The complexity of this evaluative process is fully reflected in the tasks of calculation and communication. The process is only practical when supported by computer programs. A proprietary computer program, SYSEVAL, has been written to handle calculations and produce graphic renditions of the results in color. Output formats of several kinds are selectable by content and form.

Content Options

Three options are possible for content: a Functional Assessment, Policy Assessment and System Assessment. A Functional Assessment presents results in the form of the original Function Structure, showing values, weights and criticality factors for each Function and all higher level Activities, Submodes and Modes to an all-inclusive Functional Merit Index. For Policy Assessments, the structure is similar with Defining Statements at the bottom of the hierarchy and category descriptions continuing to a peak level Policy Merit Index. System Assessments have System Elements at the lowest level with their categories at one to three levels above and then Activities (optionally) followed by Submodes and Modes to a top level System Merit Index.

Form Options

Presentations are produced in color, following either a vertical or horizontal plan. Vertical hierarchies branch from the Merit Index at the left to the base level elements on the right arranged vertically along the right side of the structure. The vertical format can be produced as a continuous hierarchy (for roll printing of a vertically oriented, left-to-right structure) or paged for report generation on 8½-inch or A4 width sheets of paper. Horizontal hierarchies branch from the top down so that base level elements are at the bottom of a roll-printed sheet scrolling from left to right. For this format, a 24-inch (60 cm) paper width is used to allow more compact description of hierarchies. The limitation imposed by maintaining a legible type size guarantees lengthy horizontal or vertical "posters" for structures with more than 100 Functions, Defining Statements or System Elements. When large numbers of elements are involved in an evaluation, the resulting hierarchies can extend to over three meters.

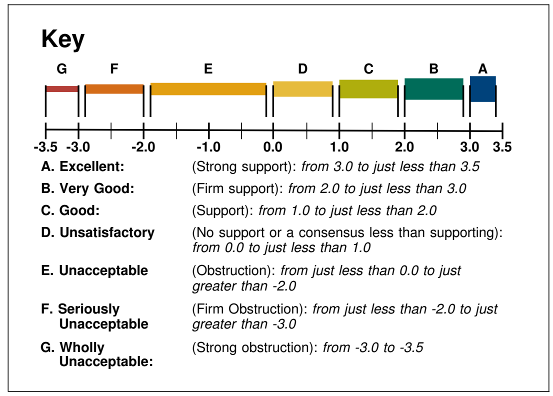


Figure 12 **Evaluation Scale with Color and Line Thickness Assignments**

Graphic Characteristics

Important aids to interpretation are color and line weight. Both are used to call attention to successes and failures within a hierarchy. Depending on the evaluation scale used, an assignment of one of five or seven colors is made to all blocks representing elements and lines linking blocks. The color spectrum is related to the evaluation scales as shown in Figure 12. Greens and blues signal successes; oranges and reds warn of problems. Yellow denotes an unsatisfactory condition, usually indicating conflict or uncertainty. The thickness of links is correlated to the values of the parent blocks they connect to offspring. Thicker lines trace success; thinner lines, problems.

Each block in a hierarchy has identification information: (1) type designation (Mode, Submode, Activity, Function, etc.), (2) number, (3) title, (4) criticality factor, (5) criticality ring, (6) weight, (7) value, and (8) value disc. These components are depicted and described in Figure 13.

Two special block types are reserved for potentially ambiguous situations (Figure 14). A red-flagged yellow block signals an entity for which there are no scores, positive or negative. If a Function or Defining Statement, that means that no System Elements support or obstruct its fulfillment. If a System Element, it means that it neither serves nor obstructs any Function. Its value will be 0.00, but that value will have been obtained by the absence of any effective score.

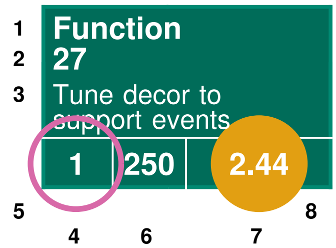
The second special block type is a green-flagged yellow block. It signals the deliberate choice of the analyst to take a category of the

System Structure out of the analysis, usually because the analyst has decided that the category is not relevant for a particular system for contextual reasons. By assigning a weight of 0 to the category, the analyst signals the program to ignore it in all hierarchical calculations. The block is shown in the output structure with its 0 weight and 0.00 value, but is flagged green.

