Validation of Organizatinal Engineering Instrumentation and Methodology

By: Robert Soltysik

Contents

Summary of the Organizational Engineering Validity Findings

(At a $p < .05$ significance or better where applicable)

	LEVEL	
	Individual	Group
Face Validity	✓	✓
Construct Validity	✓	N/A
Content Validity	✓	N/A
Convergent Validity	✓	✓
Discriminant Validity	N/A	✓
Concurrent Validity	✓	✓
Predictive Validity	✓	✓
Conclusion Validity	✓	✓
Reliability	✓	N/A

INSTRUMENT AND METHODOLOGIES

SUMMARY Organizational Engineering survey instrumentation (DecideX®, "I Opt"™), group consolidation methodologies (TeamAnalysis™), leader-group assessments (LeaderAnalysis™, OrgAnalysis™) and two person comparisons ("One-to-One"™, TwoPerson Analysis™) are addressed in this validation study.

The subject of this study is a theory specified in the books Organizational Engineering (Salton, 1996) and in *The Managers Guide to Organizational Engineering* (Salton, 2000). The theory identifies behavioral outcomes arising from strategic information processing choices. Observable behavioral effects arising from measurements made using the survey instrument (see below) have been codified in computer programs. The output of these programs creates hypotheses that this study subjects to statistical validation.

In lay terms, *validity* is a line of reasoning providing systematic evidence that the subject of validation (instrument, methodology, etc.) really works in the dimensions tested. The professional definition of validity appearing in the *Standards for Educational and Psychological Testing* (1985)—usually referred to as the APA Standards—is as follows.

> *Validity is the most important consideration in test evaluation. The concept refers to the appropriateness, meaningfulness, and usefulness of the specific inferences from the test scores. Test validation is the process of accumulating evidence to support such inferences. (p. 9)*

Both the lay and professional definitions require that the subjects of the study be defined. The instruments and methodologies addressed in this study are:

Survey Instrument

The basic data-collection instrument is a 24-question survey that is available under the trademarks of DecideX® and "I Opt"™. Evaluation of the responses is accomplished by a proprietary algorithm. A copy of the instrument appears in Appendix 4.

TeamAnalysis™ Methodology

TeamAnalysis™ is a methodology that consolidates individual instruments using proprietary algorithms to obtain an overall representation of a group of people.

LeaderAnalysis™ Methodology

LeaderAnalysis™ and OrgAnalysis™ are trademarks for a methodology that consolidates individual instruments using proprietary algorithms to obtain an overall representation of a group of people. It then contrasts the individual group members and the group as a whole to a leader. The output is an assessment of divergences and synergies.

Two-Person Analysis™

"One-to-One"™ and TwoPerson Analysis™ are trademarks for a methodology that consolidates two people. It uses proprietary algorithms to assess the probable divergences, synergies, opportunities and exposures inherent in a common, goal-directed relationship involving the people assessed.

Validation Issues

The evaluations, assessments, findings and validations made are confined exclusively to the above instruments and methodologies. Any other evaluative expressions of the overall theory of Organizational Engineering other than those specifically identified will require separate or supplemental validation.

Validation always occurs within a specified scope. An instrument that has been "validated" in a classroom is valid within that context.

It may or may not be valid in a larger context. This analysis addresses validation from data collected in field settings. It draws upon a body of data that was accumulated from all organizational types (e.g., non-profit organizations, corporations, institutions, government, etc.) and across a wide variety of industries within the United States of America. (This is specified more fully in Appendix 2.)

Similarly, validation of a tool at an individual level does not in any way imply that it is valid at a group level. In the case of this study, validation is tested on both an individual and group level where appropriate. It is incumbent upon the user to determine whether the instrument and/or methodology are validated for the purposes within which it is intended to apply.

In addition to scope, validity studies are confined by their focus. Validation of construct validity says nothing about predictive validity. In other words, a methodology may have a high construct validity but be useless in its ability to forecast outcomes into the future. This study has made an attempt to test Organizational Engineering against all accepted forms of validation. The reader can select those elements (or combination of elements) that are relevant to his or her interests.

Finally, portions of this study rely upon judgments of a panel of 50 experts. This is an accepted strategy that must be relied upon under certain conditions. The reader can review the qualifications of the panel in Appendix 3. The reader is encouraged to review this section to make certain that its composition is appropriate to the uses to which the reader intends to address.

The findings of the study are specified in the individual sections of this report. The reader should refer to those sections before making a judgment relative to their intended application.

Face Validity

SUMMARY A expert panel of 50 professionals administered 14,655 surveys and found disagreement with the survey report in less than 1% of the cases (n = 128, 0.87%). The group based TeamAnalysis™ was tested by 44 experts in 921 administrations and was found to be inaccurate in less than 1% of the cases (n = 1, 0.1%). The face validity of both the instrument and the consolidation methodology as represented by TeamAnalysis™ is judged to be very high.

Face validity of a theory refers to results that have the appearance of truth or reality (Polkinghorne, 1988). It is often considered "useless" by psychometricians because of its vagueness (Cronbach, 1971). However, in field applications exactly the opposite is the case—face validity is one of the more important aspects of an instrument.

In field applications a substantial cost increase can be expected if an instrument does not have strong face validity. Both the administrator and the respondent must devote time to reconciling differences in judgment before learning based on that instrument can occur. In addition, a portion of these contests can be expected to fail and potential benefits that might have accrued from the use of the instrument can be lost. Thus, while face validity may be unimportant to a laboratory scientist, it can be of paramount importance to the field practitioner who must navigate the shoals of budgets and limited resources.

A high face validity also allows the tool to be applied on a broad scale. Ready acceptance means that fewer resources need be provisioned for administration. This increases the potential frequency of

application. In addition, a high face validity opens the possibility of new venues (e.g., distance learning) that can substantially increase the reach of an initiative. Thus high face validity allows an organization to enjoy higher levels of the benefits accruing through the use of an instrument.

The face validity of the Organizational Engineering survey instrument was tested by referencing the panel of experts. They were asked the following questions:

1. *Approximately how many people have you given the survey instrument to?*

2. *About how many people have claimed the instrument to be grossly inaccurate?*

3. *About how many people had substantive disagreements with elements of the report?*

The responses were tabulated and are presented for review in Table 2.

Table 2

EXPERT PANEL INDIVIDUAL SURVEY RESULTS

Number of Experts Participating	50	
Total Surveys Administered (Quesion 1)	14,665	100.00%
Face Judged Grossly Inaccurate (Question 2)	52	.035%
Face Judged Somewhat Inaccurate (Question 3)	76	0.52%

The degree of face validity of the survey instrument is high. Less than 1% of the respondents found the results of the individual survey to be less than accurate on either a gross or marginal scale. The strength of this response requires no statistical test or ratio-based assessment. The reports generated by the survey instrument are

judged to have extremely high face validity.

Face validity was also tested on TeamAnalysis™ reports administered by the expert panel. TeamAnalysis is a 25 to 35 page report that consolidates individuals comprising a group. It identifies structural vulnerabilities and strengths that arise from the interaction of individual members. Sizes of the groups assessed range from 3 to 24 people with an average of 8.9 participants. An expert judgment of the face validity of the group instrument was obtained by asking the following questions:

How many TeamAnalysis assessments do you estimate you have performed? _____

In your best estimate, what proportion of these groups agreed that the *group* behavior described in the TeamAnalysis™ was:

Highly Accurate _____%
Reasonably Accurate _____%
Inaccurate _____%

The results of the consolidation of the responses are presented in Table 3. The number of experts replying did not total to 50 because some do not use the TeamAnalysis™ tool in their practice.

Table 3

EXPERT PANEL TEAMANALYSIS FACE VALIDITY RESULTS

Number of experts responding	44	
Total TeamAnalysis assessment conducted	921	100.0%
Highly Accurate	755	82.0%
Reasonably Accurate	165	17.9%
Inaccurate	1	0.1%

Of the 921 TeamAnalysis™ administrations, only one group was reported to have deemed the report inaccurate. A total of 99% of the groups judged the TeamAnalysis to be accurate, with 82% deeming it to be Highly Accurate and the 17.9% balance judging it to be Reasonably Accurate.

The number of experts participating, the large number of administrations, and high rate of positive acceptance argue strongly for assigning a high level of face validity to group based TeamAnalysis™ report.

In summary, both individual and group-based Organizational Engineering reports display a high level of face validity as measured by the responses of experts.

Construct Validity

SUMMARY Statistical evidence in the context of differential population methodology was applied to three occupational categories involving 75 distinct groups and 887 people, which were compared to a database population (N ~ 8,700). The findings are statistically significant at the .05 standard adopted in this study (p = .0152). In addition, the theory's use of only a single assumption minimizes exposures from undefined assumptions inherent in any theory. Overall, Organizational Engineering appears to meet or exceed the standards of construct validity within the discipline.

A *construct* is some postulated attribute of people, assumed to be reflected in test performance. In test validation the attribute about which we make statements in interpreting a test is a construct (Cronbach and Meehl, 1955).

> *Construct validity is ascertained by investigating...what the test score tells us about a person. [The] investigator asks, "From this theory, what hypotheses may be made concerning the behavior of individuals with high and low scores?" Inferences based on the evidence are then made concerning the theory's adequacy to account for the collected data. (Karmel, L.J. & M.O. Karmel, 1978)*

Popham offers three general types of construct validation studies (Popham, 1990). Intervention studies attempt to show that examinees will respond differently to a test after receiving some sort of treatment. This is not appropriate in this case, since Organizational Engineering does not attempt to change an individual's strategies, but rather to make use of the ones that are currently favored.

Related-measures studies show positive or negative correlations between examinees' scores on the target instrument and their scores on other measures. Since Organization Engineering is a seminal work without precedent, this is an inappropriate strategy for demonstrating construct validity. In seminal works, there is nothing with which to directly compare.

Differential population studies show that examinees representing distinctly different populations will score in predictably different ways on the instrument. This is a viable validation strategy for this study, since the theory of Organizational Engineering implies that certain styles will be favored by particular activities.

For example, information technology (IT) groups (e.g., systems analysts, programmers, software engineers, etc) share a common, highly complex environment. Success (if not survival) favors the highly structured thought-based style of Hypothetical Analyzer (HA). Therefore, Organizational Engineering theory is consistent with the testable hypothesis that groups engaged in IT are more likely to measure strongly in the disciplined, thought-based strategic style of HA than would the general population.

