CCTDI Assessment Report – Summer 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 van Gaalen, Coordinator of Academic Assessment, Academic Affairs (Joseph.VanGaalen@fsw.edu; x6965).

2 STATISTICS

During the Summer 2014 semester, 392 total tests (pre- and post-) were administered to students. Of those, 348 of which were pre-/post- paired tests and 44 tests did not have counterparts. Basic descriptive statistics of pre- and post- test scores only are shown in Table 1.

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 higher kurtosis values, meaning they have become more platykurtic (gentler peaks) than their pre-test counterpart while three have become more leptokurtic (steeper peaks) (see Figure 1 for example). All categories reflect post-test distributions with more negative skewness, meaning scores tend toward higher values with a tail towards negative values (see Figure 2 for example).

		Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity of Judgment
Pre-Test	**n**	174	174	174	174	174	174	174
	mean	35.40	41.38	49.72	44.83	41.93	45.56	41.70
	median	35	41	50	45	42	44	42
	mode	33	38	48	45	35	44	42
	max / min	60 / 22	56 / 25	60 / 33	60 / 30	60 / 27	60 / 23	60 / 23
	standard deviation	6.52	5.62	5.59	5.68	7.11	6.42	6.75
	kurtosis	0.72	0.11	-0.07	-0.14	-0.47	0.58	-0.18
	skewness	0.57	0.20	-0.21	0.18	0.25	-0.05	-0.14
Post-Test	**n**	174	174	174	174	174	174	174
	mean	36.20	41.87	49.91	46.01	43.48	47.34	41.82
	median	36	42	50	45	44	47	42
	mode	38	43	47	45	45	50	40
	max / min	60 / 21	56 / 29	60 / 29	60 / 31	60 / 21	60 / 24	60 / 21
	standard deviation	7.09	5.74	6.52	6.00	8.31	6.77	7.67
	kurtosis	0.17	-0.42	-0.03	-0.10	-0.36	0.67	-0.23
	skewness	0.46	0.11	-0.52	0.08	-0.01	-0.35	-0.21

Table 1. Pre-/Post- test scores with measured increases of the mean, standard deviation, kurtosis, and skewness in post-test results denoted with shaded cell.

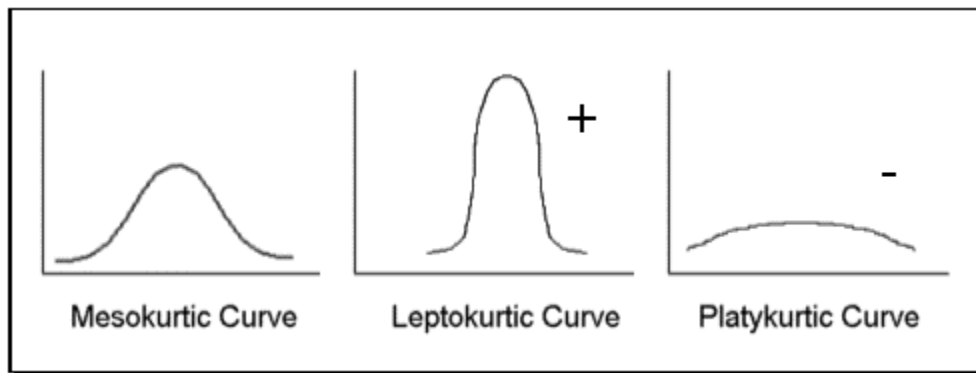Mesokurtic Curve   Leptokurtic Curve   Platykurtic Curve

Figure 1. Example of kurtosis. The normal curve (left) has a kurtosis of 0.0. A positive value, or leptokurtic distribution (center) and negative value, or platykurtic distribution (right) are depicted here in an ideal scenario (Starkweather, 2010).
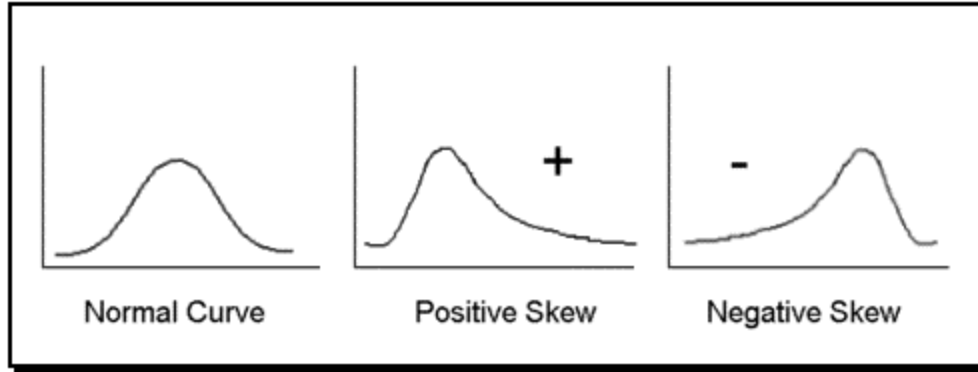
Figure 2. Example of skewness. The normal curve (left) has a skewness of 0.0. A positive value skewness (center) and negative value skewness (right) are depicted here in an ideal scenario (Starkweather, 2010).

