DOCUMENT RESUME

ED 456 165 TM 033 228

AUTHOR Wiersma, William
TITLE Continuous School Improvement Questionnaire Field Test. Technical Report.
INSTITUTION AEL, Inc., Charleston, WV.
SPONS AGENCY Office of Educational Research and Improvement (ED), Washington, DC.
PUB DATE 2001-04-00
NOTE 38p.
CONTRACT ED-01-CO-0016
AVAILABLE FROM AEL, Inc., P.O. Box 1348, Charleston, WV 25325. Tel: 800-624-9102 (Toll Free).
PUB TYPE Reports - Research (143)
EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Educational Improvement; Elementary Secondary Education; Field Tests; *Institutional Characteristics; Pilot Projects; Questionnaires; Reliability; *Teachers; Test Construction; Validity

ABSTRACT

The Continuous School Improvement Questionnaire (CSIQ) is a comprehensive inventory measuring educators' perceptions of factors that affect success with school improvement. A pilot test of the CSIQ was conducted in Spring 2000, one purpose of which was to reduce the length of the CSIQ. The instrument was reduced from 147 to 72 items. The pilot test was based on data from 274 educators, but the extensive field test reported in this monograph gathered data from 2,093 educators, primarily teachers, from 79 schools. The numbers of scores for subscales and the total scale varied, but in all cases they were less than 2,093. Pilot test results indicate that the 72-item CSIQ is highly reliable, both the total inventory and the individual subscales. For about 3 weeks, the CSIQ has moderate to high stability reliability. Mean scores on the CSIQ subscales are highly consistent and positioned about 70% of the distance from the minimum to the maximum scores. The type (level) of school appears to have a slight to modest effect on subscale and total score performance, with subjects in elementary school having the greater performance. Females tend to score higher than males, but the effect is at least partially confounded with type of school. Subjects in schools known to be in a mode of continuous improvement score higher on subscales and total score than their counterparts in the same type of schools. When considering continuous school improvement as a concept measured by the CSIQ, there appear to be six underlying constructs, which closely coincide with the six subscales of the CSIQ. (Contains a 21 references.) (SLD)

# *Technical Report:*

# *Continuous School Improvement Questionnaire Field Test*

**Dr. William Wiersma**

**April 2001**




**AEL, Inc.**
**P. O. Box 1348 • Charleston, WV 25325 • 800-624-9120**

*Technical Report:*

# *Continuous School Improvement Questionnaire Field Test*

by
Dr. William Wiersma

April 2001

AEL, Inc.
P. O. Box 1348
Charleston, WV 25325

AEL is a catalyst for schools and communities to build lifelong learning systems that harness resources, research, and practical wisdom. AEL serves as the Regional Educational Laboratory for Kentucky, Tennessee, Virginia, and West Virginia. For these same four states, it operates the Eisenhower Regional Consortium for Mathematics and Science Education. The Region IV Comprehensive Center at AEL serves North and South Carolina in addition to these states. AEL also operates the ERIC Clearinghouse on Rural Education and Small Schools.

Information about AEL projects, programs, and services is available by writing or calling AEL.


Post Office Box 1348
Charleston, West Virginia 25325-1348
304-347-0400
800-624-9120
304-347-0487 (fax)
aelinfo@ael.org
http://www.ael.org

This publication is based on work sponsored wholly or in part by the Office of Educational Research and Improvement, U.S. Department of Education, under contract number ED-01-CO-0016. Its contents do not necessarily reflect the views of OERI, the Department, or any other agency of the U.S. government.

## CONTENTS

## LIST OF TABLES



## LIST OF FIGURES

## INTRODUCTION

The Continuous School Improvement Questionnaire (CSIQ) is a comprehensive inventory measuring educators' perceptions of factors that affect success with school improvement. A pilot test of the CSIQ was conducted in Spring 2000 (Wiersma, 2000), one purpose of which was to reduce the length of the CSIQ. This was done, and the CSIQ was reduced from 147 to 72 items, 12 items for each of the 6 subscales. Reducing the number of items did not affect adversely the internal consistency reliability of the subscales or the total score. These reliability estimates were consistently high with the lowest coefficient of .89 for one of the subscales.

The pilot test was based on data from 274 educators, whereas the extensive field test gathered data from 2,093 educators, primarily teachers, who will be referred to as the subjects of the field test. Data were collected in the Fall of 2000 from faculty members in 79 schools. In general, a field test following a pilot test is completed to establish the instrument and to investigate performances on the instrument as related to variables that exist in the natural, educational setting. The purposes of the field test were

1. to determine the status of performance on the CSIQ, its subscales and total score, for a large base of subjects
2. to investigate possible differences in performance by type of school for the subjects
3. to check the consistency of reliability estimates for the subscales and total scores with estimates established by the pilot test data
4. to determine the relationships among the CSIQ subscale scores and total score, and their relationships to scores on the School Climate Questionnaire (SCQ), an inventory about perceptions of the school, many of which at least on face value appear to be related to school improvement

5. to obtain an estimate of stability reliability for the CSIQ from a test-retest of a subset of the subjects

6. to investigate as appropriate the effects of any other variables upon the CSIQ scores

The CSIQ has six subscales, appropriately named, reflecting the content of the subscale items. The subscales are listed below and, to facilitate their identification in the report and its tables, they are designated by assigned subscripts.

