New Mexico State University – Las Cruces
General Education Assessment of Student Learning Outcomes
2011-12 Final Report

IRB #7435

Purpose of the Assessment:
The purpose of this study was to assess achievement of integrated general education (GE) learning outcomes in students ranked 'junior' or 'senior' and who had completed a majority of their required lower-level GE courses. Learning outcomes were identified as they aligned with the institution's stated objectives for GE learning, competencies identified for student learning across the New Mexico Higher Education State Common Core, and the NMSU outcomes for the Baccalaureate Experience (BE).

Table 1: Learning outcomes assessed using "Tweaking Twain" assignment

Learning Outcome	NM SCC Competency	NMSU-Las Cruces GE Outcome	NMSU-Las Cruces BE Outcome
The student articulates a position.	Area I.b.	Articulates a Position	Critical Thinking; Communication
The student communicates thoughts and ideas.	Area I. c.	Communication: Rhetorical Strategies/ Reasoned Argument	Effective Communication
The student identifies diverse perspectives and viewpoints.	Area IV	Cultural Awareness	Diversity; Self-awareness
The student identifies relevant historical context.	Area V. b., c. & d.	Historical Perspective/ Knowledge	Academically Prepared

Table 2: Learning outcomes assessed using "Framing the Flood" assignment

Learning Outcome	NM SCC Competency	NMSU-Las Cruces GE Outcome	NMSU-Las Cruces BE Outcome
The student analyzes key elements of a primary document.	Area V. a.	Analysis	Critical Thinking
The student communicates thoughts and ideas.	Area I. b.	Communication	Effective Communication
The student identifies relevant distinctions in classes.	Area IV. c. & d.	Cultural Awareness	Diversity
The student demonstrates historical knowledge.	Area V. b., c. & d.	Historical Perspective/ Knowledge	Academically Prepared
The student creates unique and original work.	NA	NA	Creativity

Assessment Procedure/Process
Two student assignments and scoring rubrics were developed by the Committee for the Assessment of Student Learning in General Education (CASL-GE). One assignment was titled "Framing the Flood," and another was titled "Tweaking Twain." (See Tables 1 & 2 above for learning outcomes assessed by each

assignment.) A pilot assessment was implemented in spring 2011 using the instruments. Instruments were adjusted based on pilot data and feedback. A full assessment was implemented spring 2012.

In spring 2012, various Viewing a Wider World (VWW) courses were identified as data collection points for the assessment. VWW courses were selected because the majority of students in VWW courses are juniors or seniors, and most have completed all or a majority of their lower-level GE courses. After securing the cooperation of instructors for the selected VWW courses, members of the CASL-GE attended face-to-face classes to administer the student assignment, or met with online instructors to arrange for online students to complete the assignment. CASL-GE members were provided a script to read to students regarding the assessment and their voluntary participation. Participating students then were provided an informed consent letter requiring their signature for participation. Written scripts were provided to online instructors, as were informed consent letters. Online students who agreed to participate were required to signed letters of consent with their submitted assignment.

Table 3: View a Wider World Course Selection & Participation

Courses Identified for Participation	Delivery Type	Number of Students Enrolled	Participated Y/N
ANTH	Face-to-Face	23	Y
ANTH	Face-to-Face	27	Y
CJ	Online	unknown	Y
CJ	Online	unknown	Y
C J	Online	unknown	Y
EMD	Face-to-Face	23	Y
EMD	Online	unknown	Y
EMD	Online	unknown	N
HL S	Online	approx. 50	N
HL S	Online	approx. 50	N
HL S	Online	approx. 50	N
HL S	Online	approx. 50	N
HL S	Online	approx. 50	N
HON	Face-to-Face	12	Y
HON	Face-to-Face	13	Y
LIB	Face-to-Face	15	Y
SOC	Face-to-Face	35	Y
SOC	Online	unknown	Y

Approximately half of the students were given the "Tweaking Twain" assignment, and half were given the "Framing the Flood" assignment. Assignments were distributed randomly. In face-to-face classes, students were given 30 minutes to complete the assignment. Most students completed the assignment in 10-25 minutes. Online students were given a time period (in days) to complete the assignment as determined by the instructor. Upon completion of the assignment, student work was collected by members of the CASL-GE.