Two other graphic additions support interpretation of results. First, a lavender ring is superimposed over a 1 or 2 (critical or hypercritical) rating shown for the criticality factor on any block. This helps the evaluator to note where criticality has been set and met or not met—particularly where it has been met, because there will be a color change to red only when it has *not* been met.

The second graphic aid is a colored "value disc" printed beneath the value on any block in which a positive value has been diminished by negative scores. By its color (following the color scale), it indicates the most negative score received. As discussed earlier, the cancellation of positive and negative scores conceals the strength of scores received; the value discs call attention to that. Blocks that have overall negative values are already prominent by their orange-red colors. Positive blocks, green and blue, and yellow blocks still in the positive range, receive value discs to call attention to the existence of negative scores.

Figures 15 and 16 show a sample Functional Assessment and System Assessment from the restaurant study in horizontal format. Structures are created as pdf's, intended for purposes of legibility to be scaled and printed so that the width of blocks is approximately 1 inch (2.5 cm).



1. Type Designation
Entity represented: Function, Activity, Submode, Mode, System, Category, Defining Statement, System Element, etc.

2. Number
Sequential numbers for Functions, Defining Statements and System Elements; level + number for higher level entities (e.g., **203** = level **2**, entity **03**)

3. Title
Name or action

4. Criticality Factor
0 = non-critical; 1 = critical; 2 = hypercritical

5. Criticality Ring
Alert signal for critical and hypercritical blocks

6. Weight
Relative importance of block to other blocks under same category

7. Value
Weighted total score for block calculated from all scores given for it

8. Value Disc
Color-coded signal that negative scores have been given for the block

Figure 13 **Components of a Block**

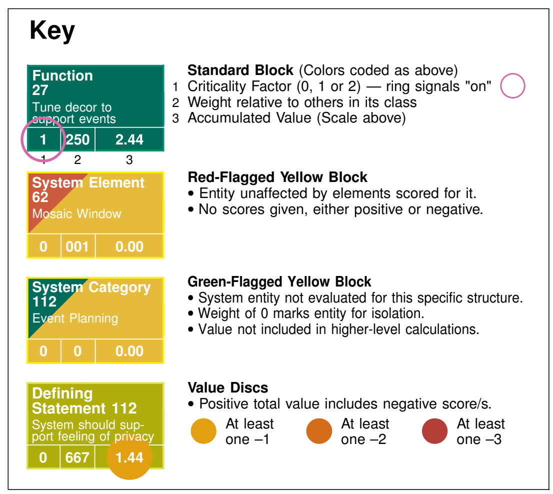


Figure 14 **Special Blocks and Graphic Interpretation Aids**

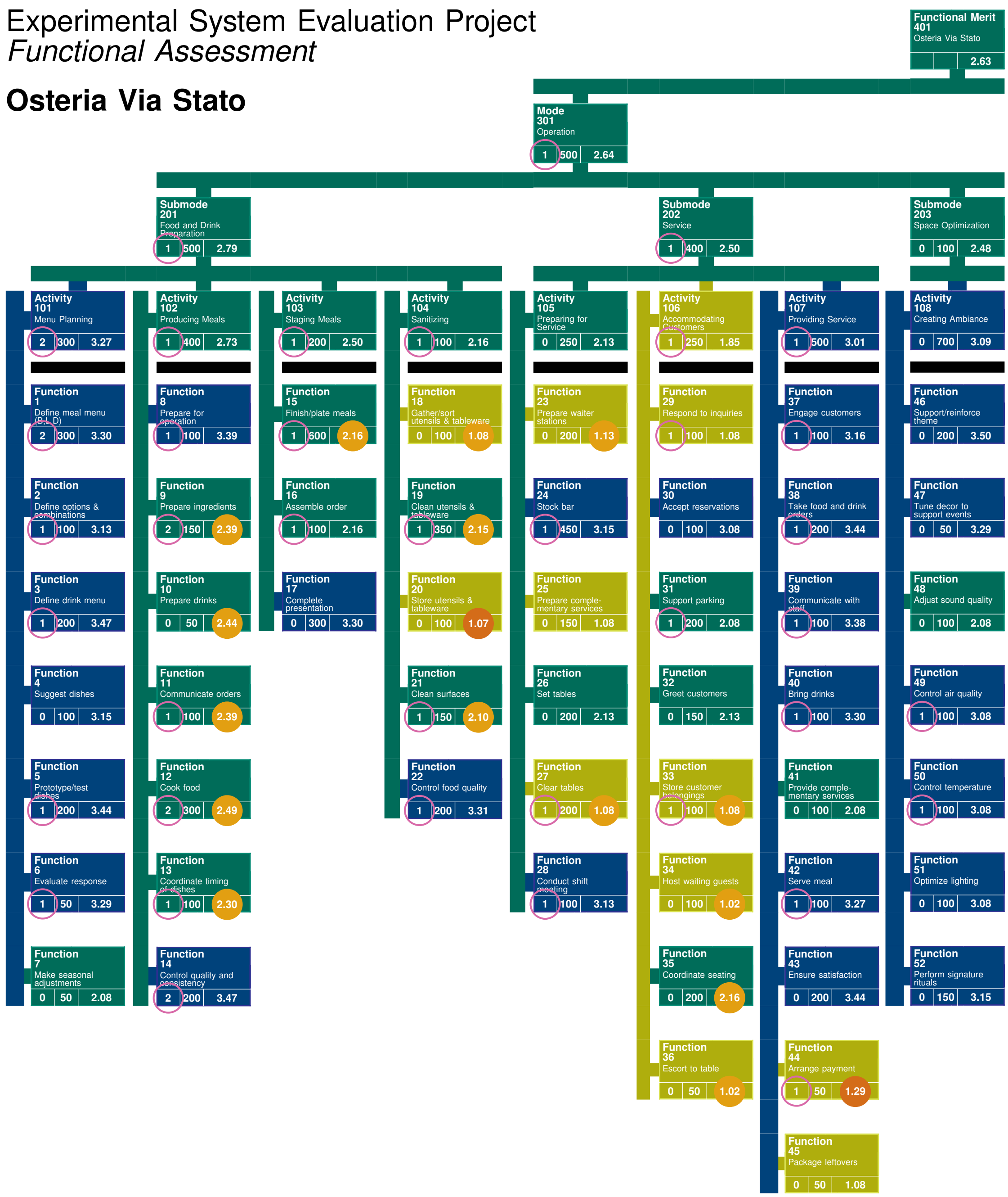


Figure 15 **Sample Functional Assessment in Horizontal Format (Left half)**

Discussion

With the cooperation of the Chicago-based Lettuce Entertain You restaurant organization, three midrange fine dining restaurants were evaluated by teams of three graduate students each from the Institute of Design. Testing was conducted to validate the evaluation model, explore potential problems associated with obtaining data and making judgments, and test the effectiveness of the hierarchical presentation scheme for communicating results. Management and staff in all restaurants were highly supportive.

Recommendations from Restaurant Tests

As expected, it was difficult for the evaluation teams to acquire the expertise advisable for high-quality independent judgments. Despite preparation with a range of good written materials on restaurant design, planning and operation, and the general knowledge available to all on the operation of restaurants, the teams quickly found that they needed the help of staff willing to