Therefore, a viable strategy is to compare the measurements of IT professionals with the rest of the population on this HA attribute. The classical test for this purpose is Student's unpaired t-test, which requires normality for each group used in the test. Well known parametric procedures such as the t-test and analysis of variance (ANOVA) require that the data be normally distributed, and that the variances of the populations involved be homogeneous. It is frequently claimed that these parametric procedures are robust in the case when these assumptions are violated. According to Thomas, Nelson, and Thomas, however:

> *Even if data are not normally distributed, researchers have often been taught that parametric statistical techniques are robust to violations of the normality assumption. Yet, there is concern among statisticians about whether parametric statistics are actually as robust to nonnormality (and heterogeneity of variance) as once thought. (Thomas, Nelson, and Thomas 1999).*

Stephens' test was employed to test the hypothesis of normality in the large (N = 8387) non-IT population. The null hypothesis of normality was rejected ($T = 6.7843$, $p < .01$), thus requiring the use of nonparametric procedures.

Rather than comparing means (as in the case of the t-test), the Mann-Whitney U test compares the medians of two groups. It is a rank-based method, requiring no assumptions other than that the measurements in the groups be independent and identically distributed. The Ansari-Bradley test was employed and found no evidence for different dispersions in the two populations ($p = 0.904$) indicating that the Mann-Whitney test (one-sided) is an appropriate nonparametric procedure.

The results of the Mann-Whitney test indicated that the population of people in the IT category, measured on the Hypothetical Analyzer attribute, differed significantly from the general population hypothesized direction indicated at the .05 alpha level ($Median_{(IT)} = 14.6$, $Median_{(population)} = 14.5$; $U = 1.50*10^6$, $p = .0152$).

Customer service offers another opportunity for a definitive test of construct validity. The customer service function involves resolving customer issues within a framework provided by the sponsoring organization. Representatives are allowed to offer certain solutions and precluded from offering others. Therefore, the theory would predict that groups engaged in customer service are likely to measure more strongly in the disciplined, action based strategic styles of Logical Processor (LP) relative to the population in general.

The Ansari-Bradley test found evidence for different dispersions in the two populations ($p = .0106$). This means that the Mann-Whitney test could not be used. Rather, a median test was employed since it does not require equality of dispersions. The test

was carried out in the following manner. The groups were pooled, and the median of the attribute was computed. A contingency table was created, with the rows corresponding to observations measured greater than or less than or equal to the median. The columns corresponded to the group membership of the observations. Fisher's exact test was then applied to this contingency table to test the null hypothesis that the medians of the two groups were equal. The results indicated that the customer service LP quality varied from the database population in the hypothesized direction at the .05 significance level ($Median_{(Customer\ Service)} = 18.7$, $Median_{(population)} = 14.5$; $p < .0001$). This finding reinforces the evidence for the construct validity of the underlying theory.

Research and development groups provide a third opportunity for contrast. R&D is charged with devising new products and methodologies. The predetermined approaches of the structured styles are clearly inappropriate for success (or survival) in this activity. Thus a testable hypothesis for this group would be that they are more likely to display salience in the unpatterned strategic style of Relational Innovator (RI) than is the population in general.

The Ansari-Bradley test found no evidence for different dispersions among the two groups ($p = .607$). The Mann-Whitney test is therefore appropriate and found a statistically significant difference in the RI dimension between the two populations in the predicted direction at the .05 alpha level ($Median_{(R\&D)} = 13$, $Median_{(population)} = 10.3$; $U = 541336$, $p < .0001$).

It should be noted that the probability of making at least one Type I error (rejecting the null hypothesis when it is, in fact, true) increases with the number of contrasts performed. A *family* of contrasts consists "of all contrasts of interest that are associated with a particular treatment or interaction" (Kirk, 1982). For purposes of assessing the current differential population study, the three foregoing contrasts were considered as a family. Consequently, the Dunn-Sidak procedure (Kirk, 1982) was employed in a effort to reduce this risk. Since the rejection level adopted in this study is $\alpha = .05$, the familywise criterion for rejection of the null hypothesis at this level for $C = 3$ contrasts is

$$\alpha_{FW} = 1 - (1 - \alpha)^{1/C} = .01695.$$

It is evident that all three results presented in the section are statistically significant by the standards of this familywise criterion.

The results of the differential population studies approach, summarized as to their focus in Table 4, has uniformly demonstrated a correspondence between the constructs of the theory and the predictions at the .05 level of significance or better. This finding provides a high degree of assurance of the construct validity of Organizational Engineering theory.

Construct validity can also be approached at a purely theoretical level. "The principle (of Occam's Razor) states that one should not make more assumptions than the minimum needed. ... Occam's razor helps (by reducing the) ... chance of introducing inconsistencies, ambiguities and redundancies" (Heylighen 1997). The more assumptions required by a theory, the weaker is the theory and the less faith that can rationally be accorded it. Essentially, each assumption can be viewed as an opportunity for error—the fewer the assumptions, the fewer are the opportunities for error.

Organizational Engineering (Salton, 1996, 2000) requires only that the reader accept the proposition that human beings are information processors. From this proposition, all of the qualities reported by the instrument are derived. The reader of the theory can apply his or her standards of logic to the acceptance or rejection of the derivations from this single premise.

Table 4

LISTING OF OCCUPATIONAL GROUPS ASSESSED

Function	Groups	People	Database
Information Technology	35	334	8387
Customer Service	30	455	8266
Research & Development	10	98	8623

An example may help clarify the above proposition. The Myers-Briggs paradigm requires that the reader accept that the human mind can be categorized into "eight possible preferences—two opposites for each of the four scales" (Hammer, 1991, p.7). While these assumptions may be true, each one offers an opportunity for error. Organizational Engineering requires only a single proposition, thus leaving fewer opportunities for masked errors.

In summary, the statistical evidence provided in the context of the differential population study provides strong evidence of construct validity at the .05 level of significance. This finding is reinforced by the minimal assumptions required by Organizational Engineering relative to alternative theories of organizational development.

Content Validity

SUMMARY Content validity is more a matter of logic than of statistics. However, a nomological net demonstrates that between 84% and 92% of the survey responses can be directly traced to specific dimensions of the underlying theory. In addition, 100% of the 50 members of the expert panel agree that the response structure incorporated in the survey is not contaminated by respondent misunderstanding. These findings suggest that the content validity is at least equal and perhaps superior to other theories within the discipline.

Content validity is concerned with sample-population representativeness (Cronbach, 1971). It is sample-oriented. Behavior is viewed as a sample when it is a subgroup of the same kind of behaviors of the larger population (Goodenough, 1949) which is the real focus of interest.

For example, computer literacy includes skills in operating systems, word processing, spreadsheet, database, graphics, the internet, and more. However, it is difficult, if not impossible, to administer a test covering all aspects of computing. Therefore, only selected tasks are "sampled" from the population of computer skills (Cronbach, 1971). It is *inferred* that the sample is representative of the larger body of skills labeled "computer literacy". This process is based on generalization—a form of inferential logic. The larger population of computer skills are "inferred" from the results of the sample.

The reason for pursuit of content validity is to insure that the judgments made on the basis of the instrument are truly appropriate to

the underlying theory or concept. In the case of the computer literacy example, omitting items such as typing and Internet skills may lead to errors of judgments that are based on the findings. In other words, content validity can be seen as concerned with the applicability of the instrument to its intended and actual use.

Organizational Engineering fundamentally differs from other theories in its area in that it is deductive in structure. It postulates that all information processors, including humans, must locate themselves at a point in the method and mode dimensions. The survey instrument is designed to locate and measure these dimensional preferences.

For example, one of the responses in the survey is "I make plans." It is set against the alternatives of "I complete the things I start," "I respond fast," and "I imagine things." Each of these responses implies a particular preference on the method-mode continuums. If the respondent makes repeated selections placing himself or herself in the same position, the theory claims justification in assigning a systematic preference to that posture.

It is reasonable to interpret content validity as pertaining to the correspondence of the survey responses to the underlying theoretical constructs. The behavioral inferences are then derived from the theory and are validated in other sections of this study (e.g., construct validity, concurrent validity, discriminant validity, predictive validity, etc.). These other portions of this study validate the behavioral correspondence between the theory and the resultant behavior or:

Theory ⟶ Behavior

Demonstrating correspondence of the content of the survey to the theory would extend the linkage. This correspondence between the theory and its representation in the survey can be tested by following a variant of Cronbach and Meel's nomological protocol (Cronbach and Meehl, 1955). They argue that theoretical constructs can be related to observables, thus creating a network of theoretical constructs, observables, and relationships. "This network would include the theoretical framework for what you are trying to measure, an empirical framework for how you are going to measure it, and specification of the linkages among and between these two frameworks" (Trochim, 1999a).

Using a variant of this procedure, the author "mapped" the survey responses back to the underlying method and mode dimensions of the theory. The intent of this exercise was to gain assurance that the responses directly related to the concepts in the underlying theory. A positive finding provides assurance of the fidelity of the survey to the theory. This, in turn, provides assurance that the judgments made on the basis of the survey represent the intended domain of the theory. The results of this codification are outlined in Table 5.

Table 5

ORGANIZATIONAL ENGINEERING NOMOLOGICAL NET

Reponses attributable to	Strategic Styles RS	LP	HA	RI
Method or Mode	12	13	12	11
Method and Mode	10	9	8	9
Inference	2	2	4	4
Percent Inference	8%	8%	16%	16%
Percent Direct	92%	92%	84%	84%

The categories of Method *OR* Mode mean that the survey response can be directly traced to either the method or mode component of the theory. For example, the response "I'm logical" directly speaks to the theoretical method concept of structure.

The category Method *AND* Mode differentiates the response on *both* the method and mode simultaneously. For example, "I react fast" speaks simultaneously to unpatterned method as a source of resolution strategy and the action mode as the direction of response.

The final category, inference, represents responses not *directly* traceable to underlying theoretical constructs. For example, the response "I'm playful" is a probable attribution to behavior using the unpatterned-action strategy of the RS strategic style. However, other interpretations are also plausible and the response was categorized as inferential.

About 84% to 92% of the responses on the survey were *directly* traced to the underlying theoretical construct as judged by the author

of this study. It is recognized that this finding is judgmental rather than definitive. However, the reader also has access to the instrument (Appendix 4) and the fully specified theory (Salton, 2000). This transparency allows the reader to construct his or her own Nomological Net and validate or dissent from these findings. While differences in specifics may arise, it is the author's opinion that the same overall results will be obtained. Therefore, it may be concluded that the following relationships have been demonstrated:

Survey Instrument ⟶ Theory ⟶ Behavior

The final step in this process involved making sure that the instrument itself was understandable to the respondents. To this end, the expert panel was asked:

In your opinion, do the responses contained in the Survey represent reasonable trade-offs that are understandable to the respondent?

