Academy of Management Review 1981, Vol. 6, No. 2, 289-296

Experiential Learning Theory and The Learning Style Inventory: A Reply to Freedman and Stumpf

DAVID A. KOLB
Case Western Reserve University

Freedman and Stumpf's critique of experiential learning theory and the Learning Style Inventory (LSI) is seriously flawed. Their judgments concerning the validity of experiential learning theory rest primarily on an analysis of the internal characteristics of the LSI, with no attention to the construct validity of that instrument; and they are made without analysis or even awareness of the considerable research literature on experiential learning. Their questions concerning the reliability of the LSI stem from a lack of understanding of the role of variability and situational adaptation in the experiential learning process. Similarly, their criticism of the forced-choice format of the LSI fails to recognize the theoretical rationale for the LSI instrument structure.

The article by Freedman and Stumpf in the July 1980 issue of *Review*, "Learning Style Theory: Less Than Meets the Eye" requires a reply in order: (1) to correct certain inaccuracies in their report, (2) to respond to their conclusion that the theory of experiential learning has little empirical support, and (3) to clarify the dialectical nature of experiential learning theory and the attendant implications for reliability studies and the structuring of the Learning Style Inventory (LSI). This rebuttal will show that Freedman and Stumpf have improperly assessed the validity of experiential learning theory by basing their judgment primarily on an analysis of the internal characteristics of the LSI, with only the most superficial review of research on the theory. Their conclusion that "independent research has not supported the theory and suggests that its normative use should be suspended" [p. 445] is based on casual scholarship and faulty reasoning. Their criticism of the reliability and structure of the LSI represents misapplications of statistical assumptions of stability and independence to a theory based on variability and interdependence.

Empirical Support for Experiential Learning Theory

The most serious misstatement of fact by Freedman and Stumpf is their assertion that "empirical evidence supporting learning style theory and the LSI has come from a single piece of unpublished research." *The Learning Style Inventory: Technical Manual* [Kolb, 1976b] describes the theory of experiential learning, the internal properties, and some validity studies of the LSI. The 1979 updating of the Bibliography of Research on Experiential Learning Theory and the Learning Style Inventory published in the *Technical Manual* lists over 60 articles and dissertations reporting research on experiential learning and the LSI. Since then, over 30 new studies have been compiled for the next manual revision. That this literature is accessible to those who seek it out is indicated by an independent comparison of the LSI with other tests that assess learning style, recently published by Kirby [1979]. Kirby gives the LSI good marks on literature citations, indicating that a "fair amount" of supportive literature exists.

The central argument of the Freedman and Stumpf paper is that experiential learning theory is invalid because the LSI is unreliable and improperly structured in a forced-choice format that biases results in favor of the theory. This argument is untenable because it is based on a notion of scientific method that fails to distinguish between a theory and the operational measures of its variables. Experiential learning theory can no more be proven invalid solely by an analysis of the internal characteristics of the LSI than it can be proven valid by such an analysis. Validation of a theory is a complex process accomplished by concurrent, predictive, and/or construct validation of operational measures of variables in the theory against *external* criteria predicted by the theory as well as by more qualitative judgments concerning the theory's ability to raise interesting and practical questions for investigation. In addition, it is preferable that variables in the theory be addressed by different methods in order to separate irrelevant method variance from variance in the construct being measured [Campbell & Fiske, 1959].

Freedman and Stumpf concentrate primarily on the internal characteristics of the LSI and hence can draw conclusions about the utility of the instrument but not about the validity of the theory on which it is based. That requires a review of the construct validity of the LSI and other operational measures of the variables in experiential learning theory. While such a review is beyond the scope of this rebuttal, a reading of the literature cited above would suggest that there is substantial empirical support for the theory of experiential learning using different operational definitions of the theory's constructs in addition to the LSI, and including replications of certain findings by independent investigators [see, e.g., Carlsson, Keane, & Martin, 1976; Clarke, Oshiro Wong, & Yeung, 1977; Fry, 1978; Gish, 1979, 1980; Griggs, 1979; Gypen, 1980; Kolb, 1981; Manring, 1979; Plovnick, 1975; Sims, 1980; Wolfe & Kolb, 1979].

