Session 2351

# A Contribution to Validation of Score Meaning for Felder-Soloman's Index of Learning Styles

**Malgorzata S. Zywno**
**Ryerson University**

Abstract

In 1988, Richard Felder and Linda Silverman developed a learning model that focuses specifically on aspects of learning styles of engineering students. Three years later, a corresponding psychometric assessment instrument, the Felder-Soloman's Index of Learning Styles, was developed. This paper offers a contribution to an ongoing validation work on the ILS instrument, based on the author's three-year study of the relationship between student learning styles and their academic achievement in hypermedia-assisted learning environment. The paper provides an analysis of psychometric properties of the ILS, based on the scores for 557 valid questionnaires collected in the study. This includes test-retest reliability, factor structure, internal reliability, total item correlation and inter-scale correlation. Construct validity is also discussed. In summary, the author supports conclusions found in the literature pointing to the ILS as a suitable psychometric tool for evaluating learning styles of engineering students. The author also concurs in recommendations for further work on validating the meaning of its scores and on improving the specific items to reduce inter-scale correlation.

## I. Introduction

### Felder Learning Style Model

The Felder model of learning styles[1, 2] focuses on aspects of learning styles significant in engineering education, and is very popular among engineering educators even though the psychometric instrument associated with the model, the Index of Learning Styles[3] (ILS), has not yet been fully validated. In brief, the model has five dimensions: Processing (Active/Reflective), Perception (Sensing/Intuitive), Input (Visual/Verbal), Understanding (Sequential/Global) and Organization (Inductive/Deductive). Felder recommends the inductive teaching method (i.e. problem-based learning, discovery-based learning), while the traditional college teaching method is deductive, i.e. starting with fundamentals and proceeding to applications. Thus the last dimension was removed from the ILS, so as not to provide incentives for a continuing use of the traditional deductive instruction[4]. The model assembles a learning preferences profile of a group of learners and provides insight into how teaching strategies can be modified to broaden their appeal to a larger cross-section of the student population. To increase the support for learners with different individual preferences, Felder advocates a multi-style approach to science and engineering education[1, 2], and incorporation of active, experiential, collaborative student-centered learning[5]. This approach has long been advocated as an effective learning environment for engineering education[6, 7, 8].

<u>Instrument Reliability and Validity</u>

Any measurement must be both reliable - measurement yields consistent, repeatable results, and valid - it measures what it is supposed to measure[9]. The first is an issue of *reliability*, the second of *construct validity*. Reliability can be estimated through *inter-rater reliability*, i.e. whether the two raters are consistent, through *test-retest reliability*, assessing the consistency of a measure from one time to another, and through *internal consistency reliability*, assessing the consistency of results across items within a test. The internal consistency of single-dimensional additive scales such as in the Felder Model, can be tested using Cronbach's alpha, a coefficient assessing how well a set of items on the scale measures a single "underlying construct"[9, 10]. The higher the score, the more reliable the generated scale is. The widely accepted social science cut-off is that alpha should be 0.70 or higher for a set of items to be considered a scale, because at $\alpha = 0.70$, the standard error of measurement will be over half of a standard deviation[10, 11]. However, lower thresholds are sometimes used in the literature. For example, Tuckman[12] states that alpha test reliability should be above 0.75 for achievement tests but only above 0.5 for attitude tests.

Construct validity refers to the degree to which inferences can be made from the operationalizations in the study to the theoretical constructs on which they were based[9]. It is established by relating a presumed measure of the construct to some behavior or manifestation that it is hypothesized to underlie[12]. Construct validity thus comprises the evidence and rationales supporting the trustworthiness of score interpretations in terms of explanatory concepts that account for both the test performance and score relationships with other variables[10].

<u>ILS Instrument</u>

While there have been many studies[13, 14, 15, 16] that used the ILS, the author located few studies that dealt with the instrument validation. Van Zwanenberg et al.[17] examined learning styles of 139 engineering students and 145 business students at two universities in Newcastle, UK, using the Felder Model. They concluded that the ILS scales had low internal reliability, with Cronbach's alpha = 0.41 to 0.65, and expressed concerns regarding robustness and construct validity of the instrument. Livesay et al.[18], in a study of 255 engineering students at Tulane University, New Orleans, found alpha to be in the range of 0.54 to 0.72. They also found relatively high test-retest reliability in repeated measurements over time, and concluded that the ILS was an appropriate and statistically acceptable tool for characterizing learning preferences. They also encouraged others to continue work on statistical validation of the ILS and to share their findings. In an unpublished study[19], Felder and Spurlin examined the ILS responses of 584 students at North Carolina State University, and found Cronbach's alpha coefficients to be in the range of 0.55 to 0.76.

## II. Methods

### Participants

The research in which ILS questionnaires were collected took place at Ryerson University, Toronto, Canada, during three consecutive offerings (2000-2002) of a course in control systems in the undergraduate Electrical and Computer Engineering program. The research dealt with efficacy of hypermedia-assisted instruction and the relationship of learning styles, hypermedia and achievement[20]. Student participation in the study was voluntary, and all participating students were asked to sign an informed consent letter. The students were not exposed to any risks or reprisals for refusal to participate in the study. Each year the participating students were asked to complete the ILS questionnaire to assess their learning styles. Participants' names were identified for cross-referencing purposes.

Some ILS questionnaires were also collected from engineering professors. They included 13 voluntary and anonymous ILS questionnaires received from participants of a workshop conducted in February 2002 for faculty members of Faculty of Engineering at Concordia University in Montreal, Quebec. Another seven anonymous questionnaires were received from engineering faculty members at Ryerson University in an attempt to assess the viability of a more general faculty survey[21]. That voluntary and anonymous survey was eventually conducted in September 2002, with 48 responses returned. Data collected from the initial 20 questionnaires was only used in the validation of the ILS instrument.