No____ Yes ___

A response of "yes" would indicate that the resultant answers were not contaminated by misunderstandings. The results of this query are provided in Table 6.

The unvarying expert judgment is that respondents understand the "trade-offs" that are asked. This lends support to a judgment that the results are uncontaminated by misunderstanding of the questions asked. This result allows the respondents to be added to the chain of reasoning.

Table 6

EXPERT PANEL ESTIMATE OF UNDERSTANDABILITY

	Number of Experts_____		
Question	Yes	No	No Response
"...understandable to the respondent?"	50	0	50
Percentage	100%	0%	0%

Respondent ⟶ Survey Instrument ⟶ Theory ⟶ Behavior

In final analysis, content validity is more of a matter of judgment rather than of statistics. "In content validity, you essentially check the operationalization against the relevant content domain for the construct" (Trochim, 1999b). In the author's view, the content validity of Organizational Engineering is at least as valid as other theories in the area, and perhaps stronger than most.

CONVERGENT VALIDITY

SUMMARY Convergent validity was tested by comparing 19 plants of the same character involving 188 people. Individuals at the 19 plants were tested on all four strategic styles, and every test and associated multiple contrast performed failed to find any differences as a result of location at the standard $p < .05$ level of significance, providing evidence for convergent validity.

Convergent validity is a facet of internal validity and has variously been considered a component of construct validity and predictive validity. The concept of internal validity can be traced to Donald Campbell who argued that to be valid, a construct must be triangulated to insure that it is, in fact, a truly distinct construct and not a ill-defined component of another variable (Campbell & Fiske, 1959). In other words, it is a way to insure the clarity of thought.

Convergent validity tests that "measures of constructs that theoretically ***should*** be related to each other are, in fact, observed to be related to each other (that is, you should be able to show a correspondence or ***convergence*** between similar constructs)" (Trochim, 1999b). Convergent validity is typically tested by comparing a particular index with another index that is external to the theory but which purports to measure the same or a similar construct. For example, a test score in addition might be correlated to another test that measures ability in subtraction. Theoretically, these abilities should be related since they are both distinct subsets of arithmetic.

In the case of Organizational Engineering, the theory is seminal and there are, in the authors opinion, no obvious correlates for other the-

ories that could be relied upon for the purposes of convergent validity. However, there are internal consistencies (the focus of internal validity) that can be tested for convergent validity.

One test of convergent validity for individuals might be inferred from the similarity or convergence of the strategic patterns across years. This test is detailed in Appendix 1, where it is shown that the database remained constant for each measure at the .05 level of significance from 1994 through 1999. In other words, it was expected that the data would converge on similar strategic patterns, and that is what, in fact, occurred.

A test of convergent validity at a group level might be obtained by testing whether organizational units that differed only by location converge on the same strategic pattern. In other words, the organizational environments have the same characteristics and should attract and retain people who subscribe to roughly similar strategic pattern preferences. Thus, there should be a convergence in strategic patterns across these facilities.

The database contained information on 19 waste treatment facilities in different geographic locations. Strategic style data from the management team of each plant was available and provided an individual N of 188, with the average plant having a management team of 9.9 people.

Using the Shapiro-Wilk test, the hypothesis of normality was rejected in at least one group out of the 19, for each of the four measures of strategic style. Therefore the use of parametric statistics (e.g., difference in means, regression, etc.) would yield unreliable results.

The four measurements of strategic style across treatment facilities were checked for equality of variances using Levene's statistic ($F_{(RS)} = 0.926$, $p = .548$; $F_{(LP)} = 0.799$, $p = .6998$; $F_{(HA)} = 1.049$, $p = .4081$; $F_{(RI)} = 0.906$, $p = .5722$) and there was no significant evidence found for different variances. Therefore, the Kruskal-Wallis test is an appropriate procedure to test whether the 19 different plants did, in fact, display similar strategic profiles as predicted by Organizational Engineering theory.

The hypothesis tested was that the plants did not differ from each other in the strategic profiles of the management team. The results

are displayed in Table 7. Each of the 4 omnibus tests (corresponding to each strategic style) failed to reject the null hypothesis, and every pairwise group comparison for each style resulted in a *p*-value of .999 or above. The failure to reject this hypothesis by the Kruskal-Wallis test provides evidence for the expected convergence of the different locations on a particular strategic profile distribution.

These findings lend credence to the convergent validity of Organizational Engineering.

Table 7

CONVERGENT OF STRATEGIC PROFILES ACROSS SIMILAR FACILITIES

Strategic Style	Kruskal-Wallis H	*P*	Projec at 95% Confidence Level
Reactive Stimulator	25.320	0.1163	NO
Logical Processor	21.943	0.2345	NO
Hypothetical Analyzer	8.111	0.9769	NO
Relational Innovator	9.974	0.9328	NO

N = 19 Plants
188 People

DISCRIMINANT VALIDITY

SUMMARY Discriminant validity was tested using an unsupervised learning method of cluster analysis. The PAM algorithm run with k = 3,887 was able to discriminate among three groups that should be different at a $p < 10^{-29}$ significance level, a level substantially in excess of the generally accepted $p < .05$ standard of significance.

Discriminant validity is the opposite side of the coin from convergent validity. Discrimant Validity examines the degree to which the operationalization is not similar to (i.e., diverges from) other operationalizations that should be dissimilar. For example " . . . to show the discriminant validity of a test of arithmetic skills, we might correlate the scores on our test with scores on tests with verbal ability, where *low* correlations would be evidence of discriminant validity" (Trochim, 1999b). In other words, the scores are not expected to be related and they are not.

Convergent validity checks whether things that a theory says should be related are related. Discriminant validity checks whether things that a theory says are unrelated are really, in fact, unrelated. Together, convergent and discriminant validity triangulate a construct like strategic style to assure that it is operationally clear in exactly what it is measuring.

Within the bounds of Organizational Engineering theory, discriminant validity applied to individual styles and patterns is not applicable. This is because all of the strategic styles and patterns are *inherently* related. This occurs because method and mode dimensions are

exhaustive and combine to form a strategic style (Salton, 2000, pp. 24-31). A high score in one style automatically requires that another style will have a lower score—a relationship is "built in." In other words, the theory requires that all "all bases are covered" on an individual level. This comprehensive coverage is one source of the theory's power.

However, on a group level discriminant validity can be demonstrated. Organizational Engineering theory implies that different strategic styles and patterns are favored in different environments. For example, as a group, brain surgeons are unlikely to display the unpatterned spontaneity of the Changer strategic pattern. Similarly, commodity traders are unlikely to display the cautious, thorough and precise characteristics of the Conservator pattern. In other words, the theory suggests that strategic patterns will be able to "discriminate" on the basis of certain occupational categories.

The three clear professional categories tested in the construct validity section provide an opportunity for testing discriminant validity. In the construct validity section it was shown that the three occupational categories differed from the general database population in a predictable manner. It did not show that they were different from each other.

To test whether the three groups would fall into distinctly different categories, it was decided to apply an unsupervised learning method of cluster analysis using the algorithm PAM, which employs a k-medioid approach (Kaufman and Rousseeuw, 1990).

Under this method, the data was processed only with the number of clusters k = 3 known to it, without knowledge of which person was in which class or even how many people are in each class. Correctly classifying the people in each group well beyond the level expected by chance is evidence of the theory's ability to "discriminate." Viewed in this manner, the test can be considered a stringent test of discriminant validity.

The PAM algorithm was run with 887 observations and four measurements per observation (corresponding to the individual's score on each strategic style). The procedure terminated with the assignment of each observation to one of three clusters (see Figures 1 and 2).

The optimal assignment of cluster assignments to occupational categories was then obtained with the ODA system (Yarnold & Soltysik, 2000) system. The assignment was:

Cluster 1 ⟶ R & D

Cluster 2 ⟶ IT

Cluster 3 ⟶ Customer Service

The *p*-values associated with the overall classification were obtained by Fisher's exact test for 3x3 tables. Since the PAM algorithm knew

Figure 1
Job Category Membership by Cluster

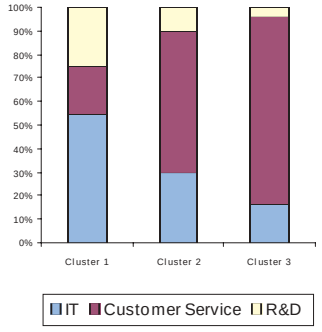
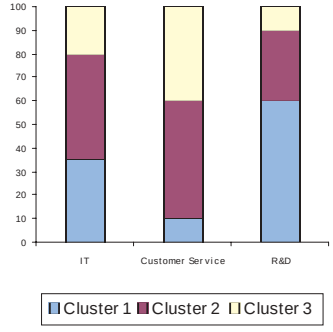


Figure 2
Cluster Membership by Job Category



only that three groups were present, and that any number of observations could have been assigned to a cluster, the correct test in this case is a contingency test without row or column marginals fixed. The use of Fisher's exact test might be questioned. It was widely held for many years that Fisher's exact probability test was limited to cross-classification tables in which both marginal frequency totals were fixed (Mielke and Berry, 1992). Research by Yates (1984) has shown this belief to be fallacious. To the contrary, Yates convincingly argued that Fisher's exact probability test is the preferred test whether both, one, or none of the marginals are fixed. Therefore, the use of Fisher's exact test is seen as appropriate for the issue at hand.

The results of the ODA analysis are summarized in Table 8. The

analysis yielded high sensitivities and predictive values for the three categories (relative to base expected values of 33.3%). The high effect strength of this analysis demonstrates three highly discriminable clusters that, in turn, correspond to the three occupational categories. In addition, the null hypothesis that the assignment of observations to categories is random was rejected by the results of Fisher's exact test for 3x3 tables.

