Results of Critical Thinking Student Learning Outcome Assessment, Fall 2013
William Loker, Dean of Undergraduate Education

Summary

In Fall 2013, the faculty teaching courses in the General Education Pathways program in Area A3, Critical Thinking (CT) agreed to a second pre-test, post-test assessment of student learning outcomes in their courses. This effort follows a similar successful endeavor in Fall 12. As in Fall 12, the Fall13 effort relied on twenty multiple choice questions drafted and approved by faculty teaching the varied courses. The questions were designed to test students' understanding of three components of CT:

(1) students will have the ability to distinguish matters of fact from issues of judgment or opinion,
(2) students will have the ability to reach well-supported factual or judgmental conclusions, and
(3) students will have the abilities to analyze, criticize, and advocate ideas.

Faculty teaching 18 sections of Area A3 courses participated: two sections of CMST 255 (n=51), one section of GEOS 104 (n=97), eleven sections of PHIL 102 (n=874) and three section of PSYC 100 (n=215). Faculty agreed to administer the pre-test during the first week of class (the first day if possible) and include the same questions on the final exam either embedded in the exam for credit or for extra credit. Valid results were obtained from 1,237 students. The following statistical tests were performed, with results as noted:

1. Analysis of Variance (ANOVA) was carried out comparing mean **pre-test scores** across sections. There was no significant difference in pre-test scores among any of the sections. (see Table 1, below.)
2. A matched-pairs t-test was carried out comparing the **learning gain difference between pre- and post-test scores**. Post-test scores (mean = 13.23 across all sections) were significantly higher than pre-test scores (mean = 10.34). (See Table 2, below.)
3. ANOVA was carried out to test the **mean learning gain across sections**. PHIL 102-01 was significantly higher than all other sections. PHIL 102-03 was significantly higher than GEOS 104.
4. ANOVA was carried out to compare **post-test scores across sections**. PHIL 102-01 was significantly higher than all other sections. PHIL 102-03 was significantly higher than GEOS 104.
5. ANOVA was carried out to compare **mean scores across <u>courses</u>** (aggregating sections for CMST, PHIL and PSYC) for all the variables mentioned: mean pre-test scores, mean difference scores and mean post-test scores. There was still no significant difference in pre-test scores across courses.

a. <u>Learning Gains</u>: PHIL 102 was significantly higher than all other courses. PSYC 100 was significantly higher than GEOS 104. (See Table 3.) When PHIL 102-01 is excluded from the analysis, PHIL 102 was significantly higher than GEOS 104, but <u>not</u> CMST 255 and PSYC 100. PSYC 100 remained significantly higher than GEOS 104. (See Table 5, below.)

b. <u>Post-test scores</u>: PHIL 102 was significantly higher than all other courses. (See Table 4.) When PHIL 102-01 was removed from the analysis, there was no significant difference in between PHIL 102 and other sections. When PHIL 102-01 was removed from the analysis, PSYC 100 students scored significantly higher on the post-test than GEOS 104 students. (See Table 6, below.)

Statistical analyses support the following conclusions:

1. All sections started with students that had similar levels of understanding of the concepts measured by the pre-test.

2. Across all sections, there was a gain in student learning of the concepts measured by the test. On average, students' scores increased by about 30% from pre- to post-test. While this is a positive outcome, it may be somewhat concerning that on the post-test, students on average only got about 65% of the questions right. The only significant differences in learning gains across sections were:
   a. PHIL 102-01 scored higher than all other sections;
   b. PHIL 102-06 scored lower than PHIL 102-01, 02, 03, 40, 44.

3. PHIL 102 students scored significantly higher than students in other courses both in terms of learning gains and post-test scores.
   a. PHIL 102-01 stood out in terms of both the average learning gain (pre vs. post-test score) and post-test score, which was significantly higher than all other sections.
   b. When PHIL 102-01 was removed from the analysis of pre- post-test learning gains, the significant difference between PHIL and PSYC and CMST disappeared. PHIL 102 remained significantly higher than GEOS 104 on learning gains. PSYC 100 was also significantly higher than GEOS 104 on learning gains.
   c. When PHIL 102-01 was removed from the analysis of post-test scores, the difference between PHIL 102 and other courses was no longer significant. PSYC 100 students scored significantly higher on the post-test than GEOS 104 students.

