DOCUMENT RESUME

ED 377 221 TM 022 391

AUTHOR Ruble, Thomas L.; Stout, David E.
TITLE A Critical Assessment of Kolb's Learning Style Inventory.
PUB DATE Jun 94
NOTE 54p.
PUB TYPE Information Analyses (070) -- Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS *Cognitive Style; *Construct Validity; Learning; *Psychometrics; *Scoring; *Test Reliability; Test Use
IDENTIFIERS *Learning Style Inventory

ABSTRACT

This paper reviews and critically evaluates the psychometric properties of Kolb's Learning Style Inventory (LSI). The LSI was developed originally in the 1970s (Kolb, 1976a) and was revised in the 1980s (Kolb, 1985). Although the LSI has been very popular, extensive evidence available in the published literature indicates that both the original and revised versions of the LSI are deficient in reliability and construct validity. It is concluded that the LSI does not provide adequate measures of learning styles and that its use in research should be discontinued. To improve understanding of the learning process, valid instruments are essential. An appendix presents three exhibits containing a learning style gird, a scoring key, and a list of investigations of the LSI. (Contains 86 references.) (Author/SLD)

A CRITICAL ASSESSMENT OF KOLB'S LEARNING STYLE INVENTORY

Thomas L. Ruble
Department of Management
Rider University
Lawrenceville, NJ 08648-3099

David E. Stout
Department of Accountancy
Villanova University
Villanova, PA 19085-1678

June, 1994

A CRITICAL ASSESSMENT OF KOLB'S LEARNING STYLE INVENTORY

TABLE OF CONTENTS

ABSTRACT

This paper reviews and critically evaluates the psychometric properties of Kolb's Learning Style Inventory (LSI). The LSI was developed originally in the 1970s (Kolb, 1976a) and was revised in the 1980s (Kolb, 1985). Although the LSI has been very popular, extensive evidence available in the published literature indicates that both the original and revised versions of the LSI are deficient in reliability and construct validity. We conclude that the LSI does not provide adequate measures of learning styles and, therefore, its use in research should be discontinued. To improve our understanding of the learning process, valid instruments are essential.

INTRODUCTION

The Learning Style Inventory (LSI) was introduced in the 1970s by David Kolb (1971; 1976a) to measure an individual's relative preferences for four different learning abilities: (1) concrete experience (CE), (2) abstract conceptualization (AC), (3) reflective observation (RO), and (4) active experimentation (AE). The LSI was based on the Experiential Learning Model (ELM), a two dimensional model for classifying learning styles corresponding to different stages in the learning process (Kolb, 1974). According to the ELM, the four learning abilities represent two separate, bipolar dimensions (CE versus AC and RO versus AE). Further, the ELM proposes that individuals tend to favor one ability on each dimension based on their heredity, experience, and environment. When the preferred abilities are combined, they define a distinct learning style: Accommodator (CE and AE), Diverger (CE and RO), Assimilator (AC and RO), or Converger (AC and AE). A diagram of the dimensions and learning styles is presented in Exhibit 1 of the Appendix.

This paper reviews and critically evaluates the psychometric properties of the LSI. The original LSI (LSI-1976), was revised in the 1980s (Kolb, 1985). Although the LSI-1976 was very popular, it had become the subject of increasing criticism. While the revised LSI (LSI-1985) represents an improvement in some areas, in other areas it accentuates problems with the original instrument.

Despite the revision of the instrument in 1985, the LSI-1976 has been used in a variety of studies published recently (e.g., Bostrom, Olfman, & Sein, 1990; Green, Snell, & Parimanath, 1990; McKee, Mock, & Ruud, 1992; Sein & Robey, 1991). Moreover, the LSI-1976 remains alive in various textbooks (e.g., Daft, 1994). Thus, our assessment begins with the LSI-1976.

The evaluation of both versions of the LSI will address three major concerns: (1) measurement problems based on ordinal and ipsative scales, (2) basic issues in reliability including internal consistency and temporal stability, and (3) construct validity. For the LSI-1985, the effects of a response-set bias also will be considered.

THE LSI-1976

To measure the four learning abilities, the LSI-1976 asks respondents to rank-order nine sets of four words, each word corresponding to one of the four abilities. Each set of four words is ranked from 1 (low) to 4 (high). The LSI-1976 is scored by summing six items in each column (three items per column were dropped from scoring and serve as distractors). The sums for each column yield scores for the four learning abilities: CE, RO, AC, and AE. The nine sets of words and the scoring key are presented in Exhibit 2 of the Appendix.

Because the ELM proposes that the learning abilities represent the opposite ends of two bipolar dimensions, the four ability scores are combined into two dimension scores (AC-CE) and (AE-RO). The dimension scores are used to locate individuals in one of four quadrants corresponding to different learning styles: Accommodator (CE and AE), Diverger (CE and RO), Assimilator (AC and RO), or Converger (AC and AE). To make the classification of learning styles, Kolb (1976b) provides norms derived from the scores of 1,933 subjects. Based on these norms, scores of +2 on the AC-CE scale puts one on the CE side and +3 puts one on the AC side. Similarly, scores of +2 on the AE-RO scale puts one on the RO side and +3 puts one on the AE side.

MEASUREMENT PROBLEMS BASED ON ORDINAL AND IPSATIVE SCALES

According to Carmines and Zeller (1979), measurement can be viewed as the process of linking theoretical constructs (e.g., learning styles) to empirical indicators (e.g., LSI scores). When the link between constructs and indicators is strong, analysis of empirical results based on these indicators can lead to inferences about the relationships among the theoretical constructs. However, when the linkage between constructs and indicators is weak or faulty, analysis of empirical results can lead to incorrect or misleading inferences about the underlying constructs. Thus, measurement issues are of fundamental importance in assessing the usefulness of the LSI. In this section, we focus on problems with the LSI-1976 based on the use of ordinal and ipsative measures.

Limitations of Ordinal Measures

Due to the ranking format of the LSI-1976, each block of four words constitutes a set of ordinal measures. Ordinal measures simply indicate a numerical order (e.g., 1-2-3-4) and are contrasted with "interval" measures. Even though the numbers in ordinal measures are equally spaced, it cannot be assumed that the items being ranked are equally spaced in terms of the respondent's preferences (Kerlinger, 1986). For example, assume that two students have differences in their use of each of the four learning abilities (i.e., CE, RO, AC, and AE). If the students are asked to rank-order the abilities based on their relative use, the ranking might yield the following:

Learning	Student 1		Student 2	
Ability	Percent Used	Rank	Percent Used	Rank
CE	80	1	32	1
RO	12	2	30	2
AC	6	3	20	3
AE	2	4	18	4

In this example, the learning abilities of the two students are quite different. However, the ranking format indicates that they are identical. The ranking measures do not discriminate, in this example, between a very strong CE mode and a marginally dominant CE mode. As noted by Pedhazur and Schmelkin (1991), information is "lost" under ordinal measures. Thus, the relationship between empirical indicators (the numerical ranks) and theoretical constructs (the respondent's preferred learning ability) is weakened and the validity of the measures is reduced.

Problems with Ipsative Measures

The ranking format of the LSI-1976 also creates "ipsative" measures. According to Hicks (1970, p. 167) ipsative measures "yield scores such that each score for an individual is dependent on his own scores on other variables, but is independent of, and not comparable with, the scores of other individuals". Thus, ipsative measures may be contrasted with "normative" measures (cf. Hicks, 1970; Kerlinger, 1986; Pedhazur & Schmelkin, 1991). Normative measures are the usual kind of measures obtained by tests (Kerlinger, 1986, p. 463). To interpret an individual's score, his/her results are compared to the mean for the group of respondents (i.e., the "norms" of the test). However, ipsative measures cannot be meaningfully interpreted relative to the group mean (Pedhazur & Schmelkin, 1991, p. 21).

Unfortunately, ipsative measures have a number of inherent psychometric limitations since they violate the assumptions of usual statistical tests. Kerlinger notes that the limitations of ipsative scales often are overlooked in research. Such is the case with Kolb's use of these measures as well as many researchers who have used the LSI.

For example, the ranking format of the LSI-1976 creates interdependence between items within each block. This item interdependence produces spurious negative correlations between the four learning abilities (cf. Kerlinger, 1986). Further, use of these correlations in estimates of reliability (e.g., coefficient alpha) or in factor analysis, can lead to serious distortions (Jackson & Alwin, 1980; Tenopyr, 1988). The problem of item interdependence is particularly relevant to factor analysis of the LSI-1976 since Kolb (1974) postulates two bipolar dimensions underlying the four learning abilities (these issues are addressed later in the section on construct validity).

A number of studies have reported the pattern of intercorrelations among the four learning ability scales. Based on the premise of two bipolar dimensions in learning, AC and CE should be negatively correlated, AE and RO should be negatively correlated, and all other correlations should be near zero. The LSI Technical Manual (Kolb, 1976b, p. 10) reports intercorrelations for a sample of 807 individuals. As expected the AC-CE and AE-RO correlations were both negative (-.57 and -.50, respectively). The other four correlations ranged from -.19 to +.13. Five of the six correlations were statistically significant and four of these were negative. Freedman and Stumpf (1978) presented intercorrelations based on a sample of over 1,100 graduate business students. They found that the AC-CE and AE-RO correlations were the strongest (-.49 and -.43, respectively) and five of the six possible correlations were negative. Without a comment on the spurious negative intercorrelations of ipsative measures, the results of these two studies were considered supportive of the bipolar ELM.

However, Lamb and Certo (1978) and Ruble (1978) compared the pattern of intercorrelations of the standard (ipsative) LSI-1976 with normative versions of the instrument using Likert scale ratings. In both cases, the patterns of intercorrelations for the standard version were similar to the results obtained by Kolb (1976b) and Freedman and Stumpf (1978). In contrast, for the normative versions of the LSI, all intercorrelations were positive. These results provide evidence that the LSI-1976 may contain an instrument bias based on the spurious negative correlations.

It must be noted that the LSI-1976 does not meet the criteria for purely ipsative measures (cf. Hicks, 1970). Because only 24 of the 36 items (words) on the LSI-1976 are scored, it can be classified as a partially ipsative instrument. An examination of the scoring format of the LSI-1976 (Exhibit 2 in the Appendix) indicates that two sets of words (#3 and #8) are purely ipsative. That is, all four words are in each set are scored. Thus, the final rank is totally determined by the previous three ranks. Another two sets of items (#7 and #9) score three of the four words while the remaining five sets of items score only two of the four words.

