



The Validity and Reliability of Spherical Dynamics' (SDI) Assessment

Background

Face Validity

W. Michael Sturm, SDI founder, created the assessment over a 25-year period of research and development. The assessment uses the method of paired comparisons, a widely accepted and well-researched psychological assessment method. The assessment was initially developed empirically to match learning and teaching styles in the education environment. The assessment has been continually refined and improved over the 25-year period to produce the current Internet delivered version. In our experience with client self reporting, the assessment and its results have shown over 90% validity and reliability as determined by the more than 30,000 administered clients in business, industry, and educational applications.

Substantive/Structural Validity

The Assessment is based on a combination of Paired Comparisons and Task Completion from a Gestalt framework. The methods of paired comparisons are based on Thurstone's Law of Comparative Judgment.

L.L. Thurstone was the first to apply mathematical statistical analysis to psychological testing. When a person is presented with a stimulus, such as, a word-pair or a picture, a sequence of events occurs. First the distal stimulus are patterned on the retina, then transferred into electrical signals and delivered to the brain. The brain then sets up mental representations using separate groups of neurons, one group for each stimulus. A complex psychophysiological transformation then occurs, which triggers a series of electrochemical events in the cells of the brain.

In the case of the SDI assessment, when an individual is presented with word pairs for comparison, a neurological mechanism triggers two mental representations—one for each of the word pairs. The representation that the individual most prefers will create the highest neurological signal strength, causing that pair to be selected.

Although an individual believes that they are consciously "choosing" one of the word pairs, in reality, it is the electrochemical process within the brain that applies certain rules and controls the response in a somewhat reflexive manner.

Thurstone's method of "paired comparisons" is psychologically measurable and a possible mathematical description of the mechanism is described in a paper, "*Probabilistic Psychophysics with Noisy Stimuli*", by Ennis and Mullen. Although the mathematical



models of the neurological mechanisms are highly complex, the SDI assessment is fast, simple to use and highly accurate.

Development of the assessment started in 1983 and there have been three previous versions of the "Assessment" comprised, respectively, of 120 comparisons, 72 comparisons and 30 comparisons.

The original versions were designated as the *Brainstreaming Inventory*. The "pencil and paper" assessment versions were administered to over 15,000 subjects from the 1980s through the middle 1990s. The current SDI assessment is a randomized on-line software implementation based on the most current 18+ word-pair comparison.

Research and development began by using a seven-category Likert scale, valued 1-7, to weigh over 600 words for their polarity. Words were given to a sample of educators, adults from many occupations and college students. Previously, the individuals from this sample had taken a battery of standardized tests that categorized them as people whose brain processes were generally more "left hemispheric" or "right hemispheric", more "internal" or "external", "concrete" or "abstract", more "rigid" or flexible", more "logical" or "intuitive" and more "personal" or "impersonal" in their processing styles. Once the words were weighted using Likert methods and statistics were analyzed, the more positively weighted words were placed into paired comparisons. The samples were tested and retested at least four more times over eighteen months using the word-pair comparisons, Thurstone methods and statistical analysis. The paired comparisons and weights became part of the assessment in subsequent forms.

The assessment uses four domain variables representing four of the five processes of Spherical Dynamics' Success/Satisfaction Cycle: Explore, Excite, Examine, and Execute. These process domain variables are categorized by their intensity strengths from a maximum of Predominant down to a minimum of Stress. For the purposes of this document we will use the designators, S, M, A, C. The S, M, A, C designators are mapped in this manner: Explore = S, Excite= M, Examine= A, Execute = C.

Reliability And Validity

Temporal Reliability/Form Equivalence

Individual SDI Assessment tests are constructed at the time they are requested using randomly selected word-pair comparison. The word-pair comparisons are selected from a database of specifically classified word-pairs representing process phases of the Success/Satisfaction Cycle.

Since the SDI assessment measures process variables dispersed categorically at the specific point in time that is administered, temporal reliability is not primary to demonstrating the external validity of this assessment. Nevertheless, as seen below, test-retest reliability is very high. A subject may change slightly over time depending on their current situation and circumstances and the amount of time they have been exposed to the specific situation and circumstances. For example, if a subject who is low in the



“Activate Phase” is forced to manage a large complex project, their “Activate Phase” score may rise a little.

Summarized results of studies conducted in the 1990’s with various clients are presented below in Table 1. The six (6) reliability studies used the statistical method of Pearson Product Moment Correlation Coefficient and a post two-tail t-test to measure the reliability and degrees of freedom variance. These results are all statistically significant at the .05 level and not patterned with lower reliability coefficients as would be expected with subjects who might show slight increases depending on life circumstances. The consistently high reliability demonstrated a decline as the sampling went from adults to college students to high school students.

