Test-Retest Reliability of the Index of Learning Styles for First-Year Engineering Students

Glen A. Livesay, Kay C Dee

Department of Applied Biology and Biomedical Engineering
Rose-Hulman Institute of Technology, Terre Haute, IN 47803

Abstract

Several different inventories and assessment instruments have been used to characterize the learning styles of engineering students. While recent studies have examined the psychometric properties of the Index of Learning Styles (ILS) - particularly the alpha reliability - there is less information available on the test-retest reliability of the ILS.

In this study, the ILS was administered to all engineering freshmen on the first day of class and again five weeks later to examine test-retest performance. Examination of the psychometric properties for each of the administrations of the ILS in terms of the alpha reliability (a measure of internal consistency of an instrument based on a single administration) revealed that the sensing/intuitor domain displayed the highest alpha reliability (0.76 in test and retest), and the lowest alpha reliability was associated with the sequential/global domain (0.48 in test). In terms of test-retest reliability (score stability over time), individual student scores in all domains were significantly correlated between the test and retest (p < 0.01). The sensor/intuitor domain displayed the highest correlation (Spearman’s rho = 0.80) and the sequential/global domain the lowest (Spearman’s rho = 0.60). On average, individual students repeated greater than 75% of their answers identically in the test and retest. These significant correlations provide additional support for the test-retest reliability of the ILS.

Introduction

There are multiple instruments available to characterize the learning styles of students, and for engineering students, the most widely utilized has been the Index of Learning Styles (ILS) originally developed by Richard M. Felder and Barbara A. Soloman [1]. We have been using the ILS as our primary assessment instrument in an ongoing evaluation of student learning styles, with the overall goal of developing improved instructional approaches and enhancing student learning. Recently, studies have examined the psychometric properties of the ILS [2,3] - in particular the reliability of the instrument as characterized using Cronbach’s alpha [4] (a measure of internal consistency of an instrument), but there is less information available on test-retest reliability, which provides a measure of the score stability of the instrument over time. Both Zywno [2] and Seery [5] (as later presented in Felder and Spurlin [6]) have recently reported on the test-retest reliability of the ILS, finding significant correlations between subsequent administrations of the instrument. Seery [5] conducted test-retest administration over a four-week
interval and obtained significant test-retest correlations (p < 0.01, Pearson’s) ranging from 0.870 to 0.725, and Zywno [2] demonstrated significant correlations (p < 0.01, Pearson’s) across all domains for an eight-month test-retest interval ranging from 0.505 to 0.683.

We have previously examined the test-retest reliability of the ILS over a test-retest interval of seven months [7], and found significant correlations in all domains at the p < 0.05 level. The relatively small sample size of that investigation, the relatively long test-retest interval, and the fact that only biomedical engineering students were involved in the test-retest aspect of the study led us to undertake a more comprehensive investigation of the test-retest reliability of the ILS for first-year engineering students. In addition, we sought to assess potential differences in the test-retest reliability of the ILS as a function of gender and to examine the distribution of domain score changes for individual students over the test-retest interval.

Methods

Index of Learning Styles
The Index of Learning Styles (ILS) characterizes student learning styles using four major domains, each characterized using two opposing descriptors, and is well-described in the literature [1,8,9] and on the Web [10]. The ILS essentially summarizes students’ self-reported preferences for receiving information visually or verbally, processing information in an active or in a reflective manner, focusing on sensory or intuitive types of information, and understanding information in a sequential or a global fashion. The ILS can be used to identify an overall preference or to describe a degree of preference (mild, moderate, strong) for a learning style, and is a valid assessment tool for the purpose of discussing teaching and learning [2,7]. With prior Institutional Review Board approval (#UT316), we administered the ILS to all students attending the first day of the Fall 2002 session of ENGR 100, the first-semester introductory engineering course at Tulane University. The ILS was administered a second time in the same course five weeks later. Student participation was voluntary and uncompensated, and study subjects were identified only by code number during data analyses. Students were not informed of the outcome of their individual questionnaires from either administration. Information regarding the domains of the ILS and their utility in discussions of engineering student learning and pedagogy was presented to the students following the second administration. At this point, students were encouraged to take the ILS using the web version [10] as part of learning more about their own preferences for learning and about steps they might take to enhance their own learning experiences.