Notice in Exhibit 2 that there is tendency to score "paired" words for the combined dimension scores. That is, RO1 is paired with AE1, CE2 is paired with AC2, CE3 with AC3 and RO3 and AE3, and so on. In fact, for 22 of the 24 words, the rankings are interdependent for the words that are combined together for the dimension scores (only CE7 and AC9 are ranked independently of other items in the AC-CE dimension scores). Thus, although the LSI-1976 is not a purely ipsative instrument, the interdependence of items is an important feature of the two combined dimension scores.

Because the LSI-1976 is not purely ipsative, the effect of the spurious negative intercorrelations will be moderated to some unknown extent. However, the studies by Lamb and Certo (1978) and Ruble (1978) suggest that the ranking format creates some idiosyncratic results that are not replicated with a normative format. In the later section examining construct validity, additional studies will confirm the presence of a method-specific biasing effect for the standard LSI-1976.

Hicks (1970) and Pedhazur and Schmelkin (1991) note that ipsative measures may be useful for studying <u>intra</u>individual hierarchies or preferences. However, ipsative measures should <u>not</u> be used for purposes of <u>inter</u>individual comparisons (Pedhazur and Schmelkin, 1991, p. 21, emphasis in original). This means that it is fallacious to use Kolb's norms to assign individuals to learning style classifications since the use of sample means or medians as "norms" for classification is an interindividual comparison (see Gordon, 1985 for other concerns about the use of Kolb's norms for classification purposes).

For example, based on the norms provided by Kolb (1976b), individuals scoring +2 on the AC-CE dimension and +2 on the AE-RO dimension of the LSI would be classified as Divergers. Divergers have been described as "best at Concrete Experience (CE) and Reflective Observation (RO)" (Kolb, 1976b). However, the scores of +2 and +2, respectively, actually show <u>intra</u>individual preferences for Abstract Conceptualization and Active Experimentation. Thus, the use of Kolb's <u>inter</u>individual norms with ipsative measures creates some disparities between the empirical indicators (scale scores) and theoretical constructs (learning style classifications). Again, these disparities reduce the validity of the classifications.

Some researchers might suggest that the use of a cut-off score is a simple matter to correct -- simply use the 0,0 points to assign individuals to classifications. In fact, this provides an easy way to overcome the problem of using interindividual norms. Each individual would then be classified according to their own baseline. This approach makes a lot of sense, given the inherent contradiction between ipsative measures and normative classification.

Unfortunately, the 0,0 baseline approach to classifying respondents creates a dilemma for Kolb. In the <u>Technical Manual</u> for the LSI (Kolb, 1976b), the two dimension scores for over 600 individuals were plotted according to the average for their undergraduate college major. The plot revealed that Business majors fell in the Accommodator quadrant; Engineers fell in the Converger quadrant; History, English, and Political Science majors were classified as Divergers; Math, Chemistry, Economics, and Sociology majors were classified as Assimilators. According to Kolb, "the distribution of undergraduate majors on the learning style grid is quite consistent with theory" (Kolb, 1976b, p. 30). However, as Bonham (1988) has noted, if the 0,0 baseline cut-off points had been used, all majors would be in the Converger quadrant. That is, if the appropriate 0,0 baseline had been used to classify learning styles, the LSI-1976 would fail to differentiate between any of the majors. Thus, it was only by the inappropriate use of interindividual norms that Kolb could claim a relationship between major and learning style.

BASIC ISSUES IN RELIABILITY

Nunnally and Bernstein (1994) have noted that there is seldom a planned effort to develop valid measures of psychological constructs. Before the basic psychometric properties of an instrument are determined, researchers often leap to studies relating the "measured" construct to other constructs (often of unknown psychometric quality themselves). This generalized description applies too well to the LSI-1976.

In the case of the LSI-1976, the basic steps in assessing reliability and validity were neglected. As a result, researchers began correlating the LSI with other constructs and occasionally they came up with "significant" results. For example, Kolb (1976b) reported correlations of an individual's learning ability scores with their preferences for different learning situations. The theory suggests that AC types would prefer different learning situations than CE types and AE types would prefer different learning situations than RO types (although specific hypotheses were not provided). Correlations of the four learning ability scores with 16 different situations (64 possible correlations) yielded 12 statistically significant correlations ranging from .15 to .34 (average r = .20, or 4% of the shared variance between learning styles and learning preferences for the 12 significant correlations). Overlooking the relatively low correlations (i.e., strength of relationships) as well as the 52 nonsignificant correlations (over 80% of those possible), Kolb was undaunted in suggesting that certain individuals "learned best" in different situations. To cite an extreme case, the CE learning ability score did not have one significant positive correlation with any of the 16 different learning situations. Nevertheless, based on the highest positive correlation available (r = .13), Kolb (1976b, p. 27) suggested that CE individuals tend to find student feedback helpful.

The point of this example is to recognize that given a large number of studies (with many subjects and many variables), researchers are going to find some statistically significant correlations simply by chance. Unfortunately, many researchers take even the weakest results as "support" for their theory or instrument. Other researchers then cite the first study without a critical evaluation. However, this process of validating instruments is fundamentally flawed (logically and empirically). Research in behavioral science needs to recognize that a well-accepted body of theory and statistical methods exists for the validation of psychological instruments.

The process of validating instruments should proceed in a certain order. Many researchers and consumers of research overlook the basic fact that <u>reliability is a necessary but not a sufficient condition for validity</u> (Pedhazur & Schmelkin, 1991, p. 81, emphasis in original; also see Nunnally & Bernstein, 1994). If an instrument does not provide consistent measurement (i.e., is not reliable), it cannot provide valid measures.

Unfortunately, there are few studies reporting basic reliability data for the LSI-1976 (Sewall, 1986). In fact, most studies using the LSI do not report reliability statistics for the specific sample studied. We believe this reporting should be a integral part of any published study and should be requested by reviewers and editors (Stout and Ruble, 1991b). We have provided a list of references that used the LSI-1976 without reporting basic reliability statistics for their samples (see Exhibit 3 in the Appendix). In many cases, the researchers probably overlooked the need for this information. However, because the psychometric properties of the LSI-1976 are so weak, in some cases the publication of the research would be in jeopardy if researchers reported the reliability statistics for their samples.

In addition to the Technical Manual for the LSI (Kolb, 1976b), we have found seven studies by independent researchers reporting basic reliability statistics for the LSI-1976. However, before evaluating the data, it is useful to consider recommended principles for assessing the reliability of instruments. Usually, the first consideration in assessing reliability is the internal consistency of the items comprising the instrument scales.

According to Carmines and Zeller (1979) and Nunnally and Bernstein (1994), coefficient alpha (Cronbach, 1951) should be the basic formula for determining the internal consistency reliability of an instrument. Coefficient alpha is based on the average intercorrelation among items as well as the number of items. Another method for assessing internal consistency is the split-half method. However, coefficient alpha is recommended due to limitations of the split-half approach. As noted by Pedhazur and Schmelkin (1991), the obtained correlation for split-halves will vary depending on how the items are divided. Thus, only one correlation coefficient is calculated out of many possible coefficients. This single coefficient may or may not provide a good estimate of the average intercorrelation of items. Thus, the split-half method should be avoided (cf. Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991).

Another method for assessing reliability is the test-retest correlation. While this method also has limitations and should not provide the primary estimate of reliability, it can be used as a useful supplement to coefficient alpha (cf. Nunnally & Bernstein, 1994; Pedhazur & Schmelkin, 1991).

Internal Consistency Reliability

With respect to internal consistency reliability of the LSI-1976, Kolb did not report the recommended coefficient alpha. Rather, Kolb used the split-half method with its inherent limitations. Spearman-Brown split-half reliability coefficients are reported in Table 2 of the Technical Manual (Kolb, 1976b, p. 15). The coefficients for the four learning abilities ranged from .55 (CE) to .75 (AC) with an average of .65. For the two dimension scores the coefficients were .74 (AC-CE) and .82 (AE-RO). A superficial look at these coefficients might be encouraging for the two combination dimension scores. However, a closer look at the procedures used to arrive at these estimates raises serious questions about the data.

First, as noted above, a major problem with split-half reliability is that the division of the items affects the correlation coefficient obtained. For each separate scale of the LSI, there are 20 possible combinations of split-halves (thus, 20 different reliability coefficients are possible). Kolb divided each scale "taking all available item statistics into consideration, and pairing items that most resemble each other and correlate most highly" (Kolb, 1976b, p. 13). For example, in computing the split-half correlation for the RO items, "observing" was placed in one half and "observation" was placed in the other half. This approach would provide the highest possible correlation and would not represent an average estimate of reliability. In contrast, coefficient alpha would provide the best estimate of the average internal consistency. Examination of coefficient alpha statistics taken from independent investigations (see Table 1, p. 15) are considerably lower than the split-half statistics provided by Kolb.

In addition, the split-half coefficients provided in the Technical Manual

(Kolb, 1976b) are Spearman-Brown Reliability Coefficients. As noted by Pedhazur and Schmelkin (1991), split-half correlations are often "stepped up" using the Spearman-Brown formula to compensate for dividing the original length of the scale in half. However, the validity of this method for inflating split-half coefficients rests on the very restrictive assumption that the two halves are "strictly parallel". That is, the scale items on both halves should be random samples from a domain of items and the items should have uncorrelated error variances (for a discussion of parallel measures, see Nunnally & Bernstein, 1994; Paunonen & Gardner, 1991; and Pedhazur & Schmelkin, 1991). A random sample of items is important in order to balance out errors which overestimate and underestimate the "true" scores.

Since Kolb intentionally divided the ability scales to maximize the split-half correlations, the items are not random samples. This post-hoc process of assigning items to halves violates the assumptions of parallel measures. The additional use of the Spearman-Brown formula to inflate the coefficients undoubtedly overestimates the internal consistency of the four separate scales. The use of coefficient alpha in the first place would preclude the need to "step up" the split-half correlations.

Kolb argues that the reliability estimates for the two dimension scores are "very reasonable" and "highly reliable indices suitable for most research applications" (Kolb, 1976b, pp. 14 and 16). However, the problem of spurious negative correlations due to ipsative measures must be considered in evaluating the coefficients of the dimension scores. The items included in the dimensions scores (AC-CE and AE-RO) are not independent across the split-halves, again violating the requirements for parallel measures. Indeed, the way the items were divided to create the split halves, six pairs of interdependent items (i.e., ranked in the same set) were placed in opposite halves and then correlated for the split-half coefficient. Thus, half of the items violated the assumption of independence for the split-half correlations. Since the dimension scores are obtained by subtracting one set of items from the other, the spurious negative intercorrelations become spurious positive intercorrelations. This yields artificially high estimates of split-half consistency and, to our knowledge, these biased effects cannot be disentangled.