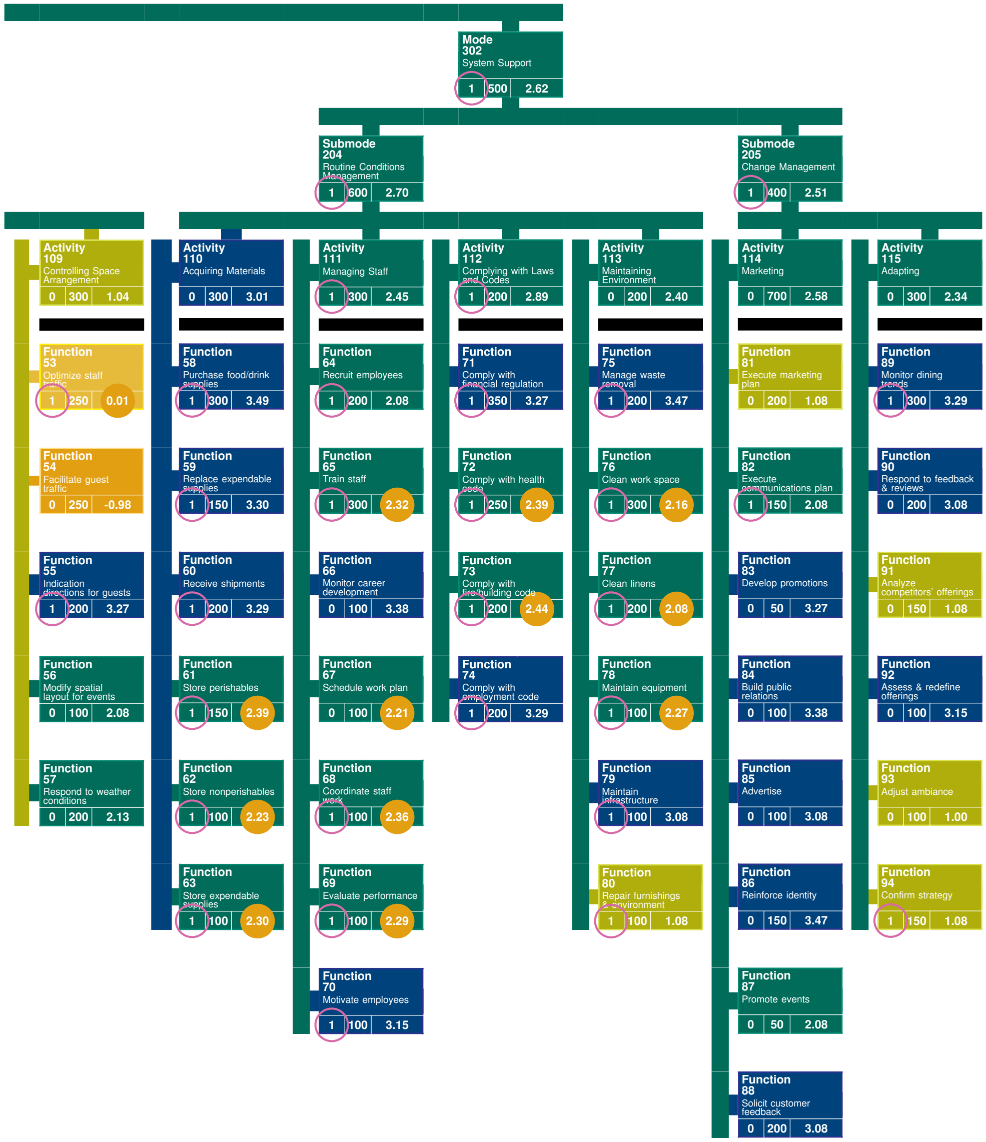


Figure 15A **Sample Functional Assessment in Horizontal Format (Right half)**

speak candidly about operations. Conventional observation techniques were not easily employed, and the value of deep system knowledge was quickly obvious. A clear recommendation was that reliable evaluations at this level of detail can best be made either by content experts or knowledgeable management conscientiously self-evaluating their systems.

Recommendations directed toward data collection and interpretation included the following, all now implemented in the process (and discussed above):

1. A 5-place scale was not granular enough to rate relationships optimally. A 7-place scale was added as an option.

2. A single level of criticality was not sufficient. Some Functions are critical, but less critical than others. Two levels of criticality (critical and hypercritical) have now been implemented. Because it is easy to miss seeing where criticality factors are "on" when there are many blocks and the value of a block is high enough to avoid triggering the criticality function, a contrasting