Demographic data (gender, intended engineering major) were obtained through the office of the Dean of Engineering and matched to student code numbers for analyses. Statistical analysis techniques used in this study included calculation of Cronbach’s alpha (a measure of internal consistency reliability obtained for a single administration), Pearson’s correlation, and Spearman’s rho correlation.

Results
Learning Styles

The ILS sets two opposing descriptors to characterize each of four learning style domains (i.e., visual/verbal, active/reflective, sensor/intuitor, and sequential/global). Although everyone learns both actively and reflectively, both visually and verbally, etc., to facilitate numerical analyses and comparisons ILS profiles are generally reported in the literature as dichotomous options – a student is thus classified as preferring either visual or verbal learning, either active or reflective learning, etc. The results of this study are presented in terms of the percent of the student population which preferred the visual, active, sensor, and global learning styles. (Subtracting these data from 100% yields the percent of each student population which preferred the verbal, reflective, intuitive, and sequential learning styles).

For the 1st administration (test) of the ILS, 192 students completed the questionnaire, and 186 students completed the questionnaire during the 2nd administration (retest). These groups of completed questionnaires yielded a total of 165 matched pairs of ILS questionnaires for use in the test-retest evaluation. Since the objective of the current study was to examine the test-retest performance of the ILS - and specifically as a function of the individual domains - all complete pairs of domains (from the test and retest) were utilized. The ILS is composed of four domains, and therefore a student who fails to answer one question has still completed three of the domains correctly. For this reason, in the following presentations, the numbers cited varies across the domains (Active-Reflective: n = 151; Sensor-Intuitor: n = 153; Visual-Verbal: n = 157; Sequential-Global: n = 143).

Since the ILS was administered to the entire freshman class of engineering students, the learning styles of a variety of planned engineering majors are contained within the data. The top three intended majors were: 25% biomedical engineering, 18% undecided engineering, 13% mechanical engineering. Assessment of differences in student learning styles as a function of intended major was not an objective of the current study, but the relative percentages of majors were examined to verify that the sample obtained in this work was representative of a ‘normal’ freshman engineering class at Tulane University. The percentage of female students included in the sample was approximately 25%, and again within the expected range for an entering engineering class at Tulane University.

1st administration of the ILS (Test)

Based on the 1st administration of the ILS, the majority of students preferred visual (83%) rather than verbal learning, and 55% of the students reported a preference for active instead of reflective learning (Figure 1). Approximately 45% of the students preferred the sensor instead of the intuitor learning style, and 46% reported a preference for global rather than sequential learning. The sensor/intuitor domain demonstrated the highest alpha reliability (alpha = 0.76), while the lowest alpha reliability was associated with the sequential/global domain (alpha = 0.48). Alpha was determined to be 0.68 for the visual/verbal domain and 0.54 for the active/reflective domain.

The learning style preferences of male and female students were generally similar for the 1st administration of the ILS, with the largest difference being in the sensor/intuitor domain: 42% of males reported a preference for the sensor rather than intuitor learning style, while 54% of
Figure 1: Learning style preferences for the overall student sample for the 1st administration (test) and 2nd administration (retest) across the four domains of the ILS.
Figure 2: Learning style preferences for the test and retest as a function of gender for the domains of the ILS.
females reported this preference (Figure 2). A larger percentage of males also reported a preference for global instead of sequential learning.

2nd administration of the ILS (Retest)

For the 2nd administration of the ILS, the majority of students preferred visual (82%) rather than verbal learning, and 42% of the students reported a preference for active instead of reflective learning (Figure 1). Students reporting a preference for the sensor instead of the intuitor learning style were 54% of the sample, and a preference for global rather than sequential learning was seen for 38% of the sample. The sensor/intuitor domain demonstrated the highest alpha reliability for the 2nd administration (alpha = 0.76), while the lowest alpha reliability was associated with the active/reflective domain (alpha = 0.57). Alpha was determined to be 0.75 for the visual/verbal domain and 0.62 for the sequential/global domain.