The Utility and Reliability of the Learning Style Inventory

Although I disagreed with some of the assumptions and data interpretations in Freedman & Stumpf's earlier paper in the *Academy of Management Journal* [1978], I applauded its emphasis on caution in the use of psychological tests for students and other laypersons. In my opinion, the public is quite naive about psychological tests and often gives test results more credibility than the scientific data merit. For this reason, the LSI has a simple, straightforward format that does not lend itself to pseudoscientific puffery. In its use, we always emphasize that the inventory is nothing more than it appears to be — the person's own self-description of how he or she learns compared with the similar self-descriptions of the normative sample. It is a nine-item self-description questionnaire. Each item asks the respondent to rank order four words in a way that best describes his or her learning style. One word in each item corresponds to one of four learning modes — Concrete Experience (sample word, *feeling*), Reflective Observation (*watching*), Abstract Conceptualization (*thinking*), and Active Experimentation (*doing*). The LSI measures an individual's relative emphasis on four learning abilities — Concrete Experience (CE), Reflective Observation (RO), Abstract Conceptualization (AC), and Active Experimentation (AE) — plus two combination scores that indicate the extent to which an individual emphasizes abstractness over concreteness (AC — CE) and the extent to which an individual emphasizes action over reflection (AE — RO). All published versions of the LSI [Kolb, 1976a; Kolb, Baker, & Gish, 1979; Kolb, Rubin, & McIntyre, 1979] stress that the inventory is only a starting point for understanding one's approach to learning that should be supported by other data about how one learns, and the *Technical Manual* states these and other limitations on use of the Inventory for selection purposes [Kolb, 1976b, p. 13].

The theory of experiential learning maintains that learning is a process involving the resolution of dialectical conflicts between opposing modes of dealing with the world — action and reflection, concreteness and abstraction. Learning styles represent preferences for one mode of adaptation over the others; but these preferences do not operate to the exclusion of other adaptive modes and will vary from time to time and situation to situation. This idea of variability seems essential, since change and adaptation to environmental circumstances are central to any concept of learning.

When it is used in the simple, straightforward, and open way intended, the LSI usually provokes an interesting self-examination and discussion that

recognizes the uniqueness, complexity, and variability in individual approaches to learning. The danger lies in the reification of learning styles into fixed traits, such that learning style types become stereotypes used to pigeon hole individuals and their behavior.

Freedman and Stumpf, contrary to experiential learning theory, do see learning styles as fixed traits. They state that "test-retest reliability for the two samples after only three weeks was rather low (median r = .50), suggesting that *the LSI is rather volatile, unlike the theoretical constructs studied*" [p. 446, emphasis added].

An emphasis on process as opposed to fixed psychological traits presents some special problems in assessing measurement error in the LSI. Concepts of split-half and test-retest reliability are most appropriate techniques for the assessment of measurement error in independent psychological traits that in theory are assumed to be fixed and unchanging. The basic learning modes assessed by the LSI, however, are theoretically *interdependent* (i.e., any action, including responding to the test, is determined in varying degrees by all four learning modes) and *variable* (i.e., the person's interpretation of the situation should to some degree influence which modes are used). Thus, even if there were no measurement error in the LSI, we would predict test-retest and split-half reliability coefficients less than 1.0.

The dialectical interdependence of the learning style modes should reduce both split-half and test-retest reliability coefficients somewhat, because few individuals in any sample would be pure types [Myers, 1962, p. 19]. Test-retest reliability coefficients should be further reduced by the hypothesized situational variations in learning style modes. An individual's learning style is conceived to be a modal orientation that varies to some degree from situation to situation. Thus an abstract person might become more concrete in viewing a painting, but still not experience it as concretely as a concrete person. In this sense, learning styles are similar to concepts of motivation, concerning which McClelland [in Atkinson, 1957, Chap. 1] has argued that traditionally reliable (i.e., stable) measures in fact have greater measurement error because they are not sensitive to changes over time.

Thus we are left with a dilemma in assessing measurement error in the LSI using reliability coefficients. While we would theoretically predict lower reliability coefficients in the LSI modes than on independent fixed psychological traits, we cannot know whether lower reliability coefficients are in fact a result of these theoretical considerations, or are simply measurement errors in the LSI. To assess measurement error, therefore, one must rely more on the construct validity of the LSI. If the LSI shows a consistent pattern of relationships with predicted dependent variables, as it does in much of the empirical literature previously cited, then that is an indicator that the inventory is to some degree accurately measuring the learning modes postulated by experiential learning theory.