Overall, our evaluation suggests that the data in Table 2 of the LSI <u>Technical Manual</u> (Kolb, 1976b) should be totally disregarded because:

1. They are based on the psychometrically flawed split-half approach. Instead, coefficient alpha estimates should have been reported.
2. The scales were divided intentionally to maximize the coefficients and do not represent an average estimate of internal consistency.
3. The coefficients were "inflated" using the Spearman-Brown formula even though the necessary assumptions were not met.
4. Problems of interdependence with ipsative measures make the dimension reliability coefficients uninterpretable.

Independent studies assessing internal consistency reliability are presented in Table 1. Most studies reported estimates of coefficient alpha. One study computed split-half coefficients which do not approach the inflated values reported by Kolb (1976b). As the table indicates, average estimates of coefficient alpha were .35, .38, .57, and .60 for an overall average of .47. These figures are far below the estimates provided by Kolb and do not approach the standards recommended for psychological instruments. Nunnally and Bernstein (1994) suggest that reliabilities of .70 will suffice <u>in the early stages of research.</u> Carmines and Zeller (1979) suggest that reliabilities of .80 should be expected for widely-used instruments. Thus, we believe the data presented in Table 1 indicate that the LSI-1976 fails to meet minimal standards for internal consistency reliability.

TABLE 1

ESTIMATES OF INTERNAL CONSISTENCY FOR LSI-1976*

STUDY	COEFFICIENT ALPHA STUDENT SAMPLES	SIZE	CE	RO	AC	AE	AVERAGE
1	MBA	1179	.40	.57	.70	.47	.54
1	MBA	412	.33	.61	.69	.51	.54
2	MBA	166	.46	.53	.59	.34	.48
3	Nursing	187	.29	.59	.52	.41	.45
4	G/UG Bus.	438	.48	.58	.52	.23	.45
5	UG Acctg.	235	.11	.56	.56	.30	.38
	TOTAL/AVERAGES	2617	.35	.57	.60	.38	.47

STUDY	SPLIT-HALF STUDENT SAMPLES	SIZE	CE	RO	AC	AE	AVERAGE
6	Adult Mgt.	102	.15	.53	.49	.41	.40

REFERENCES

1. Freedman and Stumpf (1978).
2. Ruble (1978)
3. Merritt and Marshall (1984).
4. Sims, Veres, Watson, and Buckner (1986).
5. Stout and Ruble (1991b).
6. Wilson (1986).

* NOTE: reliability coefficients for the combination scales are not reported because the assumption of independent measures is violated

Temporal Consistency Reliability

The LSI *Technical Manual* also reported a series of test-retest reliability studies (Kolb, 1976b, p. 17). The average test-retest correlations for the four learning abilities were: CE=.48; RO=.50; AC=.60, AE=.48 (average = .52). The averages for the two combination dimension scales are .49 for the AC-CE and .53 for AE-RO (average = .51). Thus, the shared variance (r^2) between tests was approximately 25%. These coefficients indicate that either the construct (learning styles) or the instrument (LSI-1976) is not stable.

Test-retest correlations from independent researchers are presented in Table 2. A total of 403 subjects took the LSI-1976 twice with intervals between tests ranging from one month to six weeks. The average for the four ability scales was .52 while the average for the two combination scales was .58. Again, these figures do not support the contention that learning styles are stable or that the LSI-1976 provides consistent measures.

TABLE 2

ESTIMATES OF CONSISTENCY OVER TIME FOR LSI-1976
(TEST-RETEST CORRELATIONS)

STUDY	STUDENT SAMPLES	SIZE	INTERVAL	CE	RO	AC	AE	AC-CE	AE-RO
1	MBA	101	5 weeks	.39	.49	.63	.47	.58	.51
2	Medical	50	1 month	.56	.52	.59	.61	.70	.55
3	G/UG Bus.	201	5 weeks	.45	.46	.53	.43	--	--
4	Adult Mgt.	51	6 weeks	.40	.77	.63	.40	.53	.61
	TOTAL/AVERAGES	403		.45	.56	.60	.48	.60	.56
	SHARED VARIANCE	(r^2)		20%	31%	36%	23%	36%	31%

REFERENCES

1. Freedman and Stumpf (1978).
2. Geller (1979).
3. Sims, Veres, Watson, and Buckner (1986).
4. Wilson (1986).

Note that the test-retest reliabilities for the combined dimension scores are not much better than the strongest of the sub-scales. These results fail to support Kolb's argument that the dimension scores are reliable indices suitable for research.

Note also that the average variance shared between tests (r^2) is approximately 27% for the ability scales and 33% for the dimension scores. This means that approximately two-thirds of the variance cannot be attributed to some stable construct. It is either situational or error. Given the low estimates of internal consistency, most of the variance is probably error. Moreover, some of the stable shared variance could be due to spurious correlations resulting from ipsative measures (cf. Tenopyr, 1988). Either way, the LSI-1976 does not provide very stable measures over time.

Reliability of the LSI-1976: A Summary

To summarize the evidence, the LSI-1976 simply is not reliable. The only data supporting the reliability of the instrument is presented in the Technical Manual. As we have noted, those data are fraught with statistical artifacts and misleading coefficients. In contrast, evidence reported by independent investigators unanimously fails to support the reliability of the LSI-1976.

CONSTRUCT VALIDITY: ASSESSMENTS OF THE LSI-1976 BY FACTOR ANALYSIS

Normally, low estimates of reliability would indicate that validity studies are unwarranted. However, several factor analytic studies of the LSI have been reported. These studies deserve attention because they reveal numerous problems and misconceptions regarding the assessment of validity of the LSI.

Factor analysis provides information on the internal structure of an instrument. This information is considered relevant to the assessment of construct validity (cf. Nunnally & Bernstein, 1994). However, there are two compelling reasons why factor analysis will yield little in the way of substantive evidence on the validity of the LSI: (1) since acceptable reliability is a necessary condition for an instrument to be valid, the low reliabilities noted above indicate no basis for performing further analysis, and (2) factor analysis of ipsative data is problematic (cf. Gruber and Carriuolo, 1991; Jackson and Alwin, 1980). We have noted that the ipsative format of the LSI causes spurious negative correlations among the items. When these correlations are used in factor analysis, the ipsative procedure will distort the results. Thus, any interpretation of support for the LSI or the ELM would be tenuous at best if based on factor analysis of ipsative measures.

Nevertheless, results of some factor analyses have been cited as providing support for both the ELM and LSI. Thus, the factor analytic investigations of the LSI-1976 must be evaluated carefully. In some cases, interpretations of factor analyses of the LSI-1976 represent the basic misunderstanding of ipsative measures noted by Kerlinger (1986). Moreover, even if the measures were not distorted by ipsative scaling, some of the researchers interpreting factor analytic results seem confused about what constitutes support for Kolb's proposed two bipolar dimensions in the ELM. Unfortunately, the distorted results and misinterpretations are cited in subsequent studies as researchers attempt to justify their own use of the LSI.

In this section, we will present the results of several factor analytic studies of the LSI-1976. On the surface, some of these results seem to provide a minimal degree of support for the ELM (less so for the LSI-1976). However, given the artificial distortions due to the ipsative measures of the LSI-1976, we believe that even this minimal "support" must be regarded as a statistical artifact.

First, consider a study by Ferrell (1983). This study has been interpreted as supportive of both the ELM and the LSI. A sample of 471 high school and community college students completed the LSI along with three other learning style instruments. In comparing the instruments, Ferrell considered (1) the "match" between item loadings and learning styles hypothesized by each model, and (2) the total variance accounted for by each instrument. One key paragraph in Ferrell is cited in support of the LSI:

> "The only instrument for which a match between factors and learning styles existed was the Kolb LSI. Items comprising the four factors extracted matched the four learning abilities as described by Kolb (1976). Results of the factor analysis of the Kolb LSI supported Kolb's conceptualization of learning style." (Ferrell, 1983, p. 36)

A closer examination of Ferrell's (1983) study, however, indicates that this conclusion is based on an incorrect understanding of Kolb's theory.

Moreover, Ferrell's interpretation of the factor analysis fails to consider the limitations of the ipsative format of the LSI. Consider the complete results Ferrell reported for the LSI:

> "Data from the Kolb LSI were interpreted as four distinct factors. These four factors had eigenvalues of 3.978, 1.765, 1.641, and 1.176 accounting for 46.5%, 20.6%, 19.2%, and 13.7% of the common factor variance, respectively. Twenty-three items loaded on a single factor and 7 items did not have salient loadings on any factor. Common factors accounted for 31.9% of the total variance." (p. 35)

Ferrell's interpretation of the results indicated that four factors matched the four learning abilities. However, Kolb (1976b, p. 3) asserts that "learning requires abilities that are polar opposites...specifically, there are two primary dimensions to the learning process." Thus, factor analysis should not extract four distinct factors but, rather, two orthogonal factors with positive and negative loadings for the appropriate items (AC versus CE and AE versus RO). The extraction of four distinct factors suggests that the learning abilities are independent rather than aligned in two bipolar dimensions. Thus, <u>Ferrell's data do not support the ELM</u>.

In addition, Ferrell treated the data as if there were no ipsative measurement problems involved. Ferrell analyzed four instruments, two of which used normative (Likert) scales and the third used a "forced-choice" normative scale. In assessing the results of the four different factor analyses, no mention was made of the ipsative problems caused by the ranking format of the LSI and no cautions were offered in interpreting the factors.

If the failure to obtain two bipolar factors with ipsative data was not enough to indicate that this study did not support the LSI-1976, consider these comments from Ferrell:

> "The percentage of total variance accounted for by the common factors ranged from 23.5% for the Dunn LSI to 41.7% for the DMI... the Kolb LSI accounted for almost 32%... Needless to say, an instrument that only accounts for 24% of the total variance should be suspect, and even 42%

is marginal... No one instrument stood out as better than the others... The implication is that either the instrument or the paradigm is lacking, perhaps both." (pp. 38-39).

Disregarding Ferrell's misinterpretation of the four independent learning abilities versus two bipolar factors and neglect of the ipsative measurement problems, the LSI-1976 was characterized as somewhere between "suspect" and "marginal" in a field with a questionable paradigm and weak instrumentation.

