

CCTDI Assessment Report – Fall 2014

Author: Joseph F. van Gaalen, Ph.D., Coordinator, Academic Assessment

1 INTRODUCTION

Florida SouthWestern State College’s Quality Enhancement Plan goal is to permit first-time-in-college students to become independent learners proficient in critical thinking. Through course completion, students will be able to demonstrate their analytical and evaluation skills. One measurement for the achievement of that goal is the use of the California Critical Thinking Disposition Inventory tests (CCTDI). FSW has identified a set criterion for defining student advancement in the Cornerstone Experience course. The results of the overall means scores of the CCTDI are expected to statistically significantly improve in the following Critical Thinking Dispositions: Truth Seeking, Open-Mindedness, Analyticity, Systematicity, Inquisitiveness, Confidence in Reasoning, and Maturity of Judgment, as measured by the CCTDI. This report is the continued assessment of the FSW QEP.

Pre-test/post-test studies in small groups provide an assessment foundation for learning and skill set adoption under given criteria. While scores do yield some error related to the target subject such as grade level or demographic, many can be accounted for in small sub-samples (individual classes). Moreover, those correlative measures that cannot be accounted for can be better understood through assessment (Cole et al., 2011).

For additional detail or further analysis not provided in this report, please contact Dr. Joseph F. van Gaalen, Coordinator of Academic Assessment, Academic Affairs (Joseph.VanGaalens@fsw.edu; x6965).

2 STATISTICS

2.1 DESCRIPTIVE STATISTICS AND SIGNIFICANCE TESTING

During the Fall 2014 semester, 4348 total tests (pre- and post-) were administered to students. Of those, 1849 of which were pre-/post- paired tests and 650 tests did not have counterparts. Pre-test and post-test mean scores for each dimension as well as the overall score are provided in Table 1. The difference in the means from pre-to-post test scores for each dimension (Truth-seeking, Open mindedness, Inquisitiveness, Analyticity, Systematicity, Confidence in Reasoning, and Maturity of Judgment) was tested for significance using a paired means t-test according to standard methods (Davis, 1973; McDonald, 2009; Siegel, 1956; Wilkinson, 1999). The results of significance testing for each dimension are also shown in Table 1.

	Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity of Judgment	Overall
Pre-test Mean	33.7	40.5	47.5	44.4	40.8	44.8	39.7	291.1
Post-test Mean	34.5	40.9	47.2	45.0	41.0	46.5	39.5	294.4
Effect size	0.06	0.03	-0.02	0.05	0.01	0.13	-0.01	0.05
p-value	6.72×10^{-9}	0.010*	0.005	2.30×10^{-9}	0.581	2.76×10^{-41}	0.028*	7.88×10^{-8}

Table 1. Comparison of mean scores for Pre/post test scores. Bold denote statistically significant difference. Positive effect sizes indicate a higher mean score for Post-test scores. *Denote marginal significance as defined by Johnson (2013).

Based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant results depicted in Table 1 may be false positives (i.e. Type I errors). These marginal results, defined as those within the 95-99% confidence level, include the Open-Mindedness and Maturity of Judgment dimensions.

Effect size was calculated using a method devised by Rosenthal and Rosnow (1991) for meta-analytical purposes in potential comparisons with other institutions (Lipsey and Wilson, 1993). The statistically significant results exhibit what Cohen (1988) would consider small effect sizes ranging from 0.01 to 0.13 (Table 1). In other words, non-overlap from Pre-test scores to Post-test scores range from approximately 0% to 8%.

The paired means t-test results indicate that for six of the seven dimensions, Truth-seeking, Open-Mindedness, Inquisitiveness, Analyticity, Confidence in Reasoning, and Maturity of Judgment we must reject the null hypothesis that the difference in the means of the pre- and post-test scores are equal to 0, and we can conclude this with a 95% confidence that the differences in scores are not solely due to chance. The other dimension, Systematicity, we cannot reject the null hypothesis that the slight increase in the mean score from pre-to-post-test scores can be a result of chance.

For the six dimensions which exhibit statistically significant increases in mean score, it can be reasonably concluded that the average increase in score of the students as a group is a result of some change in the students as a group. For the remaining dimension, this cannot be stated or quantified.

2.2 SUPPORTING EXPLORATORY DATA ANALYSIS

Since significance tests only provide information under the assumption the two groups are unchanged excepting for the learning in the classroom between pre-/post-tests, it is necessary to explore in detail each dimension using multiple standard processes for support of significance testing. In this way, the most effective assessment can be presented toward instructive improvement (Elder and Paul, 2007).

