

# CCTDI Assessment Report

## Fall 2018

Author: Joseph F. van Gaalen, Ph.D., Asst. VP, Institutional Research, Assessment & Effectiveness

### 1 INTRODUCTION

---

The Academic Success Department's 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). The Academic Success Department has identified a set criterion for defining student advancement in SLS 1515 *Cornerstone Experience*. The results of the overall means 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 began as part of the FSW QEP and continues as course level assessment for SLS 1515 *Cornerstone Experience*.

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, Asst. VP of Institutional Research, Assessment and Effectiveness ([Joseph.VanGaalens@fsw.edu](mailto:Joseph.VanGaalens@fsw.edu); x16965).

### 2 STATISTICS

---

#### 2.1 DESCRIPTIVE STATISTICS AND SIGNIFICANCE TESTING

During the fall 2018 semester, 733 total tests (pre- and post-) were administered to students. Of those, 236 of which were pre-/post- paired tests and 261 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. Four of seven dimensions exhibit improvements. 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.

The paired means t-test results indicate that for the decrease in Inquisitiveness and the increase in Confidence in Reasoning, the results are statistically significantly different. In other words, 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. For Confidence in Reasoning, and the overall score, which exhibits a statistically significant increase in mean scores, 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 dimensions, this cannot be stated. However, for Inquisitiveness, the reverse appears to be the case.

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 0.00 to 0.26 (Table 1). In other words, non-overlap from pre-test scores to post-test scores range from approximately 0% to 20%.

	Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity of Judgment	Overall
Pre-test Mean	33.9	40.3	46.9	44.0	40.0	43.8	38.7	287.4
Post-test Mean	34.0	40.4	45.8	44.1	39.9	45.3	37.8	287.2
Effect size	0.02	0.03	-0.21	0.05	-0.01	0.26	-0.14	0.00
p-value	0.805	0.649	<b>0.001</b>	0.402	0.918	<b>8.5x10<sup>-5</sup></b>	<b>0.031</b>	0.971

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.

## 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 highlights the percentage of student test scores that improved and declined. The attributes of Confidence in Reasoning and Open-Mindedness exhibit the greatest improvement percentages at 55.6% and 50.4%, respectively. Maturity of Judgment and Inquisitiveness exhibit greater declines than improvement.

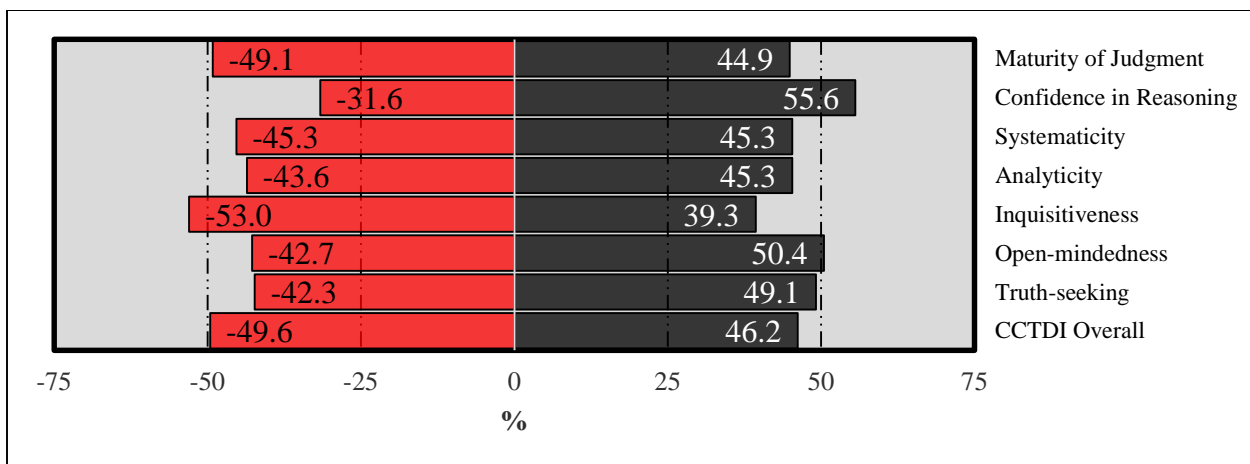


Figure 1. Percentage of students improved (black) vs. declined (red). Difference of the sum of increase and decline from 100 is the percentage of test takers that exhibited no change.

An empirical distribution (histogram) of each dimension is reported in Figures 2 through 8. Figure 2 depicts data distribution of the Truth-seeking dimension which exhibits a slight decrease in kurtosis (peakedness of the curve) coupled with a slight peak shift from 30-34 to 35-39.