The learning style preferences of male and female students were again similar, with the largest difference being in the sensor/intuitor domain: 50% of males reported a preference for the sensor rather than intuitor learning style, while 69% of females reported this preference. A larger percentage of males also reported a preference for global instead of sequential learning.

Comparison of the Test and Retest

When considering the student learning preferences as a whole between the test (the 1st day of class) and the retest (five weeks later), there was no change in the preference for visual learning instead of verbal (Figure 1). There were, however, changes in the other domains: as a group and using only single descriptors for each of the domains, the retest yielded data that indicated a preference for less active learning (down to 42% from 55%), more sensing (up to 54% from 42%) and less global (down to 38% from 46%) than the test (Figure 2). A complete breakdown of the ILS scores by domain for the test and retest is provided in Table 1. Males and females demonstrated similar changes on the whole from the test to the retest, although males showed larger reductions in their preference for active learning, and females showed larger increases in their preference for the sensor instead of intuitor learning style, and larger reductions in their preferences for global instead of sequential learning. Based on the assessment of reliability using alpha, all domains of the ILS became more ‘reliable’ (or remained the same), from the test to the retest. The largest increase was observed in the sequential/global domain, for which alpha increased to 0.62 in the retest from 0.48 in the test.

Across all domains and considering individual students, the average percentage of questions repeated identically on the ILS on the test and retest was 77%, ranging from a high of 100% to a low of 48% (Figure 3). In terms of repeatability of specific questions on the ILS, the five most highly repeated from the test to the retest were Q43 (92% of all students repeated exactly), Q37 (87%), Q39 (87%), Q23 (86%), and Q19 (85%). Four of these five questions come from the visual/verbal domain, and the fifth from the active/reflective domain. The five questions with the lowest percentages of repeats were Q33 (69% of all students repeated exactly), Q44 (69%), Q17 (68%), Q13 (68%), and Q8 (66%). Three of these questions come from the active/reflective domain, and two from the sequential/global domain.
Table 1: The aggregate listing of the ILS scores by domain for the test and retest.
Figure 3: The percentage of questions on the ILS that were answered identically in both the test and retest by individual students.
The distributions of scores in the various ILS domains were generally similar between the test and retest, all unimodal, mound-shaped, and centered near zero (except the visual/verbal domain). In all domains except the visual/verbal domain, there was also some overall trend of movement in one direction or the other. This can be seen in Figure 4, which plots the sequential/global domain for both the test and retest of the ILS, demonstrating a slight overall shift towards a more sequential learning style preference between the test and retest. Individual students' scores, however, in the sequential/global domain (and in all domains) were observed to move in both directions between the test and retest.

It is also possible to consider changes between the test and retest in terms of the number of questions in each of the domains that were repeated exactly by individual students. Within each domain, individual students repeated between eight and nine questions (out of eleven) on average, leading to average domain repeat percentages across all students that ranged from a high of 80% in the visual/verbal domain to a low of 74%, in both the active/reflective and sequential/global domains. More important, however, is the change in the individual scores students received for the various domains of the ILS from the test to the retest. The distribution of changes in the score received for the sequential/global domain from the test to the retest is shown for all students in Figure 5. The distributions of the individual student changes in domain score were relatively tight around zero across all domains. The percentage of students who experienced changes in an ILS domain score of two points or less in either direction - two being the smallest increment of change possible on the ILS domain scales - were 54% (active/reflective), 65% (sensor/intuitor), 64% (visual/verbal) and 56% (sequential/global).

**Correlation between the Test and Retest**

Pearson’s correlation coefficients were computed for paired student scores in all domains between the test and retest, and all were found to be significantly correlated at the p < 0.01 level: active/reflective = 0.640; sensor/intuitor = 0.799; visual/verbal = 0.667; sequential/global = 0.617. Test-retest correlations were also run for the females and males separately for all domains, revealing only very minor (if any) differences in the correlations as a function of gender. Although the distributions of ILS scores across all domains were generally mound-shaped and unimodal, they did not necessarily match the normal curve (particularly the visual/verbal domain, which is heavily skewed towards the visual side), so a non-parametric correlation was also run. Spearman’s rho correlation was found to be statistically significant at the p < 0.01 level for all domains: active/reflective = 0.621; sensor/intuitor = 0.798; visual/verbal = 0.694; sequential/global = 0.601. A sunflower plot of the matched scores between the test and retest for the sequential/global domain is shown in Figure 6, which illustrates the correlation of scores for the two administrations.