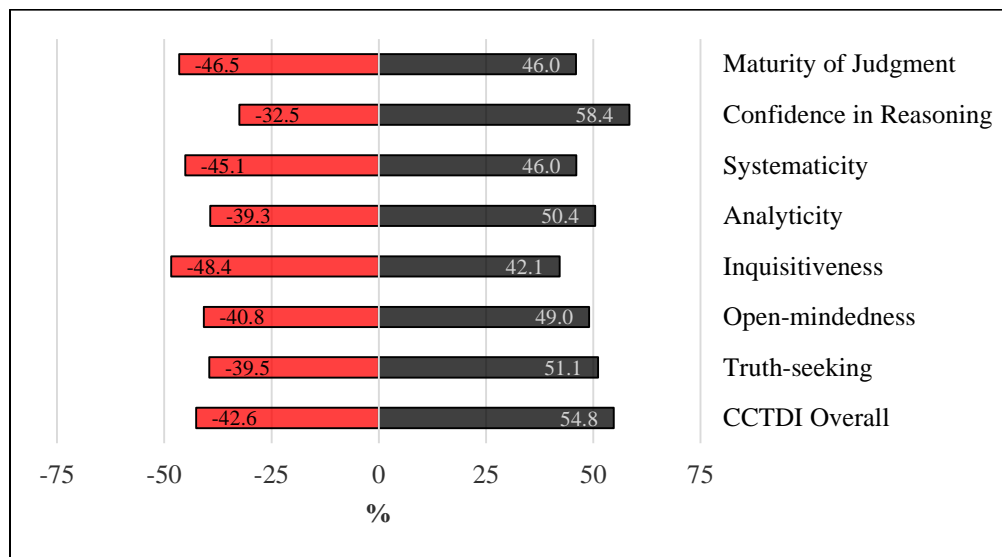


Figure 1. Percentage of students improved vs. declined. Difference of the sum of increase and decline from 100 is the percentage of test takers that exhibited no change.

Figure 1 highlights the percentage of student test scores that improved and declined. The learning dimensions of Confidence of Reasoning exhibits the greatest improvement percentage of improvement at 58.4%. By comparison, Inquisitiveness shows the least improvement at 42.1%.

An empirical distribution (histogram) of each dimension is reported in Figures 2-8. Figure 2 depicts data distribution of the Truth-seeking dimension which exhibits no shift in the mode (central peak of the data). When comparing pre-to-post scores, a net decrease in test scores in the 25-39 scoring bins is coupled with increases in scoring bins 40-44 and above.

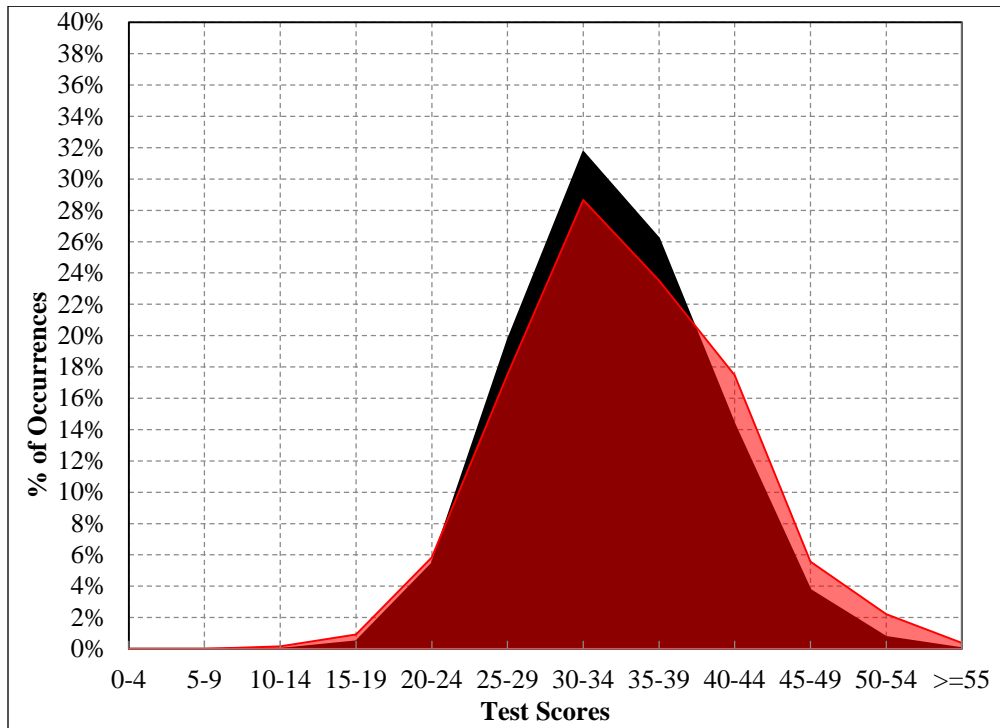


Figure 2. Truth-seeking dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 3 depicts data distribution of the Open Mindedness which exhibits no discernable shift in the mode (central peak of the data) from pre-to-post test scores. When comparing pre-/post- scores, decreases in the 35-44 scoring bins are coupled with increases in scoring bins 45-49 and above.

Figure 4 depicts data distribution of the Inquisitiveness dimension exhibiting a decrease in kurtosis from pre-to-post test scores (decreased peakedness of data). As results become more spread out in the post-test scores increases in percentages of lower scores and higher scores can be seen. As Inquisitiveness exhibits the highest pre-test scores of any of the learning dimensions, historically it is not uncommon to see little change from pre-to-post test scores in FSW studies (see Section 2.3 Longitudinal Studies).