2.1 SIGNIFICANCE TESTS

Study goals demand that 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 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 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 of Judgment
mean	0.80	0.49	0.20	1.18	1.55	1.78	0.13
standard deviation	4.92	4.87	4.53	4.06	4.89	4.72	5.33
standard error	0.37	0.37	0.34	0.31	0.37	0.36	0.40
df	173	173	173	173	173	173	173
t_{crit}	1.97	1.97	1.97	1.97	1.97	1.97	1.97
t_{obs}	2.14	1.34	0.57	3.84	4.19	4.98	0.31
p-value	0.034*	0.183	0.570	1.70×10^{-04}	4.44×10^{-05}	1.52×10^{-06}	0.755

Table 2. Difference between Pre-/Post- results with significance at the α=0.05 level. Shaded cells denote statistically significant difference. *Denote marginal significance as defined by Johnson (2013).

The paired means t-test results indicate that for four of the seven dimensions, Truth-seeking, Analyticity, Systematicity, and Confidence in Reasoning, 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 dimensions, Open Mindedness, Inquisitiveness, and Maturity of Judgment, 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. Based on the work of Johnson (2013), there is a 17-25% chance that the marginally significant results depicted in Table 4 may be false positives (i.e. Type I errors). The Truth-seeking dimension falls into this category of a marginal result, defined as those within the 95-99% confidence level, or p-value of 0.05 to 0.01.

For the four 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 three dimensions, 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 Summer 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 Systematicity. This suggests the Spring 2014 FSW student is more capable of approaching problem solving in a systematic, ordered, and disciplined method (CCTDI, 2013).

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 3 highlights the percentage of student test scores that improved and declined.
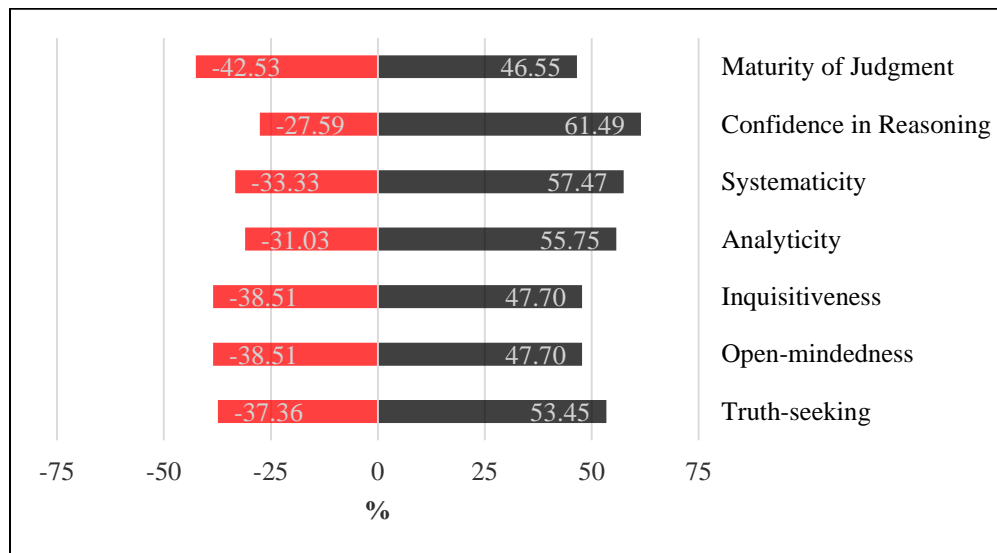

Figure 3. Percentage of students improved vs. declined. Difference of the sum of increase and decline from 100 is the percentage of test takers that exhibited no change.