Results of the four test-retest reliability studies reported in the *Technical Manual* plus those of a recent study by Geller [1979] support the conclusion that responses to the LSI are determined by variable situational factors as well as a more stable personal disposition. When these studies are arranged in a hierarchy combining time between testings and discontinuity of experience in the test-retest time period from experience immediately preceding the first administration of the LSI, we see that in general *test-retest reliabilities for the six LSI scales are highest* when the test-retest time period is short and experience in the test-retest period is highly similar to previous experience — i.e., *when there is no great change in situational circumstances* (see Table 1; the 1978 Freedman and Stumpf study is not included in this table because the discontinuity-of-experience variable could not be rated from the description provided in their paper). The studies by Geller and Plovnick [1974], because they represent the shortest time period between tests and the least discontinuity of experience, probably give about the highest test-retest reliabilities one could practically expect. Although these results would not be satisfactory for measurement of a stable psychological trait, they are more acceptable for a construct that is theoretically conceived of as situationally variable.

Split-half reliabilities for the LSI are better than the test-retest coefficients — as one might predict, since they are unaffected by situational variability. Table 2 shows split-half reliabilities obtained by applying the Spearman-Brown prophecy formula to obtained correlations between LSI scale halves for five different groups: two groups of about 50 MIT Sloan Fellows (mid-career managers attending a one-year master's program in management), a

Table 1
Learning Style Inventory Test-Retest Reliability Studies[a]

Population	Time Between Testing	Discontinuity of Experience[b]	LSI Scales CE	RO	AC	AE	AC-CE	AE-RO	n
1. U.S. students in foreign medical schools[c]	1 mo.	low	.56	.52	.59	.61	.70	.55	50
2. Boston U. senior medical students[d]	3 mos.	low	.48	.73	.64	.64	.61	.71	27
3. MIT MS students in management	3 mos.	high	.48	.51	.73	.43	.51	.48	23
4. MIT MS students in management	6 mos.	medium	.46	.34	.64	.50	.53	.51	18
5. MIT Sloan Fellows	7 mos.	high	.49	.40	.40	.33	.30	.43	42

[a]Reliability coefficients are Pearson product-moment correlations.
[b]Details of this rating are described in Geller, Lester. Reliability of the learning style inventory. *Psychological Reports*, 1979, *44*, pp. 555-561; and in Kolb, David A. *Learning style inventory: Technical manual.* Boston: McBer & Co., 1976 (rev. eds., 1978, 1979), pp. 12-18.
[c]Geller, 1979, p. 557.
[d]Plovnick, Mark. *Individual learning styles and the process of career choice in medical students.* Unpublished doctoral dissertation, MIT Sloan School of Management, 1974, p. 32.

Table 2
Spearman-Brown Split-Half Reliability Coefficients for the Learning Style Inventory

Sample	n	CE	RO	AC	AE	AC—CE	AE—RO
MIT Sloan Fellows	47	.69	.37	.65	.64	.78	.78
MIT Sloan Fellows	50	.43	.59	.81	.61	.80	.81
Active Managers	90	.61	.58	.71	.62	.78	.85
Harvard MBA's	442	.50	.63	.74	.67	.75	.86
Lesley College Undergrads	58	.48	.63	.74	.65	.82	.86
Total	687	.55	.62	.75	.66	.74	.82

miscellaneous group of 90 practicing managers, 442 Harvard MBA's, and 58 female Lesley College undergraduates.

The results show reasonable reliability coefficients for the two combination scores AC — CE and AE — RO. Coefficients of about .80 are consistent across all five samples and are on a par with most psychological self-report instruments. They are, for example, almost identical with the split-half reliabilities reported by Myers [1962] for the Jungian type-combination scales of the Myers-Briggs Type Indicator, a test that is widely used for counseling and research. The coefficients for the four basic scales are somewhat less satisfactory, with the possible exception of AC. It seems likely that these lower coefficients are reflecting genuine measurement error due to the shortness of the scales (only six scored items). Based on these results, the cautious recommendation is that researchers should rely on the combination scores AC — CE and AE — RO and use the single scales primarily for qualitative description.