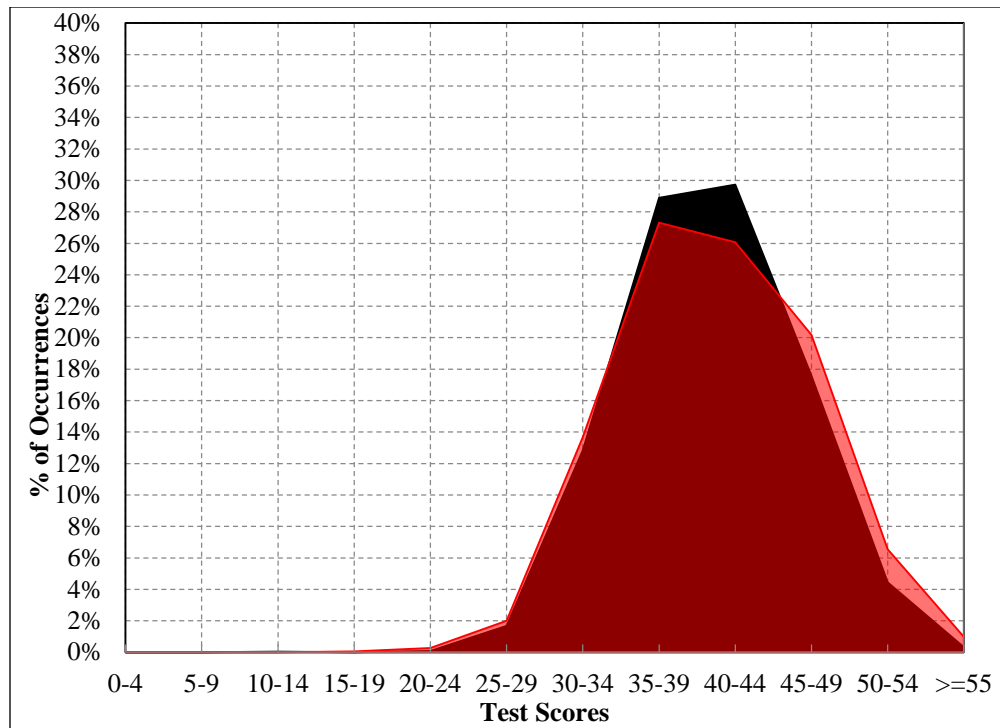


Figure 3. Open Mindedness dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 5 depicts data distribution of the Analyticity dimension which exhibits no shift in the mode (central peak of the data). When comparing pre-to-post scores, a net decrease in test scores in the 30-49 scoring bins is coupled with increases in scoring bins 50-54 and above.

Figure 6 depicts data distribution of the Systematicity dimension exhibiting a decrease in kurtosis from pre-to-post test scores (decreased peakedness of data). As with the Inquisitiveness learning dimension, results become more spread out in the post-test scores yielding increases in percentages of lower scores and higher scores. Similar cases in which systematicity did not significantly change following early testing have been reported (Beser and Kissal, 2009).

Figure 7 depicts data distribution of the Confidence in Reasoning dimension. The post-test scores exhibit a shift in modality (central tendency of data) from 40-44 scoring bin in pre-test scores to 45-49 scoring bin in post-test scores. Net decreases in post-test scores in the number of scores in the 40-44 scoring bin and below can be seen coupled with net increases in post-test scores in both the 45-49 bin and above.

Figure 8 depicts data distribution of the Maturity of Judgment dimension exhibiting a decrease in kurtosis from pre-to-post test scores (decreased peakedness of data). As with the Inquisitiveness and Systematicity learning dimensions, results become more spread out in the post-test scores yielding increases in percentages of lower scores and higher scores.