Table 1: Reliability scores for test-retest of all four phases—S, M, A, C

Organizations	Test-Retest Time	N	S [†]	M [†]	A [†]	C [†]
Seagate Software & Kinkos; Orlando, FL	2 days	248	0.91*	0.88*	0.93*	0.90*
University of Maine; Augusta	9 Weeks	121	0.85*	0.78*	0.87*	0.86*
Security Plastics, Inc; Miami Lakes, FL; San Juan, Puerto Rico; & McAllen, TX	4 Weeks	376	0.81*	0.82*	0.94*	0.91*
Growth Associates International, Inc.; Boston, MA; Portland, ME; Boca Raton, FL; & Albany, NY	7 days	456	0.85*	0.87*	0.86*	0.82*
Sample of College Students	~9 months	427	0.79*	0.73*	0.83*	0.84*
Sample of High School Students (grades 10-12)	~8 months	268	0.71	0.72	0.76	0.75

[†]Values for S (Strategizer), M (Motivator), A (Analyzer), and C (Activator) are Pearson Product Moment Correlation scores.

*T-test for significance of the correlation coefficient being greater than zero is $p \leq 0.05$

Most of the above reliability studies and the validity studies given below were conducted on behalf of the University of Maine and the State Department of Maine. Every teacher and administrator participant earned three re-certification credits to be counted towards their professional accreditation. These courses’ three credits were studied and approved by the research and teaching committees of the University of Maine and by the State of Maine’s Commissioner of Education after rigorously applied research into the courses’ content, assessment tools and teaching/learning process methods. Also, the Assessment tool was used in several school districts by their superintendents after they, their school board members and their school management teams were trained.



Internal Consistency

Reliability is a necessary condition but it not a sufficient condition for validity. The additional conditions for validity are presented below.



Construct/Content Validity

The development process for the above mentioned set of word pairs began over 25 years ago. The word-pairs were developed, chosen and subgroups were weighted using the following methodology:

In the 1980's, six different studies were conducted by clients with educational psychologist and SDI founder, W. Michael Sturm. Each study had between 150 and 300 subjects.

1. Word-pairs were chosen using a 7-category Likert scale methodology. Only word-pairs that scored 7 (complete agreement, maximum score) on the Likert scale in all prior testing of subjects were chosen.
2. The word-pairs in (1) were presented as comparisons to a known sample group of only predominant (no Stress) and only Stress (no Predominant or Preferred) in the four variables. Pearson Product Moment Correlations were calculated for these groups and the category 7 word pairs. A significant positive correlation was found with the Predominant sample groups and a significant negative correlation with the Stress sample groups ($p \leq 0.05$).
3. Overall Likert scaling was used to select the 72 word-pairs and weights that, until recently, comprised the assessment word-pair database.
4. Recently, the word-pair database went from 72 pairs to 56 pairs. This reduction was based on two year's of data gathering from the 72 word-pair Assessment outcomes. Statistical averaging and percentile frequency was done on every word-pair within specified energy intensities (Predominant, High Preferred and Preferred) of all four of the Success/Satisfaction Cycle's (S/SC) phases: Strategize (S), Motivate (M), Analyze (A) and Activate (C). Only 14 word-pairs from the highest percentile frequencies within each phase were selected. This refinement from 72 selected word-pairs to 56 selected word-pairs will be statistically analyzed for its efficacy in 2006.



External Validity

In the studies summarized in Table 2, subjects were given the SDI assessment instrument and then they were asked to complete a self-rating (Likert method: 1-7) of the short versions of the “Assessment” phase descriptors as applied to themselves or workshop colleagues. A statistical analysis was applied to the two highest phase scores and descriptor ratings in each study.

Table 2: Validity measures of SDI assessment made against self-rating

Organization	Participants	N	Highest Phase Correlation [†]
Maine State Department of Education; Augusta, ME	Educators	210	0.87*
Growth Associates workshop; Boston, MA	Lay Adults	67	0.84*
Workshops for the Maine State Department of Education; Augusta, ME	Educators	487	0.92*
Security Plastics, Inc.; Miami Lakes, FL; San Juan, Puerto Rico; McAllen, TX	Employees	376	0.85*
University of Maine; Portland, ME	Faculty & Students	214	0.82*
University of Maine; Presque Isle, ME	Workshop Attendees	86	0.91*
University of Maine; Augusta, ME	Re-certification Workshop Attendees	203	0.87*

[†]Comparisons made between highest SDI assessment ‘Phase’ and either self-ratings of the participants or ratings made by colleagues attending the workshop sessions.

