CCTDI Assessment Report – Spring 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).

2 STATISTICS

During the Spring 2014 semester, 1027 total tests (pre- and post-) were administered to students. Of those, 816 of which were pre-/post- paired tests and 209 tests did not have counterparts. Basic descriptive statistics of pre- and post- test scores only are shown in Table 1.

		Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity in Judgment
Pre-Test	**mean**	34.25	39.91	48.84	44.40	42.01	45.46	40.04
	median	34	39	49	44	42	46	40
	mode	33	38	52	45	45	47	37
	standard deviation	6.65	6.01	6.47	5.36	6.96	6.32	7.46
	Kurtosis	0.32	0.60	-0.11	0.16	-0.29	0.19	-0.44
	n	408	408	408	408	408	408	408
Post-Test	**mean**	35.36	40.75	48.87	45.60	42.81	47.19	41.10
	median	35	41	49	45	43.5	47	42
	mode	33	43	49	45	45	46	44
	standard deviation	7.67	6.17	6.96	5.65	7.21	6.81	8.19
	Kurtosis	0.09	0.26	0.18	0.05	-0.10	0.01	0.15
	n	408	408	408	408	408	408	408

Table 1. Pre-/Post- test scores with measured increases in post- test results denoted with shaded cell.

All categories of post-test scores show increases in means, although not all are significant (see discussion in 2.1 Significance Tests). The data in post-test scores reflect an increased standard deviation

(spread of data distribution) in all categories. Four categories reflect post-test distributions that have become more platykurtic (gentler peaks) than their pre-test counterpart while three have become more leptokurtic (steeper peaks).

2.1 SIGNIFICANCE TESTS

Study goals demanded significance tests be conducted to determine whether the difference in the means of pre- and post-test scores is solely due to chance. Each dimension (Truth-seeking, Open mindedness, Inquisitiveness, Analyticity, Systematicity, Confidence in Reasoning, and Maturity in Judgment) was tested for significance using a paired means t-test according to standard methods (Davis, 1973; McDonald, 2009; Wilkinson, 1999). The results of significance testing for each dimension are shown in Table 2. Additional details of the distribution of the results are explored in subsequent sections to provide further information into the variation of the pre-/post-test score relationship as foundation for potential future causal studies, if necessary.

	Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity in Judgment
mean	1.10	0.84	0.03	1.19	0.80	1.73	1.06
standard deviation	5.70	5.03	5.13	4.82	5.71	5.80	6.15
standard error	0.28	0.25	0.25	0.24	0.28	0.29	0.30
t_{crit}	1.97	1.97	1.97	1.97	1.97	1.97	1.97
t_{obs}	3.91	3.36	0.12	5.00	2.83	6.02	3.47
p-value	0.0001	0.0008	0.9078	8.721×10^{-7}	0.0049	3.9547×10^{-9}	0.0005

Table 2. Difference between Pre-/Post- results with significance at the α=0.05 level. Shaded cells denote statistically significant difference.

The paired means t-test results indicate that for six of the seven dimensions, Truth-seeking, Open mindedness, Analyticity, Systematicity, Confidence in Reasoning, and Maturity in 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 only dimension that we cannot reject the null hypothesis is Inquisitiveness, meaning 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 Inquisitiveness, this cannot be stated or quantified.

The Confidence in Reasoning learning dimension exhibits the most significant increase in mean score. Based on these results, this suggests the Spring 2014 FSW student, above all else, is more readily able to recognize their awareness to problems and think about their situations. The second most significant increase in mean score is in Analyticity. This suggests the Spring 2014 FSW student is quick to become aware of consequential thinking becoming more prepared for expected outcomes and scenarios. In contrast, the Spring 2014 FSW student does not appear to increase curiosity in learning in response to course learning as evidenced by the Inquisitiveness results.

