SmarterMeasure 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 provide students ways to develop strategies for effective writing and verbal communication and demonstrate independence through personal management, use of college resources, and positive relationships with peers, staff, and faculty. One measurement for the achievement of that goal is the use of the SmarterMeasure Learning Readiness Indicator. FSW has identified set criteria for defining student advancement in the Cornerstone Experience course through the published assessment tool SmarterMeasure.

The SmarterMeasure Learning Readiness Indicator was designed to help quantify the degree of student preparedness and breadth of skill set (SmarterServices, 2011). Through course completion, students are expected to statistically significantly improve in the following learning readiness indicators of the SmarterMeasure Student Readiness Test: Personal Attributes, Life Factors, Technology Knowledge, and Technology Competency. These indicators can serve as an effective tool for evaluating growth over time, allowing college assessment faculty and staff a means to identify strengths and weaknesses, and allow the college to compare FSW students with those across the country, if desired (Suskie, 2004). Results are compared by term in a longitudinal study in support of the 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 (Sessions A & B), 434 total tests (pre- and post-) were administered to students. Of those, between 189 and 191 of them were pre-/post- paired tests. Some students did not have pre-/post- counterparts for all readiness indicators. Depending on the readiness indicator, between 52 and 56 tests did not have counterparts. Basic descriptive statistics of pre- and post- test scores are shown in Table 1.

All readiness indicators exhibit increases in means, although not all are significant (see discussion in 2.1 Significance Tests). Personal Attributes and Technology Competency indicators post-test scores exhibit an increased standard deviation (spread of data distribution). Both technology indicators exhibit increased kurtosis, meaning they have become more leptokurtic (distribution of the scores has a steeper peak) (see Figure 1).

The Personal Attributes and Technology Knowledge indicators exhibit a negligible increase in skewness. All indicators both pre- and post-tests, except for Personal Attributes post-test scores exhibit negative skewness. For an example see Figure 2.

		Personal Attributes	**Life Factors**	**Technology Knowledge**	**Technology Competency**
Pre-Test	**n**	191	191	189	189
	mean	75.40	76.65	65.81	87.8
	median	75.00	77.00	65.63	90
	mode	75.00	76.00	64.06	100
	max / min	93.75 / 53.13	99 / 51	96.88 / 31.25	100 / 40
	standard deviation	7.85	9.34	12.96	12.11
	kurtosis	-0.10	-0.11	-0.34	1.72
	skewness	-0.16	-0.32	-0.11	-1.24
Post-Test	**n**	191	191	189	189
	mean	76.19	79.19	71.19	88.09
	median	75.00	81.00	71.88	90
	mode	71.88	81.00	71.88	100
	max / min	96.88 / 53.13	100 / 48	98.44 / 39.06	100 / 30
	standard deviation	7.89	9.14	12.20	13.5
	kurtosis	-0.19	0.35	-0.46	5.88
	skewness	0.06	-0.52	-0.10	-2.03

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.

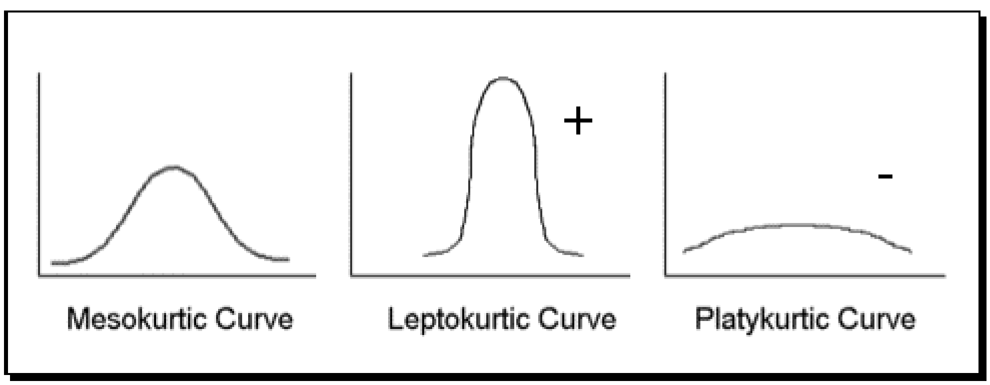

Figure 1. Example of kurtosis. The normal curve (left) has a kurtosis of 0.0. A positive value, or leptokurtic distribution (center) and negative value, or platykurtic distribution (right) are depicted here in an ideal scenario (Starkweather, 2010).