*T-test for significance of the correlation coefficient being greater than zero is $p \leq 0.05$

In other studies SDI assessment results were compared to other external measures that test related constructs. In all of the studies summarized in Table 3 below, the statistical analyses applied were the Pearson Product Moment Correlation. Significance was set at $p = 0.05$.

Table 3: Validity measures of SDI assessment made against other related measures

Organization	Participants	N	SDI Measures	Other Measures	Pearson Correlations
Growth Associates; Boston, MA	Lay Adults	67	"I-E" scores	Julian Rotter's Internal-External Test	0.82*
Growth Associates; Boston, MA	Lay Adults	67	"Left-Right" process scores	Rokeach's Open and Closed Mind Test	0.84*
Growth Associates; Boston, MA	Lay Adults	19	Related SDI Assessment Sections	Personality Assessment System of Gittinger and Saunders (PAS)	I-E = 0.79* RA-RU = 0.85* R-F = 0.83*
University of Maine; Presque Isle, ME	Educators	74	"Left-Right" process scores	Raven Progressive Matrices	0.74*
University of Maine; Augusta, ME	Educators	146	SDI Profile	Loye's HCP (Hemispheric Prediction Grid) Profile Test	0.92*
American Cancer Society; Boston, MA	Lay Adults	42	SDI Profile	Left-Brain and Right-Brain descriptors	0.88*
American Cancer Society; Boston, MA	Lay Adults	56	SDI Hemispheric Preference Scores	Miller's "BrainStyle Inventory"	0.86*

*T-test for significance of the correlation coefficient being greater than zero is $p \leq 0.05$

Conclusion Validity

Different groups of individuals were taught using the Assessment and not using the Assessment. Results of the learning were measured and statistically compared to see if the Assessment groups' learning was significantly improved in comparison to control groups (non-Assessment groups). In addition, members of teams and work performance were also studied in relation to the use of the assessment results.

The statistical method used in all of the studies summarized below is the "T- test." A probability ≤ 0.05 was set as the level of significance and was achieved in each study.

The samples, types of comparisons-predictions and validity scores were as follows:

1. At the University of Maine in Fort Kent and Presque Isle, a total of 86 educators in different training programs with a minimum of 12 trainees in each group took the Assessment tool and were taught using the Success/Satisfaction Cycle's (S/SC) phases for individual class instruction and for teaming the participants. These

educators were compared to 2 control groups (29 educator members) where the instructor did not give the Assessment nor teach the classes using the Cycle's methods. Each group was tested for pre-tested for content knowledge and post-tested for the same content using identical questions and scoring. The difference in the individuals' test scores were the measurements used to compare the results of the experiment's two groups. The result is as follows:

$t = 2.89, p \leq 0.05$ (significant difference)

2. For Growth Associates, Boston, MA, 19 six hour corporate training workshops (total of 236 participants) were tested by the Assessment and taught using the S/SC methods of teaching. Three six hour corporate training workshops not using the Assessment (total of 34 participants) and its teaching methods were given. Each workshop class was pre and post tested for content using identical questions and scoring. The difference in the individuals' test scores in the pre and post tests were the measurements used to compare the results of the experiment's two types of workshops. The result is as follows:

$t = 3.08, p \leq 0.05$ (significant difference)

3. In Boston, MA, Growth Associates gave 27 two and one half hour workshops for business people (total of 683 participants). The participants were given the Assessment, tested for content, taught using the S/SC methods of instruction and then re-tested for content. 5 similar workshops (similar content and a total of 67 business people), were given the Assessment, were identically pre-tested and post-tested for content and were taught not using the S/SC methods. The difference in the individuals' test scores in the pre and post tests were the measurements used to compare the results of the experiment's two groups. The result is as follows:

$t = 3.23, p \leq 0.05$ (significant difference)

4. 47 certification teaching courses, certified by the State Department of Education, Augusta, ME were given to educators consisting of 17- 46 participants. The educators were given the Assessment the first class and during the next two classes they were taught by the S/SC methods for teaching. 12 control classes were given the Assessment the first class but were taught without using the S/SC as a method of teaching. All classes were pre and post tested for content using identical questions and scoring. The difference in the individuals' test scores in the pre and post tests were the measurements used to compare the results of the experiment's two classes. The result is as follows:

$t = 2.99, p \leq 0.05$ (significant difference)

5. 83 engineers, sales, customer service and business people in one company (Security Plastics, Inc.) were tested using the Assessment. Then, they were placed into specific jobs, put on teams and given certain supervisors based on their S/SC and the criteria of the role process and team process fit for their position. The performance measures of success and satisfaction for all of the study's participants

after one year were measured using a verbal Likert scale (1 - 7) with a success polarity of “extremely productive” to “ready to be fired” and a satisfaction polarity of “love my work” to “hate my work” or had left their position because they were promoted or didn’t like what they were doing. 94 other employees were tested using the Assessment and were placed into jobs, on teams and with supervisors by using the company’s usual methods. They and their supervisors were also given the verbal Likert scales. The Likert scale scores were the criteria used to demonstrate the following T-test results:

Placement satisfaction by employee: $t = 3.22, p \leq 0.05^*$

Placement success by employee: $t = 2.91, p \leq 0.05^*$

Placement success by the supervisor: $t = 3.14, p \leq 0.05^*$

Placement satisfaction observed by the supervisor: $t = 3.43, p \leq 0.05^*$

*Statistically significant difference

6. 41 employees (29 of this sample were college students) at two stores of a national copy chain (Kinkos of South Florida) were given the Assessment; evaluated for overall job performance by doing each of their rotating tasks during a shift; and evaluated for job performance by doing the rotating task that the S/SC phases predicts would be the highest performance criteria. An overall rotating task performance was measured by the supervisor evaluating each employee with a verbal Likert (1 -7: Failing Performance – Excellent Performance). Overall job performance was measured by averaging all the rotating task scores together. This result of comparing the employees’ productivity doing the overall tasks and their doing the S/SC predictive tasks was as follows:

$t = 2.34, p \leq 0.05$ (significant difference)

A more practical result was that the stores’ percent of turnover dropped 34% during the 3 months that employees spent 80% of their time on the S/SC tasks and 20% rotating.

7. During a three year period at Security Plastics, Inc., in Miami Lakes, FL, 346 adults were placed on 68 teams. 47 teams were formed using S/SC input data and 21 teams were done without using S/SC input data. The performance of these 2 different types of teams were measured by the individual team leaders using a verbal Likert scale (1 -7: 1 = failing and 7 = excellent). The statistical comparison of these differently formed teams was done by a T-test.

$t = 3.16, p \leq 0.05$ (significant difference)

Generalizability

This topic is under study for the total underlying population. Early results yield an underlying multivariate normal distribution. The assessment is based on non-parametric methods but does yield indications that the resulting distribution is multivariate normal, which is common in the social and behavioral sciences.



Current Work

Specific Applications of the Assessment

A three-year study, under the combined auspices of a National Science Foundation (NSF) grant of \$800,000, a Packard Foundation grant and a James Irvine Foundation grant, has been initiated for the Watsonville Digital Bridge Academy (WDBA) at the Cabrillo Community College of California. The initial 2 phases of this project are already completed.

The SDI Assessment and S/SC methods are providing fundamental precepts for teaching, learning, teaming and communicating, as well as management training for students in college or internships; plus the Assessment and S/SC methods are assisting the mentoring adults associated with the students' supporting workforce. The NSF grant provides for disseminating the results of the assessment at WDBA across the USA.

The following results were obtained by an independent evaluator from Columbia Teachers College in 2005:

	PRE-Program (cumulative GPA for last term prior to Foundation Seminar course)	POST-Program (last cumulative grade points minus PRE grade points / last cumulative units attempted minus PRE units attempted)	
Grade Point Average Improvement	1.6696	3.0155	F=23.28, p<.0001
Retention Improvement (Units Completed Per Semester)	5.8198	10.2187	F = 42.93 P<.0001
Academy Progress Improvement (Completion of Units Attempted for Letter Grade Instead of Pass/Fail)	2.7842	8.1458	F=50.31 P<.0001
Improved Persistence (# of semesters students attempted coursework / # of semesters)	0.27798	0.51302	F=12.85 P<.0007

In 2006 newly acquired grants from NSF, the James Irvine Foundation and the Packard Foundation will be supporting the dissemination and training of college instructors to bring



this program and its methods to community colleges across California and throughout the USA.

Summary and Conclusions

The SDI Assessment has been administered to over 30,000 clients in the past 25 years with almost a 90% overall validity and reliability. The validity and reliability has been measured by numerous methods, inclusive of face validity in business and educational applications, substantive and structural validity, temporal reliability and form equivalence, internal consistency, construct and content validity and external validity.

The current SDI Assessment and its action tools of Team Process Inventory and Role Process Inventory are constantly being researched within different populations and markets for reliability, validity for human resource measures and for its customers' knowledge about its Return on Investment (ROI) . Present and future studies will be published by the researchers of Spherical Dynamics at the appropriate time.

Addendum for Perusal

The fundamental structure of the SDI Assessment is based on Thurstone's Law of Comparative Judgment, a well-known and highly respected mathematical representation of measurable psychological responses. Thurstone's method is more fully described in the paper by Ennis and Mullen, "*Probabilistic Psychophysics with Noisy Stimuli*". A copy of which is attached to this paper's email.