2.2 SUPPORTING EXPLORATORY DATA ANALYSIS

Since significance tests only provide information on the rejection of a null hypothesis and not on specific details of the changes from pre-/post-test scores, it is necessary that exploratory analyses be performed such that further information of value can be extracted if an evaluation of the program methods effects is to be quantitatively understood. Therefore, each dimension was rigorously analyzed using multiple standard processes for support of significance testing in order to most effectively apply the results toward instructive improvement, therefore allowing assessment to drive instruction as defined by Elder and Paul (2007).

Each learning dimension varied widely with respect to student-by-student pre- to post- test score. Figure 1 highlights the percentage of student test scores that improved and declined.
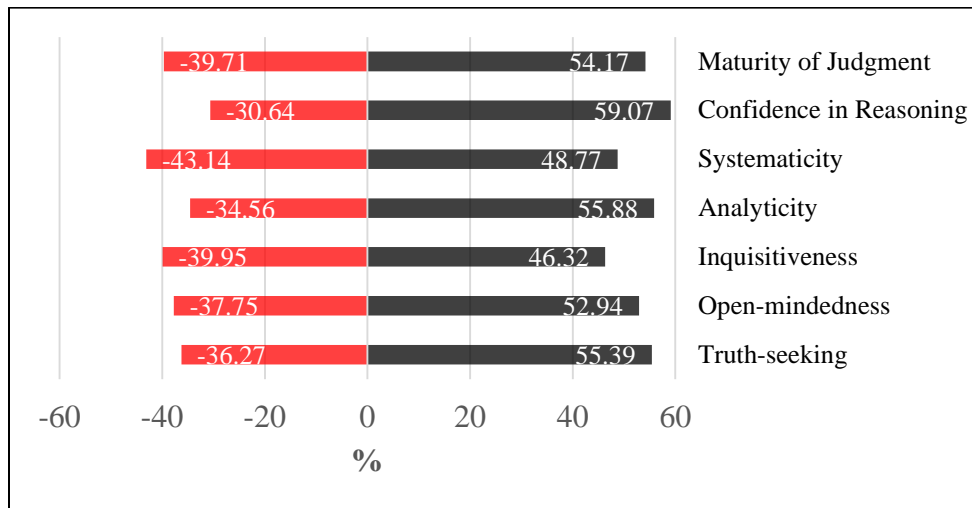

Figure 1. Percentage of students improved vs. declined. Difference of the sum of increase and decline from 100 is the percentage of test takers that exhibited no change.

The learning dimensions of Confidence of Reasoning, Truth-seeking, and Analyticity exhibit greater than 55% of students improved from pre- to post- test scores with Confidence of Reasoning exhibiting the greatest improvement percentage at 59.07%. By comparison, Inquisitiveness and Analyticity show the least amount of improvement at 46.32% and 48.77%, respectively.

Further investigation into the manner of these improvements/declines graphically describes the wide variation in kurtosis (peakedness of the distribution) and standard deviation (spread of the data) reported in Subsection 2: Statistics. An empirical distribution (histogram) of each dimension is reported in Figures 3-7 and 9, with overlain 5-point moving averages to give a rough estimate of probability density.