The learning dimensions of Confidence of Reasoning, Systematicity, 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 61.49%. By comparison, Maturity of Judgment shows the least amount of improvement at 46.55%.

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 4-10.

Figure 4 depicts data distribution of the Truth-seeking dimension which exhibits a shift in the mode (central peak of the data) from 35 to 38, pre- and post-, respectively. When comparing pre-to-post scores, a net decrease in test scores in the 30-34 scoring bin is coupled with increases in scoring bins 40-44 and above.
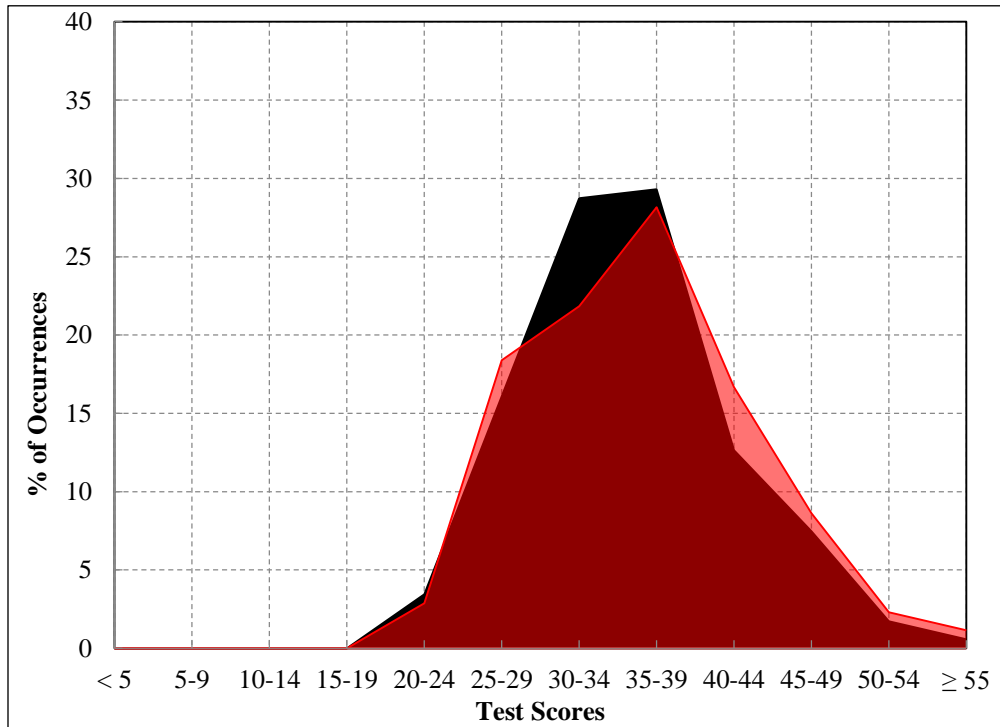