The Format of the Learning Style Inventory

Freedman and Stumpf also raise questions about the forced-choice ranking format of the LSI. The choice of this format in constructing the LSI was governed by two considerations. The first and most important was the criterion of ecological validity [Brunswick, 1943]. Because the theory of experiential learning postulates that a learning response to any life situation requires the resolution of conflicts among the four learning modes, it was reasoned that a test of learning styles should be constructed so that it also required a similar conflict among choices. The second consideration was the social desirability response set. Variation in responses to self-report instruments in Likert scale or true-false formats has been shown to be largely influenced by the tendency to rate oneself highly on items that are socially desirable [Edwards, 1953]. In constructing the LSI, we attempted to select four words for each of the nine ranking items that were of equally positive social desirability, thus controlling for this response bias.

Freedman and Stumpf, however, suggest that the LSI is biased *because of* the forced-choice format. They base their argument on two studies. The first, by Lamb and Certo [1978], compared the responses of 383 undergraduate students to the LSI and an instrument that asked subjects to rate the 24 scored LSI words independently on a seven-point Likert scale. With the LSI, they found negative correlations between AC and CE scales and the AE and RO scales similar to those of previous studies, as well as a similar pattern of negative correlations among the individual scale items. The Likert scale LSI, however, showed all positive correlations among the scales and the items. The conclusion they draw from this comparison is that the forced ranking LSI is biased because the Likert scale LSI did not show the negative correlations between abstractness and concreteness and between action and reflection postulated in experiential learning theory.

This reasoning is not persuasive. First, it is likely that it is the Likert scale LSI that is biased because of social desirability response sets. Because the LSI words were all deliberately chosen to be of equally positive social desirability, one would predict that the major variation in response would be due to this positive bias in self-rating, producing high positive intercorrelations among items. In another report of the same data [Certo & Lamb, 1980], the authors show that nearly all of the items in the Likert scale LSI load heavily on a first factor, a pattern that is consistent with other investigations of the social desirability response set in psychological tests. Although they, in neither report, show means and standard deviations for items or scales using the Likert scale LSI, I would predict, based on my own experiments with a similar format, that the self-ratings would be highly skewed toward the high end with restricted variance — a further indication of the positive social desirability response set. Second, in light of the earlier ecological validity consideration, it is difficult to see how one can argue that the assessment of how an individual will resolve conflicts among alternative orientations will be achieved more accurately by independently presenting the orientations than by directly asking the person to prioritize them. Finally, one would not argue in the first place that the internal negative correlations between the AC and CE and between the AE and RO scales are strong evidence for the validity of experiential learning theory or even for the validity of the LSI instrument. What these statistics do is describe the characteristics of the LSI to allow an assessment of

how adequately the instrument meets the assumptions of the theory it is designed to test. The LSI is designed to meet the assumptions of interdependence and dialectical conflicts among adaptive modes, not to test these assumptions. Freedman and Stumpf are correct when they say "The LSI cannot fail to support the theory on which it is based" [1980, p. 466], if by that they mean the AC—CE scales and AE—RO scales will be negatively correlated, since the forced-ranking format of the LSI will necessarily produce negative correlations among items and scales.

Testing the basic assumptions of experiential learning theory requires controlling for the built-in negative correlations in the LSI or validation of the scales against external criteria. It is possible to control for the "bias" introduced by the forced-choice format of the LSI by using data from the second Certo and Lamb study [1979] cited by Freedman and Stumpf. Certo and Lamb generated 1,000 random responses to the LSI instrument and intercorrelated the resulting scale scores. The resulting correlations measure the magnitude of the built-in negative correlations in the LSI. If these correlations are used as the null hypothesis instead of the traditional zero point to test for significance of difference, the hypothesized negative relationships between AC and CE and AE and RO can be tested with the forced-ranking effect partialed out. Thus, when Certo and Lamb's random correlations are compared to the empirical correlations obtained from 807 subjects reported in the LSI *Technical Manual*, using the formula provided by McNemar [1957, p. 148], both the AC—CE correlations and AE—RO correlations are significantly *more* negative than the random correlations (random AC—CE = -.26, empirical = -.57, p of difference < .001; random AE—RO = -.35, empirical = -.50, p of difference < .001).