Figure 2 depicts data distribution of the Truth-seeking dimension which exhibits more platykurtic post-test scores with respect to pre-test scores reported in Subsection 2: Statistics. The 0.84 point statistically significant increase in the mean from pre- to post- test coincides with slightly negative skewness of post-test scores (i.e. decreased peak of post-test scores centered between 30-40 is accompanied by an increase in the area under the curve between 43-54).
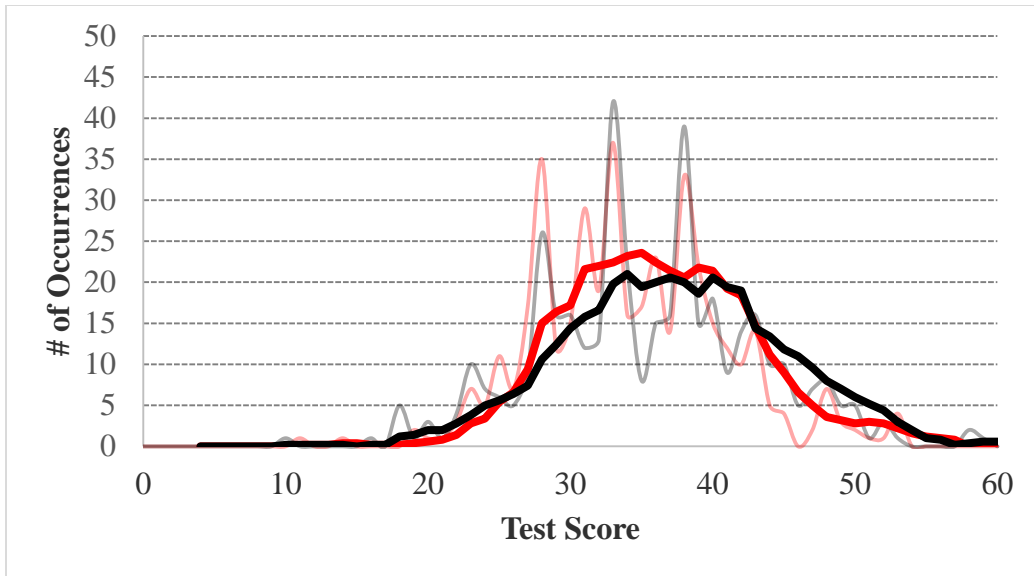
Figure 2. Empirical distribution of Truth-seeking dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.

Figure 3 depicts data distribution of the Open Mindedness dimension which exhibits more platykurtic post-test scores with respect to pre-test scores reported in Subsection 2: Statistics. The 1.1 point statistically significant increase in the mean from pre- to post- test coincides with slightly negative skewness of post-test scores (i.e. decreased peak of post-test scores centered on a range of 37-41 is accompanied by an increase in the area under the curve between 43-54).
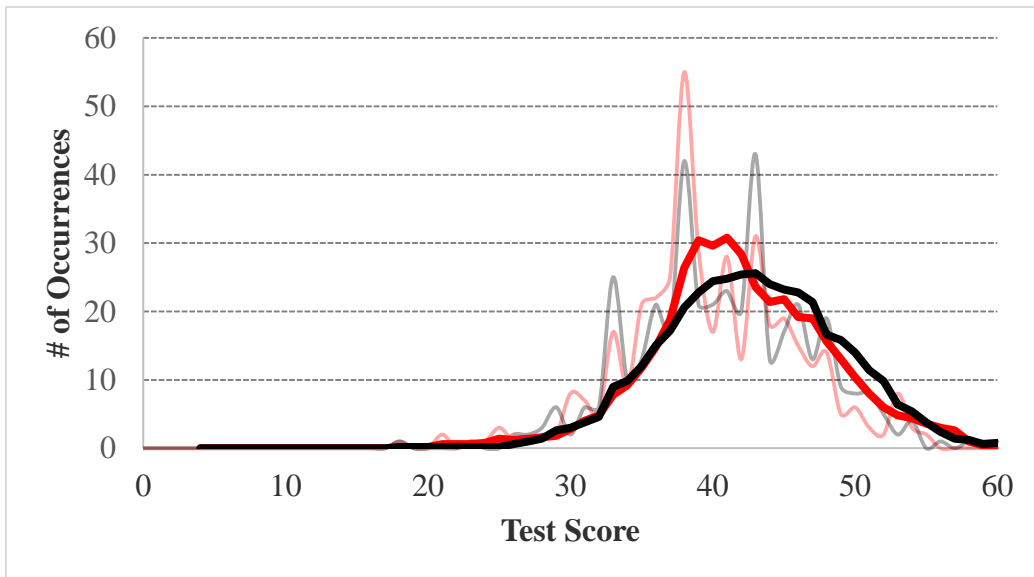

Figure 3. Empirical distribution of Open Mindedness dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.