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART[1] Learner

The items of the CSIQ are 6-point rating scale items. The response options range from 1 being "Is not present" to 6, "Is present to a high degree." The higher the response, the greater the score. Scores are generated across the items of the subscales and then totaled. Thus, subscale scores may range from 12 to 72, and the total score may range from 72 to 432.

[1]SMART is the acronym for Successful, Motivated, Autonomous, Responsible, and Thoughtful, and was one of the major elements of Quest, a four-year R&D project at AEL, Inc. (AEL, 1995 & 1998). A large number of publications from the Quest project have been published by AEL, including 12 evaluation reports, three research reports, a case study of one popular Quest process, four summative school case studies, and a final lessons learned report. See the Bibliography for the citations for these Quest publications.

## RESULTS ACROSS ALL SUBJECTS

CSIQ scores on 2,093 subjects provide an extremely large database for a field test. This, along with the type of inventory used to measure the school improvement variables, imposes certain characteristics on the database. Subjects will on occasion omit items, for such possible reasons as being unsure about a response and failing to return to the item. In order to compute total scores and to estimate internal consistency reliability, there must be responses to all items of a subscale or total scale. As a result, the numbers of scores for subscales and the total scale varied, but in all cases they were less than 2,093. As expected, the total scale had the fewest number of scores.

Subjects were generally coded according to type of school (elementary, middle, or high school). However, four schools did not fit the typical definitions of elementary, middle, or high school: K-12, PK-12, 6-12, and 9$^{th}$ grade only. Throughout this report, these schools are designated "other." For whatever reason, some schools wanted to remain anonymous and were unidentifiable as to type. Scores for subjects from these schools could be included in the overall analysis, but they could not be included in any analysis requiring type of school identification.

Four schools were identified as being in a continuously improving mode, thus they were designated as "known" schools. These schools reflected the six major components of Quest and their inclusion allowed comparisons with schools of their types. The "known" schools included three elementary and one private (Catholic) high school.

With the large number of scores, statistics generated were very stable; i.e., standard errors were very small. This field test does not meet all criteria for using inferential statistics in a

classical sense. This was not a situation of random sampling from some larger population. Yet, it was useful to compute analyses of variance and t-tests in order to compare within and between group variances and to note the relative positioning of group means. As expected with large numbers of scores, small differences in means were statistically significant. It should not be inferred that statistical significance reflects practical importance.

The descriptive statistics for the CSIQ subscales and total scale across all subjects are given in Table 1. The means for the subscales were quite homogeneous, being within a range of slightly more than 5 points on a scale that has a possible 60-point spread. This homogeneity is illustrated in the profile of Figure 1. The coefficients of variation (the standard deviation expressed as a percent of the mean) tended to be small for the subscales, ranging from about 15% to 22%. These coefficients of variation were slightly smaller than those of the pilot test for either the 15-item or the 12-item subscales.[2] The total score showed one of the lowest coefficients of variation at slightly more than 15%. Internal consistency reliability coefficients were high, all being greater than .90.

[2]See Table 7, page 17, *Technical Report: Continuous School Improvement Questionnaire Pilot Test.*

**Table 1**

**Frequencies (N), Means (M), Standard Deviations (SD), and Cronbach Alpha Reliability Estimates ($r_\alpha$) for CSIQ Subscales and Total**

| Subscale | N | M | SD | $r_\alpha$ |
|---|---|---|---|---|
| $S_1$ | 1,728 | 56.2 | 8.7 | .91 |
| $S_2$ | 1,725 | 54.8 | 10.5 | .94 |
| $S_3$ | 1,701 | 55.4 | 12.3 | .96 |
| $S_4$ | 1,666 | 55.6 | 10.5 | .94 |
| $S_5$ | 1,687 | 56.6 | 10.1 | .94 |
| $S_6$ | 1,743 | 59.7 | 9.1 | .96 |
| Total | 1,516 | 337.9 | 51.6 | .98 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