In summary, the pre- and post-test results demonstrate statistically significant learning gains across sections. As in the prior analysis of other CT concepts carried out in Fall 12, the instructor of PHIL 102-01 stands out as particularly effective in teaching CT as measured by student performance on this exam. Assuming that the exam is a valid measure of the concepts

CT concepts identified, other instructors might benefit from learning the pedagogical approaches of the PHIL 102-01 instructor.

In addressing the suitability of teaching CT concepts in a variety of courses, particularly those outside Philosophy, the results tend to support the conclusion that these concepts can be taught in a variety of disciplines. As with the prior assessment (Fall 12), it appears that *instructor effects are more powerful than course or discipline effects in shaping learning outcomes*.

It should be noted that these results are somewhat preliminary and subject to the usual caveats. In particular, different instructors approached administration of the post-test in different ways (embedded in the final versus taken for extra credit). In future iteration of assessments of this nature, it is important to achieve greater uniformity in the administration of the post-test exam to eliminate this possible source of error.

Table 1: Pre-Test Score by Subject Area Analysis

Dependent Variable: PreTest Scores

Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	10.000	.377	9.261	10.739
GEOS	10.969	.273	10.433	11.505
PHIL	10.261	.091	10.082	10.439
PSYC	10.488	.184	10.128	10.849

Table 1: Comparison of Pre-Test scores by course: There is no significant difference in the average pre-test scores by subject area, F = 7.245, p N.S.

Table 2: Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PostTest	13.230	1067	3.4217	.1048
	PreTest	10.337	1067	2.6814	.0821

Table 2. Matched-Pairs t-test results of learning gains (Pre-test versus post-test scores). Fall 2013 students that completed both the CT pre-test and post-test scored significantly higher on the CT post-test, t = 26.708, p < .001.

Table 3: Learning Gains by Section

Dependent Variable: PostPre_Change

Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	.964	.642	-.296	2.225
GEOS	.361	.373	-.371	1.094
PHIL	3.453	.122	3.214	3.691
PSYC	1.891	.258	1.385	2.396

Table 3. Analysis of Variance (ANOVA) comparing learning gains (pre-, post-test difference scores) by course. An Analysis of Variance on the above table shows a significant difference between the change scores by subject area, $F = 30.458$, $p < 001$. A further post-hoc analysis on the mean change scores shows that students enrolled in PHIL courses achieved a significantly greater increase in post-test scores over their pre-test scores than students enrolled in the other three subject areas (CMST, GEOS, PSYC). Additionally, PSYC enrolled students achieved significantly greater change scores than the GEOS students.

Table 4. Post-Test Score by Subject Area Analysis

Dependent Variable: PostTest Score

Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	11.032	.593	9.869	12.196
GEOS	11.218	.354	10.524	11.913
PHIL	13.669	.113	13.447	13.892
PSYC	12.330	.247	11.845	12.814

Table 4. Analysis of Variance (ANOVA) comparing post-test scores by course. An Analysis of Variance on the average post-test scores by subject area shows a significant difference in the average scores by subject area, $F = 10.902$, $p < .001$. A post hoc analysis of the average post-test scores by subject area shows that students enrolled in the PHIL courses have an average post-test score significantly higher than the other three subject areas (CMST, GEOS, PSYC).

Table 5: Learning Gains by Section, exclude PHIL 102-01

Dependent Variable: PostPre_Change

Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	.964	.642	-.296	2.225
GEOS	.361	.373	-.371	1.094
PHIL	1.802	.126	1.555	2.048
PSYC	1.891	.258	1.385	2.396

Table 5. Analysis of Variance (ANOVA) comparing learning gains (pre-, post-test difference scores) by course*; excluding PHIL 102-01.* An Analysis of Variance on the above table shows a significant difference between the change scores by subject area, F = 7.235, p < 001. A further post-hoc analysis on the mean change scores shows that students enrolled in PHIL and PSYC courses achieved a significantly greater increase in post-test scores over their pre-test scores than students enrolled in GEOS 100.