Another pair of factor analytic studies have been cited as supportive of the ELM. Merritt and Marshall (1984) reported two studies concerned with the development of better measures of Kolb's learning styles. In study 1, they analyzed two versions of the LSI-1976, the standard ipsative form and a normative form. They noted that:

"Ipsative measures are designed to maximize the differences between instrument scales <u>within</u> an individual... Statistically, the ipsative technique results in a <u>between-subjects</u> sum of squares of zero; therefore, the relative strengths of a respondent's preferences for the various modes cannot be compared with those expressed by other individuals... The use of ipsative scales in the Kolb instrument poses difficulties for researchers when <u>between-subjects</u> analysis is conducted." (Merritt & Marshall, 1984, p. 466, emphasis in original)

To make the comparison, they administered both versions of the LSI to a sample of 187 nursing students in Study 1. They found two bipolar factors for the ipsative items and concluded that this provided support for the ELM and the LSI. They found four independent factors for the normative items and also concluded that these results provided support for the ELM. The difference between two-factor and four-factor structures did not seem to matter in their evaluation of support and neither did the problems with ipsative scales that they discussed earlier. However, as we noted earlier, a bipolar two-factor structure would be supportive of the ELM while four independent factors would not (<u>assuming that the two bipolar factors were not the result of spurious negative intercorrelations caused by ipsative measures</u>). Moreover, since they went to the trouble to point out the limitations of the ipsative version of the LSI-1976, it would seem that the discrepancy in results (between ipsative and normative versions) would signal the need for a more critical assessment of their results.

Study 2 was conducted with different subjects to cross-validate the normative form. Again, a factor analysis extracted four factors with the normative instrument although the factor structure was less "distinctive" than that of Study 1 (Merritt & Marshall, 1984, p. 469) . Of the 17 items loading above .30 in Study 1, only 10 loaded above .30 in Study 2. Even though 13 of the 24 items failed to load above .40 on the appropriate factor, they concluded that the normative form "still tended to support construct validity of some of the LSI items..." (p. 471). However, once again, the normative version of the LSI used in Study 2 found four separate factors, not the two bipolar dimensions posited by the ELM. Thus, as before, these results should not be considered as supportive of Kolb's theory.

Ruble (1978) and Certo and Lamb (1980) also conducted factor analytic studies comparing the ipsative scales of the LSI with normative (Likert) scales. In both cases, the ipsative version indicated some minimal congruence with Kolb's two-factor theory but the normative scales indicated no support. Considering these two studies along with Merritt and Marshall (1984), it seems apparent that the alleged support for the LSI from factor analysis is "method-bound". That is, results called "supportive" occurred only with the ipsative format and not the normative format. For the normative versions of the instrument, Ruble found that only 10 of the 24 items loaded above .40 while Certo and Lamb found that two factors accounted for only 23.7% of the total variance. Certo and Lamb (1980) concluded that:

"These results seem consistent with the notion that the appearance of two bipolar learning dimensions based upon the original LSI is largely due to instrument bias. ... instrument bias within the LSI seems to have artificially created the illusion of two bipolar dimensions." (p. 6)

Wilson (1986) also factor analyzed different versions of the LSI. In this case the standard LSI-1976 was compared to: (1) a version with the words randomized within each block of four to offset the tendency to follow a pattern in responding, and (2) a version with additional words to clarify the meaning of the items. Although this study had small ns (approximately 100 per version of the LSI), Wilson had the following observation:

"On the basis of linkage and factor analysis of the data, it must be concluded that if there are four modes, the LSI does not measure them, and whatever it does measure varies with the order in which items appear in the inventory and the extent to which the inventory is elaborated." (p. 7)

Freedman and Stumpf (1978) probably provide the most comprehensive evaluation of the LSI-1976. A factor analysis of 1,179 subjects found items loading on two bipolar factors. However, as noted by Freedman and Stumpf:

"The total variance in the LSI accounted for by the two-bipolar-factor theory is only 20.6%, some of which is an artifact of the scoring method. ... The results indicate that the instrument measures rather little. What it does measure is obfuscated by an inordinate amount of error variance." (pp. 278, 281).

Freedman and Stumpf also noted that the LSI had problems with reliability which limited the validity of the instrument. Freedman and Stumpf (1980) concluded that the LSI-1976 was not valid and recommended that it should not be used in making decisions about educational practices.

In addition to the many studies using actual data, Certo and Lamb (1979) used a Monte Carlo technique to simulate random responses to the LSI. This study is cited by Freedman and Stumpf (1980) and Atkinson (1991). Apparently Certo and Lamb found that even random data supported a bipolar model.

We did not have access to the Certo and Lamb paper (presented at a regional meeting) so we ran a simulated study ourselves. We generated random responses ranking the 9 blocks of words on the LSI. Using these random responses, we did a factor analysis of the 24 items scored by Kolb. For an n of 200, we found 13 of the 24 items loaded above .30 on two bipolar factors. The two factors accounted for 15% of the total variance. A four-factor solution accounted for 28.5% of the variance and 20 items loaded above .30 on one of the factors. Note the similarity of these results based on random data to those reported by Ferrell (1983) and Merritt and Marshall (1984). Ferrell found four factors accounting for 31.9% of the total variance with 23 items loading above .30. For the standard, ipsative version of the LSI, Merritt and Marshall found 16 of 24 loading above .30. If random ipsative data can produce essentially the same results as previous studies, what can the LSI-1976 be contributing? Like Certo and Lamb (1980) we must conclude that the ipsative format forces artificial factors which appear supportive of the ELM.

CONCLUSIONS REGARDING THE LSI-1976

Some researchers have presented data they considered supportive of the ELM or the LSI-1976. However, none of the alleged support stands up to careful scrutiny. When the biasing effect of spurious negative intercorrelations is stripped away, very little is left. Many independent reviewers of the LSI-1976 have reached similar conclusions:

> "There appears to be no relationship between learning style congruence and perceived learning." (when using the LSI). "This possibly makes the task of utilizing learning styles data, from the current learning styles inventory, tenuous since extended discussions with respondents are necessary whenever interventions are needed." (Wolfe and Byrne, 1975, *Proceedings of the Second Annual Meeting of the Association for Business Simulation and Experiential Learning*, pp. 330;334)

> "...we began research on the LSI in 1976 in hopes of finding support for it. We had been using the LSI in an introductory OB course and we were meeting student resistance regarding its reliability and validity. Our research - much to our displeasure - bore out the student doubts." (Stumpf and Freedman, 1981, *Academy of Management Review*, p. 298)

"These findings suggest that it may be questionable to develop medical education programs that match instructional techniques to the personality characteristics of the audience, as identified through the use of Kolb's LSI." (West, 1982, *Journal of Medical Education*, p. 796)

"The studies that are aimed specifically at evaluating the LSI as a reliable instrument indicate some support for the learning model but at the same time unequivocally discredit the reliability of the LSI instrument." (Hunsaker, 1984, *Journal of Experiential Learning and Simulation*, pp. 150-151)

"...one must question the usefulness of the LSI as a guide to educational design decisions." (Fox, 1984, *Adult Education Quarterly*, pp. 83-84)

"the unreliability and lack of evidence for either construct and predictive validity suggests that the LSI could produce very misleading results and needs to be studied much more carefully before it should be used in any setting." (Sewall, 1986, *Educational Resources Information Center Document*)

"Thus, although Kolb's basic model of learning may be regarded as plausible, it would seem that there is a need for a more reliable and valid measure of learning styles than the LSI." (Allinson and Hayes, 1988, *Journal of Management Studies*, p. 271; 278)

"Criticisms of the Kolb LSI revolve around ...brevity and resulting lack of reliability...possibility of individual words being interpreted differently...lack of correlation with statements taken from Kolb's descriptions...possibility of response set...ranking format prevents dimensions from being independent...makes it inappropriate to factor analyze results and makes even simple correlations artificially high." (Bonham, 1988, *Lifelong Learning*, pp. 14-15)

"...in spite of wide acceptance of Kolb's LSI, little support for its validity or utility is apparent. Generally, a lack of significant relationships between learning style and other variables was revealed in research conducted with nursing students...the LSI instrument does not... warrant its current popularity." (DeCoux, 1990, *Journal of Nursing Education*, pp. 206-207)

"From the preceding survey, the LSI seems psychometrically deficient in several areas...Many researchers seem to agree little can be learned from using the LSI..it seems face validity has been the saving grace of the LSI." (Atkinson, 1991, *Measurement and Evaluation in Counseling and Development*, p. 158-159)

Thus, there has been substantial criticism of the LSI-1976 across many disciplines. Moreover, even this level of criticism may understate the actual number of studies failing to support the LSI-1976. As Curry (1990) has noted, "Given the predilection in the scholarly press toward considering positive results...the availability of negative results regarding learning style intervention likely underestimates the true proportion of negative results found across learning style investigations." (p. 52).

Despite all the criticism of the LSI-1976 from independent researchers, perhaps David Kolb said it best:

> "the LSI, because of its theoretical basis, will be of limited use for assessment and selection of individuals. ... While the LSI can potentially be a useful starting point for discussion with the individual about his learning style, any attempt to use the LSI for selection purposes without additional detailed knowledge of the person and his situation is likely to be inaccurate" (Kolb, 1976b, *Technical Manual*, p. 13).

To improve the instrument it was revised in 1985.

THE LSI-1985

In the LSI-1985, there are 12 sets of four sentence completion items. Each sentence begins with a short phrase such as "When I learn ..." or "I learn best when ...". To complete the sentence, respondents are asked to rank four possible endings, representing one of the four learning abilities.

To facilitate scoring, the format of the LSI-1985 provides all of the endings representing a given learning ability in the same column. That is, the sentence endings that correspond to the CE scale are presented in column one of the inventory, those of the RO scale in column 2, those of the AC scale in column 3, and those of the AE scale in column 4. Thus, a distinctive characteristic of the LSI-1985 is its "single-scale-per-column" format.

Similar to the procedure for the LSI-1976, a numerical ranking of 1 (least like you) to 4 (most like you) is assigned by the respondent to each of the sentence endings per set for the 12 sets. Scale scores are then calculated by summing the numerical scores found in each column. Unlike the LSI-1976, all items are scored. That is, there are no "distractor" items in the LSI-1985

As with the original LSI, an individual's learning style is determined by subtracting scores for the CE ability from the AC ability (placing one on the AC-CE dimension) and also subtracting scores on the RO ability from the AE ability (locating one on the AE-RO dimension). The point of intersection of these two dimension scores is compared to the sample norms (Kolb, 1985; Smith & Kolb, 1986) to place an individual in one of the learning style categories (Diverger, Accommodator, Assimilator, or Converger).

