



**New Mexico State University – Las Cruces  
General Education Assessment of Student Learning Outcomes  
2012-13 Final Report**

**Submitted by the  
Committee for the Assessment of Student Learning in General Education  
(CASL-GE)**

Greg Armfield – Communication Studies  
David Barney – Social Work  
Julie Fitzsimmons – Art  
Ryan Goss – Plant & Environmental Sciences  
Amal Mostafa – Math  
Cynthia Pierard – Library  
David Smith – Chemistry  
Kassia Wosick-Correa – Sociology

**October 22, 2013**

**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

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

**Assessment Procedure/Process**

A student assignment and scoring rubric were developed by the Committee for the Assessment of Student Learning in General Education (CASL-GE). The assignment was titled "Tweaking Twain." See Table 1 above for learning outcomes assessed.

A pilot assessment, focusing on Science and Math learning objectives was also implemented using instruments developed by a working subcommittee of the CASL-GE. A full assessment using the piloted instruments will be implemented spring 2014.

In fall 2012, various Viewing a Wider World (VWW) courses were randomly 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 sign letters of consent with their submitted assignment.

Table 2: View a Wider World Course Selection & Participation

Courses Identified for Participation	Delivery Type		Enrollment	Participated Y/N
	Face-to-Face	Online		
RGSC302V	X		39	Y
ANTH330V	X		23	N
ENGL341V	X		16	Y
ENGL339V	X		18	Y
SPAN364V	X		27	Y
GER 333V	X		22	Y
ECON384V	X		39	N
MKTG311V	X		45	Y
MKTG311V	X		45	Y
C EP451V	X		30	Y
E T309V	X		25	N
S WK331V	X		29	N
ECON335V		X	19	Y
EMD350V		X	26	Y
HL S301V		X	33	Y
HL S464V		X	26	Y
HL S380V		X	25	Y
HL S301V		X	29	Y

\* Pilot administered to these students

Students were given the “Tweaking Twain” assignment to respond to. In face-to-face classes, students were given the class session to complete the assignment. Most students completed the assignment in 10-25 minutes. Online students were given a 30-minute time frame to complete the assignment once it was opened. Upon completion of the assignment, student work was collected by members of the CASL-GE. Online work was collected through the Canvas learning management system by a CASL-GE member.

In April 2013, Assessment Liaisons participated in a scoring session at the NMSU Teaching Academy. An open invitation was also extended to faculty at large to participate. Twenty-one (21) faculty members attended the session. The Director of Assessment and the chair of the CASL-GE led a ‘norming session’ for scoring student work. All participants read and rated the student work individually. Then ratings were compared and discussed until consensus was reached.

The Science/Math pilot instruments were administered in one of the VWW courses, scored in similar fashion, but in a separate session attended only by CASL-GE members.

The following process was used:

- One copy of each paper was printed
- Two scoring rubrics were provided for each paper - scores ranged from 0-3 (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 attach their completed rubric to the back of the student paper. The second reader would then read the paper and repeat the process using a clean rubric.
- After the second reader scored the paper, they compared their scores to the scores given by the first reader. 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 to discuss the paper and see if they could reach greater consensus.
- Once both readers completed their evaluation, the paper and both completed rubrics were handed to a graduate student who recorded scores on an excel worksheet
- Scores on each component of the rubric were recorded, as well as composite scores for each paper

After all data was collected and recorded in the Excel spreadsheet, a request was made of the Office of Institutional Analysis to provide statistics on inter-rater reliability and range, central tendency and proportion of scores above 2.0 (competent) for all students; and for junior and senior students (as determined by Banner ID numbers). They were also asked to compare the sample population to the general population the sample was intended to represent (all juniors/seniors at NMSU).

On September 11, 2013, an Open Forum was held at the NMSU Teaching Academy to engage the campus community reviewing the findings and contributing to the discussion about assessment itself, and about how the results should be used to improve student learning.