### Study Design

In order to validate the ILS, two main analyses were conducted, a test-retest and Cronbach's alpha/factor analysis. To conduct the former, two sessions for completing the ILS had to be planned. While more students would likely have completed both questionnaires if the ILS were administered at the beginning and at the end of the course in which the study was located, the author wanted to avoid test-fatigue, to minimize the intrusiveness of the study on the students, and to ensure a reasonable spacing of the two tests. The ILS was administered in the second or third week of classes. Therefore, in a 13-week course, the time lapse would have been relatively short, approximately 10 weeks. Thus, a retest after the summer break was chosen instead, approximately eight months after the first ILS test was administered.

Table 1 illustrates how many valid ILS questionnaires were available at different points of time in the study. Its columns may require some explanation. A few questionnaires were handed in late, sometimes by several weeks. However, because of their small number and a relatively large nominal time lapse between the two ILS questionnaires (8 months), no attempt was made to track the late questionnaires in order to determine the exact time lapse in each case.

Table 1: Number of Students Who Completed the ILS Questionnaires

| Year | Total in Course | Valid ILS # 1 | Late ILS #1 | Removed | Total ILS #1 | Valid ILS # 2 | Late ILS #2 | **Wrote Both** | New after # 2 | **Total ILS** |
|---|---|---|---|---|---|---|---|---|---|---|
| 2000 | 102 | 56 | 5 | -6 | 55 | 28 | 5 | **3** | 30 | **85** |
| 2001 | 128 | 119 | 0 | -2 | 117 | 70 | 0 | **66** | 4 | **121** |
| 2002 | 137 | 120 | 9 | -1 | 128 | 59 | 0 | **55** | 4 | **132** |
| Total | 359 | 295 | 14 | -9 | 300 | 157 | 5 | **124** | 38 | **338** |

In 2000, 8 students (including 6 who filled out the ILS # 1 questionnaires) were disciplined for plagiarizing their project reports. Since their course grades were affected for reasons not relevant to the study, their data was removed from the study, as shown in Table 1. Two students in 2001 and one in 2002 dropped the course before the official deadline, but after filling out the ILS # 1, and their data was also pulled from the study. Each year approximately 30% of the course graduates left for an Industrial Internship year, and thus the pool of respondents for the second round of the ILS was always smaller, as seen in Table 1. In 2001-2002, this turnover has not significantly affected the overlap between the students who wrote the ILS # 1 and the ILS # 2, because of the high rate of participation in the former. However in 2000, the first questionnaire combined the ILS with a version of the Kolb learning style assessment. This led to a large number of invalid responses, as students generally had problems with the former, and many completed only one part of the handout, or ran out of time. Lesson learned, the Kolb questionnaire was abandoned in the subsequent years, significantly reducing the number of invalid questionnaires, which in turn contributed to high participation rates and a larger test-retest pool. Overall, n=124 samples were available for the test-retest analysis, as shown in Table 1.

In the factor and Cronbach alpha analyses, 557 valid ILS questionnaires were used. These consisted of the 338 individual questionnaires in the study, as seen in the last column of Table 1, 124 retest questionnaires, 68 questionnaires from engineering professors collected in a separate study[21], and 27 questionnaires from students originally participating in the 1999 pilot study[22]. The latter were collected during the 2000 retest from the students returning from the Industrial Internship year. Since all had previously signed consent forms, their ILS questionnaires were included in the reliability analysis. However, no attempt was made to include them in any other analyses. This was because the sample was not only small (only 28% of the 1999 class), but also not representative of their cohort, as the Industrial Internship students hailed almost exclusively from the high-achieving group.

Scoring the Scales

One of the problems encountered was scoring of the scales. The ILS scales are bipolar, with mutually exclusive answers to items, i.e. either (a) or (b). Because there is an odd number of items on each scale, if items are scored as +1 and –1, respectively, the total score on a scale from –11 to +11 shows an emerging preference for the given modality. However, the dichotomous nature of scales makes the use of standard statistic tests difficult[17]. Thus, only scales for either (a) or (b) should be considered, each consisting of 11 items. The responses were scored for the Active, Sensing, Visual and Sequential scales by assigning a value of 1 to (a) items, and 0 to (b) items. Scores for the respective opposite polarities, Reflective, Intuitive, Verbal and Global, can be found as a complement of 11 (i.e., if the average Active score is 6.5, the average Reflective score is 4.5).

## III. Results and Discussion

### Test-Retest Reliability of ILS

In estimating test-retest reliability, the same test is administered to the same or similar sample, on more than one occasion. Time elapsing between the measurements is critical. Typically the longer the time-gap is the lower the correlation. In the study, the time lapse of eight months was dictated by the classroom realities, as described above. Table 2 shows a moderate to strong correlation between the test and the retest scores.

Table 2: Pearson's Correlation of Test-Retest Scores for the ILS

| Active Scores | Sensing Scores | Visual Scores | Sequential Scores |
|---|---|---|---|
| 0.683** n=124 | 0.678** n=124 | 0.511** n=124 | 0.507** n=124 |

**Statistically significant at the 0.01 level, 2-tailed.

Van Zwanenberg et al.[17] did not conduct the test-retest analysis. Livesay et al.[18] repeatedly tested a group of engineering students (n =24) at four, seven, twelve and sixteen month intervals. They observed a linearly dropping off correlation of the repeated ILS scores, although learning preferences are thought to be a constant individual characteristic[23, 24]. However, the small sample size makes these results difficult to interpret. Pearson's correlation coefficients at seven months were slightly higher than in Table 2, at 0.73, 0.75, 0.68, and 0.60, respectively. As in Table 2, they were higher for the Active and Sensing scores than for the Visual and Sequential scores. Table 3 shows the results of Paired-Samples t-test, demonstrating no statistically significant differences between the test-retest mean scores on three of the four scales.