Figure 4. Truth-seeking dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 5 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-to-post scores, a slight net decrease in the 35-39 scoring bin is coupled with increases in scoring bins 45-49 and above.
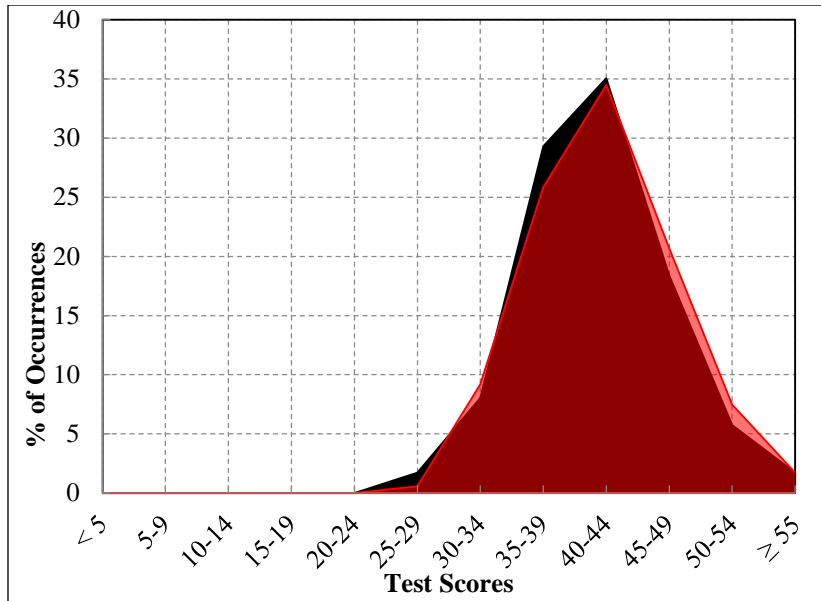
Figure 5. Openmindedness dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 4 depicts data distribution of the Inquisitiveness dimension exhibiting a decrease in kurtosis from pre-to-post test scores. When comparing pre-to-post scores, a net decrease in test scores in the 45-49 and 50-54 scoring bin is coupled with increases in scoring bins 30-34 and below, and ≥ 55, meaning post-test scores both improve and decline substantially across multiple scoring bins. This finding is consistent with a difference in mean scores of only 0.20, the second lowest of all learning dimensions (recall 2.1 for significance tests).
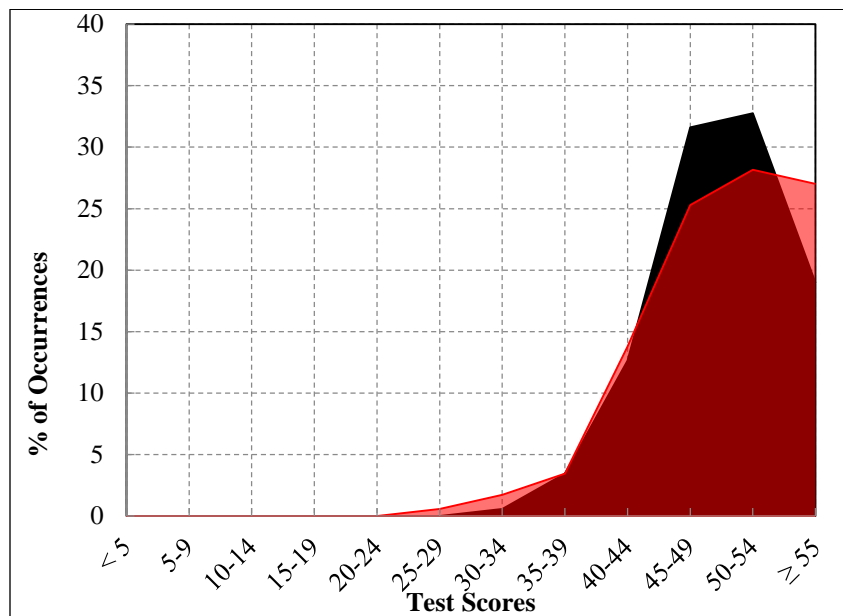

Figure 6. Inquisitiveness dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 7 depicts data distribution of the Analyticity dimension which exhibits a shift in the mode (central peak of the data) from 40-44 bin to 45-49 bin, pre- and post-, respectively. Net decreases in the number of scores in the 35-39 and 40-44 scoring bin correspond to net increases in the 45-49 scoring bin and above.