Experimental System Evaluation Project
System Assessment

Osteria Via Stato

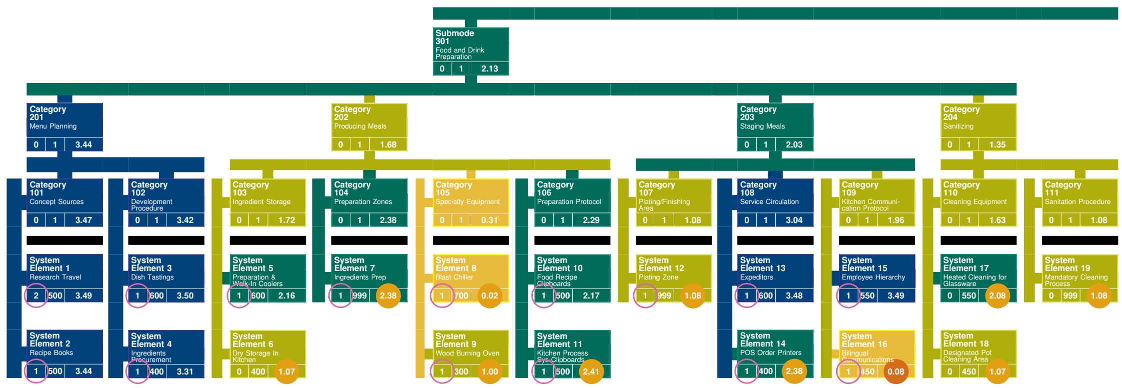


Figure 16 **Sample System Assessment in Horizontal Format (Segment 1)**

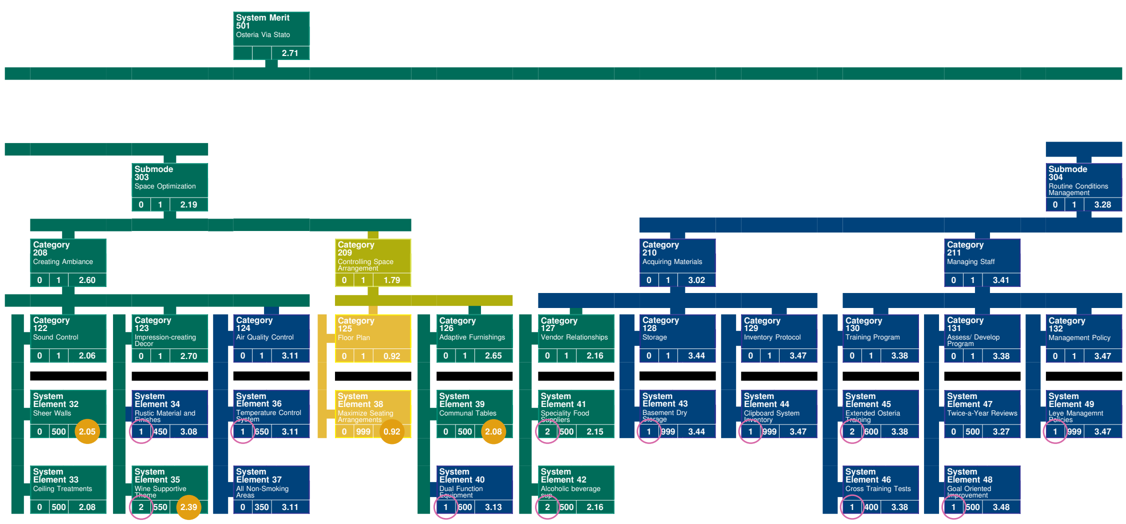


Figure 16B **Sample System Assessment in Horizontal Format (Segment 3)**

lavender ring is now superimposed over any criticality factor switched on.

3. The value 0.00 was sometimes ambiguous when given for Functions and System Elements because it can be attained in more than one way. For 0.00's attained by cancellation of equal positive and negative values, the yellow 0.00 block now remains as the correct assessment. For 0.00's attained because no positive or negative scores were given, a red-flagged yellow block was introduced (see Figure 14) signaling a Function or System Element either not served or not used. For 0.00's attained by offsetting positive and negative scores whose final value is so small that it rounds off to 0.00, the value is artificially adjusted to just show within the 3-place value window. This allows it to be seen to be numerically within its color region, in agreement with the block color triggered by the actual value.