External validation of these negative relationships comes from a recent study by Gypen [1980]. He correlated ratings by professional social workers and

Table 3
Pearson Correlation Coefficients Between Learning Style Inventory Scales and Ratings of Learning Orientations at Work[a, b]

	LEARNING ORIENTATIONS ON CURRENT JOB			
LSI Scales	Concrete Experience	Reflective Observation	Abstract Conceptualization	Active Experimentation
Concrete Experience (CE)	.49 p < .001	-.17 n.s.	-.37 p < .01	.08 n.s.
Reflective Observation (RO)	.03 n.s.	.22 p < .05	.12 n.s.	-.34 p < .01
Abstract Conceptualization (AC)	-.30 p < .05	-.04 n.s.	.27 p < .05	-.09 n.s.
Abstract Experimentation (AE)	.01 n.s.	-.09 n.s.	-.06 n.s.	.37 p < .01
Abstract-Concrete (AC—CE)	-.42 p < .001	.06 n.s.	.36 p < .003	-.07 n.s.
Active-Reflective (AE—RO)	-.02 n.s.	-.18 p < .08	-.07 n.s.	.43 p < .001

[a] $n = 58$

[b] Gypen, Jan. *Learning style adaptation in professional careers: The case of engineers and social workers.* Unpublished doctoral dissertation, Case Western Reserve University, 1980, p. 71.

comes from a recent study by Gypen [1980]. He correlated ratings by professional social workers and engineers of the extent to which they were oriented toward each of the four learning modes in their current job and their LSI scores obtained four to six months earlier. Each mode was rated separately on a seven-point scale describing the learning mode in a way that attempted to minimize social desirability bias. Table 3 shows the correlations between the subjects' LSI scores and self-ratings of their current job orientation. These results provide strong support for the negative relationship between Concrete Experience and Abstract Conceptualization and somewhat weaker support for the negative relationship between Active Experimentation and Reflective Observation.

The Gypen study and the "corrected" internal correlations among LSI scales both demonstrate empirical support for the bipolar nature of the experiential learning model that is independent of the forced-ranking method used in the LSI.

Concluding Remarks

Whether experiential learning theory is more or less than meets the eye depends in part on how carefully one looks. An evaluation of the status of experiential learning theory must await a more thorough and evenhanded review of the theoretical and empirical literature. Freedman and Stumpf's critique of the LSI is an analysis of the instrument from the perspective of a widely shared doctrine of psychological testing — namely, that above all any test must meet statistical criteria of independence and stability. I hold that this doctrine is applied inappropriately in cases such as the theory of experiential learning, in which the theory is explicitly based on assumptions of interdependence and variability. No operational measure of a theory can be used to test that theory if it is not constructed so that it is faithful to the theory's premises. Establishing better operational measures of the constructs in experiential learning theory does represent an important item for the research agenda. Important tasks on this agenda are the development of behavioral as well as self-report measures of the learning modes, assessment of the developmental dimensions of experiential learning, and the assessment of situational variability in response to environmental demands. In recent research studies, we have made significant progress on some of these issues [Kolb & Wolfe, 1980]. But there is much to be done and the efforts of our colleagues who approach this work, with whatever preconceptions, give welcome assistance in this inquiry.

REFERENCES

Anderson, Janis; & Bell-Daquilante, Cheryl A. *The influence of communication behaviors and predisipositions on individual learning style preference.* Unpublished manuscript, Department of Speech Communication, West Virginia University, 1980.

Atkinson, J. (Ed.). *Motives in fantasy, action, and society.* Princeton: Van Nostrand, 1957.

Brunswick, Egon. Organismic achievement and environmental probability. *Psychological Review*, 1943, *50*, 255-271.

Campbell, D.T.; & Fiske, D.W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, 1959, *56*, 81-105.

Carlsson, B.; Keane, P.; & Martin, J.B. R&D organizations as learning systems. *Sloan Management Review*, 1976, *17*(3), pp. 1-15.

Certo, S.; & Lamb, S. Identification and measurement of instrument bias within the learning style instrument through a Monte Carlo technique. *Southern Management Proceedings*, 1979, pp. 22-24.