Table 8

DISCRIMINANT VALIDITY USING THE K-MEDIOID APPROACH APPLIED TO DIFFERENT OCCUPATIONAL CATEGORIES

Information Technology	41.92%	36.75%
Customer Service	42.86%	71.17%
Research & Developement	54.08%	22.84%
Mean	46.28%	43.59%
Effect Strength	19.43%	15.38%

Overall Classification Accuracy = 43.74%
Overall Effect Strength = 17.40%
$p < .0001$ (from Fisher's exact test)

The findings of discriminant validity complement those of construct validity and triangulate with the findings in convergent validity. The evidence indicates that the theory can discriminate between unrelated factors as well as accurately predict the directional effect of related factors at the accepted .05 level of significance.

CONCURRENT VALIDITY

SUMMARY This dimension of validity relied upon the judgment of the expert panel of 50 professionals. Between 32 and 48 experts responded to the various instruments and methodologies tested under concurrent validity. The experts reported that in their administrations, the number of *inaccurate* reports was zero (0%). The concurrent validity of the instrumentation and methodologies is judged to be high.

Concurrent validity is a subset of criterion validity—the assessment of an instrument or methodology against some standard or "criterion." Predictive validity judges the accuracy with which criterion measures to be obtained in the future can be estimated from earlier test data. A concurrent study serves the same purpose, but obtains prediction and criterion information at approximately the same point in time.

This section tests Organizational Engineering's concurrent validity on various individual and group levels. The reader is thus provided assurance that the consolidation methodology employed as well as individual estimates are accurate and valid.

Survey Instrument

The face validity section showed an overwhelming proportion of respondents found no reason to object to their strategic style characterization (significant at the .05 level of significance). It can be argued that this is a test of concurrent validity and, at the levels regis-

tered in this study, merits weighting as a factor testifying to the concurrent validity of the technology.

However, to extend the rigor even further, the 50 member expert panel was asked for their estimates of the accuracy of the behavioral preference implied by the various strategic styles. In other words, panel members were asked for a judgment on the outcomes that are expected to flow from the strategic style preferences with the question:

> *As far as you have been able to determine, do the individual preferences reported by the instrument fairly express the respondentes' "attitudes" or "feelings" toward the subjects identified (e.g., nature of goals sought, preferred detail, work environments preferences, etc.)? No____ Yes ___*

It might be noted that the question seeks judgment on the "feelings" of the respondents. This was an attempt to cause the expert to assess underlying beliefs as well as objective behavior. In other words, the experts were asked to overlay their judgment on the respondent's judgment.

Table 9

EXPERT PANEL ESTIMATE OF CORRESPONDENCE WITH INDIVIDUAL WORK PREFERENCES

Number of experts responding	48	
Yes Agreed that preferences reported by the survey instrument were representative	**48**	**100%**
No Did not agree that preferences reported by the survey instrument were representative	**0**	**0%**

The responses given to this question are presented on Table 9. Two experts said they were themselves unable to answer the question. The balance, 48 experts, all agreed that, in their judgment, the

respondent's preferences corresponded to those that were projected to be present.

The level of agreement precludes statistical tests to demonstrate significance since statistical tests require there be members in both of the categories being compared. Attempting a statistical analysis would be the equivalent of trying to establish a failure rate on a machine in which a failure has never occurred. For purposes of this study, suffice it to say that the expert judgment reinforces the observed reaction of respondents and gives strong evidence of the concurrent validity of the survey instrument.

TeamAnalysis™

Organizational Engineering theory postulates that individual scores can be accumulated in such a manner that the overall character of the group or team as a whole can be accurately depicted. The methodology is described in the books on the subject (Salton, 1996, 2000) and is codified in a computer program that produces a report titled TeamAnalysis™.

The TeamAnalysis report describes both the current strengths and the vulnerabilities of the team or group. The presence of vulnerabilities (i.e., potential group deficiencies) in the report minimizes the "fortune teller" phenomena where agreement is obtained by focusing only on favorable attributes.

TeamAnalysis™ results are typically released to the team as a group and the content of the report is usually discussed in a group setting. The expert administering the survey is usually present at this debriefing and is positioned to judge the accuracy that the team or group accords to the results. This section of the study attempts to assess the level of group agreement with the strengths and vulnerabilities outlined in the TeamAnalysis™ with the question:

> *In your opinion, does the TeamAnalysis™ report accurately reflect the posture of the group as a whole towards the subjects considered (e.g., degree of change being sought, level of analysis desired, action orientation, etc.)?*
> *No____ Yes _____*

The expert responses given to this question are presented on Table 10. Again, 2 experts believed themselves unable to answer the question. The balance, 48 experts all agreed that, in their judgment, the TeamAnalysis™ report accurately characterized the group, as they knew it. This level of agreement again requires no statistical test to demonstrate significance. In the author's opinion, the concurrent validity of the TeamAnalysis™ report is demonstrated as viewed by the experts referenced in the study.

Table 10

EXPERT PANEL ESTIMATE OF CORRESPONDENCE WITH ACTUAL GROUP WORK PREFERENCES

Number of experts responding	48	
Yes Agreed that preferences reported by the survey instrument were representative	**48**	**100%**
No Did not agree that preferences reported by the survey instrument were representative	**0**	**0%**

OrgAnalysis™/LeaderAnalysis™

Organizational Engineering postulates individual and group information processing propensities are fractals. In other words, a group's processing patterns are directly comparable to those of an individual. Therefore, it is possible to compare an individual person such as the leader, with the group as a whole.

This methodology is outlined in the books on the subject (Salton, 1996, 2000) and codified in a computer program titled either OrgAnalysis™ or LeaderAnalysis™. The program compares the leader to the individuals in the group and to the group as a whole. The report is usually done in conjunction with a TeamAnalysis™ but is much more detailed and specific on the expected relationships of the leader with the group. The concurrent validity of this methodology was tested with the following question:

Does the interaction of the leader and the group follow the descriptions defined in the LeaderAnalysis™?
No____ Yes ____

The responses given to this question are presented on Table 11. Of the 50 experts available, 16 had not used OrgAnalysis™ /LeaderAnalysis™ technology or otherwise believed themselves unable to responsibly answer the question.

Table 11

EXPERT PANEL ESTIMATE OF THE ACCURACY OF ORGANIZATION/LEADER ANALYSIS

Number of Org/Leader Analyses	482	
Number of experts reponding	34	
Yes Agreed that the interactions followed the descriptions predicted	**34**	**100%**
No Did not agree that the interactions followed the descriptions predicted	**0**	**0%**

The balance, 34 experts had conducted 482 of these analysis and all agreed that, in their judgment, the respondent's preferences corresponded to those that were projected to be present. Once again, the absence of entries in the negative category precludes statistical tests. Also again, the author believes the results to be sufficiently strong to demonstrate the concurrent validity of the Org/LeaderAnalysis™ at a high level of certainty.

TwoPerson™/One-to-One:

The theory of Organizational Engineering can be applied to describe the relationship of two people. The report identifies the structural strengths and vulnerabilities imbedded in their joint pursuits. The presence of comparative vulnerabilities in the report once again limits the operation of the "fortune teller" phenomena.

The methodology is outlined in the books on the subject (Salton, 1996, 2000) and codified in a computer program titled either TwoPerson Analysis™ or "One-on-One" Analysis™. The concurrent validity of this methodology was tested with the following question:

> *In your best estimate, what proportion of the interactive behavior described in the TwoPersonAnalyses was*
> *Highly Accurate %*
> *Reasonably Accurate %*
> *Inaccurate %*

The results of the expert judgment on the concurrent accuracy of the TwoPerson Analysis are given in Table 12. Fully 100% of the 32 experts who had administered 1,005 of the analysis judged the results to be accurate—either at a high or reasonable level.

The existence of a "reasonably accurate" category allows us to adopt a stringent view and test the data for its concurrent validity. The "reasonably accurate" category can be combined with the "inaccurate" on the grounds that "reasonably accurate" implies a degree of inaccuracy. A two-sided, one-sample sign test can then be applied to determine if the two categories—"highly accurate" and "inaccurate"—are statistically distinct. In other words, we seek to dismiss the possibility that both categories are simply random variations within a single category (i.e., "reasonably accurate.")

The sign test resulted in $p < .0001$. Thus it is reasonable to assert that the "highly accurate" category represents a distinctly different judgment even under conditions of extreme interpretive stringency.

In the author's opinion, the overwhelming weight of expert judgment is in favor of according the TwoPerson Analysis™ or "One-on-One" Analysis™ a very high level concurrent validity.

Overall, the concurrent validity of both the individual survey instrument and the consolidation methodology appears to have met all reasonable tests of concurrent validity. Both the theory and the consolidation methodology appear well founded in terms of their ability to reflect current conditions of both individuals and groups.

Table 12

ONE-ON-ONE™/TWOPERSON ANSLYSIS™ CONCURRENT VALIDITY

Experts Reponding	32	
Number of Analyses Conducted	1,005	100%
Method Highly Accurate	990	99%
Method Reasonably Accurate	15	1%
Method Inaccurate	0	0%

PREDICTIVE VALIDITY

SUMMARY The assessment of the predictive dimension relies upon the judgment of the expert panel. Of the 50 expert professionals available, 39 believed themselves positioned to judge the predictive accuracy of the TeamAnalysis™ methodology. The experts reported that in their administrations the number of inaccurate reports was zero (0%). The predictive validity of the instrumentation and methodologies is judged to be high.

Predictive validity is a form of criterion validity. Criterion validity tests whether the relationships identified by a theory are actually evidenced in the "real world" in a way than can be objectively tested. In other words, a "criterion" is some type of standard on which a judgment of a relationship can be based.

Predictive validity is concerned with "evidence of criterion-related validity in which criterion scores are observed at a later date" (Canadian Psychological Association, 1996). This contrasts with concurrent validity, which concerns "evidence of criterion-related validity in which predictor and criterion information are obtained at approximately the same time." In other words, predictive validity is concerned with the ability of a theory to predict what will happen in the future.

The Survey instrument does not lend itself to direct test of predictive validity in field settings. This is because Organizational Engineering theory maintains that the human is a rational animal capable of changing in response to the environmental conditions he or she confronts. Therefore the individual report produced by the survey

instrument is not a "predictor" of long run behavioral preferences. Behaviors in particular circumstances also cannot be "predicted" because most people have at least some access to all of the strategic options available (Salton, 1996, p.52). The particular strategic option chosen depends upon the individual's interpretation of the particular situation.

However, an indirect test of predictive validity of the instrument is available. The theory "predicts" that strategic styles will be stable if the environment remains constant (Salton, 1996, p. 61). In other words, people are unlikely to change a successful strategic style unless their personal environment signals that the elected style may no longer be applicable. Therefore, in the absence of macroeconomic or social changes, the theory would predict that a large population of people would remain constant in their strategic style choices.