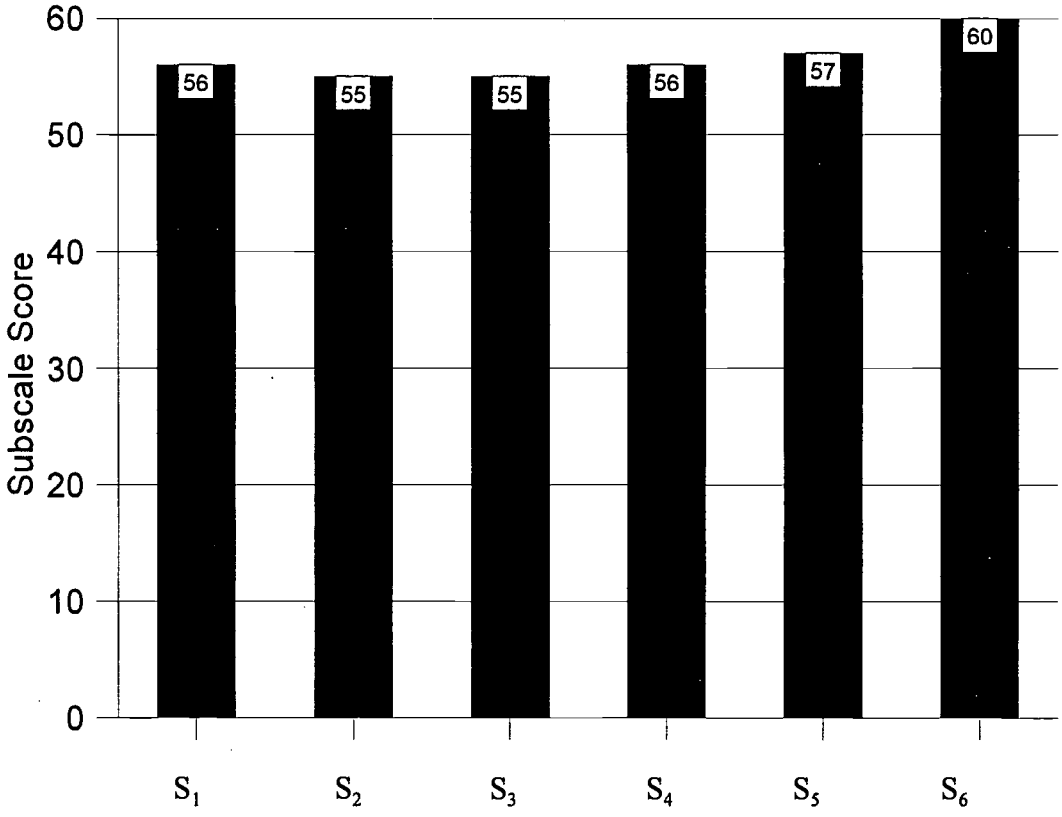


**Figure 1**

**Profile of CSIQ Subscale Means for the Field Test**

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

Note: The number at the top of each subscale bar has been rounded to a whole number.

## RESULTS BY TYPE OF SCHOOL

The data were analyzed according to type of school as defined earlier, and the scores from unidentified schools and the "known" schools were not included in this analysis. The "known" schools were omitted because they scored significantly higher (as discussed below) than their counterparts in the greater population of schools. Analyses of variance (ANOVA) were computed for the subscales and total score, with type of school (4 levels) as the independent variable. For reasons given earlier, the numbers of scores varied across the subscales, and Table 2 contains the score frequencies.

As expected, the F-ratios from the seven ANOVAs were statistically significant for the type of school effect, all beyond the .001 level of significance. The means for the subscales and total score by type of school are given in Table 3. An inspection of these means shows that the elementary subjects had the highest means for all subscales and the total score. Subsequent post hoc tests for locating the source of statistical significance showed that the elementary means were significantly greater than those of the remaining three types of schools for the total score and all subscale scores except $S_3$. For $S_3$ the elementary and middle school means were significantly greater than the high school and "other" school means.

There were four "known" schools defined above as schools in a continuously improving mode. Three were elementary schools, and one was a high school. The scores from subjects in the "known" schools were compared with those from subjects within the same type of school. Independent t-tests were then computed to test for differences between the subscale and total score means. The means are given in Table 4. All t-tests showed statistically significant

**Table 2**

**Frequencies of Subjects by Type of School for CSIQ Subscales and Total Score**

| | Type of School | | | | |
|---|---|---|---|---|---|
| **Subscale** | **Elementary** | **Middle** | **High** | **Other** | **Total** |
| $S_1$ | 538 | 175 | 361 | 119 | 1,193 |
| $S_2$ | 539 | 174 | 360 | 114 | 1,187 |
| $S_3$ | 534 | 165 | 361 | 111 | 1,171 |
| $S_4$ | 511 | 170 | 350 | 112 | 1,143 |
| $S_5$ | 533 | 167 | 357 | 107 | 1,164 |
| $S_6$ | 547 | 176 | 366 | 117 | 1,206 |
| Total | 476 | 152 | 317 | 99 | 1,044 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

**Table 3**

**Means for the CSIQ Subscales and Total Score by Type of School**

| | Type of School | | | |
|---|---|---|---|---|
| **Subscale** | **Elementary** | **Middle** | **High School** | **Other** |
| $S_1$ | 58.2 | 55.6 | 54.2 | 54.0 |
| $S_2$ | 56.5 | 52.7 | 52.3 | 54.5 |
| $S_3$ | 57.1 | 56.3 | 53.1 | 53.8 |
| $S_4$ | 57.9 | 54.2 | 53.5 | 53.0 |
| $S_5$ | 59.1 | 55.6 | 54.9 | 53.8 |
| $S_6$ | 61.1 | 59.2 | 58.3 | 58.8 |
| Total | 349.6 | 331.6 | 326.9 | 326.6 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

**Table 4**

**CSIQ Subscale and Total Score Means for "Known" (K) Schools Within Elementary and High School Types**

| | Type of School | | | |
|---|---|---|---|---|
| **Subscale** | **(K) Elementary** | **Elementary** | **(K) High School** | **High School** |
| $S_1$ | 61.1 | 58.2 | 62.3 | 54.2 |
| $S_2$ | 62.7 | 56.5 | 63.0 | 52.3 |
| $S_3$ | 57.6 | 57.1 | 61.8 | 53.1 |
| $S_4$ | 61.7 | 57.9 | 58.8 | 53.5 |
| $S_5$ | 63.9 | 59.1 | 59.3 | 54.9 |
| $S_6$ | 64.3 | 61.1 | 65.4 | 58.3 |
| Total | 370.6 | 349.6 | 367.7 | 326.9 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

differences except the means for elementary on $S_3$. For all comparisons, the "known" group had the greater means.