Figure 4 depicts data distribution of the Inquisitiveness dimension exhibiting distributions of both pre- and post- tests centered on a common score with no discernable change in kurtosis or skewness. The negligible changes seen for this learning dimension in the figure coincides with insignificant changes reported in significance tests in 2.1: Significance Tests.
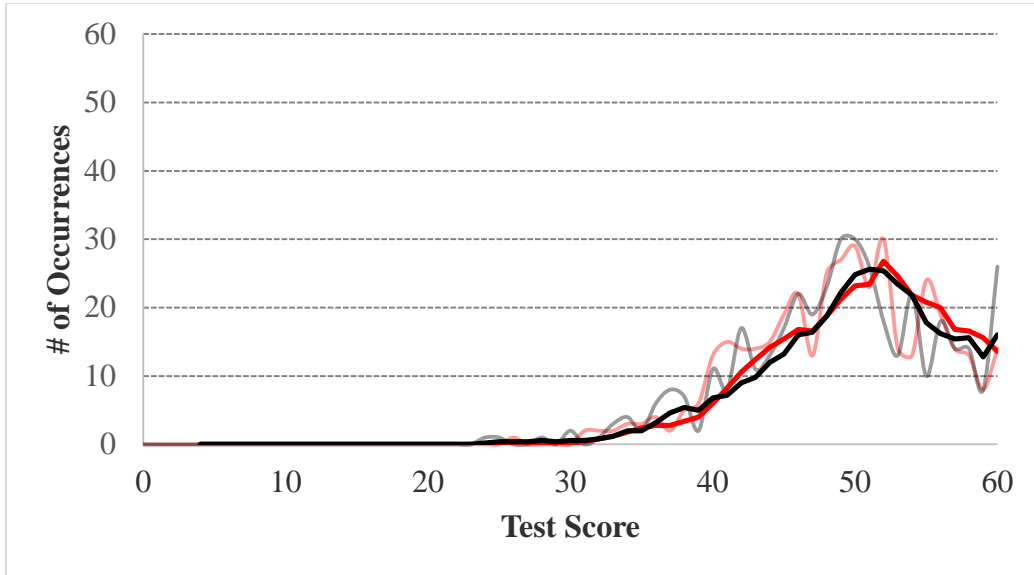

Figure 4. Empirical distribution of Inquisitiveness dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.
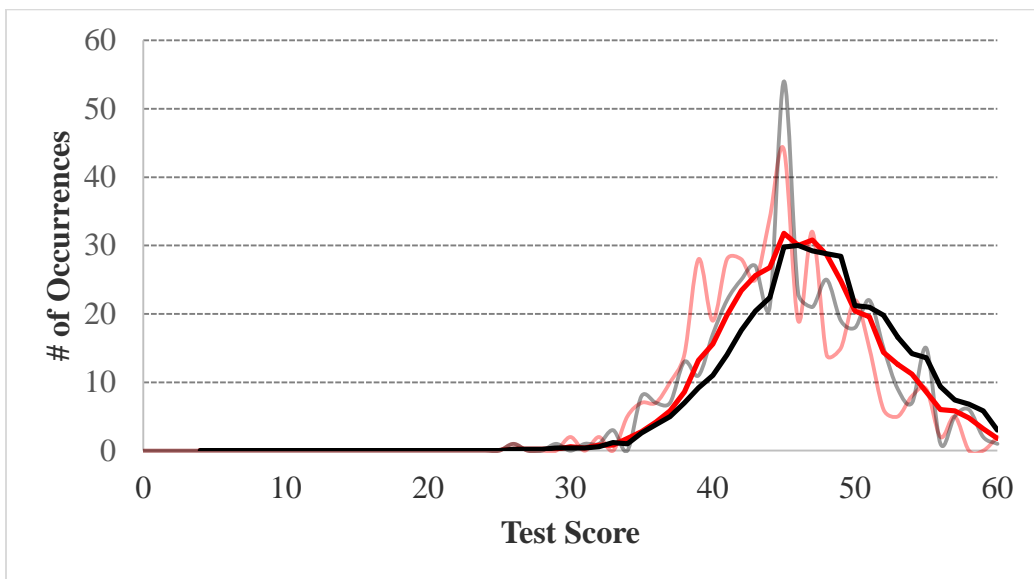

Figure 5. Empirical distribution of Analyticity dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.