However, the difference between the means of Visual scores was borderline significant ($p = 0.049$) and the correlation between the two Visual scores was the second lowest, at 0.511 (Table 2). Yet, the use of such standard statistical tools may be in fact misleading as a predictor of stability of the scales, in this case of the Visual scale. Homogeneity or heterogeneity of scores affect score reliability since a small change in raw scores leads to large changes in rankings and thus low correlation of the scales[25].

Table 3: Paired Samples Statistics of Test-Retest Scores for the ILS (n =124)

| | | Mean | N | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|---|
| Pair 1 | active score 1 | 5.99 | 124 | 2.40 | .22 |
| | active score 2 | 5.73 | 124 | 2.37 | .21 |
| Pair 2 | sensing score 1 | 6.68 | 124 | 2.66 | .24 |
| | sensing score 2 | 6.60 | 124 | 2.62 | .24 |
| Pair 3 | visual score 1 | 8.14 | 124 | 2.11 | .19 |
| | visual score 2 | 8.51 | 124 | 2.10 | .19 |
| Pair 4 | sequential score 1 | 6.00 | 124 | 2.07 | .19 |
| | sequential score 2 | 5.62 | 124 | 2.26 | .20 |

**Paired Samples Test**

| | | Paired Differences | | | | | t | df | Sig. (2-tailed) |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | Std. Deviation | Std. Error Mean | 95% Confidence Interval of the Difference Lower | Upper | | | |
| Pair 1 | active score 1 - active score 2 | .26 | 1.90 | .17 | -7.96E-02 | .60 | 1.513 | 123 | .133 |
| Pair 2 | sensing score 1 - sensing score 2 | 8.06E-02 | 2.12 | .19 | -.30 | .46 | .423 | 123 | .673 |
| Pair 3 | visual score 1 - visual score 2 | -.37 | 2.08 | .19 | -.74 | -9.56E-04 | -1.985 | 123 | .049 |
| Pair 4 | sequential score 1 - sequential score 2 | .38 | 2.15 | .19 | -4.03E-03 | .76 | 1.959 | 123 | .052 |

Indeed, because of the preponderance of strongly Visual learners in the study, over the years, the Visual scores had the highest means with the lowest standard deviations of all scales (see Table 4). Histograms for each of the four scales are shown in Figure 1, with the statistics for the distributions in Table 5. Skewness, a measure of symmetry, is the highest for the Visual and the Sensing scales, which have the longest tails and are skewed toward the high repeatability end (see Figure 1). The Visual scale yielded the highest percentage of stable responses (78%), unchanging with time on the test-retest, contrary to what results in Table 2 and Table 3 seem to suggest. As well, three out of five most stable questions were found on the Visual Scale: 107 out of 124 students (86%) answered the same way to Question 19, 105 (85%) to Question 3, and 103 (83%) to Question 43.

Table 4: Means, Standard Deviations and ANOVA Results for Time Comparisons in ILS (Different Cohorts of Students), 2000-2002

| **Year** | **Sample** | **Active Score** | | **Sensing Score** | | **Visual Score** | | **Seq. Score** | |
|---|---|---|---|---|---|---|---|---|---|
| | | **Mean** | **STD** | **Mean** | **STD** | **Mean** | **STD** | **Mean** | **STD** |
| **2000** | 85 | 6.05 | 2.33 | 6.74 | 2.52 | 8.01 | 2.28 | 6.40 | 2.14 |
| **2001** | 121 | 6.00 | 2.48 | 6.50 | 2.60 | 8.21 | 2.02 | 5.71 | 2.04 |
| **2002** | 132 | 6.05 | 2.33 | 6.26 | 2.51 | 8.02 | 2.09 | 5.87 | 2.13 |
| **00-02** | 338 | 6.03 | 2.38 | 6.46 | 2.55 | 8.09 | 2.11 | 5.95 | 2.11 |
| ANOVA stats. | | F = 0.024, df= 2, 335, p = 0.976 | | F = 0.947, df= 2, 335, p = 0.389 | | F = 0.308, df= 2, 335, p = 0.735 | | F = 2.828, df= 2, 335, p = 0.061 | |