**Schools Nested Within Type of School**

The analyses for type of school showed type to be a significant effect with the elementary subjects having the high means. For a further breakdown of the data, nested ANOVAs were computed with school being an effect nested within type of school; i.e., school does not cross type of school because any one school is of only one type. All of the F-ratios for school nested within type from the ANOVAs were statistically significant (at $\alpha = .001$). These ANOVA results show that not only were there differences among types of schools, but there was great variability among schools within a type.

In the nested analyses there were data from subjects in 57 schools, the majority of which (35) were elementary schools. There were data from subjects in 6 middle schools, 12 high schools, and 4 "other" schools. The number of subjects within a school varied greatly, from 1 to slightly over 50. The average number of subjects was slightly under 23 per school.

With so many schools, inspecting all the means and their relative positioning would be very cumbersome. Table 5 contains the minimum and maximum school means by type of school for the CSIQ subscales and total score. For most subscales and the total score, there were some wide ranges for the extreme means. However, most of these extreme means, especially on the minimum end, were outliers. The extreme means tended to be based on data from few subjects and, in some cases, only one. For those, the entire school was represented by the data from one subject.

**Table 5**

**Minimum and Maximum School Means for the CSIQ Subscales and Total Score by Type of School**

| | Type of School | | | |
|---|---|---|---|---|
| **Subscale** | **Elementary** | **Middle** | **High School** | **Other** |
| $S_1$ | 47.3 - 66.5 | 52.3 - 57.4 | 35.0 - 62.3 | 49.1 - 56.3 |
| $S_2$ | 27.0 - 67.6 | 45.3 - 60.5 | 37.7 - 63.0 | 49.7 - 59.4 |
| $S_3$ | 25.0 - 68.0 | 54.8 - 59.2 | 14.0 - 61.8 | 45.5 - 56.3 |
| $S_4$ | 36.0 - 69.2 | 48.1 - 57.8 | 42.8 - 58.8 | 45.5 - 56.3 |
| $S_5$ | 43.0 - 69.4 | 52.8 - 58.6 | 17.0 - 59.3 | 45.6 - 57.1 |
| $S_6$ | 49.0 - 68.5 | 55.5 - 62.2 | 31.0 - 65.4 | 55.0 - 60.6 |
| Total | 240.6 - 405.1 | 311.8 - 344.5 | 259.8 - 367.7 | 298.6 - 346.4 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

**Schools by Johnson Codes**

Another variable by which schools were analyzed was the Johnson Codes, a classification system of seven categories that orders schools from the most urban locale to the most rural locale.[3] ANOVAs were computed using Johnson Codes as the independent variable. The number of subjects per locale varied considerably, ranging from 32 in the most urban locale to 550 in the most rural locale.

The means for the CSIQ subscales and total score by Johnson locale code are given in Table 6. As expected, the results of the ANOVAs had all F-ratios for locale effect statistically significant at the .01 level of significance. Subsequent post hoc tests for identifying the source of the significance gave some consistent and inconsistent patterns in the positioning of the means. Those results were as follows by subscale and total score. Locales are simply given by number.

$S_1$: 1 greater than 4, 7, 3, 2, and 5; 6 greater than 4
$S_2$: 1 greater than 2, 7, 4, 3, and 5; 6 greater than 2 and 7
$S_3$: 1 greater than 5, 4, 2, 3, and 7; 6 greater than 5
$S_4$: 1 greater than 2, 7, 4, and 3
$S_5$: 1 greater than 2, 3, 7, 4, and 6
$S_6$: 1 greater than 4, 7, 2, and 3
Total: 1 greater than 2, 4, 7, 3, and 5

Code 1 schools, the most urban, had the greatest mean on all the measures, and only for one subscale, $S_5$, was that mean significantly greater than the Code 6 mean, Code 6 being next from the most rural. Above, the codes following "greater than" are listed in ascending order as to their means. Although Code 2 had the lowest mean for 4 of the 7 measures, the order of those lower means is not consistent. The order of magnitude of the means does not follow the

[3]The Johnson Codes comprise an ordinal system of categories classifying schools from the most urban (1) to the most rural (7). The source is National Center for Education Statistics (2000).