Figure 5 depicts data distribution of the Analyticity dimension exhibiting slightly more leptokurtic distribution of the post-test scores with respect to pre-test scores reported in Subsection 2: Statistics. The 1.19 point statistically significant increase in the mean from pre- to post- test coincides with slightly

negative skewness of post-test scores (i.e. area under pre-test curve is greater than post-test curve between 38-45 and less than post-test curve at test scores greater than 50).

The estimated probability density (dark red/black lines) of the Systematicity learning dimension do not immediately reflect the significance test results reported in section 2.1 (Figure 6). The slightly more leptokurtic results are clear from the empirical distribution (faded black/red lines) with a more pronounced peak in post-score data at 45 compared to adjacent data, but not clear in the estimated function (dark black/red lines), which appears platykurtic. This ambiguity is supported in Figure 1 where the difference between improved and declining scores is only 5.63% for Systematicity (48.77% improved, 43.14% declined, 8.09% no change). This is in stark contrast to the other six dimensions which exhibit an average improvement ratio of 17.48%. The disparity between these results and the positive significance tests for the dimension as reported in Subsection 2.1 could be evidence of students adapting to new learning skills and methods within the SLS 1515 course where the tests were administered. Similar cases in which systematicity did not significantly change following early testing have been reported (Beser and Kissal, 2009).
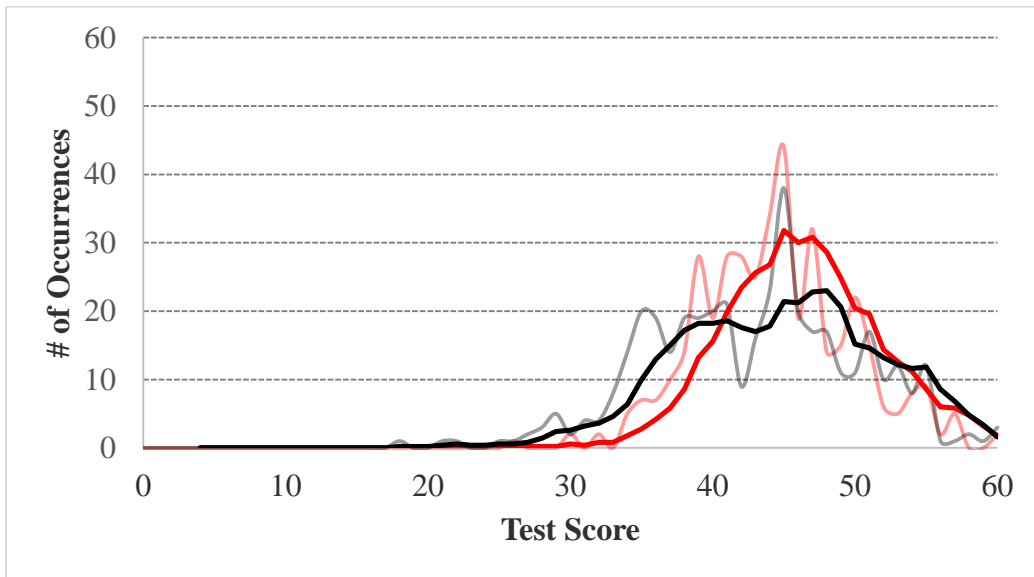

Figure 6. Empirical distribution of Systematicity dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.

Figure 7 depicts data distribution of the Confidence in Reasoning dimension exhibiting slightly more negatively skewed post-test scores (scores more heavily favor higher values). The 1.73 point mean score increase from pre-to-post-test score is the highest of all dimensions tested and coincides with fewer post-test scores ranging from 38-48 and increased post-test scores compared with pre- scores from 50-60.