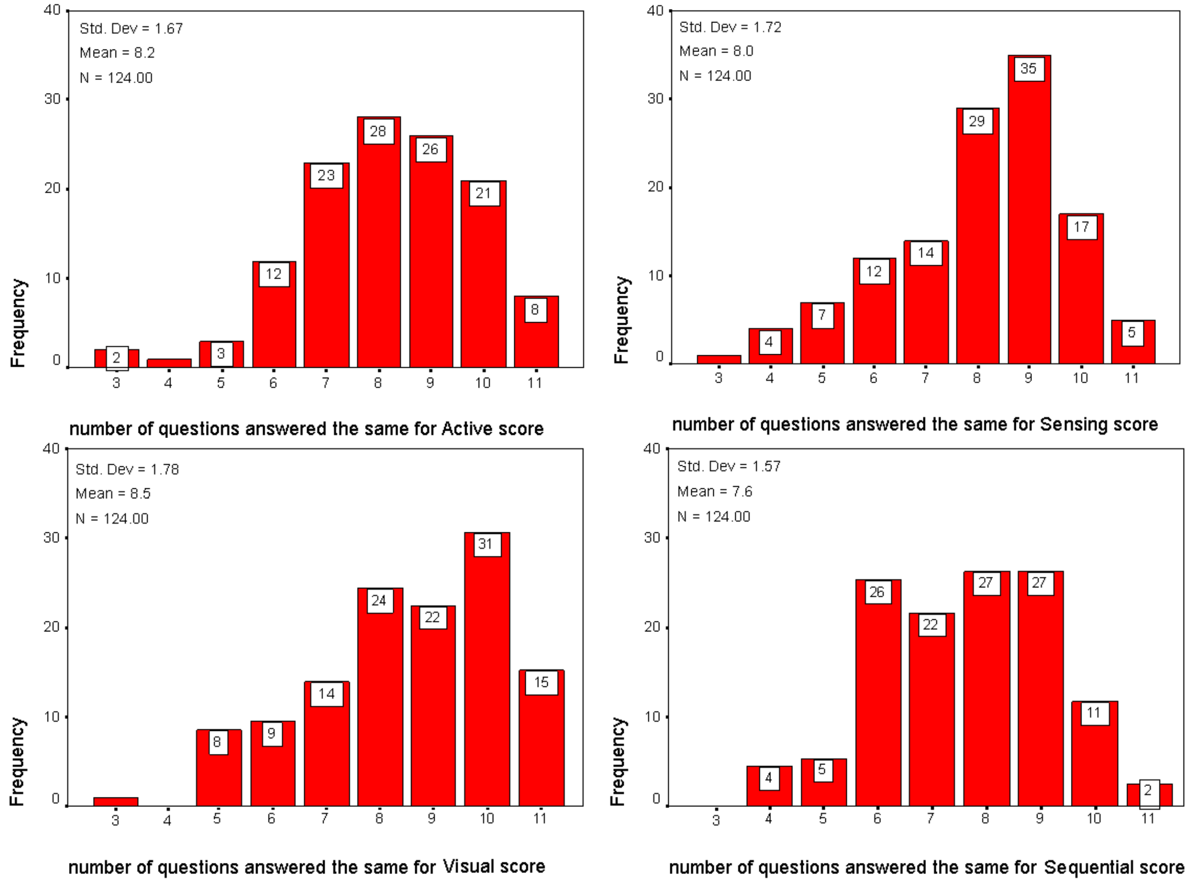


Figure 1: Histogram of Identical Responses to Items on Test-Retest

Table 5: Statistics for the Distributions of Identically Answered Questions on Test-Retest

| | Overall 44 Items | Active Scale 11 Items | Sensing Scale 11 Items | Visual Scale 11 Items | Seq. Scale 11 Items |
|---|---|---|---|---|---|
| Mean | 32.34 (73.5%) | 8.18 (74.4%) | 8.03 (73%) | 8.55 (77.7%) | 7.60 (69%) |
| STD | 4.19 | 1.67 | 1.72 | 1.78 | 1.59 |
| Skewness | -0.779 | -0.511 | -0.670 | -0.606 | -0.162 |
| Kurtosis | 1.708 | 0.361 | 0.066 | -0.233 | -0.544 |

The remaining two most stable questions were on Active and Sensing scales - 83% repeatability on Question 21 and Question 22, respectively. The Sequential scale appears to be the least stable, with the lowest average number of identically answered items (69%), and three least stable questions, Question 40 (62% repeatability), and Questions 4 and 42 at 64% repeatability rate. Kurtosis is a measure of whether the data are peaked or flat relative to a normal distribution. The Sequential scale also had the flattest distribution, i.e. the most negative kurtosis (see Table 5). Repeatability results were somewhat similar in the Livesay et al.[18] study, despite its small sample size. Out of the three most stable items (above 90% repeatability) over the 16-month period, two were on Active scale, and one on Visual scale. Out of the four least stable items (below 50%

repeatability) were thee items from Sequential scale and one item from Active scale. The larger variation of that data could be explained by a much longer time interval between tests.

Internal Consistency Reliability

Analysis of internal reliability of scales was performed on the (a) items for all 557 valid ILS questionnaires (Table 6). Cases with missing items were excluded from the analysis, and thus the number of cases shown varies. The internal reliability of the scales was found to range from 0.53 to 0.70, higher than in the Van Zwanenberg et al.[17] study. The resulting coefficients meet acceptable limits as suggested by Tuckman[12]. Average Item-Total Correlation is another measure of consistency of the scale[9]. Table 6 shows ITC coefficients for the full 11-item scales. Livesay et al.[18] provided only ITC values for the items referred to as "core", which yielded the highest reliabilities in each domain. They were respectively equal to 0.26 (10 items on Active scale), 0.44 (8 items on Sensing scale), 0.41 (6 items on Visual scale) and 0.33 (7 items on the Sequential scale). With only those "core" items considered, the ITC coefficients in Table 6 would be respectively 0.28, 0.37, 0.38, and 0.25, comparable to Livesay et al.[18].

Table 6: Internal Consistency Reliability for the ILS - Cronbach's alpha

| | **Cases** | **Items** | **Scale Mean** | **Scale Variance** | **Scale STD** | **Avg. IIC*** | **Avg. ITC**** | **Stand. $\alpha$** |
|---|---|---|---|---|---|---|---|---|
| Act-Ref | 540 | 11 | 5.7889 | 5.6177 | 2.3702 | 0.1179 | 0.264 | **0.595** |
| Sen-Int | 539 | 11 | 6.2430 | 7.0245 | 2.6504 | 0.1730 | 0.349 | **0.697** |
| Vis-Ver | 544 | 11 | 8.1801 | 4.4537 | 2.1104 | 0.1354 | 0.289 | **0.633** |
| Seq-Glo | 532 | 11 | 5.7726 | 4.7900 | 2.1886 | 0.0927 | 0.217 | **0.530** |

*IIC: Inter-Item Correlations, **ITC: Item-Total Correlations

Table 7 shows comparison of the results with other studies. The results were virtually identical when the 124 ILS questionnaires collected during the test-retest experiment and the 1999 sample (n=27) were removed. The slight difference can be attributed to the reduced statistical power.