## Table 6

## CSIQ Subscale and Total Score Means by Johnson Locale Code

| | Johnson Locale Code | | | | | | |
|---|---|---|---|---|---|---|---|
| **Subscale** | **1** | **2** | **3** | **4** | **5** | **6** | **7** |
| $S_1$ | 61.3 | 56.4 | 56.3 | 54.8 | 57.3 | 59.1 | 56.1 |
| $S_2$ | 60.8 | 53.3 | 55.5 | 55.5 | 56.5 | 59.2 | 53.7 |
| $S_3$ | 60.7 | 54.8 | 54.9 | 53.9 | 53.0 | 59.0 | 55.6 |
| $S_4$ | 61.2 | 54.8 | 56.3 | 55.7 | 58.2 | 58.5 | 55.0 |
| $S_5$ | 62.1 | 56.1 | 56.2 | 57.5 | 59.8 | 58.3 | 56.8 |
| $S_6$ | 64.5 | 60.1 | 60.8 | 58.9 | 62.0 | 61.9 | 59.4 |
| Total | 311.0 | 334.1 | 337.6 | 336.2 | 344.3 | 355.5 | 336.5 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

extent of rurality except that the most urban code consistently had the greatest mean, and that Code 6, one of the more rural codes, was positioned either second or third after Code 1 on the measures. Code 5 means fluctuated from lowest ($S_3$) to being second to Code 1 ($S_5$ and $S_6$).

**Gender Effect**

The subjects were identified by gender. However, gender is at least partially confounded with type of school because the gender split is not equal or proportional across type of school. The majority of elementary school teachers, for example, is female. Independent t-tests were computed, testing the difference between the means of male and female subjects. The numbers of subjects and their means are given in Table 7.

The results of the t-tests indicated all differences between means except the difference between the means of $S_3$ to be statistically significant at the .01 level of significance. This result was not surprising considering the statistical power in these tests. (Even the test for total score, with the fewest numbers, had 1,460 degrees of freedom.) The frequencies show that the majority of the subjects, about 80%, was female. The females consistently had the high means.

**Table 7**

**Frequencies (N) and Means (M) for CSIQ Subscales and Total Score by Gender**

| | Female | | Male | |
|---|---|---|---|---|
| **Subscale** | **N** | **M** | **N** | **M** |
| $S_1$ | 1,330 | 56.8 | 329 | 54.3 |
| $S_2$ | 1,331 | 55.2 | 327 | 53.2 |
| $S_3$ | 1,309 | 55.6 | 322 | 55.2 |
| $S_4$ | 1,281 | 56.3 | 319 | 53.5 |
| $S_5$ | 1,301 | 57.3 | 319 | 54.6 |
| $S_6$ | 1,343 | 60.2 | 329 | 58.1 |
| Total | 1,167 | 341.2 | 295 | 327.6 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

## RESULTS OF CORRELATIONAL ANALYSES

Whenever a study involves a number of variables operating in an educational context, the relationships among the variables are of interest. Of course, correlation coefficients reflect the extent of relationships among variables. For most variables, such as those measured by the subscales of the CSIQ, it is extremely unlikely to find independence among the variables, so the magnitude of correlation coefficients is of more relevance than whether or not they reflect independence. With the large number of scores in this field test, practically any correlation coefficient other than zero should be statistically significant.

### Correlations Among the CSIQ Subscales and Total Score

The correlation coefficients among the CSIQ subscales and the total score are given in Table 8. The greatest correlations were between the total score and the individual subscales. This was expected because one-sixth of the total score consists of the scores on any one subscale.

The correlations among the subscales were quite consistent, with the lowest being .51 and the greatest .81. However, these two coefficients were outliers in that there were no other coefficients less than .61, and the next greatest was .74. The coefficients indicate that most subscales in combinations of two have about 40 to 50 percent common variance. The greatest correlation was between $S_4$, Shared Goals for Learning, and $S_5$, Assessing Student Learning. The lowest correlation was between $S_3$, Sharing Leadership, and $S_6$, Enabling the SMART Learner.

**Table 8**

**Correlation Coefficients Among CSIQ Subscales and Total Score**

| Subscale | $S_2$ | $S_3$ | $S_4$ | $S_5$ | $S_6$ | Total |
|---|---|---|---|---|---|---|
| $S_1$ | .71 | .61 | .70 | .70 | .72 | .85 |
| $S_2$ | | .68 | .72 | .73 | .65 | .87 |
| $S_3$ | | | .74 | .63 | .51 | .83 |
| $S_4$ | | | | .81 | .64 | .90 |
| $S_5$ | | | | | .70 | .89 |
| $S_6$ | | | | | | .81 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

Overall, the magnitudes and patterns of the correlation coefficients were similar to those of the pilot test.[4] In fact, some of the corresponding coefficients were identical or within one or two one-hundredths of each other. These results show the stability of the relationships across different groups of educators among the variables measured by the CSIQ.

**A Measure of Stability Reliability**

Throughout its development, the internal consistency reliability of the CSIQ subscales and total score has remained high, actually very high, as already reported. In order to obtain a measure of stability reliability, a group of subjects was measured twice, allowing about a three-week interval between the "test" and the "retest." The total numbers of subjects for which there were test-retest scores and the correlations between these scores are given in Table 9.