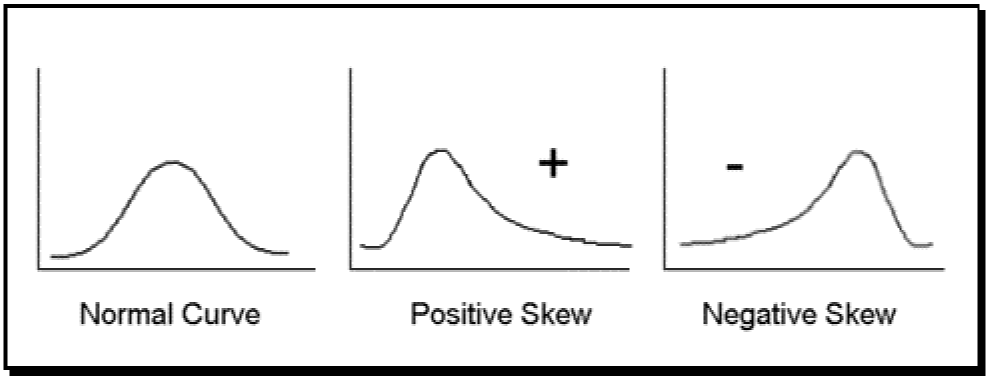

Figure 2. Example of skewness. The normal curve (left) has a skewness of 0.0. A positive value skewness (center) and negative value skewness (right) are depicted here in an ideal scenario (Starkweather, 2010).

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 learning readiness indicator (Personal Attributes, Life Factors, Technology Knowledge, and Technology Competency) 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 readiness indicator 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.

	Personal Attributes	Life Factors	Technology Knowledge	Technology Competency
Difference in mean	0.79	2.53	5.45	0.45
standard deviation	5.75	7.18	8.14	12.76
standard error	0.42	0.52	0.59	0.93
df	191	191	189	189
t_{crit}	1.97	1.97	1.97	1.97
t_{obs}	1.89	4.88	9.20	0.48
p-value	0.060	2.29×10^{-6}	7.09×10^{-17}	0.631

Table 2. Difference between Pre-/Post- results with significance at the α=0.05 level. Shaded cells of the p-value denote statistically significant difference in mean for readiness indicator.

The paired means t-test results indicate that for two of the four learning readiness indicators, Life Factors and Technology Knowledge, 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 with a 95% confidence that the differences in scores are not solely due to chance. With the remaining two learning readiness indicators, Personal Attributes and Technology Competency, we cannot reject the null hypothesis. Therefore, the increases in mean scores from pre-to-pot tests for Personal Attributes and Technology Competency can be a result of chance.

For the two indicators 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 two this cannot be stated or quantified.

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 so 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 readiness indicator 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.

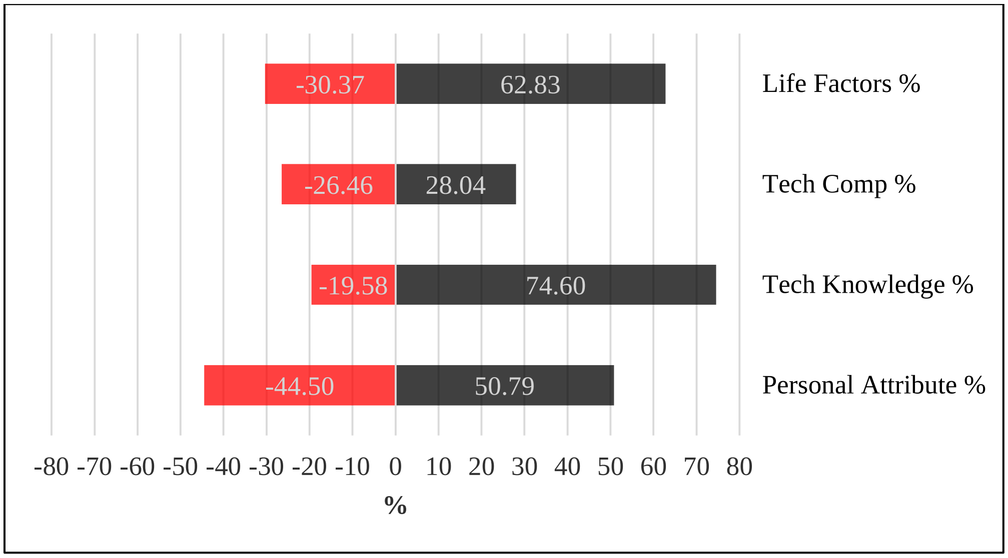

Figure 3.Percentage of students improved vs. declined. Difference of the sum of increase and decline from 100 is the percentage of test takers that exhibited no change.