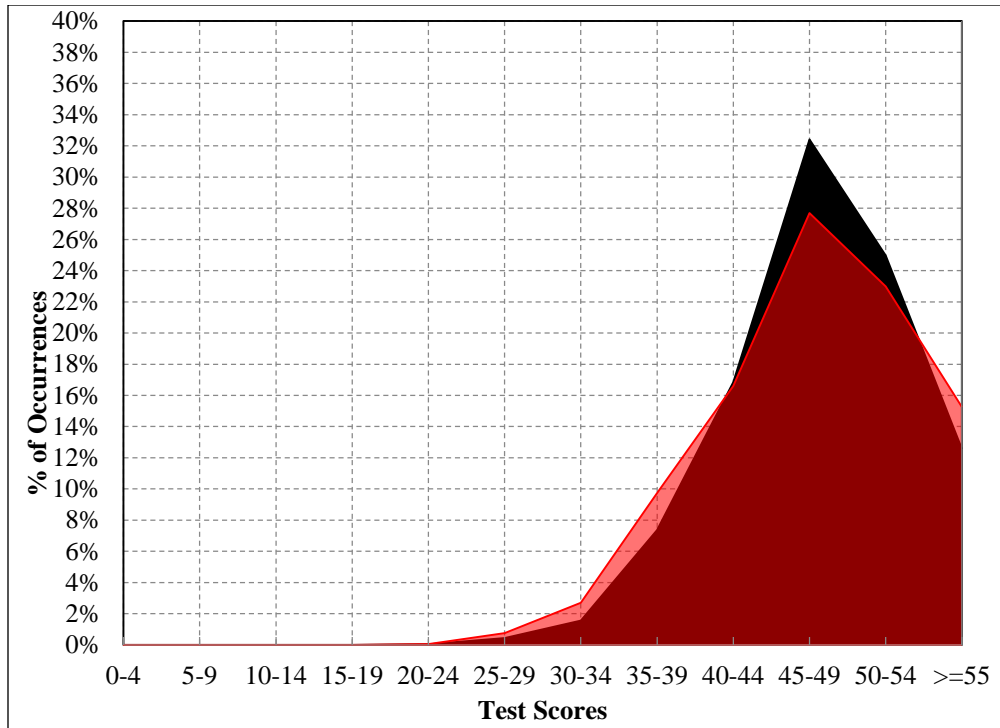


Figure 4. Inquisitiveness dimension distribution of test scores for Pre- (black) and Post- (red).

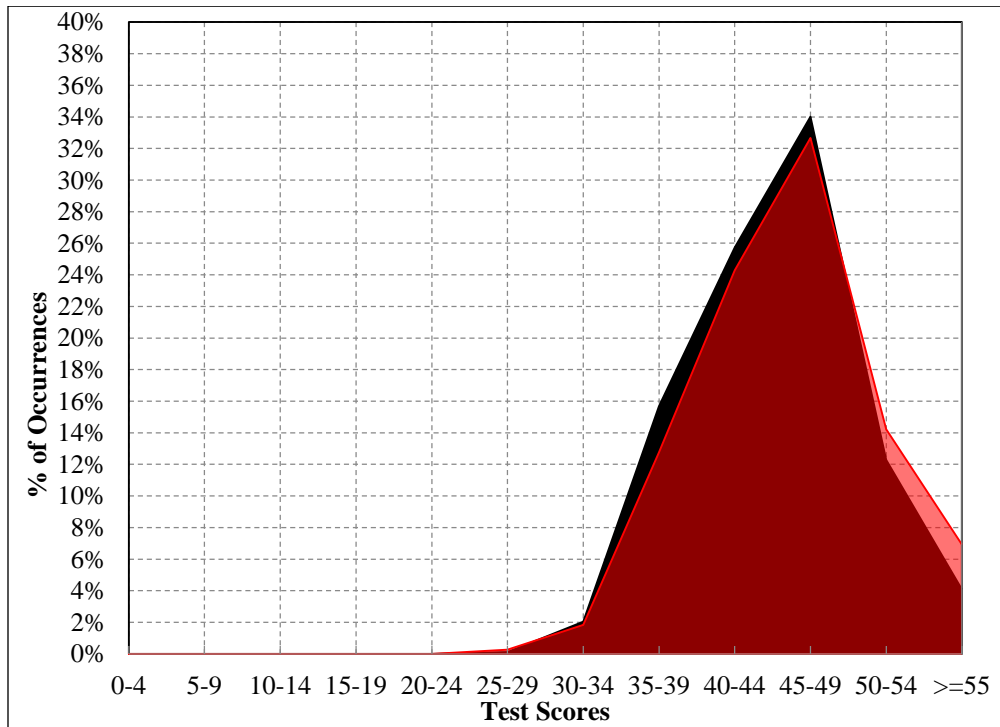


Figure 5. Analyticity dimension distribution of test scores for Pre- (black) and Post- (red).