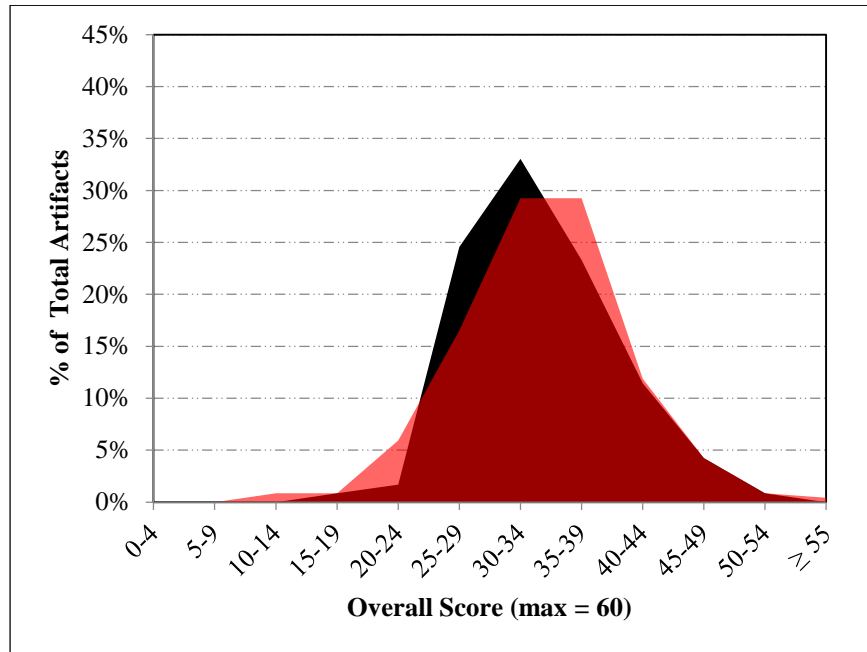


Figure 2. Truth-seeking dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 3 depicts data distribution of the Open Mindedness. Post-test results exhibit a decreased kurtosis. Scores also do not appear to exhibit any shift from the post-test compared with the pre-test.

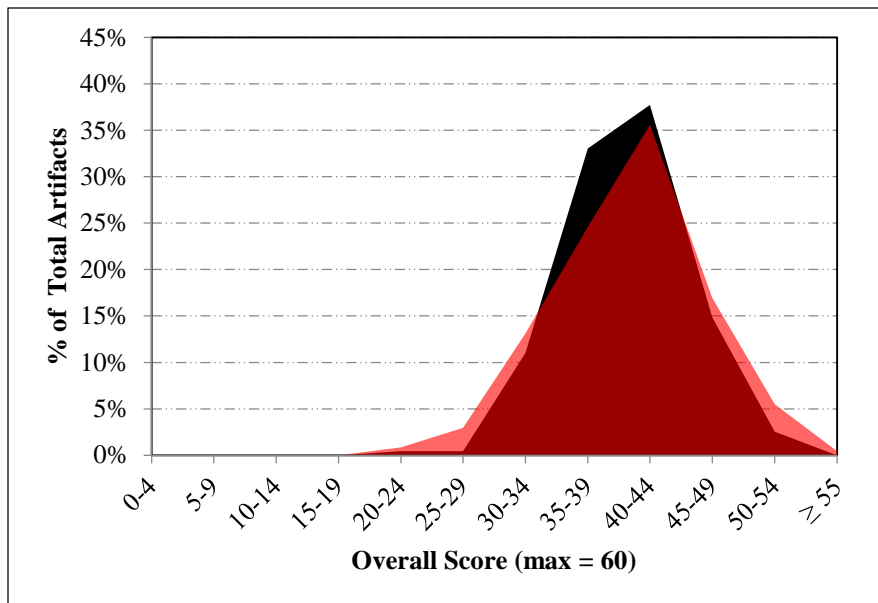


Figure 3. Open Mindedness dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 4 depicts data distribution of the Inquisitiveness dimension. Results of the post-test exhibit a decreased kurtosis (peakedness) with data tending toward lower values. As Inquisitiveness consistently

exhibits the highest pre-test scores of any of the attributes, historically it is not uncommon to see little change from pre-to-post test scores in FSW studies typically the lowest change of the seven attributes (see Section 2.3 Longitudinal Studies).