4. Using the same System Structure archetype for multiple systems sometimes led to situations in which there were no System Elements to fill a category (usually because of different emphases in the system's context of use). A provision for a weight of 0 was added to allow System categories to be marked as knowingly unfilled and not included in calculations for higher level category

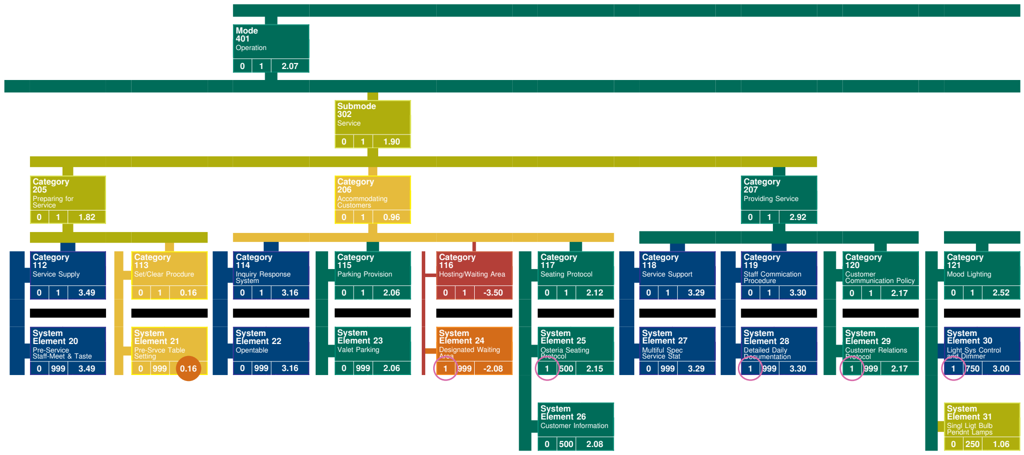


Figure 16A **Sample System Assessment in Horizontal Format (Segment 2)**

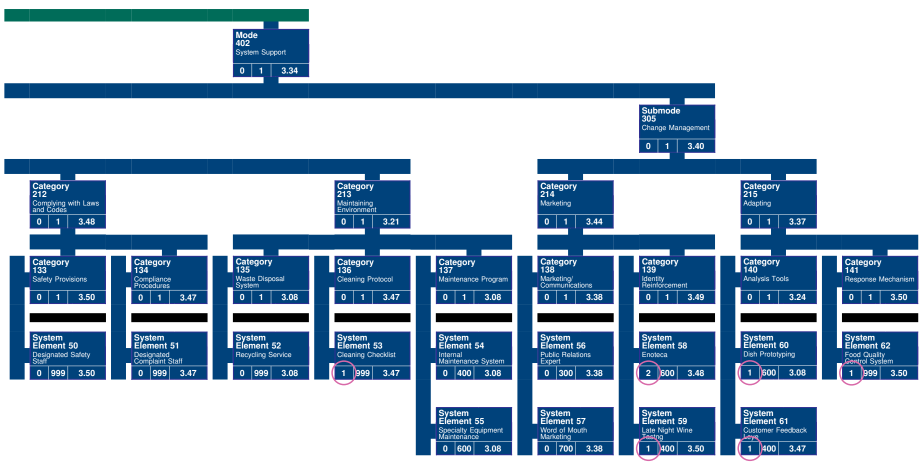


Figure 16C **Sample System Assessment in Horizontal Format (Segment 4)**

values. A green-flagged yellow block appears in the System Assessment structure (see Figure 14).

5. The effects of negative scores were directly observable in the assessment hierarchies when the color of blocks indicated a negative block value. Where negative scores simply lowered the aggregate positive value, however, it was not clear that they were responsible. To correct this, colored "value discs" were introduced on positively colored blocks where any negative score had been registered (see Figures 13 and 14).

Summary and Conclusions

Hierarchical system evaluation enables new levels of access to system performance. From the experiments reported in this paper, several insights can be drawn.

On the Good Side:

1. Hierarchical description of criteria permits analysis to be performed at levels of detail convenient to purpose. Function Structure methodology from Structured Planning provides a sound basis for both developing a structure and covering system functionality thoroughly.

2. Hierarchical description of the system as an archetypal framework similarly supports full coverage of the system as the object of evaluation. Subject to purpose and available human resources, coverage can be extended to as fine a level of detail as desired.