MEASUREMENT PROBLEMS BASED ON ORDINAL AND IPSATIVE SCALES

Because the ranking format of the LSI-1976 is retained, the basic problems of ordinal and ipsative measures also are retained. However, in the case of the LSI-1985, the problems with ipsative measures are accentuated. Whereas the LSI-1976 provided only partially ipsative measures, the LSI-1985 yields purely ipsative measures. That is, all items per set are scored so that ranking three items totally determines the score on the fourth item. Moreover, combining the AC and CE scores and AE and RO scores yields dimension scores that have every item interdependent with another item in the scale. Thus, the spurious negative correlations of ipsative scales should be even stronger than in the LSI-1976. According to Hicks (1970), the average intercorrelations should approach a limiting value according to the formula, $-1/(m-1)$, where m is the number of variables in the ipsative test. Thus, the average intercorrelation for the LSI-1985 should be approximately -.33. This average coefficient represents the strength of the negative relationship that is artificially created by the purely ipsative format.

Spurious Negative Intercorrelations

Studies reporting intercorrelations of the ability scales for the LSI-1985 confirm the presence of stronger negative relationships than those of the LSI-1976. In the User's Guide for the LSI-1985, Smith and Kolb (1986) report a pattern of all negative intercorrelations ranging from -.15 to -.42 with an average of -.29 (n=1,446). Highhouse and Doverspike (1987) found correlations of the ability scales ranging from .00 to -.45 with an average of -.27 (n=111). Ruble and Stout (1990) reported patterns of intercorrelations for both the standard LSI-1985 and a version with items "scrambled" to eliminate the single-scale-per-column format. For the standard LSI-1985 (n=312), correlations of the ability scales ranged from -.28 to -.39 with an average intercorrelation of -.33. For the scrambled version of the LSI-1985 (n=323), correlations of the ability scales ranged from -.25 to -.39 with an average intercorrelation of -.33. Thus, the revised LSI-1985 yields spurious negative intercorrelations that are characteristic of a *purely ipsative* instrument.

Disparities between Scores and Learning Style Classifications

As noted previously for the LSI-1976, it is fallacious to use *inter*individual norms to make comparisons with an individual's ipsative scores. However, as with the original LSI, individuals are assigned to learning style categories based on LSI-1985 norms. Again, the use of interindividual norms with ipsative measures creates some disparities between the empirical indicators (scale scores) and theoretical constructs (learning style classifications) thereby reducing the validity of the classifications.

For the LSI-1985, individuals scoring +3 on the AC-CE dimension and +5 on the AE-RC dimension of the LSI would be classified as Divergers. Again, Kolb (1985) asserts that the Diverger learning-style combines CE and RO abilities. However, the scores of +3 and +5, respectively, actually show preferences for AC and AE. Compared to the LSI-1976 norms, the norms for the LSI-1985 mean that individuals further away from the 0,0 baseline are assigned to categories that do not represent their learning ability preferences. Thus, the disparities between scores and learning style classifications may be even greater for the LSI-1985 than for the LSI-1976.

To add to the possible disparity, Kolb (1985) and Smith and Kolb (1986) do not use the same cut-off score for the AC-CE dimension. Thus, individuals scoring +4 on AC-CE are not assigned consistently to one side or the other of the AC-CE dimension based on the two different sets of norms.

The 1986 User's Guide also presents average scores for 21 different majors plotted into the four different learning style classifications relative to the norms. In this case, if the 0,0 baseline had been used to assign scores to learning style classifications, 18 of the 21 majors would show preferences for AC (over CE) and AE (over RO) and consequently would be classified as Convergers. Thus, if the appropriate intraindividual comparisons are made using the 0,0 baseline, the LSI-1985 would fail to differentiate between majors in science, the arts, history, business, medicine, engineering, education, and 11 other assorted majors.

It is interesting to compare the results of the 1986 distribution of majors and learning styles with those reported in the 1976 Technical Manual. In 1976, Kolb argued that the distribution of majors was consistent with the ELM. Six majors that fell into a distinct learning style in 1976 were included also in the 1986 User's Guide distribution. Of these six majors, four were placed in different quadrants in 1986. This raises the question as to what distribution of majors would provide support versus nonsupport for the theory.

RESPONSE SETS AND MEASUREMENT ERROR

Several investigators have pointed to the possible existence of a response-set bias in the LSI-1985 attributable to the single-scale-per-column format (Atkinson, 1988, 1989; Ruble & Stout, 1990, 1991; Sims et al., 1986; Smith & Kolb, 1986; Veres, Sims, & Shake, 1987). Because all the items for one learning ability are in a single column, respondents may be encouraged to be consistent as they work their way down the page rather than responding to each set of sentence completions as independent comparisons. This phenomenon was apparently present in the LSI-1976 (Wilson, 1986).

Response sets lead to systematic (nonrandom) error. As Carmines and Zeller (1979, p. 14) note, "Unlike random error, nonrandom error has a systematic biasing effect on measuring instruments ... Thus, nonrandom error lies at the heart of validity." Systematic error leads to lower validity because the empirical indicators (such as LSI-1985 scores) are representing something other than (or in addition to) what they are intended to measure. Empirical verification of the existence and measurement impact of the response set for the LSI-1985 is provided in a number of recent studies that compared the standard form to a scrambled form of the instrument.

In our own research, we compared the standard form of the LSI-1985 with a scrambled version that balanced the number of times an item from a particular scale appeared in each of the four columns of the instrument. Thus, items for each learning ability appeared three times in each column of the instrument. The format of our scrambled version is presented in Ruble and Stout (1990).

In one study (Ruble & Stout, 1990), estimates of scale consistency (coefficient alpha) were less in the scrambled version and factor structures were less well defined, suggesting the existence of a response set. In a second study (Ruble & Stout, 1991), we found lower estimates of coefficient alpha and higher test-retest correlations for the scrambled version of the LSI-1985 (we discuss these issues in more detail in the section on reliability). Further analysis of the data (Stout & Ruble, 1991a) indicated that learning style classifications were sensitive to the format of the instrument. Taken together, these results indicate a number of psychometric differences in the two versions of the LSI-1985 based on whether the order of the items included all of one learning ability in one column versus different column locations of some items for a given ability.

Veres, Sims, and Locklear (1991) also created a scrambled version of the LSI-1985 (using a random procedure to assign items to different columns). They administered the instrument three times, at eight-week intervals, to two large samples totaling over 1,700 subjects and compared measurement properties of this instrument to those of the standard form of the LSI-1985 obtained in an earlier study (Veres et al., 1987). Veres et al. (1991) found substantially lower coefficient alpha estimates of internal consistency (an average of .84 for the standard version versus .64 for the scrambled version).

We believe that the observed decreases in coefficient alpha indicate problems with the single-scale-per-column format. Coefficient alpha represents the proportion of variance in responses that can be attributed to "systematic sources" (cf. Pedhazur & Schmelkin, 1991; Kerlinger, 1986). Systematic sources include both "true" scores and systematic measurement error. The remainder of the variance can be attributed to random measurement error. Thus, an average alpha coefficient of .84 for the learning abilities of the standard LSI-1985 indicates that 84% of the variance is a combination of true scores and systematic measurement error while 16% represents random measurement error. Systematic measurement error can include effects of response sets, social desirability, method variance (i.e., self-reports), and correlations with other hypothetical constructs such as intelligence or self-esteem. Thus, because coefficient alpha includes all systematic sources of variance (including systematic error), it represents an upper limit on the reliability and validity an instrument can attain.

Similarly, the average coefficient alpha of .64 for the modified LSI-1985 used by Veres et al. (1991) indicates that 64% of the variance consists of systematic sources (true scores and systematic error) while the remaining 36% is due to random measurement error. Thus, when the systematic effects of the response set associated with the standard version of the LSI are partialled out via a scrambling of the order of items, the random measurement error increases dramatically (from 16% to 36%).

RELIABILITY OF THE LSI-1985

As we noted earlier, the evidence indicated that the LSI-1976 simply was not reliable. The evidence regarding reliability for the revised LSI-1985 is less conclusive. Clearly, the internal consistency of the four learning ability scales has improved. Part of this improvement is due to doubling the number of items per scale from six to twelve. In addition, some unknown portion of this improvement is due to the response-set bias of the single-scale-per-column format. Other aspects of reliability, such as consistency over time, remain weak.

Internal Consistency Reliability

The *Technical Manual* for the LSI-1976 reported Spearman-Brown split-half correlation coefficients to suggest that the instrument was internally consistent. As we have noted, this method yielded inflated and misleading estimates of the average internal consistency of the four learning ability scales. In contrast, the *User's Guide* for the LSI-1985 reports coefficient alpha estimates of internal consistency reliability which indicate that the revised instrument has improved in this area. The *User's Guide* indicates that the average coefficient alpha for the four learning ability scales is .79. In Table 3 we summarize nine additional independent studies which found an average coefficient alpha of approximately .82.

TABLE 3

ESTIMATES OF INTERNAL CONSISTENCY FOR LSI-1985
(COEFFICIENT ALPHA)[a]

STUDY	SAMPLES	SIZE	CE	RO	AC	AE	AVERAGE ALPHA
1	G/UG Bus	181	.76	.84	.85	.82	.82
2	Manf. Employees	230	.82	.85	.83	.84	.83
3	UG Bus.	279	.82	.84	.84	.86	.84
4	UG Bus.	312	.85	.80	.83	.81	.82
5	UG Bus.[b]	40	.81	.85	.85	.88	.85
6	UG Bus.	229	.82	.79	.81	.82	.81
7	State Employees	333	.75	.79	.81	.84	.80
8	UG	694	.81	.79	.82	.78	.80
9	UG Bus.	455	.83	.81	.85	.84	.83
	TOTAL/AVERAGES	2753	.81	.81	.83	.83	.82

REFERENCES

1. Sims, Veres, Watson, and Buckner (1986).
2. Veres, Sims, and Shake (1987)
3. Sims, Veres, and Shake (1989)
4. Ruble and Stout (1990)
5. Geiger and Pinto (1991)
6. Ruble and Stout (1991)
7. Wells, Layne, and Allen (1991)
8. Geiger and Boyle (1992)
9. Geiger, Boyle, and Pinto (1993)

NOTES

[a] coefficient alpha for the combination scales are not reported because the assumption of independent measures is violated

[b] the entries for this study represent an average of three separate administrations of the LSI-1985 to the same sample

On the surface, these coefficients appear to indicate that the LSI-1985 is a reliable instrument. However, some of this apparent consistency is due to a response-set bias (as noted above) and the possibility that the forced intercorrelations of ipsative measures inflates the estimates (cf. Tenopyr, 1988). More important, it must be remembered that <u>reliability is a necessary, but not sufficient condition for validity</u>. Thus, simply improving the internal consistency of the LSI-1985 does not warrant the conclusion that the instrument is now a consistent and valid measure of learning styles. Indeed, further analysis indicates that the LSI-1985 has not resolved many (or most) of the problems of the original LSI-1976. For example, the temporal consistency of the revised LSI-1985 has not been improved.