Figure 8 depicts data distribution of the Systematicity dimension which exhibits a slight decrease in kurtosis (smoother peak from pre-to-post test score distribution). Net decreases in the number of scores in the 30-34 through 40-44 scoring bins are coupled with net increases in post-test scores in both the 25-29 bin and below, as well as the 45-49 scoring bin, and above. This learning dimension, however, exhibited the second largest increase in mean scores so while the distribution in post-test scores appears to be two-tailed (both higher and lower scores), the bulk of the net increase is in the 45-49 scoring bins and above, thus resulting in the statistically significant increase in mean scores (recall 2.1 for significance tests). The two-tailed, yet still positively statistically 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).
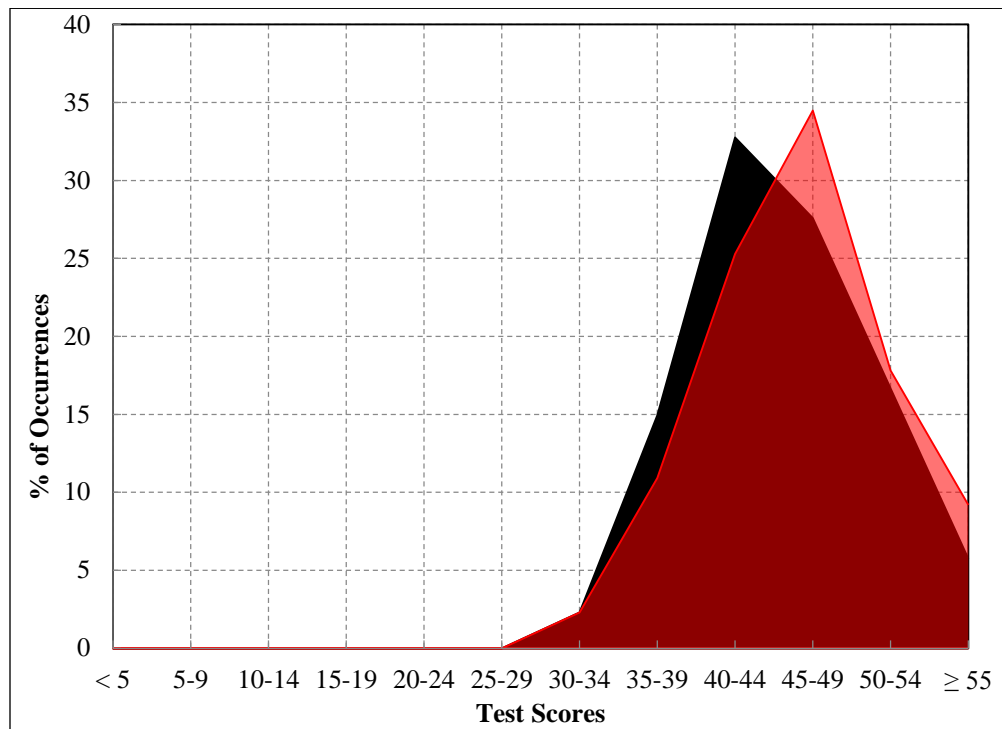

Figure 7. Analyticity dimension distribution of test scores for Pre- (black) and Post- (red).
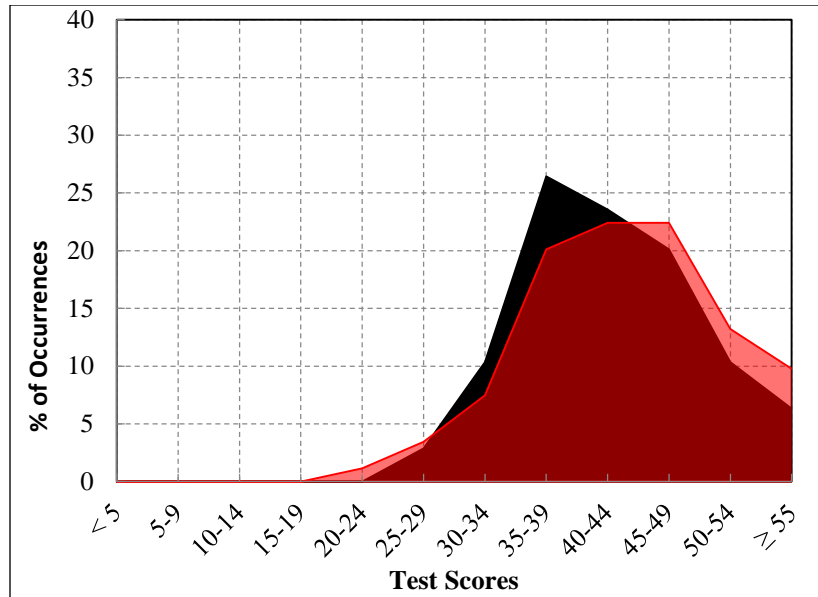
Figure 8. Systematicity dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 9 depicts data distribution of the Confidence in Reasoning dimension. The post-test scores exhibit a shift in modality (central tendency of data) from 44 in pre-test scores to 50. This shift is noticeable in the net scoring bin increases/decreases. Net decreases in the number of scores in the 40-44 through 30-34 scoring bins are coupled with net increases in post-test scores in both the 45-49 bin and above.
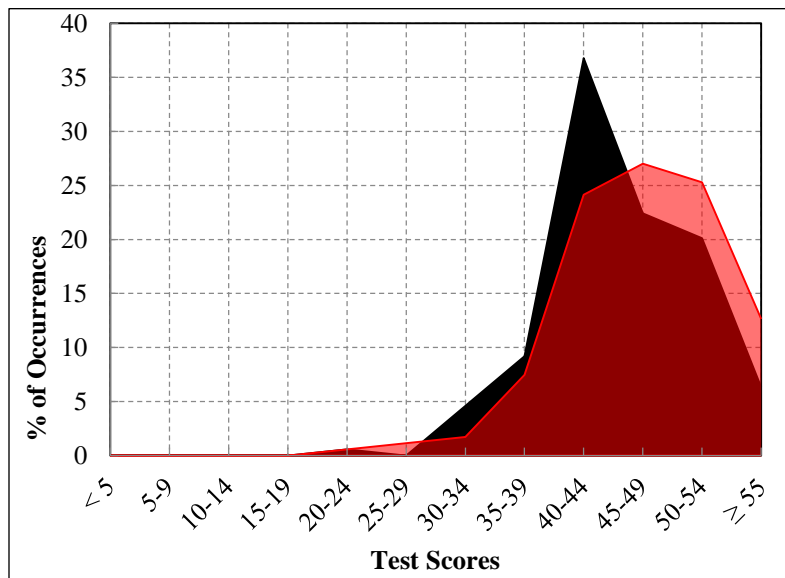