The test-retest correlations ranged from .65 to .80, which shows two results: (1) considerable stability across time for the CSIQ measures and (2) similar stability across the subscales. The Pearson-Product Moment correlation coefficient is a measure of the relative positioning of the scores being correlated and the magnitude of the differences between these scores on the two variables being correlated. As additional analyses of the test-retest scores, dependent t-tests were computed for the differences between the test-retest means. Even with the large numbers of scores (giving a lot of statistical precision), only one difference (for $S_1$) was statistically significant at the .05 level of significance. However, the difference in means was only .62. $S_4$, which had one of the larger standard deviations of the subscales, had a difference of .65 between the test and retest means, this difference being statistically significant at the .10 level.

---

[4]See Table 8, page 19, *Technical Report: Continuous School Improvement Questionnaire Pilot Test.*

**Table 9**

**Frequencies (N) and Correlation Coefficients (r) for Test-Retest of the CSIQ Subscales and Total Score**

| Subscale | N | r |
|---|---|---|
| $S_1$ | 470 | .70 |
| $S_2$ | 469 | .71 |
| $S_3$ | 459 | .80 |
| $S_4$ | 452 | .73 |
| $S_5$ | 458 | .69 |
| $S_6$ | 464 | .65 |
| Total | 384 | .77 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

The .65 difference between the test-retest means was the largest difference for the subscale means. One pair of means, those for $S_3$, had a difference of only .01. For all the subscales except $S_6$, the retest mean was the greater of the pair (however slightly), and for $S_6$, the difference was only .08. The difference in total score means was 1.68, the retest mean being the greater.

**Concurrent Validity**

The School Climate Questionnaire (SCQ) (Manning, Curtis, & McMillen, 1996) is a 10-item inventory intended to measure subjects' perceptions of the school. The items are somewhat complex, but on the surface it seems that item content is related to factors involved in school improvement. In the pilot test, the SCQ served as a measure of concurrent validity for the CSIQ. The SCQ was known to be highly reliable, and the CSIQ had a .76 validity coefficient with the SCQ.[5]

In the pilot test, the concurrent validity coefficient was computed across all subjects for the total CSIQ score, but for the field test the coefficients were computed for the CSIQ subscales and total score within the type of school, using the four school categories. This analysis provided more detailed information about possible relationships between the SCQ and CSIQ measures. The field test had many more subjects than the pilot test, enabling the correlation coefficients to remain stable within the breakdowns. The coefficients are given in Table 10.

---

[5] See pages 16 and 18, *Technical Report: Continuous School Improvement Questionnaire Pilot Test.*

**Table 10**

**Correlation Coefficients Between the SCQ and the CSIQ Subscales and Total Score by Type of School**

| | Type of School | | | |
|---|---|---|---|---|
| **Subscale** | **Elementary** | **Middle** | **High School** | **Other** |
| $S_1$ | .47 | .52 | .43 | .43 |
| $S_2$ | .54 | .45 | .64 | .65 |
| $S_3$ | .67 | .62 | .76 | .79 |
| $S_4$ | .58 | .59 | .70 | .73 |
| $S_5$ | .49 | .60 | .63 | .65 |
| $S_6$ | .40 | .24 | .38 | .48 |
| Total | .68 | .66 | .75 | .74 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

In order to be included in the computation of a correlation coefficient, subjects were required to have total scores for the measures (i.e., there could be no omitted items). Thus, the numbers of paired scores included in the correlation coefficients varied, but for most the numbers were in the hundreds, some as high as the low 300s. An internal consistency of reliability estimate was computed for the SCQ and that coefficient was .97.

The validity coefficients of Table 10 show considerable variability. The coefficients for the CSIQ total scores for the high schools and "others" were very close to the coefficient of .76 found in the pilot test. The elementary and middle school coefficients for the CSIQ total score were slightly less, but these results were quite consistent with those of the pilot test.

Even with the variability, there were some patterns among the validity coefficients. Relatively, $S_3$ and $S_4$ had consistently high coefficients across the type of school, and $S_1$ and $S_6$ had consistently low coefficients. The coefficients for $S_2$ and $S_5$ were sort of "in the middle."

There were some differences by type of school and, again, these were relatively consistent. The high school and "other" groups had very similar patterns and, overall, the higher coefficients except for $S_1$. The elementary and middle groups generally had lower coefficients than the high school and "other" groups, and, except for $S_6$, their patterns were quite similar.

The validity coefficients ranged from .24 to .79, actually a wide range, although there were only two coefficients less than .43, both for $S_6$. This result shows that the SCQ content is quite closely related to some factors measured by the CSIQ, especially $S_3$, Sharing Leadership, and $S_4$, Shared Goals for Learning. However, the SCQ content has considerably less in common with the CSIQ on other factors, particularly $S_6$, Enabling the SMART Learner, and $S_1$, Learning Culture. Another noteworthy result was the similarity in patterns for the high school and "other" school group.

**Factor Analysis of the 72-Item CSIQ**

In this field test, the concept of continuous school improvement was defined operationally by the 72 items of the CSIQ. When constructing the CSIQ, it seemed logical to generate items for six subscales that have face validity in representing the kinds of constructs involved in continuous school improvement.