The Technology Knowledge exhibits the greatest improvement percentage where 74.60% of students improved from pre- to post- test scores. By way of contrast, 19.58% of post-test scores were lower than pre-test scores, and 4.71% showed no change. Both the Life Factors and Technology Knowledge readiness indicators are reflective of their significance tests, with improvement percentages of 62.83%, and 74.60%, respectively. The Technology Competency exhibits a large majority of students displaying no change, 45.50%.

Figure 4 depicts the distribution of pre-test scores using 10-point bins. The Personal Attributes and Life Factors readiness indicators all clearly reflect modality in the 70-79 scoring range (peak of data distribution in 70-79 bin). The Technology Competency reflects a negative skewness (as shown in Section 2: Statistics). This fact, coupled with the data presented in Figure 3 where the indicator exhibited a large percentage of students showing no change is interpreted as a case where students enter the course proficient in the category and acquire little more from the course. The Technology Knowledge indicator exhibits the lowest mode, located in the 60-69 scoring bin. Recall it was Technology Knowledge that exhibited the highest percentage of improved scores and a statistically significant difference in the mean scores from pre-to-post.

Figure 5 depicts the distribution of post-test scores using 10-point bins. The Personal Attributes readiness indicator again reflects a modality in the 70-79 scoring range, while the Life Factors indicator modality has shifted into the 80-89 scoring bin. The Technology Knowledge indicator also reflects a shift to a higher scoring bin from pre-post (60-69 into 70-79) as well as a more defined central peak, although kurtosis does not reflect a large change since adjacent scoring bins are still heavily populated. A clearer review of these changes can be found in Figure 6.

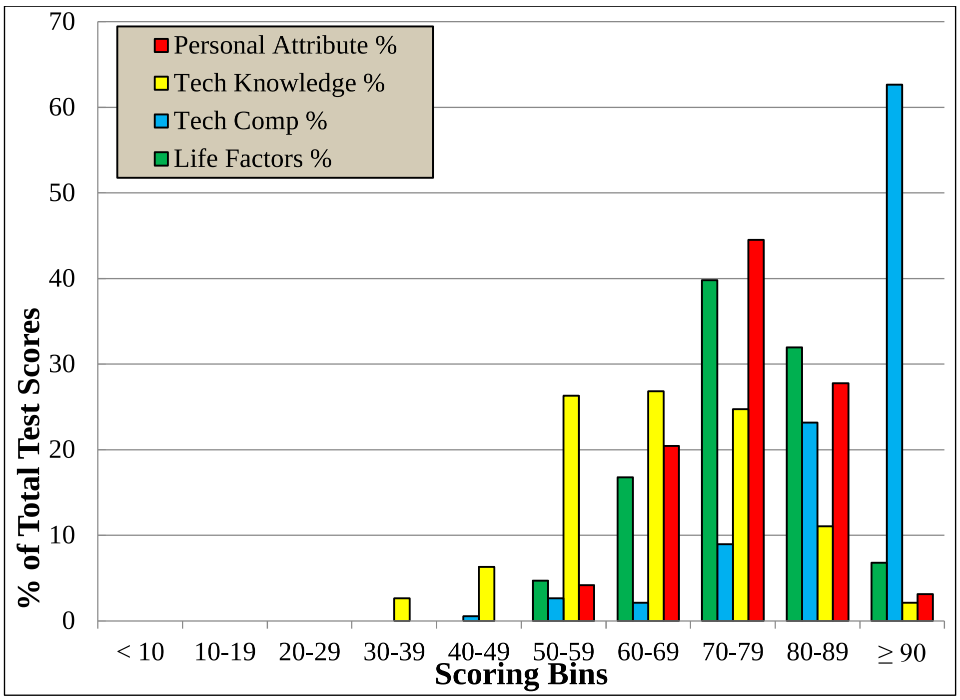

Figure 4. Histogram of Pre-Test scores data distribution binned by 10-point increments.