**Discussion**

Examination of the overall learning styles obtained for this cohort based on the test and retest as individual administrations of the ILS revealed general agreement with published reports on the learning styles of engineering students [2, 6]. As observed for other groups, the students in this study reported an overall strong preference for visual learning and a preference for sequential learning, although they also reported less active and less sensor preferences. Compared with
Figure 4: Detailed plot of the learning style preferences in the test (A) and retest (B) for the sequential/global domain of the ILS.
Figure 5: The actual change in the domain score for the sequential/global domain for individual students between the test and retest.
Figure 6: A sunflower plot of matched student scores in the sequential/global domain for the test and retest (n = 143). Each petal of a sunflower with petals represents a single case. Correlation between the test and retest for this domain was 0.60 (Spearman's rho, significant at $p < 0.01$).
previous data collected for engineering students at Tulane University, a somewhat lower percentage reported a preference for the global (vs. sequential) learning style \cite{11}. This is particularly true of the retest data in the present study, in which only 38% of students expressed a preference for the global learning style. The differences in these data from previous ILS testing of students at Tulane University may be due to differences in the student cohort examined; Dee, et al. was based on a group of engineering students that were farther along in their college careers (from freshman to senior, but the vast majority sophomore or older). Students in this study were also all from a biomedical engineering department, whereas only 25% of the students in the present group are from biomedical engineering. Differences in preferred learning style have been observed due to the engineering major of the students, although the specific reason for any differences - whether due to self-selection into a specific major, or due to instruction and pedagogical approach employed within a major - have not been determined.

The differences observed between the learning style preferences of males and females matches some trends observed by Rosati, et al. \cite{12,13}. Compared to female students, a higher percentage of male students preferred the visual learning style, and a lower percentage of male students preferred the sensor learning style in both the test and retest administrations of the ILS. Both male and female students expressed a preference for sequential learning, with more female than male students identified as sequential learners in both the test and retest, which also matches data obtained by Rosati, et al. \cite{12,13}. The data from the test (1st administration) shows a smaller proportion of students identified as sequential learners in this study than in Rosati’s work (female: 60% and male: 52% in the test, compared to Rosati’s data of female: 75% and male: 65\% \cite{12}). In the retest, however, the proportion of students expressing a preference for sequential learning increased in both females and males, up towards the range reported by Rosati \cite{12} (female: 74\% and male: 57\% in the retest).

The reliability of the ILS based on alpha (internal consistency) obtained for each of the domains in the present study for both the test and retest of the ILS are in the same range as reported previously by Zywno \cite{2}, although somewhat higher in the visual/verbal domain (test and retest) and higher across all domains than reported by Van Zwanenberg \cite{3}. It is important to bear in mind the utility and intent of the ILS before deciding ‘how good’ the alpha scores are; the ILS is intended to enable discussions of student learning and engineering pedagogy, and is not intended to be used as a predictor of academic performance or expected achievement in engineering. Specifically, there are no correct answers for the ILS, so there is not a set of learning style preferences that would be expected (meaning students would perform ‘well’ by answering in a certain way). For this reason, the alpha levels observed in the present work (and other ILS reports) are considered acceptable. The determination is alpha is the most widely utilized measure of reliability, representing the mean correlation across all possible split halves of the instrument. Since the data is being calculated from the correlation of different ‘halves’ of the instrument to one another, there is no need to conduct a second administration of the instrument, which can be difficult in practice.