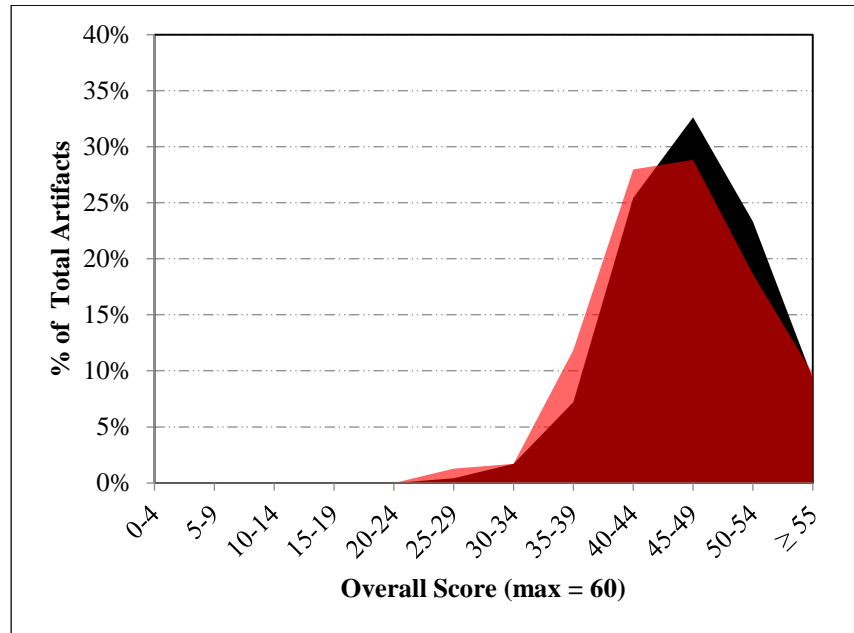


Figure 4. Inquisitiveness dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 5 depicts data distribution of the Analyticity. Results exhibit an increased kurtosis (peakedness) with data tending toward higher values. When comparing pre-to-post scores, a net decrease in test scores in the 30-34 and 50-54 scoring bins is coupled with an increase in scoring bin 45-49 and higher.

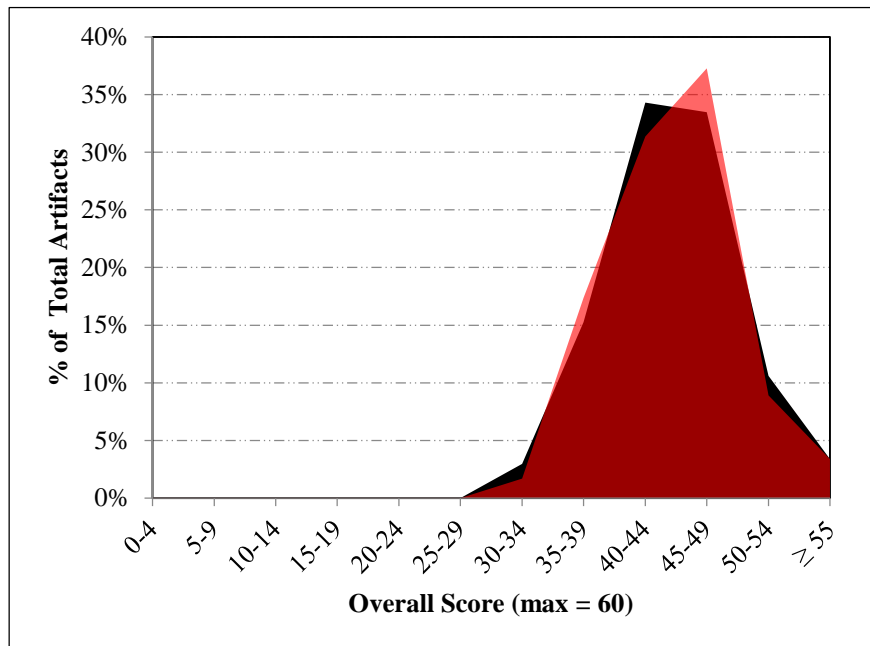


Figure 5. Analyticity dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 6 depicts data distribution of the Systematicity dimension exhibiting no discernable shift in mode (central peak of the data) from pre-test to post-test. The distribution does exhibit the formation of a bimodality at 35-39 and 45-49 from pre-test to post-test. Similar cases in which systematicity did not significantly change following early testing have been reported (Beser and Kissal, 2009).