Temporal Consistency Reliability

The ELM (Kolb, 1974) posits relatively stable individual learning styles, especially over short time intervals under similar circumstances. Thus, scores on the LSI-1985 should be reasonably consistent for individuals from one administration to another. One way to assess this consistency is to examine test-retest correlations.

Table 4 presents recent studies involving nearly 700 subjects who took the LSI-1985 twice with intervals between tests ranging from nine days to one year.

TABLE 4

ESTIMATES OF CONSISTENCY OVER TIME FOR LSI-1985 (TEST-RETEST CORRELATIONS)

STUDY	SAMPLES	SIZE	INTERVAL	CE	RO	AC	AE	AC-CE	AE-RO
1	G/UG Bus	181	5 weeks	.44	.42	.42	.62	--	--
2	Manf. Employees	201	3 weeks	.52	.46	.51	.44	--	--
3	UG	26	9 days	.57	.40	.54	.59	.69	.24
4	UG	107	1 month	.49	.72	.67	.63	.59	.71
5	UG Bus.▪	40	1 year	.17	.60	.55	.64	--	--
6	UG Bus.	139	5 weeks	.18	.46	.36	.47	.22	.54
	TOTAL/AVERAGES	694		.40	.51	.51	.56	.50	.50
	SHARED VARIANCE	(r^2)		16%	26%	26%	31%	25%	25%

REFERENCES

1. Sims, Veres, Watson, and Buckner (1986).
2. Veres, Sims, and Shake (1987)
3. Atkinson (1988)
4. Atkinson (1989)
5. Geiger and Pinto (1991)
6. Ruble and Stout (1991)

NOTES

▪ the entries for this study represent an average of three separate administrations of the LSI-1985 to the same sample

As indicated in the table, test-retest reliability coefficients for the LSI-1985 averaged approximately .50. This means that the proportion of "shared variance" in scale scores between test administrations was on the order of 25% ($.50^2$). These estimates are slightly lower than those for the LSI-1976 and indicate that the revised LSI-1985 does not show improved consistency over time.

Classification Stability

Another method for assessing temporal stability of LSI-1985 scores is to compare learning style classifications of individuals measured at different points in time. For example, are individuals classified as Convergers on an initial administration of the instrument classified similarly on a successive administration of the instrument?

Research examining classification stability indicates that the standard LSI-1985 has modest consistency at best. In a student sample, Sims et al. (1986) found only 47% were classified in the same learning style category after a five-week interval between tests. In an industry sample, Veres et al. (1987) found approximately the same results after only a three-week interval between tests. Ruble and Stout (1991) administered the LSI-1985 twice over a five-week interval to a sample of 139 undergraduate business students. We found that 56% of the subjects were classified into the same learning category upon the second administration. Finally, Geiger and Pinto (1991) administered the LSI-1985 to 40 undergraduate business students at the beginning of their sophomore, junior, and senior years. On average, 59% of these students were classified into the same categories from Time 1 to Time 2, Time 2 to Time 3, and Time 1 to Time 3. The overall average for these four studies ($n > 550$) indicates that approximately 53% of the classifications remained stable.

The 53% rate of classification agreement can be compared to chance by use of the _kappa_ statistic (cf. Siegel & Castellan, 1988, pp. 284-291). We would expect approximately 25% agreement by chance alone (due to four learning style categories). In all of the studies examining classification stability, _kappa_ coefficients indicated that the degree of agreement was statistically better than chance (see Ruble & Stout, 1992, for a correction to the _kappa_ statistics reported by Geiger and Pinto, 1991). Nevertheless, the probability that a given respondent would be classified into the same learning style category upon a second testing is only slightly better than flipping a coin. These results do not provide evidence that the LSI-1985 yields stable learning style classifications.

Temporal Stability of Modified Versions of the LSI-1985

Studies using a modified (scrambled) version of the LSI-1985 have shown some improvement in temporal stability. Our research (Ruble & Stout, 1991) found increases in test-retest correlations, but no improvement in classification stability. However, the increased test-retest correlations only reached an average coefficient of .54, indicating modest stability at best. While these results represent an improvement over the standard LSI-1985, they are not strong enough to suggest that the scrambled version yields consistent results. Moreover, although the test-retest correlations were higher for the scrambled version, a number of classification changes occurred around the means (based on Kolb's norms). Thus, compared to the standard LSI-1985, more respondents made smaller changes that resulted in classification changes (see Ruble & Stout, 1991 for further explanation of these shifts).

Veres et al. (1991) report both high test-retest correlations and high classification stability for their modified version of the LSI-1985. However, as noted by Veres et al. (1991, p. 149) "This unexpected result is difficult to explain." Given that the sentence endings in their study were distributed randomly, there may be an artifact based on the unique ordering of the alternative responses. Since the results of their study stand out from all other studies, they may simply represent an aberration that cannot be replicated. Further, it must be emphasized that in the Veres et al. (1991) study, <u>coefficient alpha registered a sizable decrease for the modified LSI-1985</u>. Thus, researchers should not be overly encouraged by the higher levels of temporal stability reported by Veres et al. (1991).

Temporal Stability of the LSI-1985: A Summary

Taken together, results pertaining to the temporal stability of measures yielded by the standard form of the LSI-1985 are disappointing. Either the instrument itself is unreliable or learning styles, as posited by the ELM, are not very stable personal characteristics (or both). If one's learning is determined primarily by the situation, the concept of "style" is misleading and an instrument to measure "style" has little value in generalizing from one situation to the next.

CONSTRUCT VALIDITY OF THE LSI-1985

Basically, there are two empirical approaches for assessing construct validity: (1) internal-structure analysis, and (2) cross-structure analysis (Pedhazur & Schmelkin, 1991). As the term suggests, internal-structure analysis focuses on the relationships of the items within the instrument itself. In contrast, cross-structure analysis examines the relationships between the measures of one instrument (e.g., the LSI-1985) and other measures of similar constructs (also see Kerlinger, 1986; Nunnally & Bernstein, 1994).

Internal-Structure Analysis: Patterns of Intercorrelations

One method of assessing the internal structure of the LSI-1985 is to examine the pattern of intercorrelations of the four scales. Given the bipolar assumptions of the ELM, the proposed opposite learning abilities should have strong negative correlations with each other (CE with AC and RO with AE) and essentially no correlation with the non-opposite scales. However, because the LSI-1985 is a purely ipsative instrument, the average intercorrelation is forced toward a moderately negative level (i.e., -.33).

With regard to the internal structure of the LSI-1985 scales, the data from three studies (Highhouse & Doverspike, 1987; Ruble & Stout, 1990; Smith & Kolb, 1986) failed to yield the expected pattern of intercorrelations. These studies all showed negative correlations of a given ability with other non-opposite abilities. In many cases, the unpredicted negative correlations were approximately equal or higher in magnitude than the predicted negative correlations. Thus, the pattern of intercorrelations from the LSI-1985 failed to support the bipolar assumptions of the ELM.

Internal-Structure Analysis: Factor Analysis

Factor analysis provides a more detailed approach to examining the internal structure of an instrument and is useful for assessing construct validity (Carmines & Zeller, 1989; Nunnally & Bernstein, 1994). In Ruble and Stout (1990), we presented early evidence regarding the factor structure of the LSI-1985. In that study, both two-factor and four-factor solutions were obtained. The two-factor solution is directly relevant to assessing the construct validity of the LSI because the ELM proposes two bipolar dimensions of learning. The four-factor solution provides supplemental information on the measurement of the four separate learning abilities posited by the ELM.

In the data set we analyzed (n=312), we found the following: (1) for the two-factor solution, AC items and CE items tended to load as separate factors while the AE and RO items did not generally load on either factor; (2) for the four-factor solution, the AC, RO, and AE items tended to load on separate factors, while the CE items did not. Most important, the results of the two-factor solution did not yield the two bipolar dimensions posited by the ELM. Thus, our factor analysis failed to support the construct validity of the LSI-1985.

Cornwell, Manfredo, and Dunlap (1991) administered the LSI-1985 to a sample of 317 respondents. Both two-factor and four factor solutions were generated from the response data. Results for the four-factor solution indicated that two of the factors were ill-defined (CE and RO). In the two-factor solution, AC and AE loaded together (contrary to expectations based on Kolb's ELM) while CE items did not load as a group on either of the two factors. Thus, the authors note that their evidence provides "little support for Kolb's two bipolar dimensions" (p. 455) and that "if the instrument is scored and interpreted in the way Kolb suggests, it may be misleading." (pp. 460-461).

Geiger, Boyle, and Pinto (1992) administered the LSI-1985 to 718 introductory accounting students and also generated two-factor and four-factor solutions. In the two-factor solution, contrary to predictions based on the ELM, CE items and RO items tended to load together, as did AC items and AE items. In the four-factor solution, only the AC items loaded together as a distinct factor. The authors conclude (p. 758) that their results "do not offer support of the construct validity of the revised LSI" which, in turn, "makes meaningful interpretation of the theorized learning abilities problematic."

More recently, Geiger, Boyle, and Pinto (1993) administered two versions of the LSI-1985 to a sample of 455 business administration students. They used the standard LSI-1985 (ipsative format) as well as a modified version (with a normative, rating format). The rating format was designed to overcome the problems of using factor analysis with ipsative measures. As in previous studies, both two-factor and four-factor solutions were obtained. For the standard version, results were similar to those reported in Geiger et al. (1992). In the two-factor solution, CE items tended to load together with RO items, while AC items tended to load together with AE items. In the four-factor solution, only the AC items loaded together strongly as a single dimension. Results for the rating version of the instrument also failed to support the hypothesized bipolar dimensions.

Taken together, four independent studies indicate that the LSI-1985 lacks a coherent structure necessary for construct validity. Further, the two-factor solutions that were obtained from these data sets yielded evidence that is not consistent with predictions based on Kolb's ELM.

Cross-Structure Analysis

The LSI-1985 has been correlated with related constructs with little success. Sims, Veres, and Shake (1989) and Goldstein and Bokoros (1992) have compared the LSI-1985 with a similar instrument, the Learning Styles Questionnaire (LSQ). In both cases, correlations between major dimensions of the LSI-1985 and the LSQ indicated relatively low levels of congruence. Goldstein and Bokoros (1992) also examined the consistency of learning style classifications between the two instruments and found that only 30 percent of the subjects were classified in equivalent styles. Highhouse and Doverspike (1987) and Baxter Magolda (1989) compared scores on the LSI-1985 with measures of cognitive style and cognitive complexity and did not find the expected relationships. Boyle, Geiger, and Pinto (1991) failed to find a relationship between the Diverger learning style and creativity as proposed by Kolb (1985).

