

Human Synergistics International Diagnostic System: Survey Reliability and Validity



Robert A. Cooke, Ph.D.

Associate Professor Emeritus
of Managerial Studies,
University of Illinois at Chicago
Director, Human Synergistics
International

Human Synergistics surveys for individual, group, and organizational development were developed by Robert A. Cooke, Ph.D., CEO of HSI, J. Clayton Lafferty, Ph.D., founder of HSI, and their colleagues. These surveys are based on HSI's Circumplex-based model of Constructive and Defensive styles and related constructs identified in proprietary theoretical models such as "How Culture Works."

Human Synergistics develops and offers, through both Internet and paper-based systems, the best in extensively researched and widely respected training and development materials. Our surveys, diagnostic inventories, and experiential exercises have been used for over 45 years by organizations throughout the world for leadership development, team building, and organizational change. HSI strives to provide clients with tools that are practical and relevant to the realities of organizational development—yet are scientifically based, valid, and reliable. Prior to releasing its surveys, HSI develops and tests multiple prototypes, carries out reliability and validity analyses, and uses the results to iteratively improve each successive version.

Additionally, to continuously test and expand the knowledge base around its surveys, HSI conducts and/or supports studies incorporating its surveys—the findings of which are published in research reports and journal articles that are available for your review. Reports and journal articles are listed at: <https://www.humansynergistics.com/resources/research-and-publications>.

Described below are the types of validity and reliability statistics presented in such studies, along with the kinds of statistics (e.g., "norms") presented in our feedback reports to facilitate the interpretation of survey results.

VALIDITY

Def. The extent to which an instrument effectively does what it is intended to do.

Addresses the question: *Does the survey measure what it is being used to measure?*

Criterion-Related Validity

Def. The extent to which scores on an instrument are in agreement with (concurrent validity) or predict (predictive validity) some given criterion or outcome.

Addresses the question: *Are the survey results related to anything important, assessed either at the present time or in the future?*

Criterion-related validity is tested by using measures of association (such as correlation, regression, and analysis of variance).

Construct Validity

Def. The extent to which an instrument adequately reflects one or more relatively abstract organizational characteristics or psychological traits (i.e., “constructs” such as personal styles or cultural norms).

Addresses the question: *Are different phenomena really being measured (and do distinct underlying constructs or dimensions drive responses to the items and explain the survey results?)*

Construct validity is tested using factor analysis techniques such as principal components analysis with varimax rotation.

Consensual Validity

Def. The extent to which there is convergence (agreement) between self-reports and descriptions-by-others on a 360° instrument.

Addresses the question: *Are self and others' reports reasonably consistent?*

Correlations between self and others' reports are used to estimate consensual validity.

Content Validity

Def. The extent to which the survey items are relevant to, and fully represent, whatever the instrument is designed to measure.

Addresses the question: *Do the items adequately capture the full content of the construct(s) being measured?*

Expert judgments are used to estimate content validity.

Face Validity

Def. The extent to which the items on an instrument superficially appear to be relevant to the instrument's purpose and measure what they're supposed to measure.

Addresses the question: *Do the items seem to make sense?*

The reactions of respondents and administrators to the instrument are used to assess its face validity.

RELIABILITY

Def. The extent to which an instrument is consistent in measuring whatever it does measure.

Addresses the question: *Can you depend on the instrument to measure things uniformly?*

Internal Consistency Reliability

Def. The extent to which the individual items or questions associated with a scale on an instrument are measuring the same thing.

Addresses the question: *Do individual respondents answer the relevant questions in a consistent manner?*

Cronbach's coefficient alpha is the most widely used statistic to estimate this type of reliability. There are different opinions regarding the minimum acceptable values of alpha, but HSI surveys and scales generally range from 0.70 to 0.90. **Note:** Coefficient Alpha does not simply reflect scale consistency or unidimensionality, as scale reliability coefficients also are a function of test length. A longer test or a scale with more items increases the reliability coefficient regardless of whether the test is consistent or not. Thus, a very high alpha value (> 0.90) may suggest that certain items are redundant and indicate that the scale length should be shortened.

Inter-Rater Agreement

Def. The extent to which there is convergence among the respondents who are describing an individual, group, or organization via a survey.

Addresses the question: *Is there consistency in the answers provided by different people?*

Inter-rater agreement or reliability can be assessed by means of analysis of variance, the *eta*-squared statistic, or multiple-item estimates of convergence.

Test-Retest Reliability

Def. The degree to which the results obtained on an instrument are consistent over time.