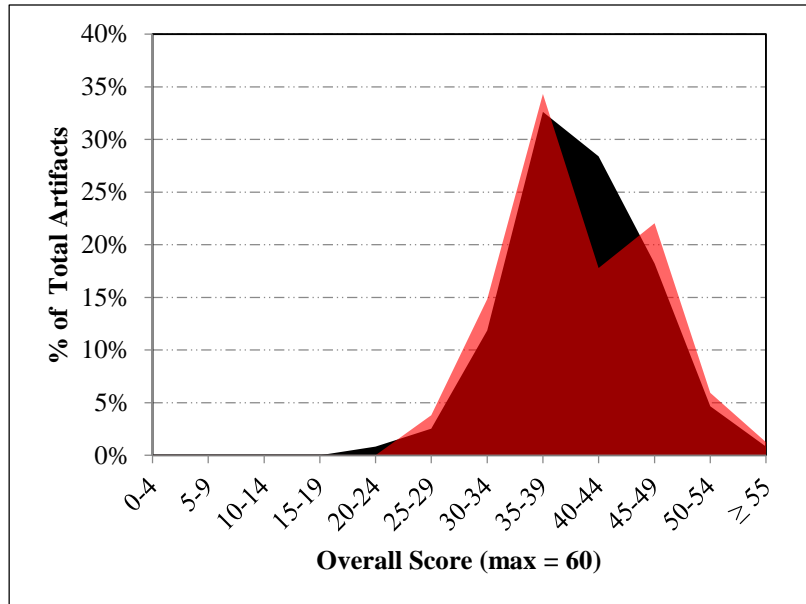


Figure 6. Systematicity dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 7 depicts data distribution of the Confidence in Reasoning dimension. The post-test scores exhibit a slight decrease in kurtosis (peakedness of curve). The post-test exhibits decreases in 30-34 and 35-39 scoring bins coupled with increases in scoring bins 45-59 and above.

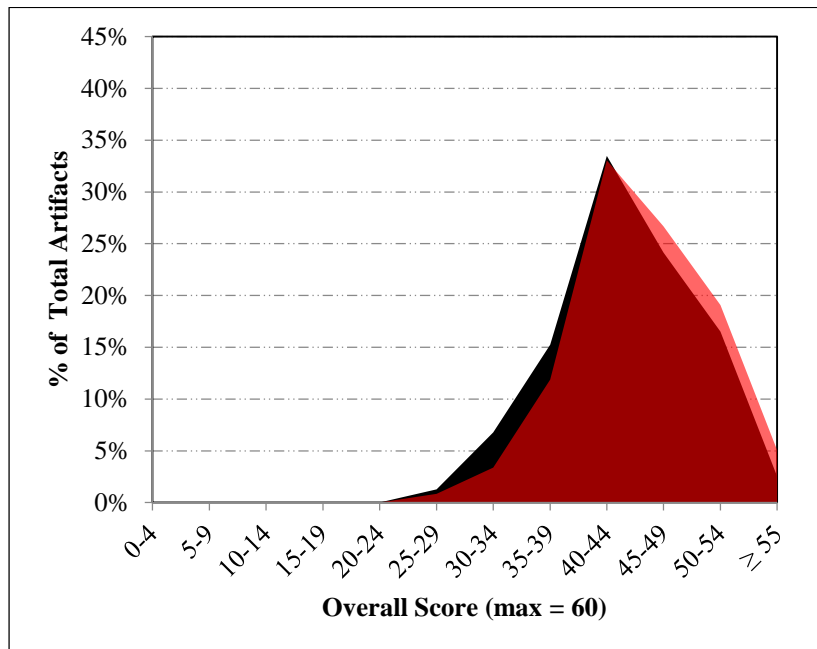


Figure 7. Confidence in Reasoning dimension distribution of test scores for Pre- (black) and Post- (red).

Figure 8 depicts data distribution of the Maturity of Judgment. The post-test distribution exhibits a decrease in kurtosis (peakedness). However, this decrease is coupled with a larger spread of data in the lower range (scoring bins 15-19 through 25-29).