Factor analysis is an analytical technique that generates artificial variables (factors) representing the one or more constructs measured by the entire inventory or test, in this case the 72 items of the CSIQ. Although factors are artificial variables, they must be defined or described in terms of the variables (72 items) on which they are based. Factor loadings, correlations between the scores on individual items and the factors, serve this purpose. Thus, a high, positive factor loading indicates that an item contributes extensively to the composition of the factor.

A desirable outcome of factor analysis is to have as many noteworthy factors as there are logical constructs underlying the concept under study. For the field test of the CSIQ with six subscales, it would be desirable to have six factors. It should be noted that the factor analysis is of the item scores, not the scores of the subscales. Because of the high intercorrelations of the subscale scores, a factor analysis of those scores would result in only one general factor.

The factor analysis computed was a Principal Axis, Rotation Varimax factoring. This is an orthogonal rotation, which means that the factors extracted are uncorrelated (independent). The factor analysis extracted eight factors (using the conventional criterion of eigenvalues of 1.0 or greater). However, there were six that may be called "primary or major factors." Each of the six subscales had 12 items. Table 11 contains the factor loadings for the 12 items and the factor

## Table 11

### Factor Loadings by Item Number and Factor/Subscale

| | Factor Number and Corresponding Subscale | | | | | |
|---|---|---|---|---|---|---|
| **Item No.** | **1/$S_3$** | **2/$S_6$** | **3/$S_2$** | **4/$S_5$** | **5/$S_4$** | **6/$S_1$** |
| 1 | .59 | .68 | .48 | .53 | .37 | .54 |
| 2 | .62 | .70 | .66 | .61 | .42 | .60 |
| 3 | .77 | .76 | .73 | .56 | .57 | .75 |
| 4 | .77 | .71 | .71 | .59 | .60 | .56 |
| 5 | .76 | .72 | .64 | .55 | .62 | .33 |
| 6 | .83 | .73 | .68 | .59 | .65 | .74 |
| 7 | .85 | .72 | .44 | .61 | .59 | .40 |
| 8 | .67 | .76 | .57 | .54 | .52 | – |
| 9 | .82 | .75 | .56 | .57 | .40 | .40 |
| 10 | .76 | .79 | .57 | .48 | .55 | .49 |
| 11 | .78 | .79 | .56 | .49 | .45 | .42 |
| 12 | .67 | .68 | – | .43 | .47 | .43 |

$S_1$: Learning Culture
$S_2$: Community of Learners
$S_3$: Sharing Leadership
$S_4$: Shared Goals for Learning
$S_5$: Assessing Student Learning
$S_6$: Enabling the SMART Learner

most closely representing the construct measured by the subscale. The desirable result is for the items of a subscale to load heavily on one factor and have low or zero loading on other factors. The accepted criterion of reporting loadings equal to or greater than .30 was used for Table 11.

To illustrate the information in Table 11, consider Factor 1. $S_3$ loads heavily on Factor 1 and the 12 factor loadings in the first column are the correlations between the 12 items of $S_3$ and Factor 1. These factor loadings range from .59 to .85, certainly substantial loadings. $S_3$ is the Sharing Leadership subscale and Item 7 of that subscale, "The school administration believes in shared leadership," is the item with the greatest factor loading. In the entire matrix of Table 11, there were only two factor loadings less than .30, and those occurred for different factors (3 and 6).

With an inventory such as the CSIQ, there very likely will be some substantial loadings across two or more factors. $S_4$, which corresponds to Factor 5, had seven loadings above .30 on Factor 1, which corresponds with $S_3$ (these multiple loadings are not reflected in Table 11). This pattern of results is reasonable because $S_3$ and $S_4$ are the "sharing" subscales, sharing leadership and goals. $S_1$ loaded heavily on Factor 6, and it also had six loadings ranging from .31 to .37 on Factor 2, the factor that corresponds to $S_6$. So, Enabling the SMART Learner and the Learning Culture have some common loadings.

As factors are extracted in order, they account for variance, but variance can be accounted for only once, so the factors extracted early will account for the large percentages of variance. That is, any variance that may be accounted for by two or more factors is awarded to the first factor extracted. The first factor accounted for 44.6% of the variance. The first six factors extracted accounted for 64.2% of the variance, which is almost two-thirds of the total. To

account for all of the variance, it is necessary to keep extracting factors until the factor space is exhausted. Each successive factor accounts for less variance than preceding factors.

All in all, the resulting factor pattern followed closely the pattern of the six subscales. Of course, this is a desirable result. The remaining two factors that had eigenvalues greater than 1.0 each had only one factor loading greater than .30, and these were less than .40. They may be ignored.

## CONCLUSIONS

The pilot test conducted prior to the field test focused primarily on the measurement characteristics of the CSIQ, while the field test focused more on variables that might affect, or be related to, performance on the CSIQ. The following conclusions are based on the results of the field test.

1. The 72-item CSIQ is highly reliable, both the total inventory and the individual subscales.

2. Within a short period of time, about three weeks, the CSIQ has moderate to high stability reliability. It should be noted that this result was based on a situation of no intervention (e.g., an intervention might be a program of school improvement). With intervention programs implemented over substantial periods (e.g., a school year), stability levels might be quite limited and change would be desirable.