Table 6. Post-Test Score by Subject Area Analysis, exclude PHIL 102-01

Dependent Variable: PostTest Score

Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	11.032	.527	9.998	12.067
GEOS	11.218	.315	10.601	11.836
PHIL	12.161	.125	11.916	12.406
PSYC	12.330	.247	11.899	12.760

Table 6. Analysis of Variance (ANOVA) comparing post-test scores by course, *excluding PHIL 102-01.* An Analysis of Variance on the average post-test scores by subject area shows a significant difference in the average scores by subject area, F = 4.377, p < .005. A post hoc analysis of the average post-test scores by subject area shows that students enrolled in the PSYC course have an average post-test score significantly higher than the students in the GEOS course. There is no significant difference between PHIL and other subject areas when PHIL 102-01 scores are excluded from the analysis.

Results of Critical Thinking Student Learning Outcome Assessment, Fall 2012
William Loker, Dean of Undergraduate Education

Summary

In Fall 2012, the faculty teaching courses in the general education Pathways program in Area
A3, Critical Thinking (CT) agreed to a pre-test, post-test assessment of student learning
outcomes in their courses. Twenty multiple choice questions were drafted and approved by
faculty. The questions were designed to test students' understanding of two components of CT:
inductive versus deductive logic and logical fallacies. Faculty teaching 18 sections (two sections
of CMST 255, one section of GEOS 104, fourteen sections of PHIL 102 and one section of
PSYC 100) in Fall 12 agreed to administer the pre-test during the first week of class (the first day
if possible) and include the same questions on the final exam either embedded in the exam for
credit or for extra credit. Valid results were obtained from 1,194 students. The following
statistical tests were performed, with results as noted:
1. Analysis of Variance (ANOVA) was carried out comparing mean **pre-test scores**
   across sections. The only significant difference was that results from the Honors
   section (PHIL 102H) were significantly higher than other sections. There was no
   significant difference in pre-test scores among other sections. (See Table 1.)
2. A matched-pairs t-test was carried out comparing the **learning gain difference
   between pre- and post-test scores**. Post-test scores (mean = 10.65 across all
   sections) were significantly higher than pre-test scores (mean = 6.98). (See Table 2.)
3. ANOVA was carried out to test the **mean learning gain** by section. PHIL 102-01
   was significantly higher than all other sections. PSYC 100 was significantly higher
   than PHIL 102-04 and PHIL 102-15. (See Table 3.)
4. ANOVA was carried out to compare **post-test scores** across sections. PHIL 102-01
   was significantly higher than all other sections. PHIL 102H was higher than PHIL
   102-04, 05, 11, 12. PSYC 100 was significantly higher than PHIL 102-04, 05. (See
   Table 4.)
5. ANOVA was carried out to compare **mean scores across <u>courses</u>** (aggregating
   sections for CMST and PHIL) for all the variables mentioned: mean pre-test scores,
   mean difference scores and mean post-test scores. There was no significant
   difference in mean pre-test scores by course.   In comparing mean difference scores
   and post-test scores, the means for GEOS were significantly lower than those for
   PHIL and PSYC. (See Tables 5-7.)

Statistical analysis supports the following conclusions:
1. All sections, with the exception of the Honors section, PHIL 102H, started with
    students that had similar levels of understanding of the concepts measured by the pre-
    test.

2. Across all sections, there was a gain in student learning of the concepts measured by the test. On average, students' scores increased by about 30% from pre- to post-test. The only significant difference in learning gains across sections were those measured in PHIL 102-01 compared to all other sections and PSYC 100 scoring better than the PHIL 102 sections showing the lowest gains. While this is a positive outcome, it may be somewhat concerning that students on average only got about 50% of the questions right.
3. PHIL 102-01also stood out in terms of the average post-test score, which was significantly higher than all other sections. The average post-test score in PSYC 100 was higher than a couple of PHIL 102 sections, and the Honors students scored higher overall on the post-test than four of the PHIL sections.
4. The only aggregate difference across <u>courses</u> was the lower learning gains in the GEOS course compared to the aggregated results from PHIL and the PSYC course. Pooling the results of the PHIL courses "washes out" the variability in post-test outcomes across sections of this class.