3. Measuring Functions or Defining Statements against all System Elements, and measuring System Elements against all Functions or Defining Statements provides a degree of comprehensiveness difficult to attain with other evaluation methods.

4. The series calculation method for establishing multivariable values for Functions, Defining Statements and System Elements returns values matched both to the measurement scale and an evaluation policy that regards component quality to be best represented by single strongest expressions of support or obstruction.

5. An evaluation model combining a standardized Function Structure with a System Structure archetype designed for specific instantiation is a powerful system comparison tool. The strengths and weaknesses of systems can be observed and compared at equivalent positions and at parallel levels of abstraction.

6. Weighting of nodes within hierarchical categories provides a mechanism for matching criteria and system descriptions to context, giving an analyst the flexibility to fit a standardized model to the subtleties of systems designed for niche environments.

On the Cautionary Side:

1. While it is true for any evaluation technique, for this process it is more important than ever that evaluators have thorough knowledge of their field. The breadth of coverage required and level of detail possible demand in-depth domain knowledge and experience for success. A team approach offers a good solution; a panel or team of specialists not only would increase the accuracy of judgment, but would reduce the individual scoring load (see below).

2. Many scoring decisions must be made. The usefulness of an evaluation is strongly related to the level of detail covered; to a certain extent, the more detail, the more thorough and better will be the results. For the three studies reported in this paper, the number of scoring decisions for each restaurant was approximately 6,000. For an evaluation with 150 criteria and 100 System Elements, the number would be 15,000. Surprisingly, the process is not as time consuming as might be expected. The scores, by far, are neutral "no effect"—as should be expected—and are readily recognized as such. But scoring does take some hours. All computation and graphic production processes (on the good side) are done by the SYSEVAL computer program.

Acknowledgments

The author thanks the Lettuce Entertain You Enterprises organization and the staffs of its restaurants for their support. In particular, thanks go to John Buchanan, President of the Lettuce Consulting Group, Mark Palicki, General Manager of Shaw's Crabhouse; Sarah Johnson, now General Manager of VTK; Ed Culleeney, General Manager of Ben Pao; and Randy Brand, General Manager of Osteria Via Stato.

Nine graduate students conducted the analyses in the test trials, developed the test data, and contributed recommendations for process improvements: Clinton Barth, Michael Beebe, Christian Bernard, Rebecca Frisch, Chunlun Lee, Taylor Lies, Rosa Manfredi, Kristina Marich and Jason Ring.

Special thanks are owed to Amgraf, Inc. maker of the MECCA2000 publishing software used by the visualization component of the SYSEVAL computer program. Its accuracy and flexibility made possible the precise graphics required in building the assessment structures. Finally, the management of Steelcase, Inc. is owed a debt of gratitude for asking the questions that initiated this research in the first place.

References

Bruno, P (Sep 3, 2004) *Chicago Sun-Times,* Weekend Plus p 22

Cross, N (2000) *Engineering design methods: strategies for product design* (3d) Wiley, Chichester

Ostrofsky, B (1977) *Design, planning, and development methodology* Prentice-Hall, Englewood Cliffs

Owen, C L (1992) Context for creativity *Design Studies* Vol 13 No 3 pp 216-228

Owen, C L (1993) A critical role for design technology *Design Management Journal* Vol 3 No 2 pp 10-18

Owen, C L (2001a) Structured planning in design: information-age tools for product development *Design Issues* Vol 17 No 1 pp 27-43

Owen, C L (2001b) Structured planning: advanced planning for business, institutions and government, *Institute of Design, Illinois Institute of Technology, Chicago*

Owen, C L (2005) Structured planning: advanced planning for business, institutions and government: notes on the process with summary pages and examples *Institute of Design, Illinois Institute of Technology, Chicago*

Pahl, G and Beitz, W (1996) *Engineering design: a systematic approach* (2d) Springer-Verlag, London

Roozenburg, N F M and Eckels, J (1995) *Product design. fundamentals and methods* Wiley, Chichester

Saaty, T L and Vargas, L G (2000) *Models, methods, concepts & applications of the analytic hierarchy process* Springer-Verlag, London

Woodson, T T (1966) *Introduction to engineering design* McGraw-Hill, New York