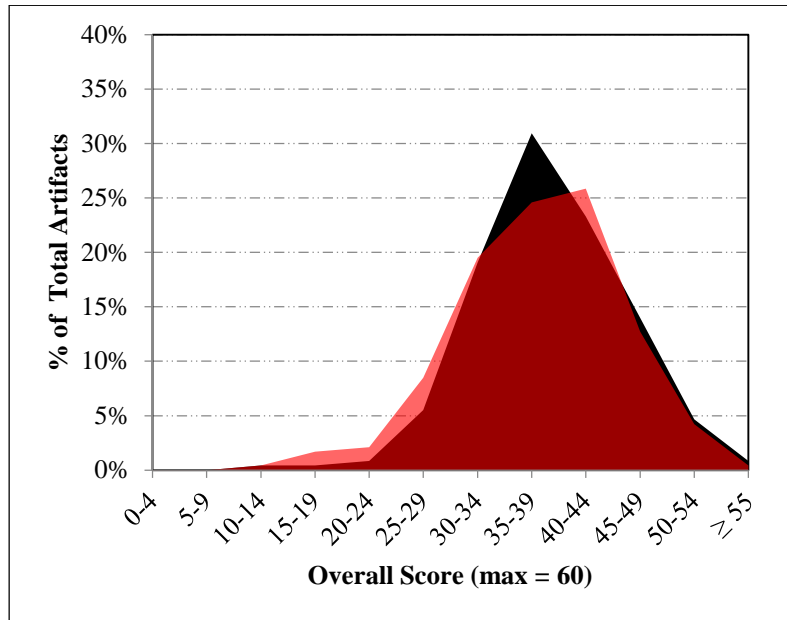


Figure 8. Maturity of Judgment dimension distribution of test scores for Pre- (black) and Post- (red).

### 2.3 LONGITUDINAL STUDY

The results of paired means t-test of pre-/post- test scores for all semesters from fall 2012 through fall 2018 are shown in Table 2. Analyticity and Confidence in Reasoning exhibit improvements from pre-test to post-test in all terms, although not all improvements are statistically significant. The Confidence in Reasoning attribute has exhibited statistically significant improvements in all but one term since implementation (14 of 15 terms). Analyticity exhibits a similar rate, with statistically significant improvement in 10 of 15 terms. The Truth-seeking attribute also demonstrates strong results with 7 of 15 terms exhibiting statistically significant improvement. Open Mindedness consistently exhibits positive changes from pre-test to post-test, although only 5 of 15 terms are statistically significant.

	Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity of Judgment
Fall '12	1.1	0.7	0.7	1	0.8	1.6	1.2
Spring '13	0.4	0.7*	0.1	1.2	0.8*	1.8	1.1
Summer '13	0.8	0.4	0.4	1.1	1.1	2.6	0.0
Fall '13	0.5	0.0	-0.6	0.4*	0.0	1.4	-0.1
Spring '14	1.1	0.8	0.0	1.2	0.8	1.7	1.1
Summer '14	0.8*	0.5	0.2	1.2	1.6	1.8	0.1
Fall '14	0.8	0.4*	-0.3	0.7	0.2	1.7	-0.2*
Spring '15	0.9	1.0	0.1	1.3	0.4*	2.3	-0.0
Fall '15 <sup>#</sup>	0.0	0.0	-0.4	0.5	-0.5	1.3	0.3
Spring '16 <sup>#</sup>	0.8	0.4	0.0	1.1	0.6	2.0	0.3
Fall '16 <sup>#</sup>	0.1	0.0	-0.1	0.5	-0.5	2.0	-0.3
Spring '17 <sup>#</sup>	-0.3	0.1	-0.2	0.8	0.9	2.1	0.0
Fall '17 <sup>#</sup>	1.5*	0.3	-0.6	0.3	0.3	0.5	1.4*
Spring '18 <sup>#</sup>	0.5	0.5	0.1	1.5	0.6	2.3	-0.6
Fall '18 <sup>#</sup>	0.1	0.1	-1.0	0.2	0.0	1.5	-0.9

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 Spring 2018. 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). <sup>#</sup>Indicate scores originate from a random sample of the full SLS 1515 population.

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 for this study has not been consistent historically (Rosenthal and Rosnow, 1991; Dunlop et al., 1996).

