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.