## **Findings**

### Inter-Rater Reliability

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

Krippendorff's Alpha is a numerical measure of the extent to which two or more raters agree with each other when assigning a score to the same artifact. The observed disagreement between evaluators is corrected by the amount of disagreement expected by chance. It ranges from 0.0 (no agreement) to 1.0 (perfect agreement). Scores above 0.667 are generally desired.

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

Table 3: Krippendorff's alpha for each category on "Tweaking Twain"

<b>Rubric Dimension</b>	<b>Krippendorff's <math>\alpha</math></b>	<b>Interpretation</b>
Articulates a Position	0.7177	Good agreement
Communication: Rhetorical Strategies/Reasoned Argument	0.6921	Good agreement
Cultural Awareness	0.4902	Moderate agreement
Historical Perspective/Knowledge	0.6373	Good agreement
OVERALL	0.7468	Good agreement

Average Scores and Their Counts

N for "All" Students = 165; N for "Junior/Senior" Students = 90

The average scores (over readers and students) are given in Table 4.

Table 4. Average category and composite scores, percentages, and counts of students scoring "2" (competent) or above.

Outcomes	Average Score (3pt scale)		Percent scoring 2.0 (competent) or above		Number (out of 165/90) scoring 2.0 (competent) or above	
	All	Jr/Sr	All	Jr/Sr	All (165)	Jr/Sr (90)
Articulates a Position	1.78	1.89	63%	67%	104	60
Communication: Rhetorical Strategies/Reasoned Argument	1.68	1.84	50%	61%	83	55
Cultural Awareness	.92	.91	8%	9%	13	8
Historical Perspective/Knowledge	.85	.90	12%	11%	19	10
Composite	1.31	1.38	13%	16%	21	14

Tables 5-9 compare all student scores and junior/senior student scores for each learning category.

Table 5: Student Articulates a Position: Percentage of students scoring in each point range from 0.5-3.0

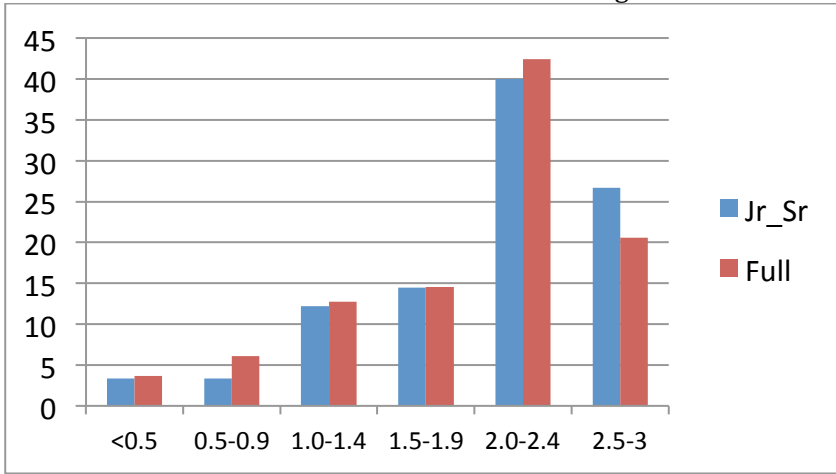


Table 6: Student Communicates Thoughts and Ideas: Percentage of students scoring in each point range from 0.5-3.0