Figure 8 provides additional information not visible in the previous figure with regard to the pre-test scores which exhibited the most improvement. Point A' on the post-test cumulative plot (black line) depicts the number of students who scored that value (46) or higher, 268. By comparison, the same score (46) on the pre-test cumulative plot (red line) shows 41 fewer students recorded a score that high

or higher (217 in total). From Figure 8, it can also be determined that no one in either pre- or post-test scored below a 25, and that 14 students scored a perfect 60 in post-test compared with only 7 on the pre-test.
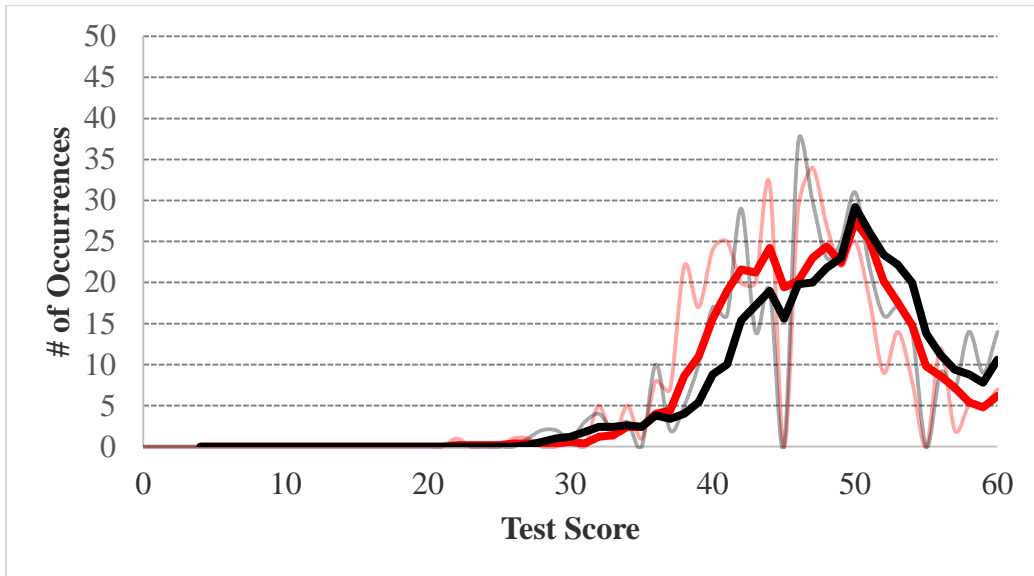

Figure 7. Empirical distribution of Confidence in Reasoning dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.
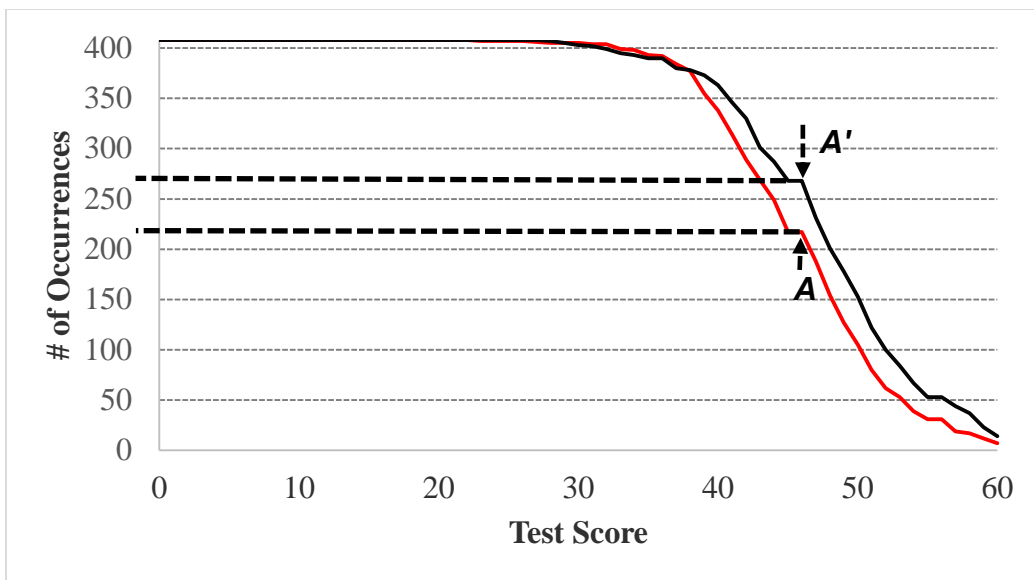