3. The mean scores on the CSIQ subscales are highly consistent and positioned about 70% of the distance from the minimum to the maximum scores. The CSIQ subscales provide adequate measurement of the constructs they measure with good location on the scale of measurement and desirable coefficients of variation.

4. Type of school, that is level of school, appears to have a slight to modest effect on subscale and total score performance, with subjects in elementary schools having the greater performance.

5. Female subjects score higher than male subjects on the subscales and total score, but the effect is at least partially confounded with type of school because the majority of educators in elementary schools are female.

6. Subjects in schools known to be in a mode of continuous improvement score higher on subscales and total score than their counterparts in the same type of schools.

7. There may be an effect of rurality-urbanality as indicated by the Johnson Codes, with the most urban schools having the greatest scores on the subscales and total score. However, there is no definite pattern overall between the extent of rurality and performance. Also, there was a limited number of scores in the most urban code, and this may be more a reflection of the specific school or schools than of the extent of urbanality.

8. For all effects, it is important to consider the size of the effect versus the practical significance of any differences. Because of the large number of subjects, statistical tests of field test data had great statistical power.

9. The CSIQ subscale scores have modest to quite high intercorrelations, thus the subscales are not independent.

10. The CSIQ total score shows moderate concurrent validity when compared with the SCQ. The concurrent validity of the subscales within type of school is low to moderate when compared to the SCQ.

11. The SCQ and the individual subscales of the CSIQ measure some similar constructs, such as those related to Sharing Leadership and Shared Goals for Learning. Learning Culture and other possible constructs are not commonly measured to any noteworthy degree.

12. When considering continuous school improvement as a concept measured by the CSIQ, there appear to be six underlying constructs, and these closely coincide with the six subscales of the CSIQ.

13. The underlying constructs of continuous school improvement as measured by the six subscales account for almost two-thirds of the variance in the item scores, and, for this type of measurement, two-thirds is a substantial portion.

14. The length of the CSIQ very likely could be reduced to 9 or 10 items per subscale without adversely affecting reliability or validity. Items deleted could be those that load on multiple factors reducing the overlap in factor loadings and, at least theoretically, making the constructs measured by the subscales slightly "cleaner."

## REFERENCES


AEL, Inc. (1995). *AEL technical proposal for operation of the Regional Educational Laboratory in the Appalachian Region for 1996-2000.* Charleston, WV: Author.

AEL, Inc. (1998). *1999 updated annual plan and budget.* Submitted to the Office of Educational Research and Improvement, U.S. Department of Education. Charleston, WV: Author.

Manning, G., Curtis, K., & McMillen, S. (1996). *Building community: The human side of work.* Cincinnati, OH: Thomas Executive Press.

National Center for Education Statistics. *1997-98 Common Core of Data: Information on public schools and school districts in the United States.* Public elementary/secondary school universe survey data. Washington, DC: Author. http://nces.ed.gov/ccd/pubschuniv.html (retrieved July 2000)

Wiersma, W. (2000). *Technical report: Continuous School Improvement Questionnaire pilot test.* Charleston, WV: AEL, Inc.

## BIBLIOGRAPHY


Cowley, K. (1999). *A study of teacher efficacy and professional learning community in Quest schools.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (2000a). *Bowman Elementary: A case study of the Quest network.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (2000b). *Bending Knee Elementary: A case study of the Quest network.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (2000c). *Tinder Elementary: A case study of the Quest network.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (2000d). *Xavier Senior High School: A case study of the Quest network.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (2000e). *Lessons learned from the Quest for quality learning communities project, 1996-2000.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999a). *Engaging with school improvement: A study of factors influencing initial and sustained involvement in the Quest network.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999b). *Evaluation of Quest elementary and high school network rally, November 1998.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999c). *Evaluation of Quest high school rally, February 1999.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999d). *Evaluation of Quest elementary school rally, February 1999.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999e). *Evaluation of Quest Scholars program, July 1999.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1999f). *Data in a Day at Xavier Senior High School: A case study.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998a). *Interview evaluation of Leadership to Unify School Improvement Efforts (LUSIE).* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998b). *Evaluation of Quest high school network Inquiry into Improvement conference, October 1997.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998c). *Evaluation of Quest elementary school network Inquiry into Improvement conference, November 1997.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998d). *Evaluation of Quest high school network rally, February 1998.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998e). *Evaluation of Quest elementary school network rally, February 1998.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998f). *Evaluation of Quest Scholars colloquium, July 1998.* Charleston, WV: AEL, Inc.

Howley-Rowe, C. (1998g). *Evaluation of Quest summer symposium, August 1998.* Charleston, WV: AEL, Inc.

Meehan, M. L., Orletsky, S. R., & Sattes, B. (1997). *Field test of an instrument measuring the concept of professional learning communities in schools.* Charleston, WV: AEL, Inc.

Parrish, P., & Howley-Rowe, C. (2000). *Evaluation of Quest summer symposium, July 1999.* Charleston, WV: AEL, Inc.


**U.S. Department of Education**
*Office of Educational Research and Improvement (OERI)*
*National Library of Education (NLE)*
*Educational Resources Information Center (ERIC)*

ERIC®

# NOTICE

## Reproduction Basis

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☒ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").