Appendix 1 demonstrates that no evidence for differences were found for the strategic profile distribution of the database population for the years 1994 through 1999. The macroeconomic and social condition of the United States, from whose population the database is primarily drawn, was roughly stable during this period. Therefore, 1995 served as an accurate "predictor" of 1996, which then served as a "predictor" of 1997, and so on. This finding can be considered evidence that the predictions made by the theory for the behavior of individuals are valid over the time period investigated.

TeamAnalysis™ is methodology that consolidates individuals to obtain representations of entire groups as single entities. The report is entirely mechanical and requires no knowledge of a group's purpose or circumstances beyond the strategic style profiles of the members. Therefore, the report can be seen as direct extension of the Organizational Engineering theory at a group level.

A fundamental difficulty of assessing the predictive validity of the TeamAnalysis™ report is that it is prescriptive. After specifying the strengths and vulnerabilities of a group, the report offers suggestions on how vulnerabilities might be offset and strengths magnified. To the extent that the group accepts these recommendations, the "natural" structural inclinations of the group are voided or redirected.

Further complicating matters, different groups adopt the recommendations to different degrees. Even those recommendations accepted are often modified to better accommodate local conditions that are unknown by the computer generating the report.

While a definitive judgment based on objective data is unattainable in field settings, an informed judgment can be made. Some teams substantially ignored the recommendations and proceed following their "natural" strategic inclinations. In these teams, the structural conditions outlined in the TeamAnalysis™ would be "predicted" to persist. Thus, these groups can be seen as offering an opportunity for predictive validation.

In addition, the recommendations made in the analysis are definitive. They typically define what is being proposed, why it is proposed, and the probable outcome of implementing the recommendation. Therefore, a skilled observer is positioned to make a reliable judgment on the predictive validity of the methodology, even in the case where recommendations are adopted by the team.

The foregoing observations suggest that a reliance on the judgment of the expert panel would be well founded. Its members are physically present in the field setting and are often present at group meetings. They are positioned to judge the degree of adoption of the recommendations. They are also typically situated to witness the behavior of the group over time. Therefore, expert judgments on the predictive validity of the TeamAnalysis™ can be reasonably relied upon as indicative for purposes of this study.

To assess the predictive validity of the TeamAnalysis™ the members of the expert panel were asked this question:

> How good would you say the TeamAnalysis™ report predicted the behavior of the group into the future? Was it:
>
> *Highly Accurate* _____
> *Reasonably Accurate* _____
> *Inaccurate* _____

Of the total of 50 experts available, 11 believed themselves unable to make accurate judgments based on nonuse, their positioning or

because of their degree of participation with the groups involved. The findings from this inquiry among the 39 experts positioned to make a judgment are displayed in Table 13.

Table 13

EXPERT PANEL EVALUATION OF TEAMANALYSIS™ PREDICTIVE VALIDITY

TeamAnalyses Administered		
Experts Responding	39	100%
Judging Method Highly Accurate	31	79%
Judging Method Reasonably Accurate	8	21%
Judging Method Inaccurate	0	0%

The results of the expert judgment on the predictive accuracy of the TeamAnalysis are self-evident. Fully 100% of the experts judged the results to be accurate either on a "highly" or "reasonably" accurate basis.

Even a stringent view of the data produces strong evidence of predictive validity. The "reasonably accurate" category can be combined with the "inaccurate" on the grounds that "reasonably accurate" implies a degree of inaccuracy. A statistical test can then be applied to determine if the two categories—"highly accurate" and "inaccurate"—are distinct. In other words, we seek to dismiss the possibility that both categories are simply random variations within a single category (i.e., "reasonably accurate").

A two-sided, one-sample sign test of these responses yielded p = .0003. Thus it is reasonable to assert at the .001 rejection level that the "highly accurate" category represents a distinctly different judgment even under conditions of extreme interpretive stringency. In other words, it can be reasonably assumed that the 31 people who judged the results "highly accurate" saw something systematically different than the 8 people who judged the results "reasonably accurate."

Under conditions of the tests applied, the predictive validity of Organizational Engineering was found to obtain on both an individual and group level at the .05 significance level in those cases where such statistical estimates could be applied. The absolute level of agreement between the experts testifies to the existence of a systematically high degree of predictive validity.

Conclusion Validity

SUMMARY The large number of individuals (N = 8,721) and groups (1,003) encompassed by the study provide assurance of generalizability. The statistical tests performed were shown to fully satisfy the proper criteria (e.g., identical dispersions, equality of variances, etc.) minimizing exposures based on statistical power. In addition, the cross-validation across multiple dimensions of validity amplifies the assurance of the validity of the underlying theory and its expression in instrumentation and methodology. In the author's judgment, the theory and methodology fully meet the standards of validity as applied within the discipline of organizational development.

"Conclusion validity is the degree to which conclusions we reach about relationships in our data are reasonable." (Trochim, 1999c). As interpreted by this author, conclusion validity represents the summarization of the various tests conducted in other parts of this study.

The first threat to conclusion validity is reliability (Trochim, 1999d). Effectively, this threat involves measures that have too much variability to be trusted. Appendix 1 provideds evidence that the individual survey report produces consistent results over a six year period.

Statistical power is seen as another threat to conclusion validity. Trochim recommends a large sample size as one means of offsetting this threat (Trochim, 1999d). In this case, the sample of 8,721 individuals and 1,003 groups is seen as very large by the standards typically applied within the discipline. At this level, it is unlikely that an

increase in sample size will add any significant statistical power to the results obtained.

Trochim suggests "raising the alpha level" as another method of decreasing the threat arising from statistical power (Trochim, 1999d). In this study, the "worst case" alpha level was set at the .05 rejection level and the data often tested better than this well-accepted standard. In other words, this study required (at minimum) that chance be responsible for the results obtained in only 5 of 100 cases. This conforms to the well-accepted standard within this and other disciplines.

Trochim sees a final aspect of the threat of statistical power as the "effect size" (Trochim, 1999d). "Effect size is a ratio of the signal of the relationship to the noise in the context" (Trochim, 1999d). Since the "noise" is already at minimal levels (i.e., the reliability is high), the only other method of improving "effect" is to make the signal more salient. This may be a viable strategy in experimental settings where the degree of "treatment" can be manipulated. However, this study is based on field data that cannot be manipulated. In the author's judgment, the clarity of the findings obtained in this context negates the need for any such enhancement.

Poor implementation is seen by Trochim as another threat to conclusion validity (Trochim, 1999). Within this category is the misapplication of statistical methods. For example, many studies in this area employ parametric statistics without testing the data for normality of distribution or other requirements imbedded within the statistical method selected. This is usually done on the grounds that the parametric statistic used is "robust." Whether it is "robust" enough for the issue at hand is controversial and left to speculation among those interested in the subject.

In this study, an effort was made to test the data against all of the assumptions imbedded within the elected statistic. Nonparametric statistics were used when appropriate and the assumptions upon which they rest were tested before they were employed. The reader need not put reliance on an undefined "robustness" in assessing the findings of this study.

From a statistical vantage point, the use of an expert panel to provide judgments on various aspects of validity is perhaps the least secure of the elements of the study. In effect, this represents "secondhand" data and is subject to the vagaries of human judgment. However, the large panel size, the high qualifications of the participants, and the strong internal consistency of the judgments lend great confidence in those validity elements that rely upon their judgments.

In final analysis, any statistical study is confronted with the possibility of two basic types of error.

Type I: Concluding that there is no relationship when in fact there is one.

Type II: Concluding that there is relationship when in fact there is not one.

There is no way of providing 100% assurance that both of these errors have been completely avoided in this or any other statistical study. This is why results are typically framed in terms of probabilities. This is also why any validity study can be considered a form of argument to which the reader is the final judge (Cronbach, 1984).

The multiplicity of forms of validity tested, the large sample size, the size and quality of the expert panel, and the rigor applied in the statistical assessments should provide the reader with a high level of confidence in both the theory and its associated methodology. In the author's judgment, the theory and methodology fully meet the standards of validity as applied within the discipline of organizational development.

Appendix 1

RELIABILITY

SUMMARY Reliability is technically not a form of validity. The reliability of the instrument was tested for all pairwise combinations for the years 1994 through 1999 (15 individual contrasts) using the Kruskal-Wallis test. In all cases, the findings confirmed reliability by showing that differences in the data between years could not be established. The survey instrument is judged reliable by the accepted standards of the discipline.

"In research, the term reliability means 'repeatability' or 'consistency'."(Trochim, 1999e) The definition of reliability implies that it has two distinct components. Traditional reliability measures seek to establish reliability based on multiple scorings of individual respondents (i.e., repeatability component of Trochim's definition). This is appropriate since many of these tests were applied in the validation of constructs that were hypothesized to be fixed components of a human being (e.g., "introversion").

Organizational Engineering theory, however, posits that people operate within ranges on the underlying method and mode scales—a "built-in" source of variation that is imbedded in the theory. In addition, the theory postulates that people are responsive to their environment. It proposes that people will change their strategic approach (i.e., their method/mode range election) in response to personal environmental changes (Salton, 2000, pp. 53-59). Since reliability can be considered an element of construct validity (Moss,

1994), it would be an error to apply the traditional repeatability tests of reliability. To do so would violate and invalidate the underlying construct that the test attempts to validate. In other words, procedures such as Test-Retest reliability cannot be used without undermining the very foundation of the validity study itself.

Consistency is the second component of the definition of reliability. "We judge the reliability of the instrument by estimating how well the items that reflect the same construct yield similar results." (Trochim, 1999f) Applying the same measure to different subjects and obtaining the same expected result can thus be interpreted as evidence of consistency.

One method of determining this type of consistency is a redundancy strategy. Here the respondent is repeatedly asked the same question at multiple points in an instrument. The responses can then be compared and the degree of correlation viewed as an index of consistency within the instrument. This method is typically employed to ferret out deceptive responses.

The redundancy method is not applicable to the Organizational Engineering survey instrument. The respondent is not answering a simple question. Rather, he or she is expressing a preference for one response versus three other alternatives.

For example, in one place the respondent is presented a selection of "I respond fast" and in another "I react fast." However, each of these selections is set off against different optional alternatives generated by the other potential elections on the method and mode dimensions. Thus it is not inconsistent for a respondent to elect the "I respond fast" option in one case and reject "I react fast" in another. It is merely a statement of preferences among the alternatives provided. Thus the specification of the instrument precludes the use of the traditional consistency measures based on redundancy. If applied, they are likely to yield a false negative, since they presume that the same thing is being asked multiple times.