Figure 9. Confidence in Reasoning dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 10 depicts data distribution of the Maturity of Judgment dimension. When comparing pre-to-post scores, a net decrease in test scores in the 40-44 and 45-49 scoring bin is coupled with increases in scoring bins 25-29 and below, and 50-54 and above, meaning post-test scores both improve and decline

substantially across multiple scoring bins. This finding is consistent with a difference in mean scores of only 0.13, the lowest of all learning dimensions (recall 2.1 for significance tests).

Figure 11 depicts the difference from pre-to-post test scores data distribution using 3-point mean difference bins. For example, a change from pre-to-post of 0 points is included in the percentage in the (-1)-1 (-1, 0, 1) scoring bin. Mode shifts for Analyticity, Systematicity, and Confidence in Reasoning learning dimensions are readily apparent (mode appears to center between the (-1)-1 and the 2-4 mean difference bins). This is consistent with significance testing which yielded positive results for all three. Truth-seeking was the fourth dimension which was found to be statistically significant, although from the mean difference distribution this is not readily apparent. The mean score difference in this dimension is supported by a subtle change in skewness towards the negative as modality remains the same.
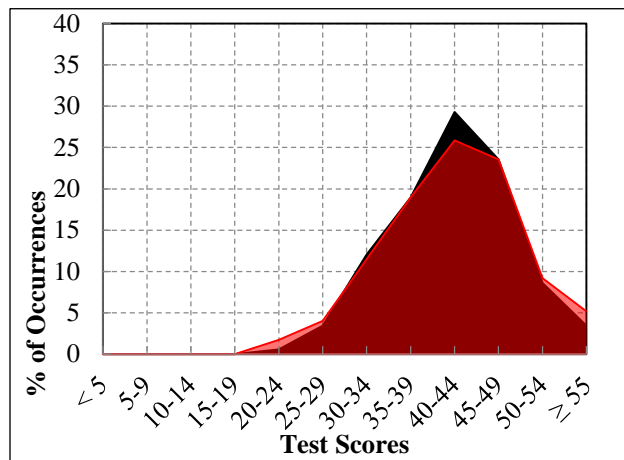

Figure 10. Maturity of Judgment dimension distribution of test scores for Pre- (black) and Post- (red).
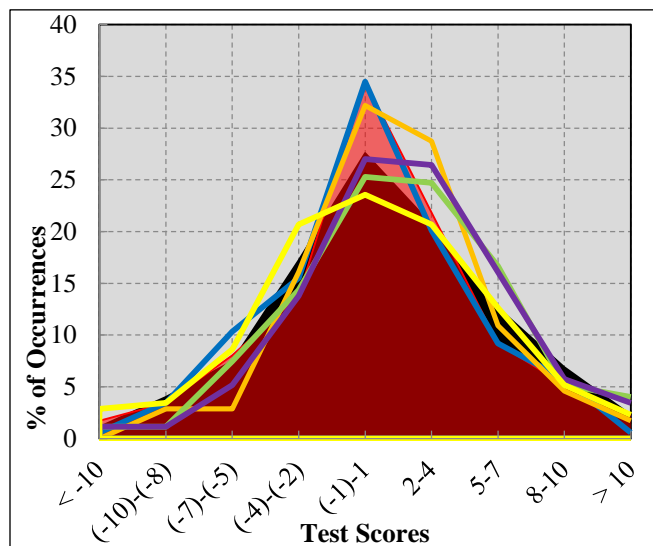

Figure 11. Distribution of mean score differences from pre-to-post test scores for all dimensions with 3-point scoring bins centered on 0 (-1 to 1). Truth-seeking (black area), Open-mindedness (red area), Inquisitiveness (blue line), Analyticity (orange line), Systematicity (green line), Confidence in Reasoning (purple line), and Maturity of Judgment (yellow line).