	Truth-seeking	Open Mindedness	Inquisitiveness	Analyticity	Systematicity	Confidence in Reasoning	Maturity of Judgment
Fall 2012 $t_{crit} = 1.97$	t(365)=4.00, p<0.05	t(365)=2.67, p<0.05	t(365)=2.40, p<0.05	t(365)=4.18, p<0.05	t(365)=2.81, p<0.05	t(365)=5.97, p<0.05	t(365)=3.73, p<0.05
Spring 2013 $t_{crit} = 1.97$	t(204)=1.09, p=0.275	t(204)=2.24, p=0.026*	t(204)=0.24, p=0.813	t(204)=3.46, p=0.0007	t(204)=2.08, p=0.039*	t(204)=5.28, p<0.001	t(204)=2.89, p=0.004
Summer 2013 $t_{crit} = 1.98$	t(145)=1.71, p=0.090	t(145)=0.94, p=0.347	t(145)=0.95, p=0.345	t(145)=2.92, p=0.004	t(145)=2.69, p=0.008	t(145)=5.79, p<0.001	t(145)=0.03, p=0.980
Fall 2013 $t_{crit} = 1.97$	t(859)=2.69, p=0.007	t(859)=0.07, p=0.941	t(859)=3.10, p=0.002	t(859)=2.26, p=0.024*	t(859)=0.05, p=0.963	t(859)=7.71, p<0.001	t(859)=0.54, p=0.590
Spring 2014 $t_{crit} = 1.97$	t(407)=3.91, p=1.09x10 <sup>-4</sup>	t(407)=3.36, p=8.56x10 <sup>-4</sup>	t(407)=2.83, p=0.907	t(407)=5.00, p=8.72x10 <sup>-7</sup>	t(407)=2.83, p=0.005	t(407)=6.02, p=2.95x10 <sup>-9</sup>	t(407)=3.47, p=5.79x10 <sup>-4</sup>
Summer 2014 $t_{crit} = 1.97$	t(173)=2.14, p=0.034*	t(173)=1.34, p=0.183	t(173)=0.57, p=0.570	t(173)=3.84, p=1.70x10 <sup>-4</sup>	t(173)=4.19, p=4.44x10 <sup>-5</sup>	t(173)=4.98, p=1.52x10 <sup>-6</sup>	t(173)=0.31, p=0.755
Fall 2014 $t_{crit} = 1.97$	t(1848)=5.82, p=7.88x10 <sup>-8</sup>	t(1848)=2.57, p=6.74x10 <sup>-9</sup>	t(1848)= -2.78, p=0.010	t(1848)=6.01, p=0.005	t(1848)=0.55, p=2.30x10 <sup>-9</sup>	t(1848)=13.80, p=2.77x10 <sup>-41</sup>	t(1848)= -2.20, p=0.028*
Spring 2015 $t_{crit} = 1.97$	t(764)=4.36, p=7.44x10 <sup>-12</sup>	t(764)=5.46, p=1.51 x10 <sup>-5</sup>	t(764)=0.51, p=0.608	t(764)=7.58, p=9.82x10 <sup>-14</sup>	t(764)=2.26, p=0.024	t(764)=11.79, p=1.41x10 <sup>-29</sup>	t(764)= -0.06, p=0.954
Fall 2015 $t_{crit} = 1.97$	t(268)=0.03, p=0.974	t(268)=0.08, p=0.936	t(268)= -1.11, p=0.178	t(268)=1.61, p=0.108	t(268)= -1.65, p=0.101	t(268)=4.11, p=5.26x10 <sup>-5</sup>	t(268)=0.80, p=0.422
Spring 2016 $t_{crit} = 1.97$	t(148)=1.91, p=0.058	t(148)=1.18, p=0.239	t(148)=0.11, p=0.912	t(148)=2.93, p=0.004	t(148)=1.35, p=0.180	t(148)=5.13, p=9.05x10 <sup>-5</sup>	t(148)=0.78, p=0.436
Fall 2016 $t_{crit} = 1.97$	t(274)=0.34, p=0.737	t(274)= -0.16, p=0.870	t(274)= -0.37, p=0.710	t(274)=1.62, p=0.106	t(274)= -1.43, p=0.154	t(274)=6.52, p=3.46x10 <sup>-10</sup>	t(274)= -0.96, p=0.336
Spring 2017 $t_{crit} = 1.98$	t(140)= -0.51, p=0.608	t(140)= -0.30, p=0.766	t(140)= -0.42, p=0.678	t(140)=1.85, p=0.067	t(140)=1.78, p=0.078	t(140)=4.20, p=4.65x10 <sup>-5</sup>	t(140)= -0.04, p=0.972
Fall 2017 $t_{crit} = 1.97$	t(175)=2.22, p=0.028*	t(175)=0.53, p=0.599	t(175)= -0.88, p=0.382	t(175)=0.61, p=0.546	t(175)=0.38, p=0.702	t(175)=0.80, p=0.427	t(175)=1.99, p=0.049*
Spring 2018 $t_{crit} = 1.97$	t(139)=1.19, p=0.234	t(139)=1.11, p=0.271	t(139)=0.12, p=0.906	t(139)=3.77, p=0.0002	t(139)=1.11, p=0.268	t(139)=4.96, p=2.01x10 <sup>-6</sup>	t(139)= -1.16, p=0.247
Fall 2018 $t_{crit} = 1.97$	t(233)=0.25, p=0.805	t(233)=0.46, p=0.649	t(233)= -3.23, p=0.001	t(233)=0.84, p=0.402	t(233)= -0.10, p=0.918	t(233)=4.00, p=8.53x10 <sup>-5</sup>	t(233)= -2.18, p=0.31

Table 3. Additional significance testing statistics for attributes 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 is a graphical representation of the difference in mean scores of attributes across all semesters. The consistently strong improvement in scores from pre-to-post tests of the Confidence in Reasoning attribute is clearly visible (see consistent red zones in Figure 9). Analyticity exhibits the second strongest improvement over time followed by Truth-seeking.