Table 7: Internal Consistency Reliability Comparisons

| **Study** | **N** | **Active Scale $\alpha$** | **Sensing Scale $\alpha$** | **Visual Scale $\alpha$** | **Sequent. Scale $\alpha$** |
|---|---|---|---|---|---|
| Newcastle, UK, Van Zwanenberg et al.[17] | 279 | 0.51 | 0.65 | 0.56 | 0.41 |
| Tulane, LA, Livesay et al.[18] | 255 | 0.56 | 0.72 | 0.60 | 0.54 |
| North Carolina, Felder & Spurlin[19] | 584 | 0.70 | 0.76 | 0.69 | 0.55 |
| Ryerson, Canada | 557 | **0.60** | **0.70** | **0.63** | **0.53** |
| Ryerson, Canada* | 406 | **0.60** | **0.69** | **0.61** | **0.50** |

* Test-Retest Data and 1999 Sample Excluded.

Next, factor analysis was performed. The number of factors extracted using Kaiser's criterion (eigenvalues less than 1.0) was 14, accounting for 54.1% of the total variance. Using the "scree plot" test, in which components are ignored beyond the place where the smooth decrease of eigenvalues appears to level off to the right of the plot, the number of extracted factors was equal to 5, accounting for 28.9% of the total variance. The corresponding scree plot is shown in Figure 2. The first method (Kaiser criterion) sometimes retains too many factors, while the second (scree test) sometimes retains too few, however, both do quite well under normal conditions, that is, when there are relatively few factors and many cases.

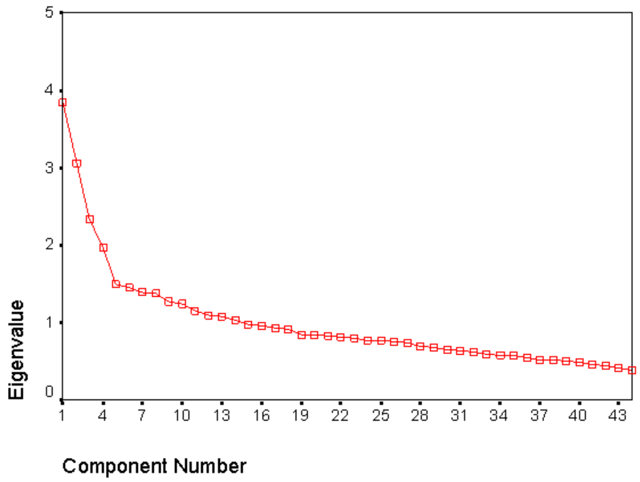

Figure 2: Scree Plot for Factor Analysis on ILS Scores (n =551)

Both methods were used here and results were checked for consistency. Initially, an unrotated solution was obtained. The extracted factors were identified with those items on the ILS that loaded highly (> 0.30). The distribution of high loading items for the Kaiser method (14 factors) is shown in Table 8.

Table 8: Distribution of High Loading Items, Kaiser Method, Unrotated

| **Factors:** | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** | **12** | **13** | **14** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Active | 3 | | 7 | 2 | | | | | | | 1 | 1 | | 1 |
| Sensing | 10 | | | 1 | | | | 3 | 2 | | | | | 2 |
| Visual | 1 | 8 | | 1 | | | 1 | | 1 | 2 | | | | |
| Sequential | 2 | | | 2 | 3 | 1 | | | 1 | | 1 | 1 | 1 | |

The results offer support for the relative orthogonality of three of the four scales, with items from the Sensing scale predominantly loading Factor 1, items from the Visual scale predominantly loading Factor 2, and items from the Active scale predominantly loading Factor 3. However, items from the Sequential scale load several factors. For comparison, the scree method was used, reducing the number of extracted factors to 5. To discern the patterns better, the solution was rotated, but because of the suspected overlap, instead of the orthogonal Varimax rotation, a non-orthogonal Oblique rotation was used, with results shown in Table 9. A clear pattern can now be seen, with the Sequential scale loading predominantly Factor 5, but also overlapping somewhat with the Sensing scale.

The resulting component correlation matrix is shown in Table 10. All correlation factors are negligible, except for the correlation between Component 1, mostly loaded by the items from the Sensing scale, and Component 5, mostly loaded by the items from the Sequential scale.

Table 9: Distribution of High Loading Items, Scree Method, Oblique Rotation

| **Factors:** | **1** | **2** | **3** | **4** | **5** |
|---|---|---|---|---|---|
| Active | 2 | | 7 | | |
| Sensing | 9 | | | 2 | 2 |
| Visual | | 8 | | | |
| Sequential | | | | 2 | 6 |

Table 10: Component Correlation Matrix, 5 Factors Extracted, Oblique Rotation

**Component Correlation Matrix**

| Component | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.000 | 9.43E-02 | 6.86E-02 | -1.3E-02 | .216 |
| 2 | 9.43E-02 | 1.000 | 9.82E-02 | -8.6E-03 | -7.7E-02 |
| 3 | 6.86E-02 | 9.82E-02 | 1.000 | -2.6E-02 | 2.20E-02 |
| 4 | -1.3E-02 | -8.6E-03 | -2.6E-02 | 1.000 | 5.95E-02 |
| 5 | .216 | -7.7E-02 | 2.20E-02 | 5.95E-02 | 1.000 |

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Next, direct inter-scale correlation was considered. In order to assess separate qualities, the inter-scale correlation should be minimal. Table 11 shows Pearson's correlation coefficients computed between scores on the ILS scales. Three of the four scales had negligible inter-scale correlation. However, a weak correlation ($r = 0.334$) was observed between the Sensing and Sequential scores. This is consistent with the results of the factor analysis. Van Zwanenberg et al.[17] also found the overlap between Sensing and Sequential scales, as well as the inter-scale correlation between these two.