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 Summer 2014 are shown in Table 3. With the exception of Inquisitiveness, 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. 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. 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.

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, Fall 2013 Analyticity, and Summer 2014 Truth-seeking.

	Fall '12	Spring '13	Summer '13	Fall '13	Spring '14	Summer '14
Truth-seeking	1.09	0.40	0.81	0.48	1.10	0.80*
Open Mindedness	0.71	0.72*	0.40	0.01	0.84	0.49
Inquisitiveness	0.70	0.08	0.38	-0.56	0.03	0.20
Analyticity	1.01	1.15	1.12	0.37*	1.19	1.18
Systematicity	0.78	0.78*	1.05	0.01	0.80	1.55
Confidence in Reasoning	1.60	1.77	2.64	1.44	1.73	1.78
Maturity of Judgment	1.24	1.09	0.01	-0.11	1.06	0.13

*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 Summer 2014. Shaded cells indicate statistically significant differences in the mean at the 95% confidence level. *Denote marginal significance as defined by Johnson (2013).*

Figure 12 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 (grey line) is clearly visible. The consistent improvement of the Analyticity (blue line) is also apparent. The marginally significant result of Fall 2013 can be seen here as a noticeable drop with respect to other semesters.

By comparison, the learning dimension with minimal results across all semesters is Inquisitiveness. Shown graphically, this dimension has the lowest mean score difference across all but one semester (Summer 2013) and exhibited a noticeable decline in the Fall 2013 semester. 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 $t_{crit} = 1.97$
Truth-seeking	$t(365)=4.00$, $p<0.05$	$t(204)=1.09$, $p=0.275$	$t(145)=1.71$, $p=0.090$	$t(859)=2.69$, $p=0.007$	$t(407)=3.91$, $p=1.09\times10^{-4}$	$t(173)=2.14$, $p=0.034*$
Open Mindedness	$t(365)=2.67$, $p<0.05$	$t(204)=2.24$, $p=0.026*$	$t(145)=0.94$, $p=0.347$	$t(859)=0.07$, $p=0.9414$	$t(407)=3.36$, $p=8.56\times10^{-4}$	$t(173)=1.34$, $p=0.183$
Inquisitiveness	$t(365)=2.40$, $p<0.05$	$t(204)=0.24$, $p=0.813$	$t(145)=0.95$, $p=0.345$	$t(859)=-3.10$, $p=0.002$	$t(407)=2.83$, $p=0.907$	$t(173)=0.57$, $p=0.570$
Analyticity	$t(365)=4.18$, $p<0.05$	$t(204)=3.46$, $p=0.0007$	$t(145)=2.92$, $p=0.004$	$t(859)=2.26$, $p=0.024*$	$t(407)=5.00$, $p=8.72\times10^{-7}$	$t(173)=3.84$, $p=1.70\times10^{-4}$
Systematicity	$t(365)=2.81$, $p<0.05$	$t(204)=2.08$, $p=0.039*$	$t(145)=2.69$, $p=0.008$	$t(859)=0.05$, $p=0.963$	$t(407)=2.83$, $p=0.005$	$t(173)=4.19$, $p=4.44\times10^{-5}$
Confidence in Reasoning	$t(365)=5.97$, $p<0.05$	$t(204)=5.28$, $p<0.001$	$t(145)=5.79$, $p<0.001$	$t(859)=7.71$, $p<0.001$	$t(407)=6.02$, $p=2.95\times10^{-9}$	$t(173)=4.98$, $p=1.52\times10^{-6}$
Maturity of Judgment	$t(365)=3.73$, $p<0.05$	$t(204)=2.89$, $p=0.004$	$t(145)=0.03$, $p=0.980$	$t(859)=-0.54$, $p=0.590$	$t(407)=3.47$, $p=5.79\times10^{-4}$	$t(173)=0.31$, $p=0.755$