In April 2012, an open invitation was extended to faculty at large to participate in a scoring session in the NMSU Teaching Academy. Fifteen (15) faculty members attended the session. The Director of Assessment and the chair of the CASL-GE facilitated a 'norming session' for scoring student work. All participants read

and rated the first paper individually. Then ratings were compared and discussed until consensus was reached. The first assignment used for the norming session was a "Tweaking Twain" assignment. The second assignment used in this process was a "Framing the Flood" assignment. During discussion of the scores for the second assignment, readers determined the "Framing the Flood" assignment and rubric still needed to be adjusted to accurately measure the desired student learning, and should therefore not be considered in this assessment cycle.

The following process was used:

- One copy of each paper was printed
- Two scoring rubrics were attached to each paper - scores ranged from 1-4 (0=no evidence, 1=emerging, 2=competent, 3=skillful) on 4 criteria (Tweaking Twain)
- Each paper included either the student's Banner ID number (provided by the student) or an assigned student number (if Banner ID was not provided by the student) - no other identifying information was included on the student paper (e.g. course, name, instructor, rank, etc.).
- Each reader was assigned a 'reader number' – reader numbers were written on the rubric as papers were scored
- Student papers were randomly distributed among faculty readers
- Each paper was read and scored by two independent readers using the rubric provided
- For each paper, one faculty member would read the paper, write their assigned reader number on the rubric, write their assigned scores on the rubric, and turn their completed rubric over – the second reader would then read the paper and repeat the process using a clean rubric
- Once the second reader completed their evaluation, the paper and both completed rubrics were handed to a graduate student who recorded scores on an excel worksheet
    - If a significant discrepancy between scores was identified (more than 1 point difference on any independent component) between the two readers, the two readers were asked to discuss the paper and reach greater consensus
    - After discussion (if needed), the rubrics were once again returned to the recorder, who input the final scores for both rubrics into the spreadsheet
- Scores on each component of the rubric were recorded, as well as composite scores for each paper

Findings

Inter-Rater Reliability

Krippendorf's alpha as used to measure Inter-Rater Reliability (IRR) for each dimension of the rubric. Krippendorf's alpha was used because:

- It is preferred for content analysis (rendering judgments of text-based information)
- It handles small data sets
- It accommodates missing data
- It uses bootstrapping for more precise measurement

Krippendorf's alpha measures agreement of evaluators who rate a set of items into distinct and mutually exclusive categories. The observed disagreement between evaluators is corrected by the amount of disagreement expected by chance. (Poesio and Artstein 2005) The value of alpha can range from -1.0 (complete disagreement) to 0 (unreliability of measurement) to 1.0 (complete agreement). The higher alpha is in the positive direction, the greater level of agreement between evaluators.

In general measures of agreement, the following guidelines are given:

0.8 and 1	Very good agreement
0.6 and 0.79	Good agreement
0.4 and 0.59	Moderate agreement
0.2 and 0.39	Fair agreement
0.0 and 0.19	Poor agreement

Fall 2012, graduate students in statistics (AST 553, under the supervision of Dr. Gould) were given data with identifying student information removed, and asked to provide statistical analysis. The following includes information provided in their final report.

Table 4: Krippendorff's alpha for each category on "Tweaking Twain"

Rubric Dimension	Krippendorf's α	Interpretation
Articulates a Position	0.561	Moderate agreement
Communication: Rhetorical Strategies/Reasoned Argument	0.548	Moderate agreement
Cultural Awareness	0.613	Good agreement
Historical Perspective/Knowledge	0.536	Moderate agreement

According to Krippendorff (2004, pp. 241-243), "social scientists commonly rely on data with reliabilities $\alpha \geq .800$, consider data with $0.800 > \alpha \geq 0.667$ only to draw tentative conclusions, and discard data whose agreement measures $\alpha < 0.667$". Our measures of agreement are less than 0.65 for the 4 categories. Therefore the scores from the 2 readers are not as consistent. These scores may also be interpreted as being equivalent to Spearman's rank correlation (Hayes and Krippendorf 2007), which in all cases above would be considered as a moderate linear relations of the ranks. Thus, average scores this study used are not entirely reflective of the students' learning. Perhaps, more training of the readers (reviewers) regarding the rules for scoring the current answer sheets is needed. For instance, before scoring these answer sheets, let the two readers score some samples. Identify where the differences are in scoring and come to a common agreement of the protocol to be used in the future.