By comparison, the attribute with minimal results across all semesters is Inquisitiveness (see consistent blue zones in Figure 9). This dimension has the lowest mean score difference across 12 of 15 semesters during the study and exhibited a decline from pre-to-post scores in the fall 2013, fall 2014, fall 2015, fall 2016, spring 2017, fall 2017, and fall 2018 semesters.

In terms of variability, Systematicity exhibits the widest variation of any indicator, as high as +1.6 in summer 2014, and as low as -0.5 in both fall 2013 and fall 2016, a range of 2.1. The Inquisitiveness attribute also exhibits substantial variability from as high as +1.4 in summer 2013 to as low as -1.0 in fall 2018, a range of 2.3. Analyticity is the most stable of all attributes with a range of just 1.3 from highest (+1.5) in spring 2018 to lowest (+0.2) in fall 2018.

	Open					Confidence in	Maturity in
	Truth-seeking	Mindedness	Inquisitiveness	Analyticity	Systematicity	Reasoning	Judgment
Fall 2012	1.1	0.7	0.7	1.0	0.8	1.6	1.2
Spring 2013	0.4	0.7	0.1	1.2	0.8	1.8	1.1
Summer 2013	0.8	0.4	0.4	1.1	1.1	2.6	0.0
Fall 2013	0.5	0.0	-0.6	0.4	0.0	1.4	-0.1
Spring 2014	1.1	0.8	0.0	1.2	0.8	1.7	1.1
Summer 2014	0.8	0.5	0.2	1.2	1.6	1.8	0.1
Fall 2014	0.8	0.4	-0.3	0.7	0.2	1.7	-0.2
Spring 2015	0.9	1.0	0.1	1.3	0.2	2.3	0.0
Fall 2015	0.0	0.0	-0.4	0.5	-0.5	1.3	0.3
Spring 2016	0.8	0.4	0.0	1.1	0.6	2.0	0.3
Fall 2016	0.1	0.0	-0.1	0.5	-0.5	2.0	-0.3
Spring 2017	-0.3	0.1	-0.2	0.8	0.9	2.1	0.0
Fall 2017	1.5	0.3	-0.6	0.3	0.3	0.5	1.4
Spring 2018	0.5	0.5	0.1	1.5	0.6	2.3	-0.6
Fall 2018	0.1	0.1	-1.0	0.2	0.0	1.5	-0.9

Scale (+/-) from Pre-test to Post-test

-3.0
0.0
3.0

Figure 9. Colormap comparison of the difference in mean scores from pre-to-post tests across semester by attribute. Color bar represents maximum range of changes in the mean from pre-test to post-test with strongest positive changes in the mean denoted by darker reds (max = +3) and strongest negative changes in the mean denoted by darker blues (max = -3).



The attributes of the CCTDI exhibit a wide range of scores from pre-test to post-test. In some cases, the post-test scores of one attribute are lower than the pre-test scores of another. A review of this comparison can be helpful in determining if there is further room for improvement in that attribute as well as if minimal improvements are a result of an already strong attribute score. Figure 10 compares time-averaged pre-test mean scores from fall 2012 through the present to time-averaged post-test mean scores. Note that the highest pre-test scoring attribute, Inquisitiveness, also exhibits the smallest improvement from pre-to-post test with an improvement that barely registers on the graph. By comparison, the second highest pre-test attribute, Confidence in Reasoning, exhibits the largest improvement. For more clarity, see Figure 11 which displays time-averaged pre-test scores with time-averaged improvements from pre-test to post-test.