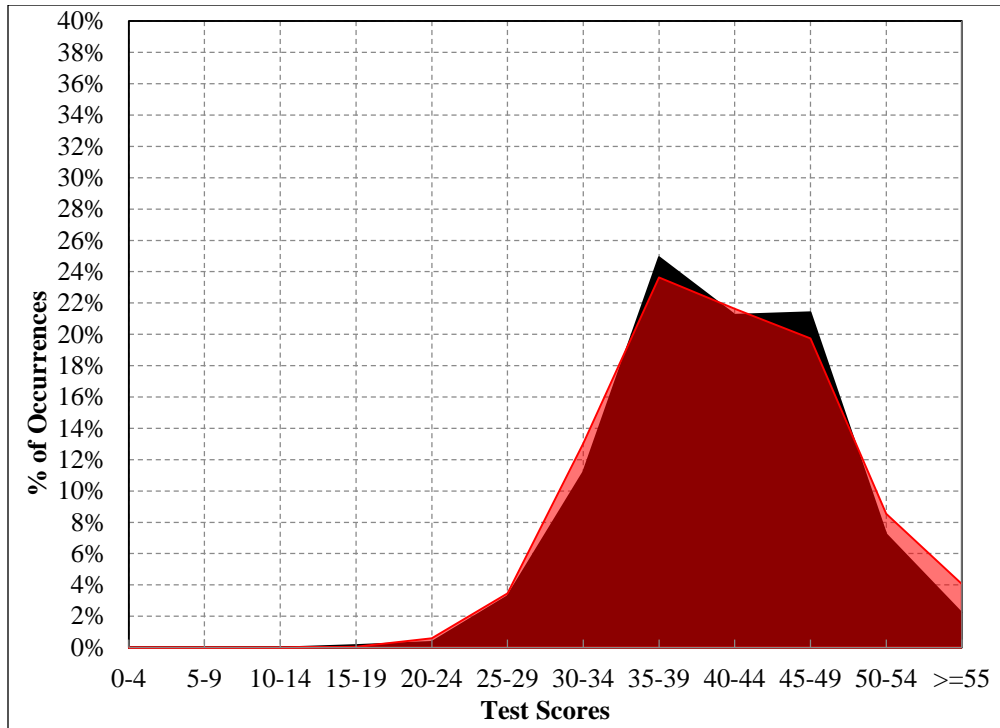


Figure 6. Systematicity dimension distribution of test scores for Pre- (black) and Post- (red).

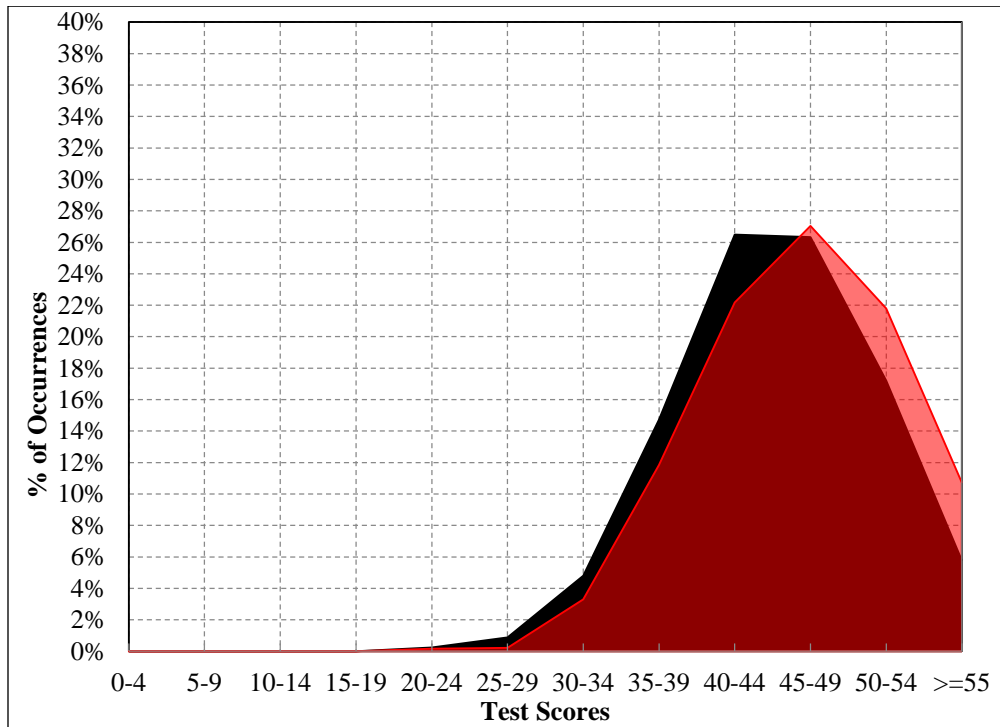


Figure 7. Confidence in Reasoning dimension distribution of test scores for Pre- (black) and Post- (red).

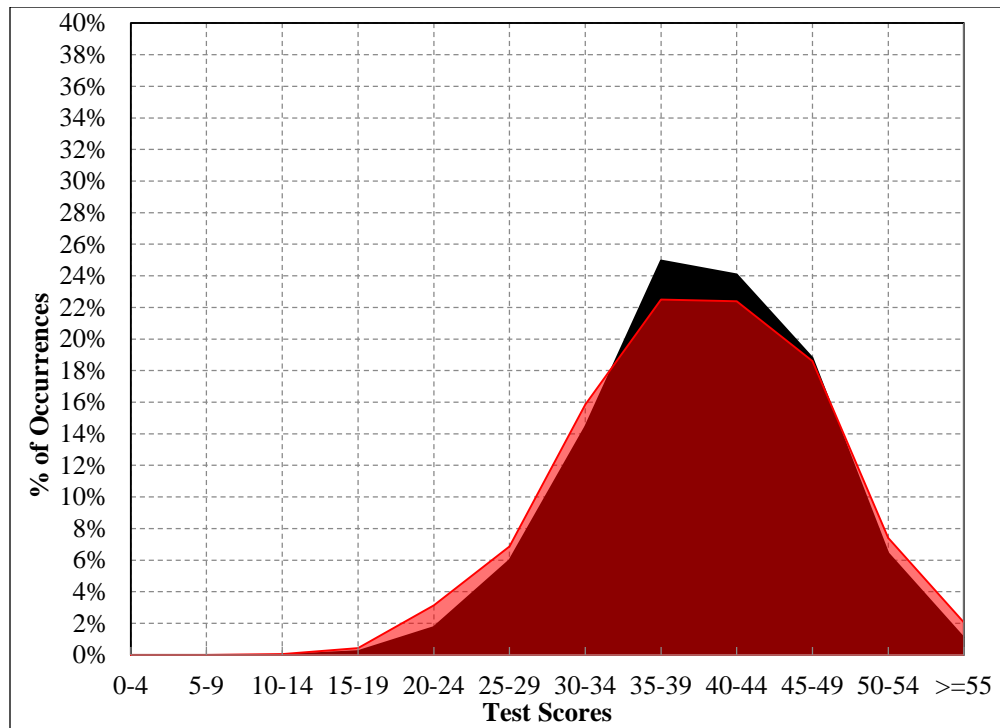


Figure 8. Maturity of Judgment dimension distribution of test scores for Pre- (black) and Post- (red).