Hayes, A.F., and K. Krippendorff. 2007. Answering the call for a standard reliability measure for coding data. Communication methods and measures 1:77-89.

Krippendorff, K. 2004. Content analysis: an introduction to its methodology. Second Edition, Thousand Oaks, CA, Sage.

Average Scores and Their Counts

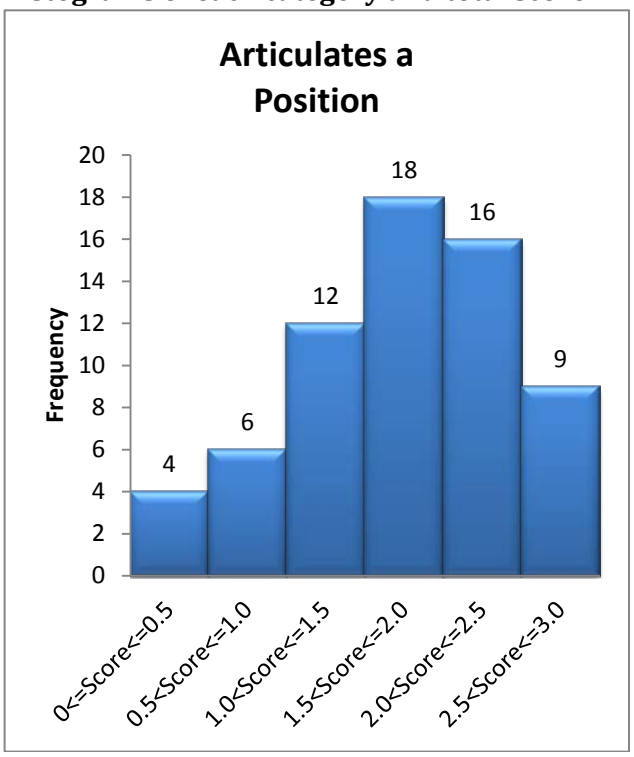
N = 68

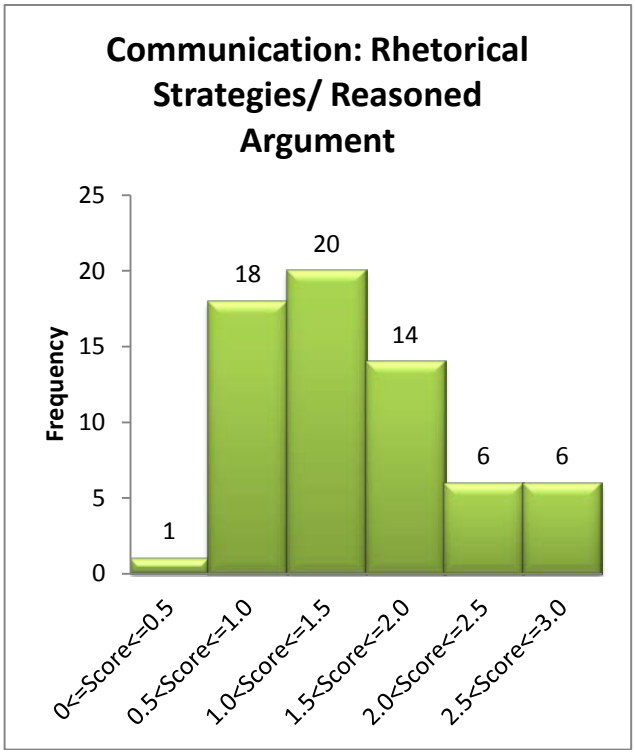
The average scores (over readers and students) are given in Table 5. As you can see, the scores decrease as you move across outcomes. The overall average score across all categories is 1.54. The highest composite score was 2.875 (by 1 student), and the lowest composite score was 0.5 (also by 1 student).