Figure 6 depicts a plot of the difference in data distribution from pre-to-post test scores. By combining the 10-point scoring bins into three larger scoring bins the changes can be more readily understood. Since a change in one bin means an opposite effect in one of the other scoring bins, with three bins instead of ten, a reduction, for example, in the number of scores falling in one scoring bin must mean an increase in an adjacent bin. The difference between expected change in the adjacent bin and actual change is the result of the input from the third bin. This type of analysis is good for locating large net shifts in data distribution.

The Technology Knowledge indicator exhibits a net decline of 13.16% in the < 60 scoring bin. With no other change, the 60-79 bin would exhibit a net increase of 13.16%. However, the 60-79 scoring bin exhibits a net increase of only 1.05%, the remainder, 12.11% increase is recorded in the ≥ 80 bin. To be clear, this is not to suggest the improved scores from the < 60 bin shifted through the 60-79 bin and directly to the ≥ 80 bin. These changes may also be a result of replacement in the bins. In other words, a shift of score A into the 60-79 bin is accompanied by a shift of score B from the 60-79 bin to the ≥ 80 bin.

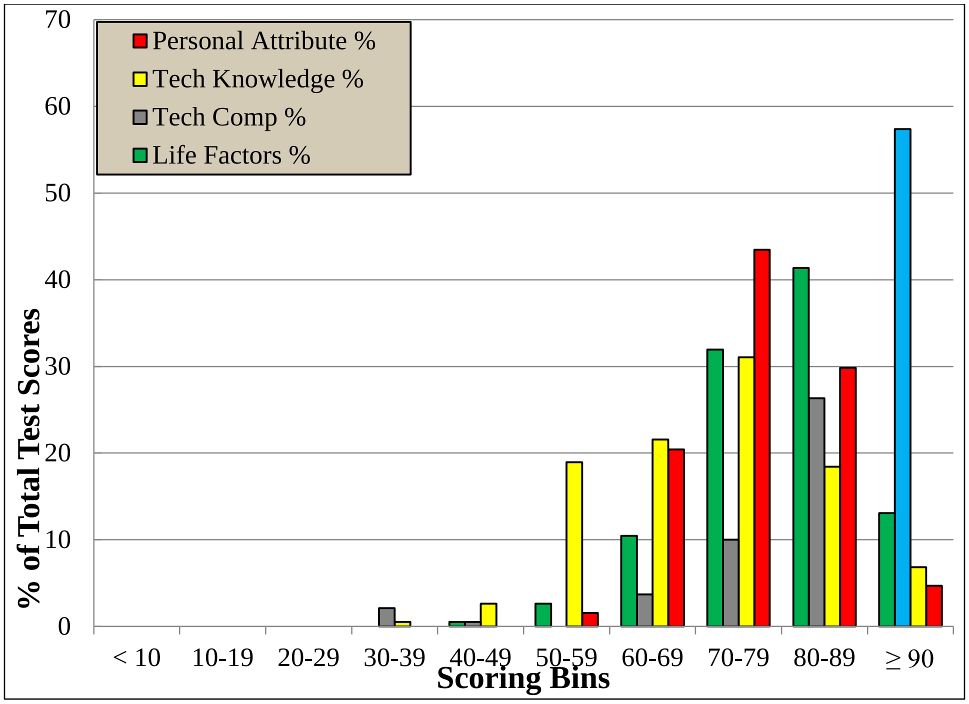

Figure 5.Histogram of Post-Test scores data distribution binned by 10-point increments.