2.3 COMPARISON WITH PREVIOUS FSW ASSESSMENTS

The results of paired means t-test of pre-/post- test scores for all semesters from Fall 2012 through Fall 2014 are shown in Table 2. All learning dimensions with the exception of the Inquisitiveness exhibit statistically significantly different means from pre-to-post test scores in at least four of the seven semesters in which studies were conducted. In Spring 2013, the t-test showed improvement in five of seven learning dimensions. Summer and Fall 2013 exhibited the weakest scores with both semesters exhibiting statistically significant improvement in three of seven dimensions and in Fall 2013, and Inquisitiveness showing a statistically significant decrease from pre-/post- test scores. In Spring 2014, six of seven dimensions exhibit statistically significant improvement. The remaining dimension, Inquisitiveness, exhibited no discernable change and could not be ruled out as chance. For the Summer 2014 semester, four of seven learning dimensions exhibited statistically significant increases in mean score. In Fall 2014, four of seven dimensions exhibit statistically significant improvement. Throughout the study, Analyticity and Confidence in Reasoning have consistently exhibited the largest improvement in scores from pre-/post-test scores. Both dimensions have exhibited a statistically significant increase in mean score across all semesters, although according to Johnson (2013) Fall 2013 Analyticity may be a false positive.

	Fall '12	Spring '13	Summer '13	Fall '13	Spring '14	Summer '14	Fall '14
Truth-seeking	1.1	0.4	0.8	0.5	1.1	0.8*	0.8
Open Mindedness	0.7	0.7*	0.4	0.0	0.8	0.50	0.4*
Inquisitiveness	0.7	0.1	0.4	-0.6	0.0	0.2	-0.3
Analyticity	1.0	1.2	1.1	0.4*	1.2	1.2	0.7
Systematicity	0.8	0.8*	1.1	0.0	0.8	1.6	0.2
Confidence in Reasoning	1.6	1.8	2.6	1.4	1.7	1.8	1.7
Maturity of Judgment	1.2	1.1	0.0	-0.1	1.1	0.1	-0.2*

Table 2. SLS 1515 CCTDI Pre-/Post- test results mean difference. Comparison of significance test results for mean difference of pre-/post-test scores for Fall 2012 through Fall 2014. Shaded cells indicate statistically significant differences in the mean at the 95% confidence level. Red text denotes decrease from pre-to-post. *Denote marginal significance as defined by Johnson (2013).

Table 3 provides additional information regarding the paired means t-test including the observed t-statistic (t_{obs}) and probability of difference due to chance (p-value) with respect to the degrees of freedom for each study. No effect size is reported since the method of calculation has not been consistent historically (Rosenthal and Rosnow, 1991; Wilson et al., 1995; Dunlop et al., 1996).

Figure 9 is a graphical representation of the difference in mean scores of learning dimensions across all semesters. The consistently strong improvement in scores from pre-to-post tests of the Confidence in Reasoning learning dimension (pink bar) is clearly visible. By comparison, the learning dimension with minimal results across all semesters is Inquisitiveness (blue bar). This dimension has the lowest mean score difference across all but one semester (Summer 2013) and exhibited a decline from pre-to-post scores in both the Fall 2013 and Fall 2014 semesters. Systematicity exhibits the widest variation of any indicator, as high as +1.55 in Summer 2014, and as low as 0.01 in Fall 2013.