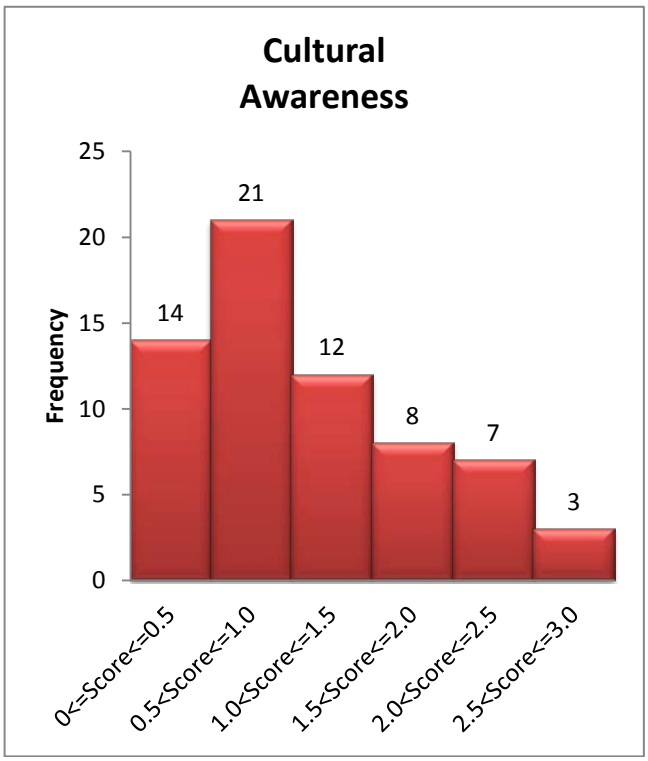
Table 5. Average category and composite scores, counts and percentages of students scoring "2" (competent) or above.

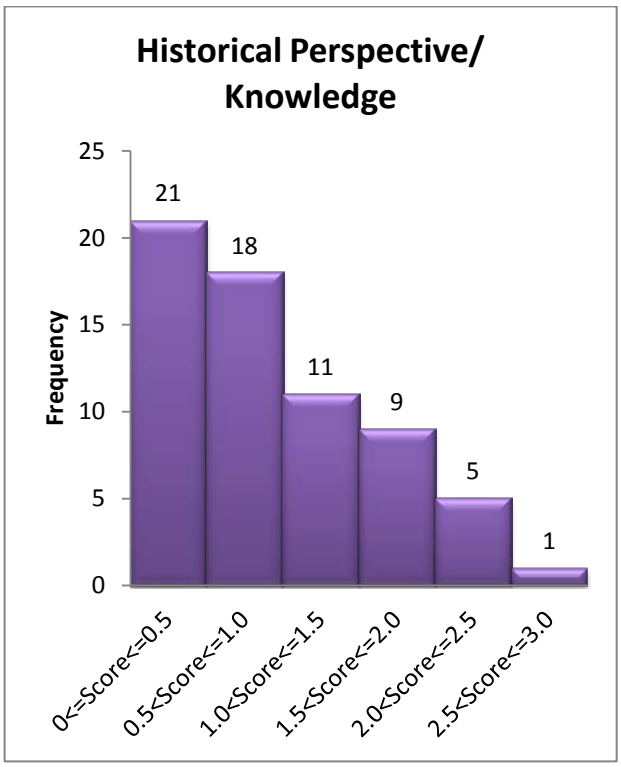
Outcomes	Average Score	Percent scoring 2.0 (competent) or above	Number (out of 68) scoring 2.0 (competent) or above
Articulates a Position	1.98	66.15%	43
Communication: Rhetorical Strategies/ Reasoned Argument	1.68	40.00%	26
Cultural Awareness	1.32	27.69%	18
Historical Perspective/ Knowledge	1.15	23.08%	15
Composite	1.54	27.69%	18