Clarke, D.; Oshiro, S.; Wong, C.; & Yeung, M. *A study of the adequacy of the learning environment for business students in Hawaii in the fields of accounting and marketing.* Unpublished manuscript, School of Business, University of Hawaii, 1977.

Edwards, A.L. The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology*, 1953, *37*, 90-93.

Freedman, R.; & Stumpf, S. What can one learn from the learning style inventory? *Academy of Management Journal*, 1978, *21*, 275-282.

Freedman, Richard; & Stumpf, S. Learning style theory: Less than meets the eye. *Academy of Management Review*, 1980, *5*, 445-447.

Fry, R. *Diagnosing professional learning environments: An observational scheme for matching learner style with situational complexity.* Unpublished doctoral dissertation, MIT Sloan School of Management, 1978.

Geller, Lester. Reliability of the learning style inventory. *Psychological Reports*, 1979, *44*, 555-561.

Gish, Glen. *Integrated learning: A process model.* Unpublished manuscript, Department of Organizational Behavior, Case Western Reserve University, 1979.

Gish, Glen. *Adult development and adaptation: An empirical test of the experiential learning theory and adaptive flexibility.* Unpublished doctoral dissertation, Case Western Reserve University, 1980.

Griggs, Walter. *On the measurement of the learning press of technical work environments.* Unpublished doctoral dissertation, Case Western Reserve University, 1979.

Gypen, Jan. *Learning style adaptation in professional careers: The case of engineers and social workers.* Unpublished doctoral dissertation, Case Western Reserve University, 1980.

Kirby, Patricia. *Cognitive style, learning style, and transfer skill acquisition.* Columbus, Ohio: National Center for Vocational Education, 1979.

Kolb, David A. *The learning style inventory: Self-scoring test and interpretation booklet.* Boston: McBer, 1976. (a)

Kolb, David A. *Learning style inventory: Technical manual.* Boston: McBer & Company, 1976 (rev. eds., 1978, 1979). (b)

Kolb, David A. Disciplinary inquiry norms and student learning styles: Diverse pathways for growth. In A. Chickering (Ed.), *The modern American college.* San Francisco: Jossey-Bass, 1981.

Kolb, David A. *Experience, learning, and development: The theory of experiential learning.* Englewood Cliffs, N.J.: Prentice-Hall, in press.

Kolb, David A.; Baker, Richard; & Gish, Glen. *Personal learning guide.* Dallas: Richard Baker, 1979.

Kolb, David A.; Rubin, Irwin; & McIntyre, James. *Organizational psychology: An experiential approach* (3rd ed.). Englewood Cliffs, N.J.: Prentice-Hall, 1979.

Kolb, David A.; Wolfe, Donald M. *Professional education and career development in social work and engineering: A cross-sectional study of adaptive competencies in experiential learning: Final report.* NIE Grant No. NIE-G-77-0053, 1980. Available from the first author.

Lamb, Steven: & Certo, Samuel. The learning styles inventory and instrument bias. *Academy of Management Proceedings,* 1978, pp. 28-32.

McNemar, Q. *Psychological statistics.* New York: Wiley, 1957.

Manring, Susan. *Career patterns of technically trained professionals: A person/environment interaction model.* Unpublished doctoral dissertation, Case Western Reserve University, 1979.

Myers, I.B. *The Myers-Briggs type indicator.* Princeton, N.J.: Educational Testing Service, 1962.

Plovnick, Mark. *Individual learning styles and the process of career choice in medical students.* Unpublished doctoral dissertation, MIT Sloan School of Management, 1974.

Plovnick, Mark. Primary-care career choices and medical student learning styles. *Journal of Medical Education,* 1975, *50,* 849-855.

Sims, Ronald. *Preparation for professional careers and changing job roles: An assessment of professional education.* Unpublished manuscript, Department of Organizational Behavior, Case Western Reserve University, 1980.

Wolfe, Donald M.; & Kolb, David A. Career development, personal growth, and experiential learning. In D. Kolb; I. Rubin; & J. McIntyre (Eds.), *Organizational psychology: A book of readings.* Englewood Cliffs, N.J.: Prentice-Hall, 1979, 535-563.

David A. Kolb is Professor of Organizational Behavior at Case Western Reserve University, Cleveland.

Received 9/10/80