In summary, the pre- and post-test results demonstrate statistically significant learning gains across sections. PHIL 102-01 stands out as particularly effective in teaching CT as measured by student performance on this exam. Assuming that the exam is a valid measure of the concepts identified (inductive versus deductive arguments and logical fallacies) instructors might benefit from learning the pedagogical approaches of that instructor. In addressing the suitability of teaching CT concepts in a variety of courses, particularly those outside Philosophy, the results tend to support the conclusion that these concepts can be taught well in a variety of disciplines as evidenced by the performance in the PSYC 100. The approach taken in the Geosciences course might benefit from revision given the significantly lower learning gains and post-test scores in this course. As always, it is difficult to separate out instructor effects from curriculum per se. The widely varying results across the PHIL102 sections tend to indicate that the approach taken by individual instructors may outweigh the disciplinary basis of the course in affecting positive learning outcomes.

It should be noted that these results are somewhat preliminary and subject to the usual caveats. In particular, different instructors approached administration of the post-test in different ways (embedded in the final versus taken for extra credit). For the next round of assessment, to focus on aspects of CT not measured in this iteration, it is important to achieve greater uniformity in the administration of the post-test exam to eliminate this possible source of error.

Table 1. Mean Pre-test scores by section

2. Course

Dependent Variable: PreTest

| Course | Mean | Std. Error | 95% Confidence Interval | |
			Lower Bound	Upper Bound
CMST 255-01	7.519	.451	6.634	8.403
CMST 255-02	8.000	.451	7.115	8.885
GEOS 104	6.855	.316	6.235	7.474
PHIL 102-01	6.788	.134	6.525	7.050
PHIL 102-02	6.976	.210	6.563	7.389
PHIL 102-03	7.258	.214	6.839	7.678
PHIL 102-04	6.672	.215	6.251	7.094
PHIL 102-05	6.100	.331	5.450	6.750
PHIL 102-10-44	6.968	.243	6.491	7.444
PHIL 102-11	7.234	.293	6.660	7.809
PHIL 102-12	7.180	.300	6.592	7.769
PHIL 102-14	6.670	.221	6.235	7.104
PHIL 102-15	7.060	.331	6.410	7.710
PHIL 102-40	7.563	.338	6.899	8.226
PHIL 102-41	7.109	.346	6.431	7.786
PHIL 102-43	6.404	.342	5.734	7.075
PHIL 102H	10.071	.443	9.203	10.940
PSYC 100-01	6.946	.272	6.412	7.480

Dependent Variable: PreTest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	425.246[a]	17	25.014	4.555	.000
Intercept	51482.687	1	51482.687	9374.599	.000
Course	425.246	17	25.014	4.555	.000
Error	7869.637	1433	5.492		
Total	79212.000	1451			
Corrected Total	8294.884	1450			

Note: Pre-test scores differ significantly at p 0.01 level. PHIL 102H pre-test scores are significantly higher. No other significant differences.

Table 2. Paired Samples t-test, Pre-test versus Post-test scores

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	PostTest	10.615	1194	4.1629	.1205
	PreTest	6.978	1194	2.3743	.0687

Paired Samples Test

		t	df	Sig. (2-tailed)
Pair 1	PostTest - PreTest	30.095	1193	.000

Note: Difference Scores/Learning Gains (Post-test – Pre-test) differ significantly at <p 0.01 level.

Table 3. Paired Samples t-test, Pre-test versus Post-test scores, by Section

2. Course

Dependent Variable: PostPre_Change

Course	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST 255-01	3.118	.677	1.789	4.446
CMST 255-02	1.167	.806	-.414	2.748
GEOS 104	1.750	.403	.960	2.540
PHIL 102-01	9.046	.165	8.721	9.370
PHIL 102-02	2.327	.263	1.812	2.843
PHIL 102-03	2.227	.266	1.705	2.749
PHIL 102-04	.562	.285	.004	1.121
PHIL 102-05	2.116	.426	1.281	2.951
PHIL 102-10-44	2.052	.318	1.428	2.676
PHIL 102-11	.615	.547	-.459	1.689
PHIL 102-12	.625	.570	-.493	1.743
PHIL 102-14	2.147	.276	1.605	2.689
PHIL 102-15	.087	.582	-1.055	1.229
PHIL 102-40	1.957	.412	1.149	2.764
PHIL 102-41	1.826	.412	1.019	2.634
PHIL 102-43	1.976	.436	1.120	2.831
PHIL 102H	2.080	.558	.985	3.175
PSYC 100-01	4.050	.360	3.343	4.757

Tests of Between-Subjects Effects

Dependent Variable: PostPre_Change

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	11634.784[a]	17	684.399	87.833	.000
Intercept	3227.339	1	3227.339	414.183	.000
Course	11634.784	17	684.399	87.833	.000
Error	9163.464	1176	7.792		
Total	36588.000	1194			
Corrected Total	20798.248	1193			

Note: Difference Scores/Learning Gains (Post-test – Pre-test) by Section differ significantly at <p 0.01 level. See text for explanation.