Table 11: Correlations between Scale Scores on the ILS (n =557)

**Correlations**

| | | active score | sensing score | visual score | sequential score |
|---|---|---|---|---|---|
| active score | Pearson Correlation | 1.000 | .176** | .080 | .069 |
| | Sig. (2-tailed) | . | .000 | .060 | .105 |
| | N | 557 | 557 | 557 | 557 |
| sensing score | Pearson Correlation | .176** | 1.000 | .107* | .323** |
| | Sig. (2-tailed) | .000 | . | .012 | .000 |
| | N | 557 | 557 | 557 | 557 |
| visual score | Pearson Correlation | .080 | .107* | 1.000 | -.086* |
| | Sig. (2-tailed) | .060 | .012 | . | .042 |
| | N | 557 | 557 | 557 | 557 |
| sequential score | Pearson Correlation | .069 | .323** | -.086* | 1.000 |
| | Sig. (2-tailed) | .105 | .000 | .042 | . |
| | N | 557 | 557 | 557 | 557 |

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Construct Validity

In terms of construct validity, the most illuminating in regard to score meaning are studies of differences over time, across groups, and settings[10]. Table 4 compared the means of Active, Reflective, Visual and Sequential scores collected from the consecutive cohorts of engineering students enrolled in ELE639 during the study. ANOVA statistics show no significant differences between the means of the scales in the consecutive years, supporting the construct validity of the ILS.

Next, since several studies were identified that used the ILS to assess learning styles, a comparison of style distributions in those studies is provided in Table 12. As it shows, engineering students are consistently found to be predominantly Active, Sensing, Visual and most are also Sequential. Table 4 and Table 12 thus support *convergent validity* of the ILS scores, as engineering students at different times and in different places share many characteristics hypothesized by the model.

There is also support for *discriminant validity* of the ILS, with significant differences in scores for populations with different characteristics. Van Zwanenberg et al.[17] administered the ILS to 135 engineering student and 145 business students. In ANOVA analysis they found statistically significant differences at 0.05 level between the two populations in the mean scores on Active-Reflective and Sequential-Global scale, and at 0.001 level on the Visual-Verbal scale, with the business students significantly more Verbal, Global and Reflective than engineering students were. Others found significant differences in distributions of learning styles of students and faculty[26, 27, 28]. In this study, the author compared the learning styles of engineering students with the learning styles of engineering faculty at Ryerson University, as shown in Table 13. There were statistically significant differences between the two populations in the mean scores on Active-Reflective, Sensing-Intuitive and Sequential-Global scale, with engineering faculty predominantly Reflective, Intuitive, Visual and Global.

Table 12: Frequencies of Felder Learning Styles Among Engineering Students

| Study | No. | Active | Sensing | Visual | Seq. |
|---|---|---|---|---|---|
| University of Western, Ontario, Canada[13] | 858 | 69% | 59% | 80% | 67% |
| University of Michigan, Michigan[14] | 143 | 67% | 57% | 69% | 71% |
| Tulane University, Alabama[18] | 255 | 60% | 58% | 85% | 50% |
| University of Technology, Kingston, Jamaica[15] | 33 | 55% | 60% | 70% | 55% |
| University of Sao Paulo, Sao Paulo, Brazil[16] | 351 | 60% | 74% | 79% | 50% |
| Newcastle, UK, Van Zwanenberg et al.[17] | 135 | YES* | YES* | YES* | YES* |
| Ryerson, Canada, this study, 2000-2002 | 338 | 61% | 65% | 88% | 63% |

* In their study of 135 engineering students in Newcastle, Van Zwanenberg et al. (2000) indicated that the students were more Active, Sensing, Sequential, and considerably more Visual than Reflective, Intuitive, Verbal and Global. However, they provided only mean scores for scales, instead of percentage distributions.

Table 13: Means, Standard Deviations and ANOVA Results for Comparisons between Students and Professors in ILS Scores

| Population | Sample | Active Score | | Sensing Score | | Visual Score | | Seq. Score | |
|---|---|---|---|---|---|---|---|---|---|
| | | Mean | STD | Mean | STD | Mean | STD | Mean | STD |
| **Students** | 338 | 6.03 | 2.38 | 6.46 | 2.55 | 8.09 | 2.11 | 5.95 | 2.11 |
| **Professors** | 68 | 4.88 | 2.15 | 4.75 | 2.88 | 8.01 | 2.15 | 4.99 | 2.22 |
| ANOVA statistic | | F = 13.603, df= 1, 404, p = 0.000*** | | F = 24.547, df= 1, 404, p = 0.000*** | | F = 0.064, df= 1, 404, p = 0.801 | | F = 11.540, df= 1, 404, p = 0.001** | |

**Satistically significant at 0.01 level, 2-tailed; ***Statistically significant at 0.001 level, 2-tailed.

## IV. Conclusions and Recommendations.

This paper contributes to the ongoing work on the ILS validation. Test-retest analysis of the ILS scores suggested a strong to moderate reliability of all scales. The internal reliability of the scales ranged from 0.53 to 0.70, comparable to the Livesay et al.[18] study, and higher than in the Van Zwanenberg et al.[17] study. Cronbach alpha coefficients met acceptable limits[12] and correlational and factor analysis suggested that the model scales assess separate qualities, as theoretically predicted. There was some evidence of an overlap between Sensing-Intuitive and Sequential-Global domains, also observed in the other studies[17, 18].

While Livesay et al.[18] concluded that the ILS was an appropriate and statistically acceptable tool for characterizing learning preferences, Van Zwanenberg et al.[17] concluded that the ILS scales had low internal reliability and expressed concerns regarding robustness and construct validity of the instrument. The seeming contradiction has more to do with the differing views on the use of the ILS than with the actual results. Livesay et al.[18] used the ILS as a tool for characterizing learning preferences, consistent with the intentions of the model's author, who advocates the use of the model to provide both the students and the instructor with an insight into how they approach the learning/teaching process[1, 2, 3]. This awareness can then be used as a scaffold to expand the range of strategies.