The data obtained for the test-retest correlations in the present work over an interval of five weeks matches well with existing data for intervals ranging from four weeks \cite{5} to eight months \cite{2} (Table 2). Correlations of the test and retest were found to be statistically significant for all domains of the ILS (p < 0.01) in this study, but only the sensor/intuitor domain displayed a
correlation as high as that observed by Seery\cite{5}. Since the test-retest administrations in the present work and that of Seery\cite{5} were conducted over similar intervals they might be expected to be more similar. At the same time, all measures of the reliability of an instrument are a function of the instrument itself as well as the population that is tested, so the differences may be due to different populations. In the present work, first-year engineering students were examined, and, in fact, these were ‘first-day’ engineering students at the time of the test, so it is perhaps not surprising that the score stability may be somewhat lower for this population. The current test-retest correlations are somewhat higher in the sensor/intuitor, visual/verbal, and sequential/global domains when compared to studies reporting longer test-retest intervals (seven months \cite{7} and eight months \cite{2}). This fits with the general trend towards lower correlations over longer test-retest intervals, although ideally, a set of test-retest (and retest and retest) correlations over a range of intervals for the same population would be collected. This, however, is difficult to do in practice; we’ve tried it for intervals up to 16 months, but the loss of participants at any time point does compromise the overall sample size. The active-reflective domain in the present work, however, demonstrated a correlation that was lower than found for much longer test-retest intervals \cite{2,7}, although this correlation was still statistically significant. Again, this may be due to differences in the population tested.

<table>
<thead>
<tr>
<th>Interval</th>
<th>V-V</th>
<th>A-R</th>
<th>S-I</th>
<th>S-G</th>
<th>n =</th>
<th>Significant?</th>
<th>Correlation</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 weeks</td>
<td>0.870</td>
<td>0.804</td>
<td>0.787</td>
<td>0.725</td>
<td>46</td>
<td>All, p &lt; 0.01</td>
<td>Pearson's</td>
<td>Seery\cite{5}</td>
</tr>
<tr>
<td>5 weeks</td>
<td>0.667</td>
<td>0.640</td>
<td>0.799</td>
<td>0.617</td>
<td>varies by domain</td>
<td>All, p &lt; 0.01</td>
<td>Pearson's Spearman’s</td>
<td>Current Study</td>
</tr>
<tr>
<td></td>
<td>0.694</td>
<td>0.621</td>
<td>0.798</td>
<td>0.601</td>
<td>n = 153</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.68</td>
<td>0.73</td>
<td>0.78</td>
<td>0.60</td>
<td>40</td>
<td>All, p &lt; 0.05</td>
<td>Pearson's</td>
<td>Livesay\cite{7}</td>
</tr>
<tr>
<td>8 months</td>
<td>0.511</td>
<td>0.683</td>
<td>0.678</td>
<td>0.505</td>
<td>124</td>
<td>All, p &lt; 0.01</td>
<td>Pearson's</td>
<td>Zywno\cite{2}</td>
</tr>
</tbody>
</table>

**Table 2**: Comparison of test-retest reliability (score correlation) for the domains of the ILS for a range of test-retest intervals.

If one focuses on the responses of individual students, it can be seen that the vast majority of questions on the ILS were answered identically by students on the test and retest (approximately 76% overall, Table 3). This matches well with data reported for an 8-month test-retest interval reported by Zywno\cite{2}, although the percentages are generally higher in the present work. As score correlation is generally found to decrease with increases in the test-retest interval, the percentage of questions answered identically by individual students would also be expected to be lower for longer test-retest intervals. The visual-verbal domain was found to display the highest percentages of questions repeated by students in both the present work and that of Zywno\cite{2}, and the sequential-global domain was found to have the lowest percentage of questions repeated.

<table>
<thead>
<tr>
<th>Interval</th>
<th>V-V</th>
<th>A-R</th>
<th>S-I</th>
<th>S-G</th>
<th>n =</th>
<th>Overall %</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 weeks</td>
<td>80.4 %</td>
<td>73.9 %</td>
<td>78.3 %</td>
<td>73.9 %</td>
<td>varies by domain</td>
<td>76.7 %</td>
<td>Current Study</td>
</tr>
<tr>
<td></td>
<td>n = 157</td>
<td>n = 151</td>
<td>n = 153</td>
<td>n = 143</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8 months</td>
<td>77.7 %</td>
<td>74.4 %</td>
<td>73.0 %</td>
<td>69.1 %</td>
<td>124</td>
<td>73.5 %</td>
<td>Zywno\cite{2}</td>
</tr>
</tbody>
</table>