Histograms of each category and total score

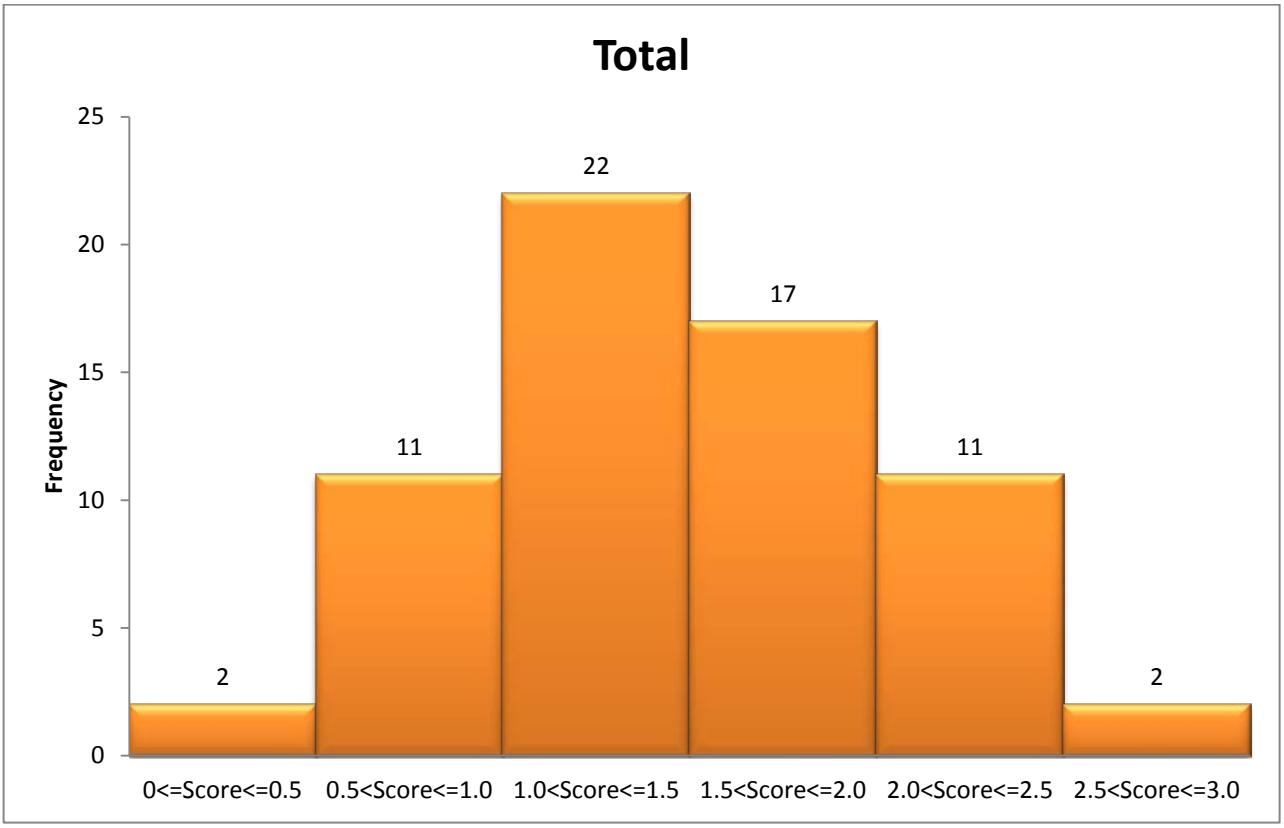








Legend:
0<= Score<=0.5: 0.5 and below
0.5<Score<=1.0: above 0.5, up to and inlcuding 1.0
1.0<Score<=1.5: above 1.0, up to and inlcuding 1.5
1.5<Score<=2.0: above 1.5, up to and inlcuding 2.0
2.0<Score<=2.5: above 2.0, up to and inlcuding 2.5
2.5<Score<=3.0: above 2.5, up to and inlcuding 3.0



Legend:
0<= Score<=0.5: 0.5 and below
0.5<Score<=1.0: above 0.5, up to and inlcuding 1.0
1.0<Score<=1.5: above 1.0, up to and inlcuding 1.5
1.5<Score<=2.0: above 1.5, up to and inlcuding 2.0
2.0<Score<=2.5: above 2.0, up to and inlcuding 2.5
2.5<Score<=3.0: above 2.5, up to and inlcuding 3.0

Table 6: Averaged student scores (2 readers) by category and composite score

Student ID	Average Score for Each Student				
	Articulates a Position	Communication: Rhetorical Strategies/ Reasoned Argument	Cultural Awareness	Historical Perspective/ Knowledge	Composite Score
1	2.5	1.5	2.5	1.5	2
2	3	3	2	2	2.5
3	2.5	2.5	1.5	1.5	2
4	2.5	1.5	1.5	1.5	1.75
5	3	2.5	2	2.5	2.5
6	1	1	0	0	0.5
7	3	3	1.5	2.5	2.5
8	1.5	1	0	0	0.625
12	1	1.5	2	2	1.625
13	3	3	2.5	3	2.875
14	2	2	1	0	1.25
15	3	3	2	1	2.25
16	2	2	3	2	2.25
17	2	1.5	0.5	0.5	1.125