Of course, since these cross-structural studies are correlational, either instrument could be deficient and account for a lack of positive results. In the case of the LSQ, it has some psychometric deficiencies of its own (Sims et al., 1989). On the other hand, the validity of the instrument used by Highhouse and Doverspike (1987) is well documented in the psychological literature.

It is important to note that cross-structural analysis is not conducted in isolation. Rather, it is performed as part of a wider process of attempting to validate research instruments. This sequence should begin with an assessment of reliability (e.g., internal consistency, test-retest), then turn to an internal-structure analysis, and finally to a cross-structural analysis. Within the context of the results of the first two steps, the results of the third step become more meaningful (informative). Indeed, it is difficult, perhaps impossible, to fully interpret the results of a cross-structural analysis without the context of other related measurement results.

Construct Validity: Summary

Evidence published to date fails to provide support for the construct validity of the LSI-1985. This lack of support is not surprising, given a number of measurement problems with the instrument.

CONCLUSIONS

We believe that the conclusions are clear and inescapable: Kolb's LSI does not provide adequate measures of learning styles. Independent researchers have conducted dozens of studies based on thousands of respondents and repeatedly have failed to find support for the requisite psychometric properties of the LSI. Thus, we believe that the use of the LSI in research should be suspended since the instrument lacks validity.

Further, use of the LSI for educational purposes also is likely to create a misleading impression that exaggerates its results. Use of the LSI suggests that the scores have some "scientific" value. However, we believe that the use of this instrument is unlikely to yield information beyond that attainable from simply asking people to pl--- themselves in a particular quadrant. Thus, use of the LSI is unnecessary to explore the implications of experiential learning theory and engage in self-inquiry. Individuals can "cross-validate" their own self-indicated learning style and develop personal learning strategies without the "excess baggage" of completing and scoring the LSI.

To overcome the psychometric problems of the LSI, some researchers are working on the development of normative measures of Kolb's learning styles (e.g., Romero, Tepper, & Tetrault, 1992). Certainly, new measures of learning styles are necessary for continuing research in this area.

More important than developing new measures of Kolb's four learning abilities, however, the evidence seems to suggest that even the basic model of learning (ELM) must be reconsidered. For example, apparently the learning abilities do not align in two bipolar dimensions as posited by Kolb. From the many factor-analytic studies of both the 1976 and 1985 instruments, it would appear that a revision of the ELM is warranted.

REFERENCES

Allinson, C. W., & Hayes, J. (1988). The Learning Styles Questionnaire: An alternative to Kolb's inventory? Journal of Management Studies, 25, 269-281.

Atkinson, G. (1988). Reliability of the Learning Style Inventory -- 1985. Psychological Reports, 62, 755-758.

Atkinson, G. (1989). Kolb's Learning Style Inventory--1985: Test-retest deja vu. Psychological Reports, 64, 991-995.

Atkinson, G. (1991). Kolb's Learning Style Inventory: A practitioner's perspective. Measurement and Evaluation in Counseling and Development, 23, 149-161.

Atkinson, G., Murrell, P. H., & Winters, M. R. (1990). Career personality types and learning styles. Psychological Reports, 66, 160-162.

Baker, R. E., Simon, J. R., & Bazeli, F. P. (1986). An assessment of learning style preferences of accounting majors. Issues in Accounting Education, 1, 1-12.

Baldwin, B. A., & Reckers, P. M. J. (1984). Exploring the role of learning style research in accounting education policy. Journal of Accounting Education, 2, 63-76.

Baxter Magolda, M. B. (1989). Gender differences in cognitive development: An analysis of cognitive complexity and learning styles. Journal of College Student Development, 30, 213-220.

Biberman, G., & Buchanan, J. (1986). Learning style and study skills differences across business and other academic majors. Journal of Education for Business, (April), 303-307.

Bonham, L. A. (1988). Learning style instruments: Let the buyer beware. Lifelong Learning, (11), 12-16.

Bostrom, R. P., Olfman, L., & Sein, M. K. (1990). The importance of learning style in end-user training. MIS Quarterly, 14, 101-119.

Boyatzis, R. E., & Renio, A. (1989). Research article: The impact of an MBA programme on managerial abilities. Journal of Management Development, 8,5, 66-77.

Boyle, E. J., Geiger, M. A., & Pinto, J. K. (1991). Empirical note on creativity as a covariate of learning style preference. Perceptual and Motor Skills, 73, 265-266.

Brown, H. D., & Burke, R. C. (1987). Accounting education: A learning-styles study of professional-technical and future adaptation issues. Journal of Accounting Education, 5, 187-206.

Carmines, E. G., & Zeller, R. A. (1979). Reliability and Validity Assessment. Sage University Paper Series on Quantitative Applications in the Social Sciences, #07-017, Beverly Hills, CA: Sage Publications.

Certo, S. C., & Lamb, S. W. (1979). Identification and measurement of instrument bias within the Learning Style Inventory through a Monte Carlo technique. Proceedings of the Southern Management Association.

Certo, S. C., & Lamb, S. W. (1980). An investigation of bias within the Learning Styles Inventory through factor analysis. Journal of Experiential Learning and Simulation, 2, 1-7.

Collins, J. H., & Milliron, V. C. (1987). A measure of professional accountants' learning style. Issues in Accounting Education, 2, 193-206.

Cornwell, J. M., Manfredo, P. A., & Dunlap, W. P. (1991). Factor analysis of the 1985 revision of Kolb's Learning Style Inventory. Educational and Psychological Measurement, 51, 455-462.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16, 297-334.

Curry, L. (1990). A critique of the research on learning styles. Educational Leadership, 48, 50-56.

Daft, R. L. (1994). Management. Fort Worth: Dryden Press.

DeCoux, V. M. (1990). Kolb's Learning Style Inventory: A review of its applications in nursing research. Journal of Nursing Education, 29, 202-207.

Ferrell, B. G. (1983). A factor analytic comparison of four learning-styles instruments. Journal of Educational Psychology, 75, 33-39.

Fox, R. D. (1984). Learning styles and instructional preferences in continuing education for health professionals: A validity study of the LSI. Adult Education Quarterly, 35, 72-85.

Freedman, R. D., & Stumpf, S. A. (1978). What can one learn from the Learning Style Inventory? Academy of Management Journal, 21, 275-282.

Freedman, R. D., & Stumpf, S. A. (1980). Learning Style Inventory: Less than meets the eye. Academy of Management Review, 5, 445-447.

Geiger, M. A., & Boyle, E. J. (1992). Learning styles of students and instructors: An analysis of course performance and satisfaction. The Accounting Educators' Journal, 4, 86-101.

Geiger, M. A., Boyle, E. J., & Pinto, J. K. (1992). A factor analysis of Kolb's revised Learning Style Inventory. Educational and Psychological Measurement, 52, 753-759.

Geiger, M. A., Boyle, E. J., & Pinto, J. K. (1993). An examination of ipsative and normative versions of Kolb's revised Learning Style Inventory. Educational and Psychological Measurement, 53, 717-728.

Geiger, M. A., & Pinto, J. K. (1991). Changes in learning style preference during a three-year longitudinal study. Psychological Reports, 69, 755-762.

Geller, L. M. (1979). Reliability of the Learning Style Inventory. Psychological Reports, 44, 555-561.

Goldstein, M. B., & Bokoros, M. A. (1992). Tilting at windmills: Comparing the learning style inventory and the learning style questionnaire. Educational and Psychological Measurement, 52, 701-708.

Gordon, O. I. (1985). Caution: Interpreting LSI can be injurious to your conclusions. Lifelong Learning, 9, 14-15.

Gordon, V. N., Coscarelli, W. C., & Sears, S. J. (1986). Comparative assessments of individual differences in learning and career decision making. Journal of College Student Personnel, (May), 233-242.

Green, D. W., & Parker, R. M. (1989). Vocational and academic attributes of students with different learning styles. Journal of College Student Development, 30, 395-400.

Green, D. W., Snell, J. C., & Parimanath, A. R. (1990). Learning styles in assessment of students. Perceptual and Motor Skills, 70, 363-369.

Gruber, C. P., & Carriuolo, N. (1991). Construction and preliminary validation of a learner typology for the Canfield Learning Styles Inventory. Educational and Psychological Measurement, 51, 839-855.

Hayden, R. R., & Brown, M. S. (1985). Learning styles and correlates. Psychological Reports, 56, 243-246.

Hicks, L. E. (1970). Some properties of ipsative, normative, and forced-choice normative measures. Psychological Bulletin, 74, 167-184.

Highhouse, S., & Doverspike, D. (1987). The validity of the Learning Style Inventory-1985 as a predictor of cognitive style and occupational preference. Educational and Psychological Measurement, 47, 749-758.

Hudak, M. A., & Anderson, D. E. (1990). Formal operations and learning style predict success in statistics and computer science courses. Teaching of Psychology, 17, 231-234.

Hunsaker, J. S. (1984). The experiential model and the Learning Style Inventory: An assessment of current findings. Journal of Experiential Learning and Simulation, 2, 145-152.

Jackson, D. J., & Alwin, D. F. (1980). The factor analysis of ipsative measures. Sociological Methods and Research, 9, 218-228.

Kerlinger, F. N. (1986). Foundations of Behavioral Research, New York: Holt, Rinehart and Winston.

Kolb, D. A. (1971). Individual learning styles and the learning process. M.I.T. Sloan School Working Paper No. 535-71, 1971.

Kolb, D. A. (1974). Management and the learning process. In D. A. Kolb, I. M. Rubin, & J. M. McIntyre, Organizational Psychology: A Book of Readings, 27-42, Englewood Cliffs, NJ: Prentice-Hall.

Kolb, D. A. (1976a). Learning Style Inventory. Boston: McBer and Co.

Kolb, D. A. (1976b). Learning Style Inventory: Technical Manual. Boston: McBer and Co.

Kolb, D. A. (1985). Learning Style Inventory: Self-scoring Inventory and Interpretation Booklet. Boston: McBer and Co.

Lamb, S. W., & Certo, S. C. (1978). The Learning Styles Inventory (LSI) and instrument bias. Proceedings of the 38th Annual Meeting of the Academy of Management, 28-32.

Markert, R. J. (1986). Learning style and medical students' performance on objective examinations. Perceptual and Motor Skills, 62, 781-782.