	Fall '12 $t_{crit} = 1.97$	Spring '13 $t_{crit} = 1.97$	Summer '13 $t_{crit} = 1.98$	Fall '13 $t_{crit} = 1.97$	Spring '14 $t_{crit} = 1.97$	Summer '14 $t_{crit} = 1.97$	Fall '14 $t_{crit} = 1.97$
Truth-seeking	t(365)=4.00, p<0.05	t(204)=1.09, p=0.275	t(145)=1.71, p=0.090	t(859)=2.69, p=0.007	t(407)=3.91, p=1.09x10 ⁻⁴	t(173)=2.14, p=0.034*	t(1848)=5.82 p=7.88x10 ⁻⁸
Open Mindedness	t(365)=2.67, p<0.05	t(204)=2.24, p=0.026*	t(145)=0.94, p=0.347	t(859)=0.07, p=0.9414	t(407)=3.36, p=8.56x10 ⁻⁴	t(173)=1.34, p=0.183	t(1848)=2.57 p=6.74x10 ⁻⁹
Inquisitiveness	t(365)=2.40, p<0.05	t(204)=0.24, p=0.813	t(145)=0.95, p=0.345	t(859)=-3.10, p=0.002	t(407)=2.83, p=0.907	t(173)=0.57, p=0.570	t(1848)=-2.78 p=0.010
Analyticity	t(365)=4.18, p<0.05	t(204)=3.46, p=0.0007	t(145)=2.92, p=0.004	t(859)=2.26, p=0.024*	t(407)=5.00, p=8.72x10 ⁻⁷	t(173)=3.84, p=1.70x10 ⁻⁴	t(1848)=6.01 p=0.005
Systematicity	t(365)=2.81, p<0.05	t(204)=2.08, p=0.039*	t(145)=2.69, p=0.008	t(859)=0.05, p=0.963	t(407)=2.83, p=0.005	t(173)=4.19, p=4.44x10 ⁻⁵	t(1848)=0.55 p=2.30x10 ⁻⁹
Confidence in Reasoning	t(365)=5.97, p<0.05	t(204)=5.28, p<0.001	t(145)=5.79, p<0.001	t(859)=7.71, p<0.001	t(407)=6.02, p=2.95x10 ⁻⁹	t(173)=4.98, p=1.52x10 ⁻⁶	t(1848)=13.80 p=2.77x10 ⁻⁴¹
Maturity of Judgment	t(365)=3.73, p<0.05	t(204)=2.89, p=0.004	t(145)=0.03, p=0.980	t(859)=-0.54, p=0.590	t(407)=3.47, p=5.79x10 ⁻⁴	t(173)=0.31, p=0.755	t(1848)=-2.20 p=0.028

Table 3. Additional significance testing statistics for learning dimensions including observed t-stat (t_{obs}), probability of difference due to chance (p-value), degrees of freedom (df), and critical t-stat. In some cases, earlier reports did not include p-value when p<0.05 or in later studies, p<<0.001 and are indicated where applicable. Mean difference of pre-/post-test scores are reported in Table 3. *Denote marginal significance as defined by Johnson (2013).