Parallel forms reliability is an accepted strategy in the social sciences and is used to test the consistency and repeatability of the instrument simultaneously. The parallel forms strategy typically involves applying two instruments purporting to measure the same things to the

same population and comparing the results (Trochim 1999f). A variation the parallel forms methodology can be obtained by applying the same instrument to various populations. Highly correlated results could imply the existence of an underlying consistency sufficient to validate the reliability of the instrument.

In this study, the database was segregated into people who had been administered the survey in the years 1994, 1995, 1996, 1997, 1998, and 1999, creating six separate populations. The variation of the parallel forms test considers that these subsets are "samples" of a larger underlying population. If the underlying population suffered no major environmental changes, it would be expected that the underlying strategic preferences would remain constant.

The years 1994 through 1999 are similar in terms of their macroeconomic and social conditions. Thus, without large-scale dislocations, it is expected that the average personal environment of the underlying population was constant over this time period. Therefore, a testable hypothesis based on the consistency component of the reliability criteria can be stated as:

Null: The strategic postures of the population for years 1994, 1995, 1996, 1997, 1998, and 1999 are statistically indistinguishable.

A failure to reject the null hypothesis would serve as evidence of the reliability of the survey instrument. The instrument would have yielded consistent results over a long time period. The use of six years greatly strengthens the assertion of validity since this range of years provides fifteen opportunities for rejection (1994 vs 1995, 1994, vs 1996, etc.).

The choice of the variable to represent the strategic posture is informed by the underlying theory. "Strategic patterns are most useful in characterizing lengthy streams of decisions and overall strategic postures. Strategic styles are generally more useful in predicting transactional characteristics of individual or shorter streams of decisions." (Salton, 2000, p.86)

Since the hypothesis seeks to test whether the overall postures of the population remain stable over years, the single most appropriate representation of the strategic posture for purposes of testing the reliability hypothesis is the dominant strategic pattern—a combination of the person's primary and secondary strategic style.

The choice of the test statistic to employ in the validation is governed by the character of the data being addressed. For each year, a hypothesis that the strategic pattern characteristics followed a normal distribution was tested using the Shapiro-Wilk test (in the case of the year 1997, Stephens' test was used because the size exceeded the limits of the Shapiro-Wilk test). All six patterns tested resulted in rejection of these null hypotheses at the .01 significance level. In other words, normal distribution of the data could not be assured and statistical tests based upon that normality assumption could not be understood to produce reliable results.

The Kruskal-Wallis test is a nonparametric alternative to one-way ANOVA and is a straightforward generalization of the Mann-Whitney U test for two independent samples. A significance criterion of .05 was chosen for this experiment. The null hypothesis is stated as:

Null: The database populations are drawn from the same underlying population and this population has remained stable.

Table 14

DATABASE NORMALITY TESTS

Year	*N*	Test	Statistic	*p*
1994	158	Shapiro-Wilk	W = .8975	.0001
1995	1082	Shapiro-Wilk	W = .9269	.0001
1996	1891	Shapiro-Wilk	W = .9217	.0001
1997	2453	Stephens	D = .1176	.01
1998	1866	Shapiro-Wilk	W = .9197	.0001
1999	1271	Shapiro-Wilk	W = .9192	.0001

The Kruskal-Wallis procedure requires approximate equality of variance over all groups. Therefore, Levene's test was used to test the hypothesis that all years had equal variance for the measure of pattern. The test obtained Levene's statistic F = 1.113, with a corresponding *p*-value of 0.351. Thus the hypothesis was not rejected, and no significant evidence of different variances was found.

The Kruskal-Wallis test was then applied to the sample population. The overall test obtained a value of the Kruskal-Wallis test statistic $H = 1.809$, with a significance level p of 0.875, failing to reject the overall null hypothesis at the .05 level. No significant evidence was found of differences in the measure of pattern over the different years.

Table 15 contains the results of the multiple comparison of mean ranks across all possible pairs of years. The Tukey-Kramer procedure (Kirk, 1982, pp. 119-120), a well-known a *posteriori* method for evaluating pairwise comparisons, was used to control Type I error. The failure to reject the null hypothesis (that there was no difference in mean rank) in all fifteen of the year comparisons, along with the failure to reject the overall null hypothesis, provides strong evidence of the reliability of the survey instrument. This judgment is strengthened even further when the large number of observations in each pair is considered.

The statistical tests conducted using a database of respondents provide strong evidence that the survey instrument is valid on the dimension of consistency over time.

Table 15

PAIRWISE MEAN RANK COMPARISONS

	Years	q	p	Reject H_0
1.	1994 vs. 1995	0.397	.999	No
2.	1994 vs. 1996	0.751	.999	No
3.	1994 vs. 1997	0.808	.999	No
4.	1994 vs.1998	0.656	.999	No
5.	1994 vs.1999	0.905	.999	No
6.	1995 vs.1996	0.745	.999	No
7.	1995 vs.1997	0.891	.999	No
8.	1995 vs.1998	0.537	.999	No
9.	1995 vs.1999	1.027	.999	No
10.	1996 vs.1997	0.134	.999	No
11.	1996 vs.1998	0.241	.999	No
12.	1996 vs.1999	0.388	.999	No
13.	1997 vs.1998	0.390	.999	No
14.	1997 vs.1999	0.289	.999	No
15.	1998 vs.1999	0.604	.999	No

Appendix 2

DATABASE

The author of this study was provided with an unedited database of strategic style scores collected in conjunction with the preparation of group based analyses. The database contains 8,721 individual observations.

Since the data was collected in conjunction with group reports, the observations had been segregated into individual organizational entities. There are a total of 1,003 individual groups represented in the database.

The team names cited in Table 16 were provided by the groups requesting the analysis, and not all groups analyzed provided such identification. For example, some groups were submitted "blind" by the requesting organization as a test the underlying technology before they committed to adopting the technology on a wider basis.

The data was sourced from all regions of the United States and include representation from a variety of industries, as illustrated in Table 17.

The organizational levels represented in the database also include a variety of organizational positions and functions. As was the case with team names, the titles are not universally cited or recorded within the database. A sample listing of titles of respondent is provided in Table 18.

As with most studies of this type, the sample is purposeful. However, the large numbers of observations on both an individual and group level provide a high degree of confidence in the generalizability of the findings. In addition, the organizational status range represented in the sample provides assurance that the phenomenon is not local to a particular organizational level.

Table 16

EXAMPLES OF TEAM NAME/PURPOSE

Executive Committee	QS 9000 Team
Surgical Team	Institute Leadership Team
University Housing	Telemarketing Devel. Group
Supermarket Operations	Telephone Customer Service
“Business Optimization” Project Team	Board of Directors
Production & Surveillance Team	Warehousing and Distribution
Lease & Contract Administration	Payrool Department
Central Seismic Processing	Proj. Mgt. Consultants Team
Field-Safety Team	Plant Management
Accounting & Scheduling	Chemical Research Team
Business Analysis	New Product Committee
Midwest Management Team	Megastore Team
Human Resources	Creative Services
Body Interior Management	Board of Commissioners
Materials-Technology Team	Geosciences Admin. Team
Risk Management	Architectural Engineering
Executive Team	Cutom Mfg. Team
Solar Team	EEOC Operating Office
Product-Marketing Mangers Team	Museum Sr. Staff
Electric Regulatory Affairs	Adult Education Faculty
Dept. 470 Packers	Vice President Ops Team
Federal Tax Team	Org. Effectiveness Group
Strategy and Plans	Marine Construction Team
Solvents Team	Publications Staff
Legal Staff	Retail Clothing Store
Union Mangement	Radio Station Selling Team
Ice Cream Sales and Marketing	Plant Managers
Diversity Center	Secretarial Team
Systems-Integration Team	Claims Processing Team
Client/Vendor Team	“As-Is” Team
Sensor Engineering Team	Rate Investigations Unit

Table 17

INDUSTRIES/AREAS INCLUDED IN DATABASE

Insurance	Hospitals
Manufacturing	Banking
Electrical Utilities	Gas Utilities
Telecommunications	Nursing Homes
Marketing	Warehousing and Distribution
Sales	Tool & Die
Laboratories	Federal Agencies
State Goverment	City Goverment
Fast Food Chains	Chemicals
Consultants	Non-Profit Organizations
Universities	Middle Schools
Construction	Charitable Organizations
Religious Organizations	Temporary Services
Waste Management	Aerospace
Radio	Newspapers
Retail Stores	Engineering Firms
Textiles	Publishing
Advertising	Design
Information Technology	Housing Authority
Joint Ventures	Accounting Firms
Pharmaceuticals	Automobile OEM
Grocery Chains	Printers
Oil Exploration/Distribution	Logistics (e.g., Trucking)
Training	Greeting Cards
Remanufacturing	Appliances

The large size and wide sample distribution is judged to provide a high level of assurance of the representativeness of individuals and groups in organized environments. However, reader purposes will have to determine its specific applicability.

Table 18

EXAMPLES OF TITLES OF RESPONDENTS

President	Sr. Vice President
Managing Partner	Vice President
Director	Adminstrator
Manager	Supervisor
Chair	Team Leader
Consultant	Union Steward
Hourly Worker	Scientist
Analyst	Accountant
Foreman	Engineer
Doctor	Nurse
Programmer	Owner
Lawyer	Officer
Teacher	Professor
Salesman	Librarian
Technician	Clerk
Artist	Electrician
Designer	Account Rep.
Architect	Systems Analyst
Draftsman	Trainer
Biologist	Chemist
Chef	General Manager
Psychiatrist	Auditor
General Counsel	Operator

Appendix 3

EXPERT PANEL

The panel of experts is a form of nonrandom sampling. "Expert sampling involves the assembling of a sample of persons with known or demonstrable experience and expertise in some area" (Trochim, 1999g)

The use of experts is appropriate where the views of the expert are the best (or only) way of acquiring the information necessary for judgment. For example, teams represent a nonlinear system (an area sometimes referred to as "Chaos Theory") with members interacting unpredictably with each other and with an external environment that is constantly changing. Under these conditions, field-based judgments on the predictive validity of an instrument or methodology in field settings can be made no other way.

