

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 provided 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.