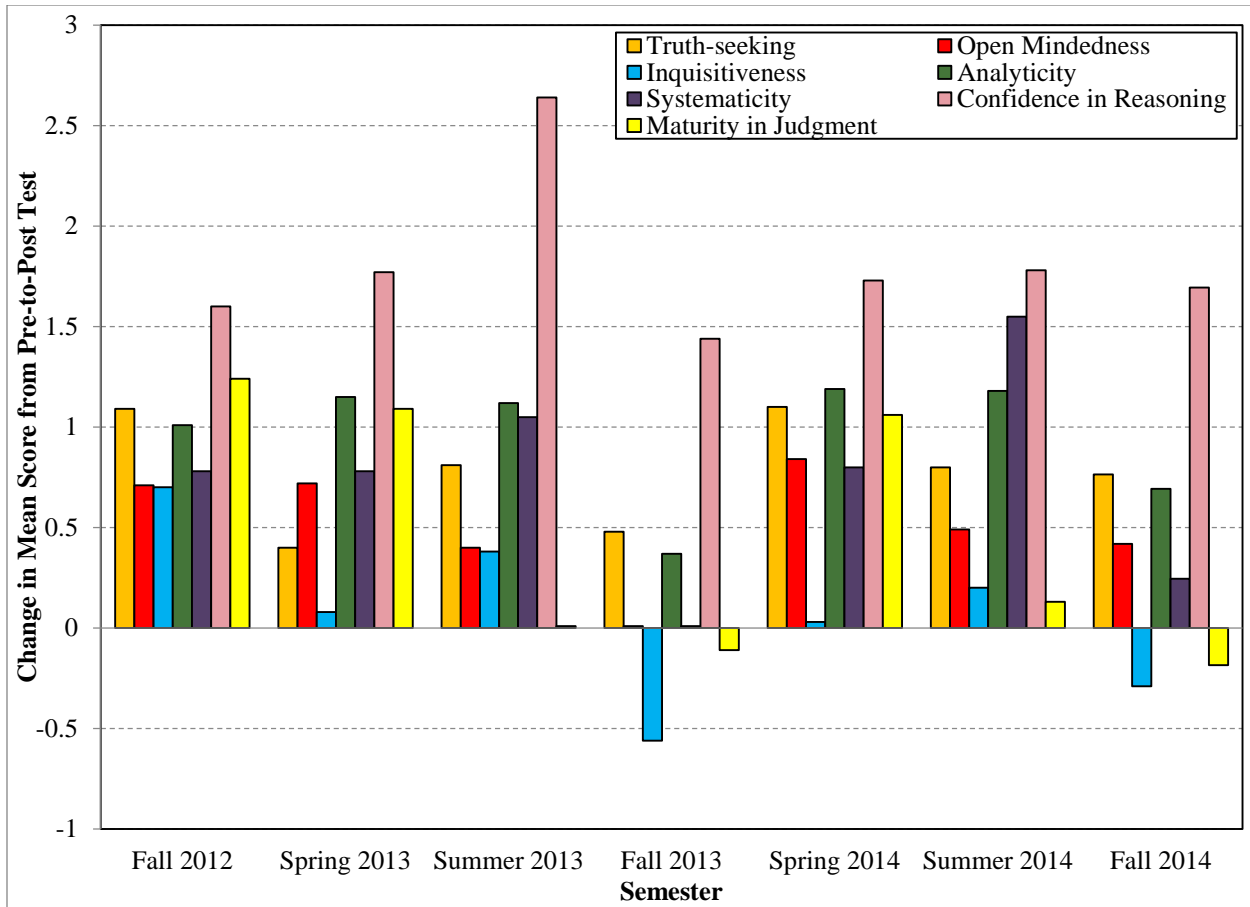


Figure 9. Comparison of the difference in mean scores from pre-to-post tests across semester by learning dimension. Orange (Truth-seeking), red (Open-mindedness), blue (Inquisitiveness), green (Analyticity), purple (Systematicity), pink (Confidence in Reasoning), and yellow (Maturity in Judgment).

3 CONCLUSIONS

In Florida SouthWestern State College’s QEP assessment, students are expected to improve in the seven Critical Thinking Dispositions: Truth Seeking, Open-Mindedness, Analyticity, Systematicity, Inquisitiveness, Confidence in Reasoning, and Maturity of Judgment based on exposure to these topics in the SLS1515 Cornerstone Experience course. These dispositions are measured using the California Critical Thinking Disposition Inventory tests (CCTDI) and analysis between pre-test (first three weeks of semester testing date) and post-test (last three weeks of semester testing date) is performed.

A drilldown of the results of the CCTDI Fall 2014 assessment are as follows:

1. The means of post-test scores exhibit statistically significant improvement over means of pre-test scores in the Truth-Seeking, Open Mindedness, Analyticity, Confidence in Reasoning, and Maturity of Judgment learning dimensions.
2. The Systematicity learning dimension exhibits no statistically significant change between pre-test and post-test scores.
3. The Inquisitiveness learning dimension exhibits a statistically significant decline from pre-test to post-test scores.

4. The means of the post-test combined scores of all seven dimensions exhibits a statistically significant improvement over means of pre-test scores.
5. In a comparison of students who improved from pre-to-post and those that didn't, the Truth-Seeking, Open Mindedness, Analyticity, Systematicity, and Confidence in Reasoning learning dimensions exhibit a greater population of students who improved from pre-to-post test scores than declined.
6. In a comparison of students who improved from pre-to-post and those that didn't, both the Inquisitiveness and Maturity of Judgment learning dimensions exhibit a greater population of students who declined from pre-to-post test scores than improved.
7. In a longitudinal study, since comparisons began in Fall 2012, Analyticity and Confidence in Reasoning have consistently exhibited the largest improvement in scores from pre-/post-test scores. Both dimensions have exhibited a statistically significant increase in mean score across all semesters.
8. In the same longitudinal study, since comparisons began, the learning dimension with minimal results across all semesters is Inquisitiveness. This dimension has the lowest mean score difference across all but one semester (Summer 2013) and exhibits a decline from pre-to-post scores in both the Fall 2013 and Fall 2014 semesters.

4 REFERENCES

- Beser, A., and Kissal. 2009. Critical Thinking Dispositions and Problem Solving Skills Among Nursing Students. *DEUHYO Ed*, 2(3), 88-94.
- Cole, R., Haimson, J., Perez-Johnson, I., and May, H. 2011. Variability in Pretest-Posttest Correlation Coefficients by Student Achievement Level. NCEE Reference Report 2011-4033. Washington, DC: National Center for Education Evaluation and Regional Assistance, U.S. Department of Education.
- CCTDI. 2013. 2013 CCTDI Manual. San Jose, California: Insight Assessment, Division of California Academic Press.
- Cohen, J. 1988. *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Earlbaum Associates, Hillsdale, NJ.
- Davis, J.C. 1973. *Statistics and Data Analysis in Geology*. John Wiley & Sons, New York, New York, 564 pp.
- Dunlop, W.P., Cortina, J.M., Vaslow, J.B., and Burke, M.J. 1996. Meta-analysis of experiments with matched groups or repeated measures designs. *Psychological Methods*, 1, 170-177.
- Elder, L, and Paul, R. 2007. Consequential Validity: Using Assessment to Drive Instruction. In: *Foundation For Critical Thinking*. Retrieved from <http://www.criticalthinking.org/pages/consequential-validity-using-assessment-to-drive-instruction/790>.
- Johnson, V. 2013. Revised Standards for Statistical Evidence. *Proceedings of the National Academy of Science*, 110(48), 19313-19317.
- Lipsey, M.W. and Wilson, D.B. 1993. The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *American Psychologist*, 48, 1181-1209.

- McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.
- Rosenthal, R. and Rosnow, R.L. 1991. Essentials of behavioral research: Methods and data analysis. McGraw Hill, New York, 692 pp.
- Siegel, S. 1956. Nonparametric statistics for the behavior sciences. McGraw-Hill, New York, New York, 312 pp.
- Wilkinson, L. 1999. APA Task Force on Statistical Inference. Statistical Methods in Psychology Journals: Guidelines and Explanations. American Psychologist 54 (8), 594–604.