While, in general, conventional statistical tests cannot be relied upon when using a panel of experts, the structure of the panel can affect the degree of certainty that might be assigned to conclusions derived from their judgments. The factors that might affect this certainty include:

1. *The number of experts. The greater the number of experts, the more likely it is that discrepant conditions will be revealed.*

2. *Independence of experts. Group processes tend to modify understandings and judgments. The greater the degree of independence of experts from each other and from any common institution, the less likely it is that group processes will taint judgments.*

3. *Judgment of experts. The quality of judgment is relative to the issue being addressed. In this case, the objective is the validity of the instrument and methodology in field settings. The judgment desired is judgment in practice. This is best demonstrated by depth of experience of the experts in field settings. The greater the depth of experience, the greater can be the reliance placed on their judgments.*

4. *Knowledge of experts. The appropriate responses to questions require an understanding of context. In this case, the context is a validity study, and a general academic understanding increases confidence in the judgments rendered.*

The remainder of this section is devoted to outlining the qualifications of this expert panel relative to the factors identified above.

Number of Experts

Certain, but not all, sections of this study draw heavily upon the judgments of a panel of experts. The expert panel is composed of people who have used Organizational Engineering technology in field settings. The panel consists of 50 individuals, all of whom are mature, fully independent practitioners participating in organizational environments on an ongoing basis.

The experts were approached via telephone and were read the questions that are quoted in the various sections of this paper. Pretests of the form indicated that dichotomous choices provided the most accurate judgments in certain cases and proportionate estimates the best in others. The questions were kept to a minimum to insure the widest possible participation. The responses of the experts were consolidated and, where appropriate, tested using well recognized statistical procedures and methodologies.

There is no generally accepted test for the adequacy of the number of experts to be included in a panel. However, a panel of 50 experts can generally be considered large and is, in the author's judgment, entirely sufficient for the purposes of this study.

Independence of Experts

Independence of experts is enhanced if the experts reside in distinct environments. Drawing experts from a single environment increases the probability that the influences generated within that common environment can create a bias based on information flows. For example, journals, newsletters and other publications that circulate within a particular area can sensitize experts to certain variables and obscure others. This, in turn, can cloud judgments and potentially bias overall results.

The organizational context of the experts used in this study draws on a variety of different societal segments. These include:

Table 19

ORGANIZATIONAL DISTRIBUTION OF EXPERTS

Universities	2	4%
Corporations	30	61%
State/Federal Agencies	4	8%
Consulting Firms	13	27%
TOTAL	49	100%

The varieties of organizations represented in this expert panel suggest that it is unlikely that the experts will be exposed to systematic judgmental bias based on limitations of organizational settings.

A second dimension of independence is the nonassociation of the individual experts with each other in their general environments. A systematic exposure to common influences (e.g., all working within one industry) would enhance the probability of bias arising from common processes that may tend to create a similarity of judgment. This similarity, should it exist, could most likely create a vulnerability directed toward a false positive. In other words, it could lead to a judgment that a condition existed which, in fact, does not.

The expert panel in this study is drawn from a wide variety of independent organizations. Many of these organizations are well known. Given this condition, a specification of their independence from each other is deemed unnecessary. These organizations include:

Table 20

SAMPLE LISTING OF EXPERT AFFILIATIONS

Aerospace Corporation	Agrilink Foods (Birdseye, etc.)
Alstom Inc.	American Greetings
Ameritech	Banta Books
BGF Industries	Blue Cross/Blue Shield
Briggs & Stratton	Caterpillar
Coastal Corporation	Cummins Engine
Discover Card	Emeritus Healthcare
Ernst & Young	Estee Lauder
Federal Aviation Admin.	General Motors
Honeywell	Lord Corporation
Marsh Supermarkets	Mastercraft Fabrics
Norfold Redev. Housing Auth.	Oklahoma Gas & Electric
SBC Communications	Shell Oil
Smithsonian Institute	Tampa Electric
University of Oklahoma	Univerisity of Michigan
USF Holland	Whirlpool Corporation

The thirteen consulting firms (fourteen individuals) represented in the sample range from single practitioners to firms of approximately 30 professionals. The practice areas involved include highly specialized orientations (e.g., quality systems) through generalists who are handle a variety of assignments. The client base to which they have applied Organizational Engineering methodologies extends the scope of firms to include areas such as specialty chemicals, life insurance, hospitals and other areas not represented on the listing in Table 20.

There is no generally accepted statistical test demonstrating the participant independence of the expert panel. However, a review of

Tables 19 and 20 suggests that it is unlikely that the expert panel would be subject to biases arising from common perspective generated by similarities in local environments.

Judgment of Experts

This study attempts to validate the instrument and methods of Organizational Engineering in field settings. The groups that are represented in the organizational sample confront actual issues within the context of complex organizational entities. The behavior and decisions of these groups have material consequences, often of a long-term nature, to the individual group members as well as to the larger organizations of which these groups are a part. Thus confidence in the judgments of the experts rests, to some substantial degree, on their ability to understand and interpret these interrelated and often volatile factors.

There is no single, generally accepted index of judgment. However, occupational position can be seen as a reasonable surrogate. The experience of experts can be seen as being evidenced by their positions within the organizations to which they are affiliated. This position implies endorsement by superiors and co-workers of both expertise and judgment.

It is noteworthy that the largest segment of the panel is made up of positions of Director/Manager. This suggests a proximity to the field setting of the subjects. The participation of experts in the professional category reinforces confidence that the judgments being reported are based on actual exposure to the subjects.

Table 21

OCCUPATIONAL POSITIONS OF EXPERTS

Corporate Officers (VP and above)	2	4%
Directors/Managers	24	48%
Professionals	10	20%
Consultants	14	28%
TOTAL	50	100%

The status and distribution of the expert panel indicates that the responses incorporated into the validity study can reasonably be judged to be untainted by inexperience or remoteness from the subjects being studied. The occupational levels attained suggest that the experts are highly experienced and are unlikely to be deceived or misdirected in their judgments.

Knowledge of Experts

The responses to questions are typically informed by the perceived context within which they are asked. In this case, the context is that of a validity study. A knowledgeable individual can be expected to understand that the appropriate responses should be more rigorous than, for example, questions framed in the context of informal information collection.

In this case, panel members were informed at the onset that the questions being asked were for the purposes of a formal validity study. The ability of the experts to understand the significance of this context can thus be considered an issue. The formal educational attainment of the panel can reasonably be assumed to be positively correlated to an understanding of the validity context. Therefore, the educational status of the panel experts was collected. The results are:

Table 22

EDUCATIONAL ACHIEVEMENTS OF EXPERTS

Ph.D.	5	10%
Master's Degree	32	64%
Bachelor's Degree	9	18%
Some College	4	8%

To some extent Table 22 understates the attainment of the panel. For example, several experts classified under the category of Master's Degree actually had multiple Master's degrees. Similarly, several experts classified in the category of Bachelor's degrees had done graduate work beyond the Bachelor's level.

The educational distribution of the expert panel gives confidence that the experts are capable of appropriately weighting their responses in terms of the validity context in which their answers were to be applied. It is unlikely that experts misunderstood the questions they were being asked due to the context in which they were to be applied.

Summary of Expert Panel Qualifications

The qualifications of the members of the expert panel used in this summary are seen as adequately fulfilling the specifications outlined at the beginning of this section, namely:

1. *The size of the panel appears sufficient to insure a repre sentation of diverse viewpoints.*

2. *The independence of the experts appears to be sufficiently demonstrated by the variety of organizations to which they are affiliated and by the different segments within which these organizations operate.*

3. *The judgment of the experts is suggested by the positions held. It is unlikely that these positions could be obtained without a degree of considered judgment being displayed over a long period.*

4. *The knowledge of the expert panel is attested to by the high level of educational attainment.*

Therefore, in the opinion of this author, the quality of the judgments rendered by the expert panel can reasonably be accepted as accurate estimates of the qualities and conditions called for by the questions asked.

Appendix 4

Organizational Engineering Survey

A
1. I complete things I start
2. I respond fast
3. I make plans
4. I imagine things

B
1. I plan before I act
2. I do things that are different
3. I change easily
4. I like clear instructions

C
1. I react fast
2. I like to have others finish what I start
3. I do things that are new and different
4. I get things done

D
1. I see into the future
2. I like things clear and direct
3. I am an organizer
4. I change ideas a lot

E
1. I have complicated ideas
2. I think of new ways to do things
3. I solve things pretty easily
4. I like things to be easy to understand

F
1. I follow directions
2. I predict what's going to happen
3. I'm quick to respond
4. I have many ideas

G
1. I pay attention to every detail
2. I have quick solutions
3. I like things my way
4. I like to follow directions

H
1. I know what I want to do
2. I know how I want to get things done
3. I'm pretty good at planning details
4. I give suggestions faster than others

I
1. I like to take chances
2. I like to follow the rules
3. I find and fix problems
4. I get into things totally

J
1. I like my own ideas best
2. It a easy for me to stay on task
3. I'm very careful
4. I sometimes do things before I think them through

K
1. I take chances
2. I adjust easily
3. I don't like changes
4. I make things happen

L
1. I like to analyze
2. I like to get things decided
3. I am easily distracted
4. I like to see ideas grow

M
1. I really don't like rules
2. I like things "just right'"
3. I like to get things done
4. I sometimes forget details

N
1. I forget things easily
2. I pay close attention to details
3. I go along with the crowd
4. I get others going

O
1. I like things to be exact
2. I'm playful
3. I get unusual ideas that I need to explain
4. I like to follow a schedule and be on time

P
1. I like directions
2. I like to invent things
3. I like adventure
4. I want to be exact

Q
1. I use things at hand to solve problems
2. I look for more than one way to solve things
3. If things get tough, I'll change ideas
4. I like to get things done the way they are supposed to get done

R
1. I like to start things
2. I tell others what I think
3. I get things done
4. I don't always know how things are going to end up

S
1. I decide things easily
2. I stir up action
3. I'm steady as a rock
4. I'm “out of synch” with others

T
1. I am thoughtful and deliberate
2. I like to think about lots of things
3. I don't like interruptions
4. I like to look at different ways to get things done

U
1. I'm careful
2. I like fast pace
3. I like to complete all the details
4. I see unusual connections between things

V
1. I'm a daredevil
2. I'm interested in getting results
3. I'm logical
4. I ignore details

W
1. I like to be in the "here and now"
2. I think about how things might be in the future
3. I like facts
4. I act on the spur of the moment

X
1. I like things to be clear and easy to understand
2. I can predict things in the future
3. I do things according to a “system”
4. I like things to happen “right now”

Appendix 5

Robert Soltysik
Summary of Qualifications

Publications/Formal Presentations:

Books:

Yarnold, P.R., and Soltysik, R.C. (2000). *Optimal Data Analysis*. American Psychological Association, in press.