**Table 3**: Average percentage of questions repeated identically on the test and retest by individual students as a function of domain.
The distributions of domain score changes for individual students were observed to be more normal (statistically) than the distributions of the scores (test or retest) themselves across all domains. There was no observed effect of gender on these domain score changes; both male and female populations demonstrated changes in both directions in all domains, and all distributions were tightly clustered around zero (representing no change), as seen in Figure 6. Owing to the way in which the ILS is scored, the smallest increment of change possible in the domain scores between the test and retest is two, so one approach to characterize the amount of change observed for individual students between the test and retest is to consider the percentage of students who demonstrated changes of two or less (meaning changes of +2, 0, or -2). For the overall population, the percentage of students who experienced changes in an ILS domain score of two points or less in either direction were 54% (active/reflective), 65% (sensor/intuitor), 64% (visual/verbal) and 56% (sequential/global). In other words, the majority of students experienced changes in individual domain scores of less than one increment in either direction over the four week test-retest period.

In addition to the potential for real changes in student learning style preferences between test and retest, and other variability associated with testing (mood effects, random error, etc.), the interval utilized in the test-retest administrations of an instrument can also have an effect on the observed score stability. Of course, there is no agreed ‘best’ interval over which to conduct these types of investigations and both very short and very long test-retest intervals have their challenges. If the interval between test and retest is too short, then the memory of the test may unduly influence performance on the retest and therefore affect the determined test-retest reliability. In contrast, if the interval is too long, then real changes in the student performance (in this case, their learning style preferences) may be occurring which would again compromise the apparent test-retest reliability. For assessments of personality/preference, which include instruments like the ILS, an interval of approximately four weeks appears to be accepted as appropriate. As an aside, the original test-retest interval in this work was intended to be three weeks, but a pair of hurricanes on successive Wednesdays forced this study to a five-week test-retest interval.

Perhaps most interesting is the change observed in the same student population between the test and retest administrations of the ILS. Overall, students reported the same general preference for the visual learning, but were less active, more sensing, and less global in the retest as compared with the test. It seems expected that learning style preferences of engineering students might change to some degree throughout a college career, since as faculty we would expect student experiences and skills (e.g. problem solving, etc.) to develop with time (we hope!). However, for test-retest administrations within the same semester, large changes in the learning style preferences would likely not be expected. At the same time, the 1st administration in the present work was conducted on the very first day of college for these students, and the 2nd administration after five weeks of the ‘traditional’ freshman engineering curriculum (e.g. chemistry, physics, calculus, and little contact with the engineering faculty). In fact, the present test-retest data may represent information collected during one of the periods of largest change in the career of a college student - the first few weeks of college life.
Conclusions

Test-retest administration of the ILS to a cohort of first-year engineering students (all disciplines) revealed statistically significant correlations in all domains for individual student scores, over a five-week interval. Although some differences in the test-retest changes were observed between males and females, there was no observed effect of gender on the test-retest correlations for all domains. The test-retest correlations ranged from 0.617 to 0.798 (Spearman’s rho) and support recent data from Zywno [2] and Seery [5] that indicate the Index of Learning Styles demonstrates good test-retest reliability (stability of scores over time) over a range of administration intervals. In addition, the current study also provides initial data on the distribution and expected range of changes in ILS domain scores for individual students that might be expected over a relatively short test-retest interval.

Acknowledgements

We gratefully acknowledge the patience and good will of the participating students. We also acknowledge Dr. Eric A. Nauman for assistance with ILS administration and data entry; Mr. Toby Morris for assistance collecting demographic data; and Dr. Andrew Martinez for allowing the ILS administrations during his ENGR 100 course.

References

Biographical Information

GLEN A. LIVESAY is an Associate Professor of Applied Biology and Biomedical Engineering. He teaches undergraduate and graduate courses on biomechanics, capstone design, experimental design and data analysis, and experimental biomechanical testing of soft tissues.

KAY C. DEE is an Associate Professor of Applied Biology and Biomedical Engineering. Her teaching portfolio includes undergraduate and graduate courses on biology, biomaterials, cell/tissue engineering, cell/tissue biomechanics, science fiction and bioethics, and teaching engineering. One of her first teaching awards was an Apprentice Faculty Grant from the ERM division of ASEE - which changed everything.