Figure 8. Cumulative distribution plot of Confidence in Reasoning dimension. Red line is pre-test cumulative scores. Black line is post-test cumulative scores. At point A (pre-test), 217 students scored 46 or higher. At point A' (post-test), 268 students scored 46 or higher.

Figure 9 depicts data distribution of the Maturity in Judgment dimension exhibiting more negatively skewed post-test scores (scores more heavily favor higher values), and a visible post-test central peak

value of 44, compared with pre-test 40. The 1.06 point mean score increase from pre-to-post-test score is coincident with fewer post-test scores ranging from 30-42 and increased post-test scores compared with pre- scores from 45-58.
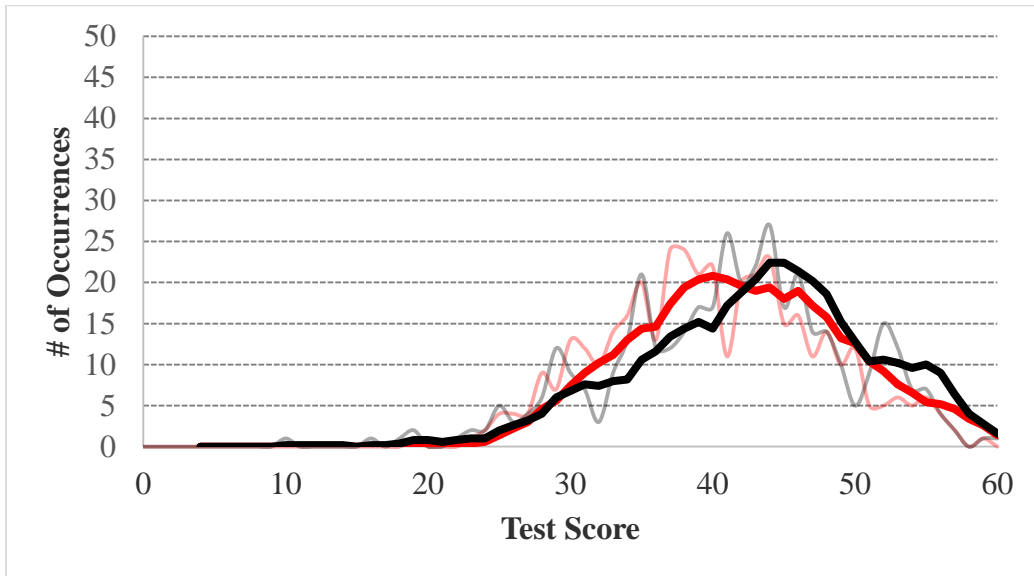

Figure 9. Empirical distribution of Maturity in Judgment dimension. Faded red is a continuous representation of pre-test scores, dark red is a 5-pt moving average. Faded black is a continuous representation of post-test scores, black is a 5-pt moving average.

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 Spring 2014 are shown in Table 3. All learning dimensions mean scores are reported and exhibit statistically significant results in at least three of the five semesters over the course of the study thus far. Fall 2012 exhibited statistically significant increases in all dimensions between the pre- and post-test administrations. Spring 2013, the t-test showed improvement in five of seven learning dimensions. Summer and Fall 2013 exhibited the weakest scores with both exhibiting statistically significant improvement in three of seven dimensions and in Fall 2013, 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. Throughout the study, Analyticity and Confidence in Reasoning have consistently exhibited the largest improvement in scores from pre-/post-test scores and have both always exhibited a statistically significant increase in mean score, although according to Johnson (2013) Fall 2013 Analyticity may be a false positive.