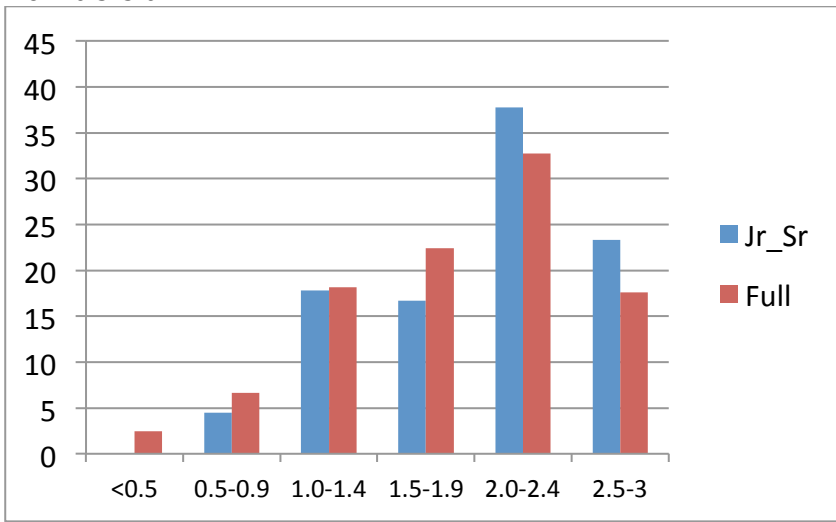


Table 7: Student Communicates Thoughts and Ideas: Percentage of students scoring in each point range from 0.5-3.0

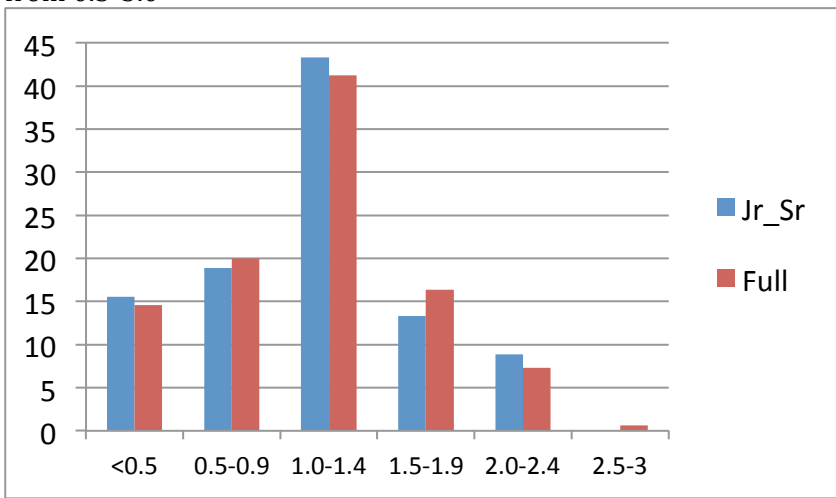


Table 8: Student Identifies Relevant Historical Context: Percentage of students scoring in each point range from 0.5-3.0

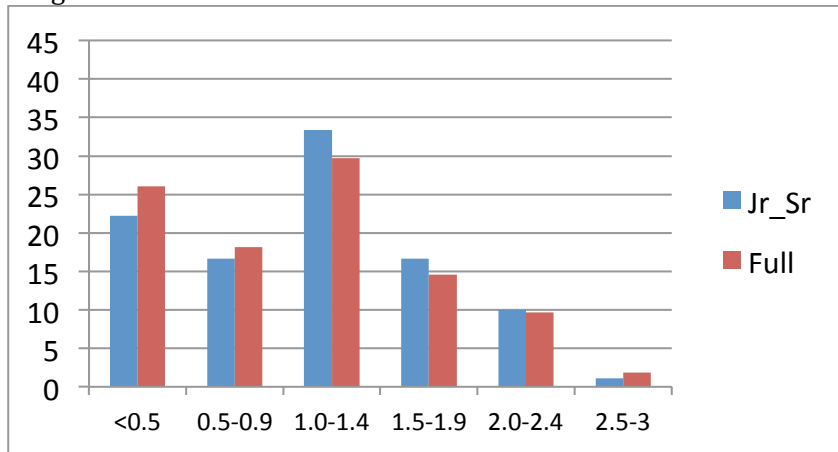
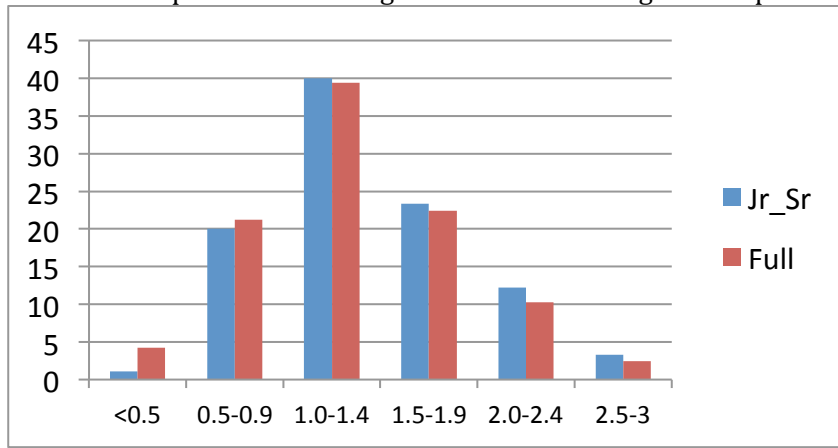