Addresses the question: *Does the instrument assess things in a stable way over time and can it be trusted to measure change?*

This type of reliability is assessed by comparing pre- and posttest survey data through the use of correlations, t-tests, or analyses of variance.

NORMS AND BENCHMARKS

Def. Statistics that supply a frame of reference and give meaning to the results obtained on an instrument.

Answers the question: *Can the scores for the different constructs being measured be compared to one another? (More technically, are the distributions of raw scores for different measures sufficiently similar to permit meaningful comparisons between those scores?)*

Raw Score

Def. The initial (unadjusted) results obtained in scoring an instrument.

Answers the question: *What is the average or total of an individual's responses to the questions associated with a specific scale?*

Scale scores are derived by simply computing the mean or sum of the responses to Likert-type questions (i.e., items with 1 to 5 response option). However, unless the distributions of raw scores for different variables are similar (i.e., their means and standard deviations are about the same), such comparisons can be misleading. Distributions often differ and, when this is the case, raw scores cannot be used to answer questions about whether an individual's tendencies along one trait (or characteristic, style, preference, etc.) are stronger or weaker than along other traits. One of the following approaches is required.

Normed Score

Def. The expression of an obtained raw score in terms of its position within a larger group of scores (e.g., scores on the scale for 100 different respondents).

Answers the question: *What does the individual's score really mean? Is it high or low compared to others' scores on the same scale—and high or low compared to the individual's scores on other scales?*

Raw scores are converted to percentile scores (or another type of standardized score) to make the above comparisons feasible and meaningful. A percentile score is the percentage of scores in the larger reference group that are equal to or lower than the individual's obtained raw score.

Historical Average Score

Def. The score along an item or scale based on the responses of previous respondents in a research sample. For many surveys, this score is represented either by the mean (average score) or median (50th percentile score) across respondents for each item and scale.

Answers the question: *How does an individual's or group's score compare to the scores of others who previously participated in this survey?*

Raw scores for an individual or organization are compared to this value to determine whether their scores are about the same, higher, or lower.

Benchmark Score

Def. The mean score on a scale for individuals, groups, or organizations that are performing relatively well (above average) on the factor being measured or an outcome associated with it.

Answers the question: *How does the score obtained by a specific individual, group, or organization compare to the scores of those who are relatively effective?*

Individuals, groups, and organizations compare their raw scores to benchmark scores to go beyond the Historical Average, interpret their scores against a higher standard, and set more challenging goals for development and improvement.

BIBLIOGRAPHY

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Anastasi, A., & Urbina, S. (1997). *Psychological testing* (7th ed.). Prentice Hall/Pearson Education.
- Cooke, R. A., & Szumal, J. L. (1993). Measuring normative beliefs and shared behavioral expectations in organizations: The reliability and validity of the Organizational Culture Inventory. *Psychological Reports*, 72, 1299-1330. <https://doi.org/10.2466/pr0.1993.72.3c.1299>
- Geisinger, K. F., Bracken, B. A., Carlson, J. F., Hansen, J. C., Kuncel, N. R., Reise, S. P., Rodriguez, M. C., & American Psychological Association. (2013). *APA handbook of testing and assessment in psychology*. American Psychological Association. <https://doi.org/10.1037/14048-000>
- Kerlinger, F. N. (1986). *Foundations of behavior research, 3rd edition*. Holt, Rinehart and Winston.
- Salkind, N. J. & Frey, B. B. (2022). *Tests & measurement for people who (think they) hate tests and measurement*. Sage Publishing.
- Schuler, R.A. & Youngblood, S.A. (1986). *Effective personnel management, 2nd edition*. West Publishing.
- Szumal, J. L., Boglarsky, C. A., & Cooke, R. A. (2021). Thinking and behavioral styles as described by self versus others: A replication and extension with male and female managers. *Management Review Quarterly*, 72, 677-706. <http://doi.org/10.1007/s11301-021-00210-3>
- The Psychological Corporation (1986). *Tests, products, and services for business, industry, and government*. Harcourt Brace Jovanovich.
- Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, 2, 53-55. doi:10.5116/ijme.4dfb.8dfd <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4205511/>
- The reliability and validity of HSI's surveys have been extensively tested by our Research & Development team as well as by independent researchers who have used our materials in their studies. For a list of, and links to, relevant articles in peer-reviewed journals and other publications, please visit our web site at:
- <https://www.humansynergistics.com/resources/research-and-publications>.