Table 4 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 calculated since measurement units (test score) are typical of the field and therefore already meaningful (Wilkinson, 1999). Based on the work of Johnson (2013), there is a 17-25% chance the marginally significant results depicted in Table 4 may be false positives. These marginal results, defined as those within the 95-99% confidence level, include Spring 2013 Open Mindedness and Systematicity, and Fall 2013 Analyticity.

	Fall '12	Spring '13	Summer '13	Fall '13	Spring '14
Truth-seeking	1.09	0.40	0.81	0.48	1.10
Open Mindedness	0.71	0.72*	0.40	0.01	0.84
Inquisitiveness	0.70	0.08	0.38	-0.56	0.03
Analyticity	1.01	1.15	1.12	0.37*	1.19
Systematicity	0.78	0.78*	1.05	0.01	0.80
Confidence in Reasoning	1.60	1.77	2.64	1.44	1.73
Maturity in Judgment	1.24	1.09	0.01	-0.11	1.06

*Table 3. 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 Spring 2014. Shaded cells indicate statistically significant differences in the mean at the 95% confidence level. *Denote marginal significance as defined by Johnson (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 $t_{crit} = 1.97$
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=1.09×10^{-4}
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=8.56×10^{-4}
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
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=8.72×10^{-7}
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
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=2.95×10^{-9}
Maturity in 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=5.79×10^{-4}

*Table 4. 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).*

3 CONCLUSIONS

In Florida SouthWestern State College's QEP assessment, students are expected to statistically significantly improve in the seven Critical Thinking Dispositions: Truth Seeking, Open-Mindedness, Analyticity, Systematicity, Inquisitiveness, Confidence in Reasoning, and Maturity in Judgment, and are measured using the California Critical Thinking Disposition Inventory tests (CCTDI).

For Spring 2014, the paired means t-test results indicate that for six of the seven dimensions, Truth-seeking, Open mindedness, Analyticity, Systematicity, Confidence in Reasoning, and Maturity in Judgment, we can conclude with a 95% confidence that the improvement in mean scores are not solely due to chance. Although there was a slight increase in mean score for Inquisitiveness, we cannot conclude this is Inquisitiveness, meaning the difference may simply be due to chance.

Based on the results, it can, with reasonable certainty be stated that the Spring 2014 FSW student, is more readily able to recognize their awareness to problems and think about their situations and is quick to become aware of consequential thinking becoming more prepared for expected outcomes and

scenarios. In contrast, the students do not appear to increase curiosity in learning in response to course learning as evidenced by the Inquisitiveness results.

In comparison with earlier studies, it can be shown that since the study began in Fall 2012, all learning dimensions have exhibited statistically significant results in at least three of the five semesters through Spring 2014 with Fall 2012 exhibited statistically significant increases in all dimensions. Summer and Fall 2013 exhibited the weakest scores with both exhibiting statistically significant improvement in only three of seven dimensions and in Fall 2013, Inquisitiveness showing a statistically significant decrease. FSW students consistently show significant improvement in Confidence in Reasoning in all studies since Fall 2012, again supporting the Spring 2014 conclusion that the FSW student is more readily able to recognize their awareness to problems and think about their situations.

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.

Davis, J.C. 1973. Statistics and Data Analysis in Geology. John Wiley & Sons, New York, New York, 564 pp.

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.

McDonald, J.H. 2009. Handbook of Biological Statistics (2nd ed.). Sparky House Publishing, Baltimore, Maryland.

Wilkinson, L. 1999. APA Task Force on Statistical Inference. Statistical Methods in Psychology Journals: Guidelines and Explanations. American Psychologist 54 (8), 594–604.