On the other hand, Van Zwanenberg et al.[17] hypothesized that the ILS could be used to predict academic performance and failure rates based on the model theoretical assumptions. Those are that lack of congruence between learning styles of students and the subject matter and prevalent teaching methods affects motivation and may result in poorer performance and higher failure rates[1, 2]. Having found no correlation between the learning styles and academic performance based on cumulative data across approximately 12 courses, nor between the learning styles and failure rates, Van Zwanenberg et al.[17] attributed the failure to confirm the hypothesis to the lack of psychometric robustness of ILS, and raised questions regarding conceptualization of the model. However, they also allowed for a possibility that the broader concept underlying the model, that of the "matching" or supporting different styles by a variety of methods may be correct, insofar as the multiplicity of teachers in the courses provide learners with a variety of learning experiences.

The author believes that this is the central problem with attempts to use the ILS as a predictor of academic performance. Unlike ability, learning styles have weak correlation with achievement[29]. A statistically significant correlation between learning styles and performance based on a cumulative criterion (Grade Point Average) was found by the author[20, 22], most likely because the GPA represented a relatively homogenous, traditional learning environment. The key factor that alleviates differences in performance among different learning style modalities seems to be instructional method[20, 22], rather than individual instructor traits[17]. Since Van Zwanenberg et al.[17] did not provide any discussion of the learning environment, it was impossible to verify whether the "matching" concept was correct.

Multiple studies, with results detailing different samples and data sets are required in order to assess the validity of any particular instrument[10, 25]. This also applies to the ILS, which has not been well researched yet. The literature on psychological assessment points out that to base decisions with far reaching consequences, such as predicting success rates, restricting valid learning opportunities, counselling, etc., on a little-researched instrument may be misguided and potentially unfair[25, 30]. Thus more studies are needed first in the discussion of the reliability of the ILS. Further work on the internal reliability of the instrument as suggested[18], may also address the issue of the relative weakness of the Sequential scale. Since guiding good teaching practice was the primary objective of the model, Felder[19] argues that as long as the inter-scale correlation is not significant, at which point the two scales become redundant, they lead to separate implications about what constitutes good teaching, and thus the model is acceptable. This is an acceptable argument, as long as the use of the ILS is for the general guidance only.

Within such context of assisting good teaching practice, validation of the ILS should rely more heavily on its *construct validity*. Style distributions, mean values and standard deviations show enduring similarities between different cohorts of engineering students in the program over time, supporting convergent construct validity. The style distributions in the study were also consistent with several other studies of engineering students where the ILS was used[13, 14, 15, 15]. There is also evidence of discriminant construct validity, with the ILS showing significant differences between the student sample population and the instructor sample population. More support for construct validity is provided by the previous study[20], where there were significant differences in distributions of learning styles between students at higher and lower levels of academic ability in a traditional teaching environment. In that study[20], students with learning styles not congruent with the traditional college teaching method were found disproportionately concentrated in the previously lower-achieving category, PBM. This observation supports assertions of the Felder Model that such students may be disenfranchised and more at risk of failure[1, 2].

Finally, while longer questionnaires such as MBTI and Kolb's LSI typically yield higher Cronbach's alpha measures for collected data, based on author's experiences, their usefulness in a classroom setting may be limited. The author observed that any voluntary survey that took longer than 10 minutes were much less likely to be completed and returned, by students and faculty alike. As well, when the Kolb's LSI I was administered, on a trial basis, together with the Felder-Soloman LSI to students in the 2000 study[20], many kept asking questions regarding the meaning of the words they were supposed to rank. It happened again in 2001. Moreover, many, instead of ranking words, simply chose one, despite repeated explanations of instructions. This suggested that the students were having trouble understanding the wording used in the questionnaire, making any subsequent results questionable. This may be specific to the demographic sample of students in the study. However, should such observations be typical of other engineering students, the clarity of the ILS may help explain in part its popularity. The web-based, self-scoring version of the questionnaire (at: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSpage.html) gets approximately 100,000 hits per year and has been translated into several languages.

In conclusion, this paper concurs with Livesay et al.[18] that the ILS is a suitable psychometric tool to assess the learning styles of engineering students. However, work on its evaluation should continue.

Bibliography


1. Felder, R.M. & Silverman, L.K., Learning and Teaching Styles in Engineering Education, Journal of Engineering Education, Vol. 78, No. 7, pp. 674-681, (1988).
2. Felder, R. M., Matters of Style, AEEE Prism, December Issue, pp. 18-23, (1996).
3. Felder, R.M. & Soloman, B.A., Index of Learning Styles Questionnaire, North Carolina State University, (2001). Online at: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS-a.htm
4. Felder, R.M. Richard Felder's Responses to Frequenly Asked Questions about the ILS, (2001). Online at: http://www2.ncsu.edu/unity/lockers/users/f/felder/public/ILSdir/ILS-faq.htm
5. Felder, R.M., Felder, G.N. & Dietz, E.J., A Longitudinal Study of Engineering Student Performance and Retention. V. Comparisons with Traditionally taught Students, Journal of Engineering Education, Vol. 87, No. 4, pp. 469-480, (1998).
6. Johnson, D.W., Johnson, R.T. & Smith, K.A., Active Learning: Cooperation in the College Classroom, Edina, MN: Interaction Book Company, (1991).
7. Smith, K.A. & Waller, A., Afterward: New Paradigms for College Teaching. In: W. Campbell & K. Smith (Eds.), New Paradigms for College Teaching, Edina, MN: Interactions Book Company, (1997).
8. Wankat, P.C., Felder, R.M., Smith, K.A., & Oreovicz, F.S., The Scholarship of Teaching and Learning Engineering, In: M.T. Huber & S. Morreale (Eds.), Disciplinary Syles in the Scholarship off Teaching and Learning: Exploring the Common Ground, Chapter 11. AAHE/Carnegie Foundation for the Advancement of Teaching, Washington, DC, (2002).
9. Trochim, W., Research Methods Knowledge Base, 2nd ed. Cornell University, (1999). Online at: http://trochim.human.cornell.edu/kb/.
10. Messick, S., Validity of Psychological Assessment: Validation of Inferences from Person's Responses and Performances as Scientific Inquiry into Score Meaning; American Psychologist, Vol. 50 (9), pp. 741-749, (1995).
11. Nunnaly, J., Psychometric Theory, New York: McGraw-Hill, (1978).
12. Tuckman, B.W., Conducting Educational Research, 5th edition, Belmont, CA: Wadsworth Group, (1999).