18	2	1.5	1	1	1.375
19	2	1.5	1	0.5	1.25
20	1.5	2	1	2	1.625
21	3	2	1	1	1.75
22	2	1	0.5	0.5	1
23	2	2	1.5	2	1.875
24	2.5	1.5	1.5	0.5	1.5
25	2.5	2.5	1.5	1	1.875
26	2.5	1.5	2.5	2.5	2.25
27	1	1	0.5	0.5	0.75
28	1.5	1.5	0.5	1.5	1.25
29	3	2	1	0	1.5
30	1.5	1	2	2.5	1.75
31	2	2	3	2	2.25
32	2.5	2	1.5	0.5	1.625
33	1.5	1.5	0.5	0.5	1
34	2.5	1.5	1.5	1.5	1.75
35	2.5	1.5	0	0.5	1.125
36	0.5	1	1	0.5	0.75
37	2	1	1	1	1.25
38	0.5	1	1.5	1.5	1.125
39	1.5	1.5	0.5	0.5	1
40	0.5	0.5	0	0	0.25
41	3	2	2.5	1	2.125
42	2	1.5	1.5	1.5	1.625
43	1	1	1	1	1
44	1.5	1	0	0.5	0.75
45	2.5	2.5	2	2	2.25
46	3	2	2.5	2	2.375
47	2.5	2	2	2	2.125
48	2	2	1.5	1.5	1.75
49	0.5	1	0.5	0.5	0.625
50	2.5	2.5	3	2.5	2.625
51	2.5	1.5	1	1	1.5
52	1.5	1	1	1	1.125
53	2	1.5	1	1	1.375
54	1.5	1.5	1	1	1.25
55	2.5	3	1	1.5	2
56	1	3	2.5	1.5	2
57	2	1	1	1	1.25
58	2	1	0.5	0	0.875
59	2	1	1	1.5	1.375
60	2	2	1	1	1.5
61	1.5	1	1	1	1.125
62	1	1.5	1.5	1	1.25
63	2	1.5	0.5	1	1.25
64	1.5	1	2	1	1.375
65	2.5	2	2.5	1	2
66	2	1.5	1	0	1.125
67	2.5	2.5	1	0.5	1.625
68	1.5	1	1	0.5	1

Conclusion
Of the 68 student assignments that were evaluated for learning on four desired outcomes of the GE experience, only on one outcome, "articulates a position" did at least 50% of students meet the desired performance level of "competent" or above. On this outcome, 43 of the 68 students included in the study (66%) met the minimum desired level of performance (competent or skillful). For each of the remaining three outcomes, 40% of students or less scored competent or above (40% on "communication: rhetorical strategies/reasoned argument"; 28% on "cultural awareness"; and 23% on "historical perspective/knowledge"). Just under 28% of students (18 total) achieve an overall composite score that indicated a performance of competent or above on the desired outcomes.

Discussion
It is clear that students did not perform at the level we, as faculty members and as an institution, would like them to. Complicating factors include the low N, given our large student population. The low N was related to challenges with the "Framing the Flood" assignment and rubric identified during the scoring session. All student papers that addressed this assignment were discounted. Additionally, challenges with garnering widespread student participation from online courses were encountered, which had a significant impact on the number of student products that were collected. These issues are already being addressed by the CASL-GE for future iterations of this process.

For the most part, faculty in the scoring session were satisfied with the assessment instruments (assignment and rubric). Never-the-less, these will be revisited to consider if further refinement is required. It may be that refinement of the instruments could help to increase inter-rater reliability. The CASL-GE is also considering ways in which the norming session could be improved to increase inter-rater reliability. It is notable however, that although agreement is in the moderate to good range, this is up substantially from the prior year (pilot assessment), where the agreement was poor to fair.

Notwithstanding the constraints of the assessment, it is clear that students are not, as a whole, performing at the desired level to indicate success in achieving GE learning outcomes. It is in the best interest of the campus community to consider what we can learn from this assessment, what we can continue to learn in the future, and what we can do, collectively, to increase student success in GE learning.

Next Steps
A Results Forum will be held at the NMSU Teaching Academy on February, and input from that discussion will be included in an updated version of this report.