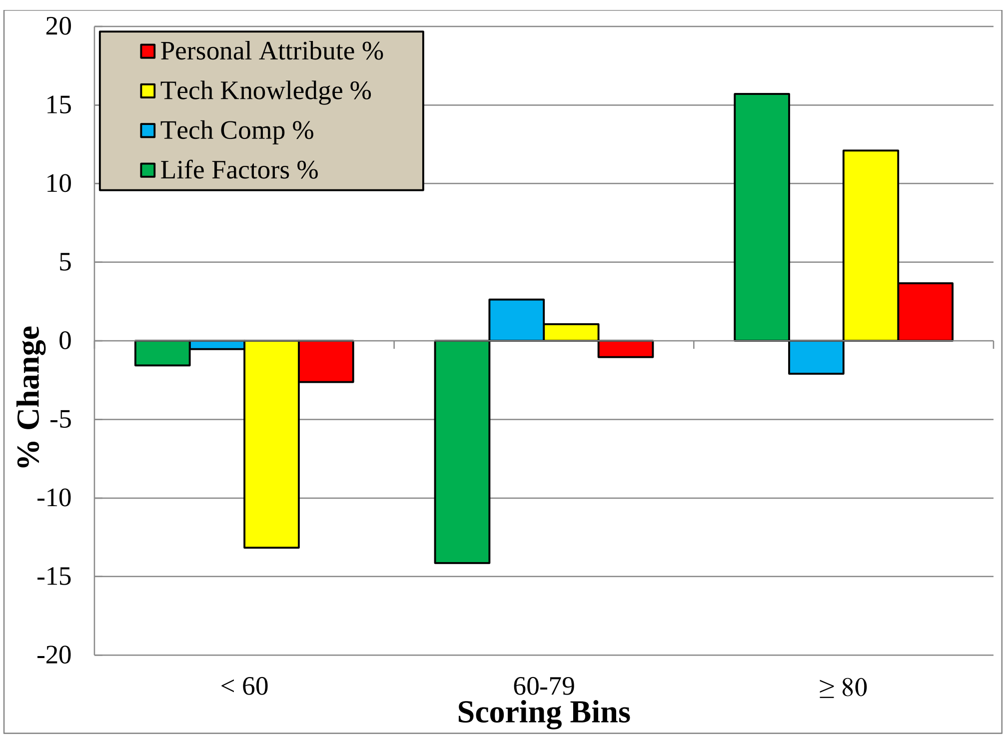

Figure 6.Histogram depicting change from pre-to-post test scores.

The Life Factors indicator exhibits a net decline in the number of scores in the 60-79 bin (14.14%). This is accompanied by a slight net decline in the < 60 bin of 1.57% and a net increase in scores in the ≥ 80 bin (15.71%).

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. All SmarterMeasure readiness indicators mean scores are reported and exhibit statistically significant results in at least one indicator every semester over the course of the study thus far. Technology Knowledge has consistently had the strongest increases across the study, with all six semesters exhibiting a statistically significant increase, the largest of which occurs in the most recent semester, Summer 2014.

	Fall '12	Spring '13	Summer '13	Fall '13	Spring '14	Summer '14
Personal Attributes	-0.17	-0.98*	0.82	-0.45	-0.32	0.79
Technology Knowledge	3.77	4.22	3.06	3.07	3.89	5.45
Technology Competency	-0.07	2.02*	2.97	-1.00	2.11	0.45
Life Factors	-0.42	0.57	2.08	-0.20	0.24	2.53

*Table 3. SmarterMeasure 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).*

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). Further, retrospective power analyses using observed effect sizes are not helpful here since the values are not necessarily meaningful to the significance of the study (Thomas, 1997). To this end, the significance test results are reported with detailed descriptive statistics as support (see Section 2). 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 (i.e. Type I errors). These marginal results, defined as those within the 95-99% confidence level, include Spring 2013 Personal Attributes and Technology Competence.

	Fall '12 $t_{crit} = 1.97$	Spring '13 $t_{crit} = 1.97$	Summer '13 $t_{crit} = 1.98$	Fall '13 $t_{crit} = 1.96$	Spring '14 $t_{crit} = 1.97$	Summer '14 $t_{crit} = 1.97$
Personal Attributes	t(292)=-0.44, p=0.66	t(200)=-2.01, p=0.0464*	t(166)=1.81, p=0.072	t(779)=-1.89, p=0.0594	t(430)=-0.90, p=0.371	t(191)=1.89, p=0.060
Technology Knowledge	t(292)=7.04, p<<0.001	t(193)=6.37, p<<0.001	t(166)=4.74, p<<0.001	t(775)=8.91, p<<0.001	t(425)=7.72, $p=9.57x10^{-14}$	t(189)=9.20, $p=7.09x10^{-17}$
Technology Competency	t(292)=-0.09, p<<0.001	t(194)=2.29, p=0.023*	t(166)=3.17, p=0.002	t(775)=-1.84, p=0.066	t(426)=3.28, $p=8.71x10^{-4}$	t(189)=0.48, p=0.631
Life Factors	t(292)=-0.79, p<<0.001	t(200)=0.86, p=0.390	t(166)=3.57, p=0.0005	t(779)=-0.71, p=0.478	t(431)=0.63, p=0.533	t(191)=4.88, $p=2.29x10^{-6}$