13. Rosati, P., Specific Differences and Similarities in the Learning Preferences of Engineering Students, Session 12c1, Proceedings of the 29th ASEE/IEEE Frontiers in Education Conference, San Juan, Puerto Rico, (1999).
14. Montgomery, S., Addressing Diverse Learning Styles through the Use of Multimedia: Material and Energy Balance, Proceedings of the 25th ASEE/IEEE Frontiers in Education Conference, Atlanta, GA, USA, (1995).
15. Smith, N.G., Bridge, J. & Clarke, E., An evaluation of students' performance based on their preferred learning styles. In: Z.J. Pudlowski (Ed.), Proceedings of the 3rd UNESCO International Centre for Engineering Education (UICEE) Global Congress on Engineering Education, Glasgow, Scotland, UK, pp. 284-287, (2002).
16. Kuri, N.P. & Truzzi, O.M.S., Learning Styles of freshmen Engineering Students, Proceedings of the 2002 International Conference on Engineering Education (ICEE), Manchester, UK, (2002).
17. Van Zwanenberg, N., Wilkinson, L J., & Anderson, A., Felder and Silverman's Index of Learning Styles and Honey and Mumford's Learning Styles Questionnaire: How do they compare and do they predict academic performance? Educational Psychology,Vol. 20 (3), pp. 365-381, (2000).
18. Livesay, G., Dee, K., Felder, R., Hites, L., Nauman, E. & O'Neal, E., Statistical Evaluation of the Index of Learning Styles, Session 2430, 2002 ASEE Annual Conference and Exposition, Montreal, Quebec, Canada, (2002).
19. Felder, R.M., Hoechst Celanese Professor at North Carolina State University, Raleigh, North Carolina, Email communication, October 27, (2002).
20. Zywno, M.S., Instructional Technology, Learning Styles and Academic Achievement, Proceedings of 2002 ASEE Annual Conference and Exposition, Session 2422, Montreal, Quebec, Canada, (2002).
21. Zywno, M.S., Engineering Faculty Teaching Styles and Attitudes toward Student-Centered and Technology-Enabled Teaching Strategies, Accepted to be presented at the 2003 ASEE Annual Conference and Exposition, Nashville, TN, (2003).
22. Zywno, M.S., & Waalen, J. K., Analysis of Student Outcomes and Attitudes in a Technology-enabled Control Education at Ryerson - a Case Study, Global Journal of Engineering Education, Vol. 5, No.1, pp. 49-56, Australia, (2001). Online at: http://www.eng.monash.edu.au/uicee/gjee/vol5no1/Zwyno&Waalen.pdf
23. Gregorc, A.F., Learning/Teaching Styles: Potent forces behind them, Educational Leadership, Vol. 36, pp. 234-236, (1979).
24. Dunn, R., Giggs, S.A., Olson, J., Beasley, M. & Gorman, B.S., A meta-analytic validation of the Dunn and Dunn Model of learning preferences, The Journal of Educational Research, Vol. 88 (6), pp. 353-362, (1995).
25. Thompson, B. & Vacha-Haase, T., Psychometrics *Is* Datametrics: The Test Is Not Reliable, Educational and Psychological Assessment, Vol. 60 (2), pp. 174-195, (2000).
26. Montgomery, S. & Groat, L.N., Student Learning Styles and Their Implications for Teaching, Occasional Paper, Center for Research on Learning and Teaching, University of Michigan, (1998). Online at: http://www.crit.umich.edu/occ10.html.
27. Nulty, D. & Barrett, M. Transitions in Students Learning Styles, Studies in Higher Education, Vol. 21 (3), pp. 333-346, (1996).
28. Wetzel, K.C. & Harmeyer, K.M., Success in Low-Level and High-Level Mathematics Courses in Undergraduate Engineering University as a Correlate to Individual Learning Style, Proceedings of the 27th ASEE/IEEE Frontiers in Education Conference, Pittsburgh, PA, USA, (1997).
29. Ayersman, D.J., Reviewing the Research on Hypermedia-Based Learning, Journal of Research on Computing in Education, Vol. 28 (4), pp. 500-525, (1996).
30. Yates, G.C., Cognitive Styles and Learning Strategies: Understanding Style Differences in Learning and Behaviour/Thinking Styles, Educational Psychology, Vol. 19 (1), pp. 103-105, (1999).

MALGORZATA S. ZYWNO

Malgorzata Zywno is a Professor in the Department of Electrical and Computer Engineering at Ryerson University in Toronto, Canada. She received her M.Eng. Degree in Electrical Engineering from the University of Toronto in 1990 and is currently in a doctoral program at Glasgow Caledonian University in Glasgow, Scotland, UK, expecting to complete her degree in Spring 2003. Her teaching and research interests include control systems, system identification, recruitment and retention strategies for women in engineering and more recently, technology-aided pedagogy. In 2002 Professor Zywno received the 3M Teaching Fellowship, the highest award for academic teaching in Canada. She has also won nine international awards and citations for her research and publications on the issues of technology-mediated teaching, including 2002 ASEE Annual Conference PIC V Best Paper Award and Best Paper Overall Award, iNEER Achievement Award, and the UICEE Silver Badge of Honour. Professor Zywno is a member of the ASEE, of the IEEE, and a registered Professional Engineer in the Province of Ontario, Canada.