Articles:

Soltysik, R.C., and Yarnold, P.R. (1998). Unit-weight MultiODA: Maximum accuracy discriminant analysis with unit-valued attrib ute coefficients. *Applied Psychological Measurement*, 22, 393.

Soltysik, R.C., and Yarnold, P.R. (1994). Univariable optimal discrimi nant analysis: One-tailed hypotheses. *Educational and Psychological Measurement*, 54, 646-653.

Soltysik, R.C., and Yarnold, P.R. (1994). The Warmack-Gonzalez algo rithm for linear two-category multivariable optimal discrimi nant analysis. *Computers and Operations Research*, 21, 735-745.

Soltysik, R.C. (1997). Review of *Organizational Engineering* by Gary J. Salton. Interfaces, 27, 108-109.

Yarnold, P.R., and Soltysik, R.C. (1991). Theoretical distributions of optima for univariate discrimination of random data. *Decision Sciences*, 22, 739-752.

Yarnold, P.R., and Soltysik, R.C. (1991). Refining two-group multivari able classification models using univariate optimal discriminant analysis. *Decision Sciences*, 22, 1158-1164.

Yarnold, P.R., Martin, G.J., Soltysik, R.C., and Nightingale, S.D. (1993). Androgyny predicts empathy for trainees in medicine. *Perceptual and Motor Skills*, 77, 576-578.

Yarnold, P.R., Hart, L.A., and Soltysik, R.C. (1994). Optimizing the classification performance of logistic regression and Fisher's dis criminant analyses. *Educational and Psychological Measurement*, 54, 73-85.

Yarnold, P.R., Soltysik, R.C., and Martin, G.J. (1994). Heart rate vari ability and susceptibility for sudden cardiac death: An example of multivariable optimal discriminant analysis. *Statistics in Medicine*, 13, 1015-1021.

Yarnold, P.R., Soltysik, R.C., McCormick, W.C., Burns, R., Lin, E.H.B., Bush, T., and Martin, G.J. (1995). Application of multi variable optimal discriminant analysis in general internal medi cine. *Journal of General Internal Medicine*, 10, 601-606.

Yarnold, P.R., Soltysik, R.C., and Bennett, C.L. (1997). Predicting in-hospital mortality of patients with AIDS-related Pneumocystis carinii pneumonia: An example of hierarchically optimal classifi cation tree analysis. *Statistics in Medicine*, 16, 1451-1463.

Yarnold, P.R., Soltysik, R.C., Lefevre, F. and Martin, G.J. (1998). Predicting in-hospital mortality of patients receiving cardiopul monary resuscitation: Unit-weighted MultiODA for binary data. *Statistics in Medicine*, 17, 2405-2414.

Kanter, A.S., Spencer, D.C., Steinberg, M.H., Soltysik, R., Yarnold, P.R., and Graham, N.M. (1999). Supplemental vitamin B and progression to AIDS and death in black South African patients infected with HIV. *Journal of Acquired Immune Deficiency Syndrome*, 3, 252-3.

Presentations:

Soltysik, R.C, and Yarnold, P.R. Fast solutions to optimal discrimi nant analysis problems. Invited presentation at *TIMS/ORSA Joint National Meeting*, Orlando, 1992.

Soltysik, R.C, and Yarnold, P.R. Special purpose optimal discriminant analyses. Invited presentation at *TIMS/ORSA Joint National Meeting*, Orlando, 1992.

Soltysik, R.C., and Yarnold, P.R. Software for MultiODA. Invited pre sentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Soltysik, R.C., and Yarnold, P.R. Software for multisample regression analysis. Invited presentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Soltysik, R.C., and Yarnold, P.R. Optimal discrimination with an ordered class variable. Invited presentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Yarnold, P.R., Soltysik, R.C., Curry, R.C., and Martin, G.J. Resident selection based on application information and mixed integer programming. Presentation at *Annual Meeting of the Society of Behavioral Medicine*, San Francisco, 1989.

Yarnold, P.R., and Soltysik, R.C. Statistical distributions underlying optimal discriminant analysis. Invited presentation at *TIMS/ORSA Joint National Meeting*, Orlando, 1992.

Yarnold, P.R., and Soltysik, R.C. Optimal discriminant analysis as an alternative to conventional statistical models. Invited presenta tion at *TIMS/ORSA Joint National Meeting*, Orlando, 1992.

Yarnold, P.R., and Soltysik, R.C. The optimal discriminant analysis paradigm. Invited presentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Yarnold, P.R., and Soltysik, R.C. Software for UniODA. Invited pre sentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Yarnold, P.R., and Soltysik, R.C. Multisample optimal discrimination. Invited presentation at *TIMS/ORSA Joint National Meeting*, Chicago, 1993.

Yarnold, P.R., and Soltysik, R.C. Multivariable optimal discriminant analysis: An optimal analog to multivariate analysis of (co)vari ance. Invited presentation at *TIMS/ORSA Joint National Meeting*, Boston, 1994.

Kanter, A.S., Spencer, D., Steinberg, M., Soltysik, R., and Yarnold, P.R.. Supplemental vitamin B complex associated with delay in progression to AIDS / death in South African patients infected with HIV. *Program and Abstracts of the 5th Conference on Retroviruses and Opportunistic Infections*, Chicago, 1998.

Technical Reports:

Soltysik, R.C., and Yarnold, P.R. An expert system for residency admission decision making. Northwestern University Medical School, Chicago, 1990.

Experience:

1991-Present: Consultant
Mathematics, Statistics, Operations Research

University Assignments/Affiliations:
Northwestern University Medical School,
Statistics Consultant, 1996-Present

Northwestern University Medical School,
Research Assistant Professor, 1989-1996

Commercial Assignments/Affiliations:
Intelligent Medical Objects
Statistics Consultant, 1996-1998

R.R. Donnelley and Sons Co.,
Management Science Consultant, 1991-1995

Various Assignments with organizations such as:
New York Times Magazine Group
R.R. Donnelley Logistics Services

1982-1991
Various technical, programming and systems positions with:
SEI Corp.
Information Resources, Inc.
CBS, Inc., Columbia House Division
Center for Research in Marketing

Education:

University of Illinois at Chicago
M.S. 1982, Industrial and Systems Engineering

Southern Illinois University
B.S. 1976, Mathematics

Bibliography

American Educational Research Association, American Psychological Association and National Council on Measurement in Education. (1985). *Standards for Educational and Psychological Testing*. American Psychological Association, Washington, D.C.

Campbell, D.T. and Fiske, D.W. (1959). Convergent and discriminant valida tion by the multi-trait-multi-method matrix. *Psychological Bulletin*, 56, 81-105.

Canadian Psychological Association. (1996). *Guidelines for Educational and Psychological Testing*. First edition 1987.
URL: http://www.cpa.ca/guide9.html#1

Cronbach, L. and Meehl, P. (1955). Construct validity in psychological tests, *Psychological Bulletin*, 52, 4, 281-302.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike (ed.). *Educational Measurement* (2nd Ed.). Washington, D. C.: American Council on Education.

Cronbach, L.J. (1984). *Essentials of Psychological Testing*. 4th Edition. New York: Harper & Row, pp. 154-157.

Goodenough, F. L. (1949). *Mental testing: Its History, Principles, and Applications*. New York: Rinehart.

Hammer, Allen L (ed.) (1991). *Introduction to Type. A description of the theo ry and applications of the Myers-Briggs Type Indicator*. Palo Alto. Consulting Psychologists Press, Inc.

Heylighen F. (1997): Occam's Razor: In: F. Heylighen, C. Joslyn, and V. Turchin (eds.): *Principia Cybernetica Web.* Principia Cybernetica, Brussels, *URL*: < http://pespmc1.vub.ac.be/ADAPYST.html > .

Karmel, L.J. and Karmel, M.O. (1978). *Measurement and Evaluation in the Schools*. 2nd Edition. New York: Macmillan.

Kaufman, L. and Rousseeuw, P.J. (1990), *Finding Groups in Data: An Introduction to Cluster Analysis*. New York: John Wiley.

Kirk, Roger E. (1982). *Experimental Design: Procedures for the Behavioral Sciences*, 2nd Edition, Pacific Grove, CA: Brooks-Cole

Mielke, P.W. and Berry, K.J. (1992). Fisher's exact probability test for cross-classification tables. *Educational and Psychological Measurement*, 52, 97-101.

Moss, P. A. (1994). Can there be validity without reliability?, *Educational Researcher*, 23, 5-12.

Polkinghorne, D. E. (1988). *Narrative Knowing and the Human Sciences*. Albany: State University of New York Press.

Popham, W. (1990). *Construct Validation Strategies*. Englewood Cliffs: Prentice Hall; 107-119.

Salton, Gary J. (1996) *Organizational Engineering: A New Method of Creating High Performance Human Structures*. Ann Arbor, Professional Communications, Inc.

Salton, Gary J. (2000) *Managers Guide to Organizational Engineering*. Amherst, MA: HRD Press, Inc.

Thomas, J.R., Nelson, J.K., and Thomas, K.T. (1999). A generalized rank-order method for nonparametric analysis of data from exercise science: A tutori al. *Research Quarterly for Exercise and Sport*, 70, 1-11

Trochim, William M. *The Research Methods Knowledge Base*, 2nd edition. Internet WWW page, version current as of April 09, 1999, at
a) *URL*: < http://trochim.human.cornell.edu/kb/nomonet.htm >
b) *URL*: < http://trochim.human.cornell.edu/kb/convdisc.htm >
c) *URL*: < http://trochim.human.cornell.edu/kb/concval.htm >
d) *URL*: < http://trochim.human.cornell.edu/kb/concimp.htm >
e) *URL*: < http://trochim.human.cornell.edu/kb/reliablt.htm >
f) *URL*: < http://trochim.human.cornell.edu/kb/reltypes.htm >
g) *URL*: < http://trochim.human.cornell.edu/kb/sampnon.htm >

Yarnold, P.R., and Soltysik, R.C. (2000). *Optimal Data Analysis*. Washington D.C.: American Psychological Association, in press.

Yates, F. (1984). Tests of significance for 2 x 2 contingency tables. *Journal of the Royal Statistical Society*, Series A, 147, 426-463.