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).
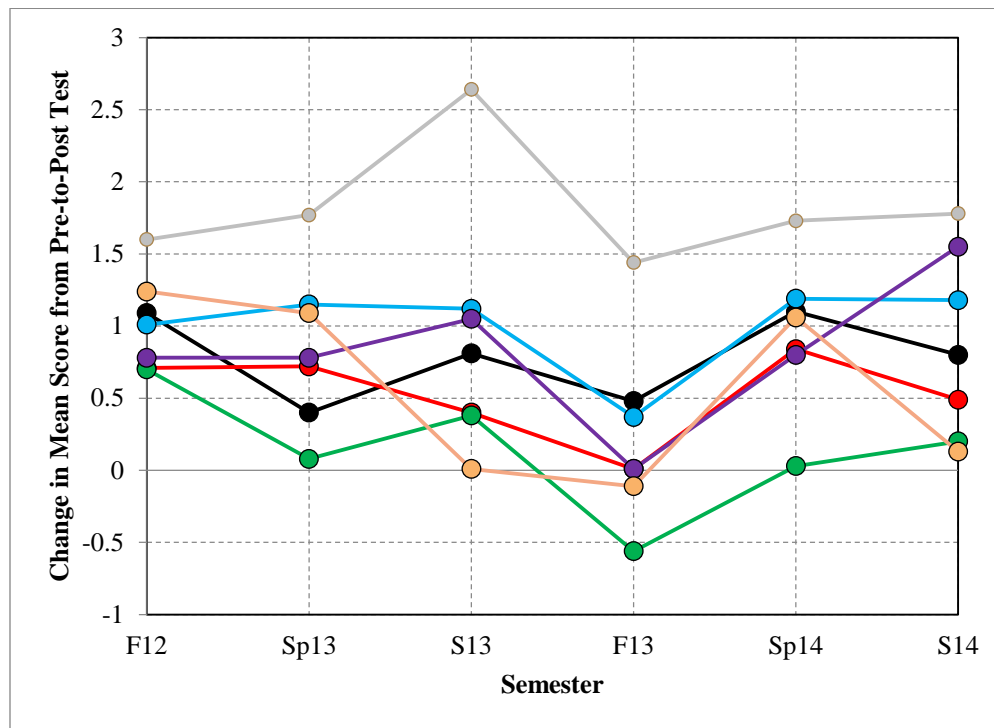

Figure 12. Comparison of the difference in mean scores from pre-to-post tests across semester by learning dimension. Black (Truth-seeking), red (Open-mindedness), green (Inquisitiveness), blue (Analyticity), purple (Systematicity), grey (Confidence in Reasoning), and orange (Maturity of Judgment).

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 of Judgment, and are measured using the California Critical Thinking Disposition Inventory tests (CCTDI).

For Summer 2014, the paired means t-test results indicate that for four of the seven dimensions, Truth-seeking, Analyticity, Systematicity, and Confidence in Reasoning, we can conclude with a 95% confidence that the improvement in mean scores are not solely due to chance. Although there were increases across the remaining dimensions (Open-mindedness, Inquisitiveness, and Maturity of Judgment), 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 Summer 2014 FSW student, is more readily able to recognize their awareness to problems and think about their situations and more capable of approaching problem solving in a systematic, ordered, and disciplined method (CCTDI, 2013). Moreover, distributions of learning dimensions consistently exhibit net growth in approximately one-third of students who score at or near the mode score. In other words, approximately 10-15% of all students show a net growth.

In comparison with earlier studies, it can be shown that, with the exception of Inquisitiveness, since the study began in Fall 2012, all learning dimensions have exhibited statistically significant results in at least three of the five semesters through Summer 2014 with Fall 2012 exhibited statistically significant increases in all dimensions. Summer and Fall 2013 exhibited the weakest scores with both yielding 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.

CCTDI. 2013. 2013 CCTDI Manual. San Jose, California: Insight Assessment, Division of California Academic Press.

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.

Siegel, S. 1956. Nonparametric statistics for the behavior sciences. McGraw-Hill, New York, New York, 312 pp.

Starkweather, J. D. 2010. Introduction to Statistics for the Social Sciences. In: Research and Statistical Support. Retrieved from http://www.unt.edu/rss/class/Jon/ISSS_SC/.

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