Marshall, E. A. (1985). Relationship between client-learning style and preference for counselor approach. Counselor Education and Supervision, (June), 353-359.

McKee, T. E., Mock, T. J., & Ruud, T. F. (1992). A comparison of Norwegian and United States accounting students' learning style preferences. Accounting Education, 1, 321-341.

Merritt, S. L., & Marshall, J. C. (1984). Reliability and construct validity of ipsative and normative forms of the Learning Style Inventory. Educational and Psychological Measurement, 44, 463-472.

Mielke, D. E., & Giacomino, D. E. (1989). Learning styles of public accountants: Measurement and applicability to practice management. Advances in Accounting, 7, 241-256.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric Theory. New York: McGraw-Hill.

Paunonen, S. V., & Gardner, R. C. (1991). Biases resulting from the use of Variables in psychology. Psychological Bulletin, 109, 520-523.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, Design, and Analysis. Hillsdale, NJ: Lawrence Erlbaum Associates.

Pigg, K., Busch, L., & Lacey, W. (1980). Learning styles in adult education: A study of county extension agents. Adult Education, 30, 233-244.

Reading-Brown, M. S. & Hayden, R. R. (1989). Learning styles--liberal arts and technical training: What's the difference? Psychological Reports, 64, 507-518.

Romero, J. E., Tepper, B J., & Tetrault, L. A. (1992). Development and validation of new scales to measure Kolb's (1985) learning style dimensions. Educational and Psychological Measurement, 52, 171-180.

Ruble, T. L. (1978). A psychometric analysis of Kolb's Learning Styles Inventory. Proceedings of the 5th Annual Meeting of The Association for Business Simulation and Experiential Learning, 193-198.

Ruble, T. L., & Stout, D. E. (1990). Reliability, construct validity, and response-set bias of the revised Learning Style Inventory. Educational and Psychological Measurement, 50, 619-629.

Ruble, T. L., & Stout, D. E. (1991). Reliability, classification stability, and response-set bias of alternate forms of the Learning Style Inventory. Educational and Psychological Measurement, 51, 481-489.

Ruble, T. L., & Stout, D. E. (1992). Changes in learning-style preferences: Comments on Geiger and Pinto. Psychological Reports, 70, 697-698.

Sein, M. K., & Bostrom, R. P. (1989). Individual differences and conceptual models in training novice users. Human-Computer Interaction, 4, 197-229.

Sein, M. K., & Robey, D. (1991). Learning styles and the efficacy of computer training methods. Perceptual and Motor Skills, 72, 243-248.

Sewall, T. J. (1986). The Measurement of Learning Style: A Critique of Four Assessment Tools, University of Wisconsin-Green Bay, Green Bay, WI., (ERIC Document Reproduction Service No. ED 267 247).

Siegel, S., & Castellan, N. J., Jr. (1988). Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill Book Company.

Sims, R. R., Veres, J. G., III, & Shake, L. G. (1989). An exploratory examination of the convergence between the Learning Styles Questionnaire and the Learning Style Inventory II. Educational and Psychological Measurement, 49, 227-233.

Sims, R. R., Veres, J. G., III, Watson, P., & Buckner, K. E. (1986). The reliability and classification stability of the Learning Style Inventory. Educational and Psychological Measurement, 46, 753-760.

Smith, D., & Kolb, D. A. (1986). User's Guide for the Learning Style Inventory. Boston: McBer and Company.

Stout, D. E., & Ruble, T. L. (1991a). A reexamination of accounting student learning styles. Journal of Accounting Education, 9, 341-354.

Stout, D. E., & Ruble, T. L. (1991b). The Learning Style Inventory and accounting education research: A cautionary view and suggestions for future research. Issues in Accounting Education, 6, 45-57.

Stumpf, S. A., & Freedman, R. D. (1981). The Learning Style Inventory: Still less than meets the eye. Academy of Management Journal, 6, 297-299.

Tenopyr, M. L. (1988). Artifactual reliability of forced-choice scales. Journal of Applied Psychology, 73, 749-751.

Togo, D. F., & Baldwin, B. A. (1990). Learning style: A determinant of student performance for the introductory financial accounting course. Advances in Accounting, 8, 189-199.

Veres. J. G., III, Sims, R. R., & Locklear, T. S. (1991). Improving the reliability of Kolb's revised Learning Style Inventory. Educational and Psychological Measurement, 51, 143-150.

Veres. J. G., III, Sims, R. R., & Shake, L. G. (1987). The reliability and classification stability of the Learning Style Inventory in corporate settings. Educational and Psychological Measurement, 47, 1127-1133.

Wells, J. B., Layne, B. H., & Allen, D. (1991). Management development training and learning styles. Public Productivity & Management Review, 14, 415-428.

West, R. F. (1982). A construct validity study of Kolb's learning style types in medical education. Journal of Medical Education, 57, 794-796.

Wilson, D. K. (1986). An investigation of the properties of Kolb's Learning Style Inventory. Leadership and Organization Development Journal, 7, 3-9.

Wolfe, D. E., & Byrne, E. T. (1975). Research on experiential learning: Enhancing the process, Proceedings of the Second Annual Meeting of the Association for Business Simulation and Experiential Learning, April, Bloomington, IN, 325-335.

Wunderlich, R., & Gjerde, C. L. (1978). Another look at the learning style inventory and medical career choice. Journal of Medical Education, 53, 45-54.

Zakrajsek, D. B., Johnson, R. L., & Walker, D. B. (1984). Comparison of learning styles between physical education and dance majors. Perceptual and Motor Skills, 58, 583-588.

APPENDIX

EXHIBIT 1

LEARNING STYLE GRID

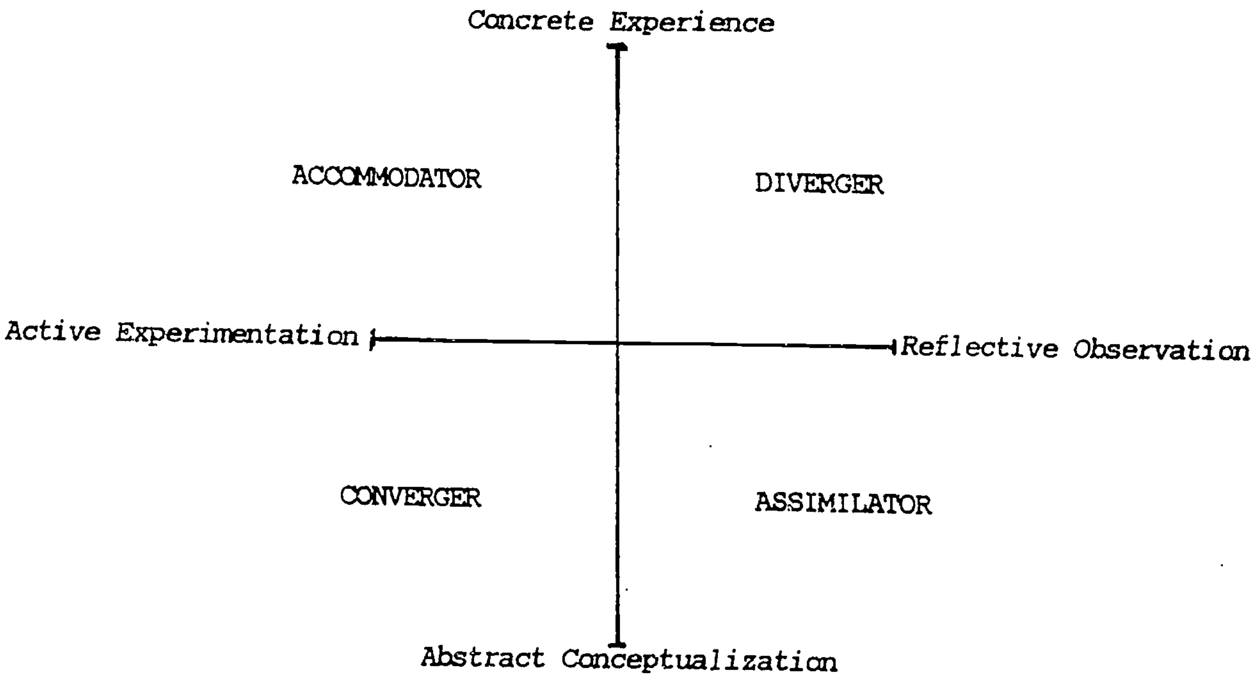

EXHIBIT 2

SETS OF WORDS AND SCORING KEY FOR LSI-1976

SET	CE	RO	AC	AE
1	discriminating	tentative	involved	practical
2	receptive	relevant	analytical	impartial
3	feeling	watching	thinking	doing
4.	accepting	risk-taker	evaluative	aware
5	intuitive	productive	logical	questioning
6	abstract	observing	concrete	active
7	present-oriented	reflecting	future-oriented	pragmatic
8	experience	observation	conceptualization	experimentation
9	intense	reserved	rational	responsible

Note: Underlined words are scored, the others serve as distractors

EXHIBIT 3

EMPIRICAL RESEARCH USING LSI-1976 WITHOUT REPORTING RELIABILITY DATA FOR THE SAMPLE STUDIED

The following studies (with over 5,000 subjects) did not report basic reliability statistics for the samples under investigation. In most cases, the justification for using the LSI-1976 was a previous study or Kolb's manual. However, without reliability data for the sample studied, the researchers, readers, and reviewers have no basis for judging the validity of the research. For these studies, it is virtually impossible to determine whether or not the results support the LSI OR ELM.

1. Atkinson, Murrell, and Winters (1990)
2. Baker, Simon, Bazeli (1986)
3. Baldwin and Reckers (1984)
4. Biberman and Buchanan (1986)
5. Bostrom, Olfman, and Sein (1990)
6. Boyatzis and Renio (1989)
7. Brown and Burke (1987)
8. Collins and Milliron (1987)
9. Ferrell (1983)
10. Fox (1984)
11. Gordon, Coscarelli, and Sears (1986)
12. Green and Parker (1989)
13. Green, Snell, and Parimanath (1990)
14. Hayden and Brown (1985)
15. Hudak and Anderson (1990)
16. Markert (1986)
17. Marshall (1985)
18. McKee, Mock, and Ruud (1992)
19. Mielke and Giacomino (1989)
20. Pigg, Busch, and Lacey (1980)
21. Reading-Brown and Hayden (1989)
22. Sein and Bostrom (1989)
23. Sein and Robey (1991)
24. Togo and Baldwin (1990)
25. West (1982)
26. Wunderlich and Gjerde (1978)
27. Zakrajsek, Johnson, and Walker (1984)