*Table 4. Additional significance testing statistics for SmarterMeasure readiness indicators including observed t-stat (tobs), probability of difference due to chance (p-value), degrees of freedom (df) (listed as t(x), 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 7 is a graphical representation of the difference in mean scores of readiness indicators across semesters. The consistently strong improvement in scores from pre-to-post tests of the Technology

Knowledge readiness indicator (yellow line) is clearly visible. By comparison, the smallest improvement of this indicator (Summer 2013) is still greater than any improvement from any indicator across all semesters. Since Fall 2013, consistent improvement can be seen in the Personal Attributes and Life Factors indicators. Technology Competency exhibits the widest variation of any indicator, as high as +2.97 in Summer 2013, and as low as -1.00 in Fall 2013.

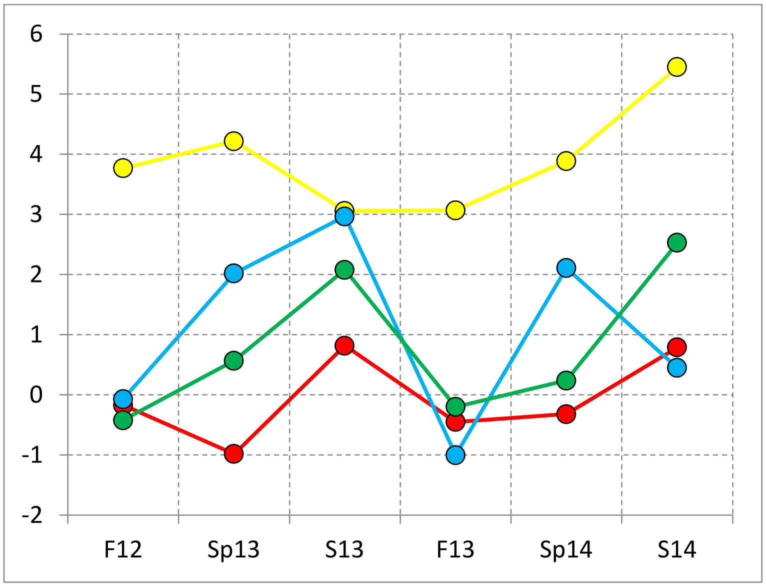

Figure 7. Comparison of the difference in mean scores from pre-to-post tests across semester by readiness indicator. Red (Personal Attributes), yellow (Technology Knowledge), blue (Technology Competency), green (Life Factors).

3 CONCLUSIONS

In Florida SouthWestern State College's QEP assessment, students are expected to statistically significantly improve in the following learning readiness indicators of the SmarterMeasure Student Readiness Test: Personal Attributes, Life Factors, Technology Knowledge, and Technology Competency.

For Summer 2014, the paired means t-test results indicate that for two of the four readiness indicators, Technology Knowledge and Life Factors, we can conclude with a 95% confidence that the improvement in mean scores is not solely due to chance. There were slight increases in means Personal Attributes and Life Factors; however, we cannot conclude that this difference is not simply due to chance.

Based on the results, it can, with reasonable certainty, be stated that the Summer 2014 FSW student is more proficient in Technology Knowledge and Life Factors upon completion of the Cornerstone course. Further, most FSW students actually enter the Cornerstone course with a strong background in technology. In the case of Technology Competence, the indicator exhibited a large percentage of students showing no change from pre-to-post tests and 63.00% of pre-test scores were greater than or equal to 90. Students do not appear to make any gains in Personal Attributes or Technology Competency.

In comparison with earlier studies, it can be shown that since the study began in Fall 2012, only Personal Attributes has exhibited no statistically significant increases across the ongoing study period. The Life

Factors and Technology Competency indicators have exhibited irregular results, with statistically significant results reported for both positive and negative changes from pre-to-post test scores. The FSW student has shown statistically significant growth with each semester of the Cornerstone course in Technology Knowledge. The FSW student also tends to enter the course already proficient in Technology Competency, often showing little growth over the course of the semester as a result.

4 REFERENCES

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.

SmarterServices. 2011. 2011 Student Readiness Report. Deatsville, Alabama: unknown.

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/.

Suskie, L. 2004. Assessing Student Learning. Anker Publishing Co., Inc., Bolton, Massachusetts, 331 pp.

Thomas, L. 1997. Retrospective power analysis. Conservation Biology, 11(1), 276-280.

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