Table 4. Mean Post-test scores by Section.

2. Course

Dependent Variable: PostTest

Course	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST 255-01	10.294	.691	8.938	11.650
CMST 255-02	8.733	.736	7.290	10.177
GEOS 104	8.490	.407	7.691	9.289
PHIL 102-01	15.844	.163	15.525	16.163
PHIL 102-02	9.188	.264	8.671	9.705
PHIL 102-03	9.348	.266	8.826	9.869
PHIL 102-04	7.333	.282	6.780	7.887
PHIL 102-05	8.000	.430	7.157	8.843
PHIL 102-10-44	9.286	.311	8.676	9.896
PHIL 102-11	7.731	.559	6.634	8.827
PHIL 102-12	7.720	.570	6.602	8.838
PHIL 102-14	8.786	.269	8.257	9.314
PHIL 102-15	8.167	.582	7.025	9.308
PHIL 102-40	9.333	.411	8.526	10.140
PHIL 102-41	8.935	.420	8.110	9.759
PHIL 102-43	8.395	.435	7.543	9.248
PHIL 102H	12.000	.549	10.924	13.076
PSYC 100-01	11.016	.362	10.306	11.726

Dependent Variable: PostTest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	12066.563[a]	17	709.798	87.373	.000
Intercept	61944.591	1	61944.591	7625.143	.000
Course	12066.563	17	709.798	87.373	.000
Error	10114.042	1245	8.124		
Total	164711.000	1263			
Corrected Total	22180.605	1262			

Note: Post-test scores differ significantly at $<p$ 0.01 level. See text for explanation.

Table 5. Mean Pre-test score by Course (Aggregated by Department)

2. Course Subject

Dependent Variable: PreTest

Course Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	7.759	.325	7.121	8.397
GEOS	6.855	.322	6.223	7.487
PHIL	6.967	.067	6.835	7.099
PSYC	6.946	.278	6.401	7.491

Tests of Between-Subjects Effects

Dependent Variable: PreTest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	33.784[a]	3	11.261	1.973	.116
Intercept	15955.465	1	15955.465	2794.732	.000
Subject	33.784	3	11.261	1.973	.116
Error	8261.099	1447	5.709		
Total	79212.000	1451			
Corrected Total	8294.884	1450			

Note: Pre-test scores do not differ significantly at the 0.05 level.

Table 6. Mean Difference Score/Learning Gains), Post-test – Pre-test scores, by Course (aggregated by Department)

2. Course Subject

Dependent Variable: PostPre_Change

Course Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	2.310	.772	.796	3.825
GEOS	1.750	.600	.573	2.927
PHIL	3.735	.128	3.484	3.986
PSYC	4.050	.537	2.997	5.103

Tests of Between-Subjects Effects

Dependent Variable: PostPre_Change

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	242.363[a]	3	80.788	4.677	.003
Intercept	1923.993	1	1923.993	111.382	.000
Subject	242.363	3	80.788	4.677	.003
Error	20555.885	1190	17.274		
Total	36588.000	1194			
Corrected Total	20798.248	1193			

Note: Difference scores (Learning Gains) differ significantly at the 0.003 level. See text for explanation.

Table 7. Mean Post-test score by Course (Aggregated by Department)

2. Course Subject

Dependent Variable: PostTest

Course Subject	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
CMST	9.563	.737	8.116	11.009
GEOS	8.490	.596	7.321	9.659
PHIL	10.725	.125	10.481	10.969
PSYC	11.016	.530	9.977	12.055

Tests of Between-Subjects Effects

Dependent Variable: PostTest

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	280.201[a]	3	93.400	5.369	.001
Intercept	23056.428	1	23056.428	1325.457	.000
Subject	280.201	3	93.400	5.369	.001
Error	21900.404	1259	17.395		
Total	164711.000	1263			
Corrected Total	22180.605	1262			

Note: Post-test scores differ significantly at the 0.01 level. See text for explanation.