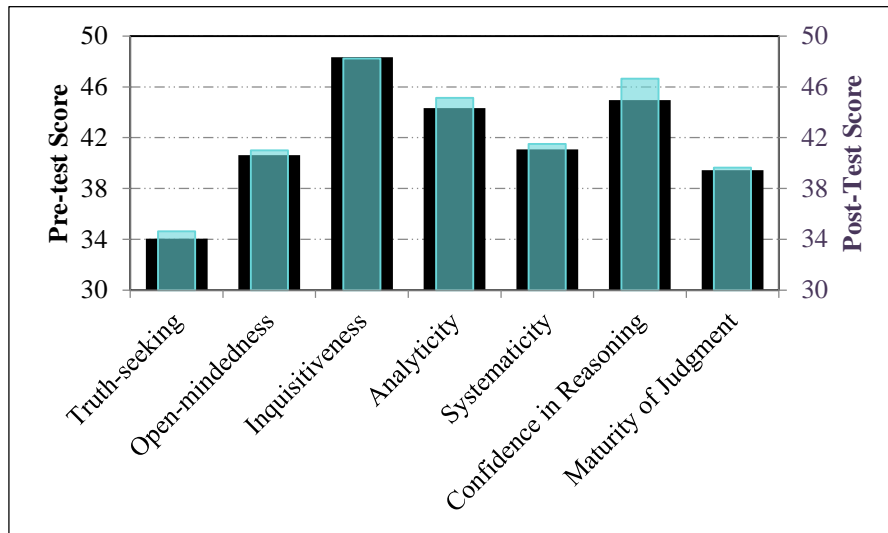


Figure 10. Average through time of the CCTDI mean scores from pre-test (black) to post-test (aqua) from fall 2012 through the present.

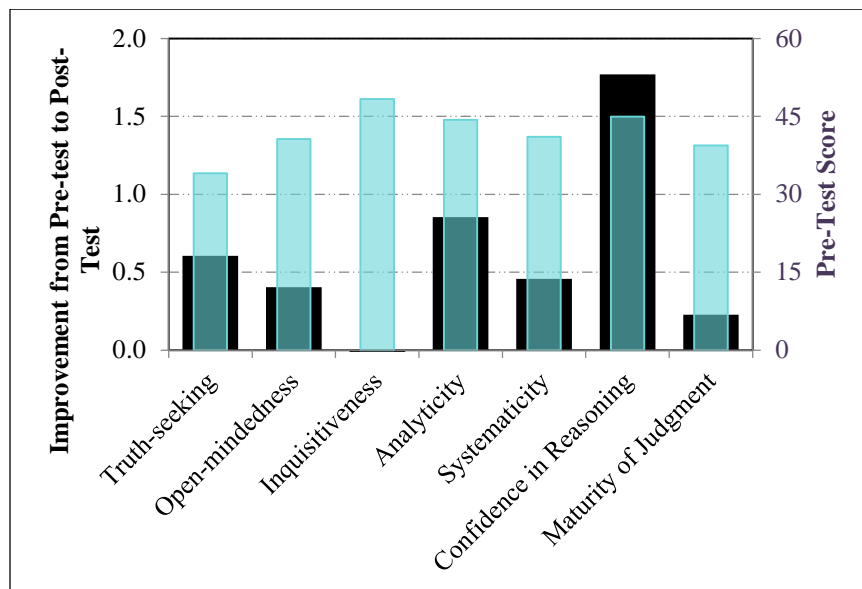


Figure 11. Time-averaged CCTDI mean improvement from pre-to-post test (black) compared with time-averaged mean scores from pre-test (aqua).

### 3 CONCLUSIONS

---

In the SLS 1515 *Cornerstone Experience* course-level 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 as a result of the 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 2018 assessment are as follows:

1. The means of post-test scores exhibit statistically significant improvement over means of pre-test scores in the Confidence in Reasoning attributes, although four of seven attributes exhibit improvement.
2. In a comparison of students who improved from pre-to-post and those that didn't, 5 of 7 attributes exhibit a greater population of students who improved from pre-to-post test scores than declined. Confidence in Reasoning and Open-Mindedness exhibit the greatest improvement percentages at 55.6% and 50.4%, respectively.
3. In a longitudinal study, since comparisons began in fall 2012 the three attributes that have consistently exhibited the largest improvement in scores from pre-/post-test scores are 1) Confidence in Reasoning, 2) Analyticity and 3) Truth-seeking, in that order. Both Confidence in Reasoning and Analyticity have exhibited increases in mean score across all semesters.
4. In the same longitudinal study, the attribute with minimal results across all semesters is Inquisitiveness and exhibits a decline from pre-to-post scores in fall 2013, fall 2014, spring 2015, fall 2015, fall 2016, spring 2017, fall 2017, and fall 2018.
5. In the same longitudinal study, a comparison between time-averaged pre-test mean scores from fall 2012 through the present to time-averaged post-test mean scores exhibits the highest pre-test scoring attribute, Inquisitiveness, also exhibits the smallest improvement from pre-to-post test with an improvement that barely registers on the graph. Meanwhile, the second highest pre-test attribute, Confidence in Reasoning, exhibits the largest improvement.

### 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* (2<sup>nd</sup> 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.