Table 9: Composite: Percentage of students scoring in each point range from 0.5-3.0



### Representativeness of Sample

When comparing students who took this assessment to those who did not, analysis was restricted to juniors and seniors (who, it is assumed, have completed most of their general education requirements). We used a Chi-Square test of homogeneity to compare these two groups and found no statistically significant differences between them for the following categories: race/ethnicity, citizenship, class, first generation status, minority status, primary campus, gender, student population, or underrepresented minority status. The primary college of students taking the assessment may not have been representative (p-value 0.0966), with engineering students possibly underrepresented. Full-time/part-time status was not representative (p-value 0.0088), with part-time students underrepresented in the assessment.

### **Discussion**

#### Inter-Rater Reliability

One of the goals from the Spring 2012 assessment was to improve inter-rater reliability. This was approached by strengthening the norming session, and ensuring faculty were clear about the need to assign scores based on rubric content, and not personal preference. Comparative results from Spring 2012 and Spring 2013 are provided in Table 10. We are pleased that the inter-rater reliability improved in three of the four categories, and that overall the inter-rater reliability falls into the “Good Agreement”

category. We are somewhat perplexed about the decreased level of inter-rater reliability in the Cultural Awareness category. Discussion at the Open Forum on September 11, 2013 suggested that this was likely indicative of a weakness in either or both the rubric, as it defines this category, and/or the assignment itself – that the assignment does not clearly align with the rubric for this particular category of learning.

Table 10: A comparison of Analysis of Inter-Rater Reliability from Spring 2012 to Spring 2013

Rubric Dimension	Spring 2012		Spring 2013 All Students		Spring 2013 Jrs/Srs	
	K's $\alpha$	Interpretation	K's $\alpha$	Interpretation	K's $\alpha$	Interpretation
Articulates a Position	0.561	Moderate agreement	0.669	Good agreement	0.718	Good agreement
Communication: Rhetorical Strategies/Reasoned Argument	0.548	Moderate agreement	0.600	Good agreement	0.692	Good agreement
Cultural Awareness	0.613	Good agreement	0.489	Moderate agreement	0.490	Moderate agreement
Historical Perspective/ Knowledge	0.536	Moderate agreement	0.655	Good agreement	0.637	Good agreement
Overall	--	--	0.710	Good agreement	0.747	Good agreement

### Student Performance

For “All” Students: Of the 165 student assignments that were evaluated for learning on four desired outcomes of the GE experience, At least 50% of students scored competent or greater in “articulates a position” and “communicates thoughts & ideas”. On Articulates a Position, 63% (compared to 66% last year) met the minimum desired level of performance (competent or skillful). On Communicates Thoughts & Ideas, 53% (compared to 40% last year) met or exceeded the desired level of performance. For the other outcomes, only 8% scored competent or above on Cultural Awareness, and 11.5% on Historical Perspective. Just under 13% of students (21 total), achieved an overall composite score that indicated a performance of competent or above on the desired outcomes.

For Juniors/Seniors: Of the 90 student assignments that were evaluated for learning on four desired outcomes of the GE experience, At least 60% of students scored competent or greater in “articulates a position” and “communicates thoughts & ideas”. On Articulates a Position, 66% met the minimum desired level of performance (competent or skillful). On Communicates Thoughts & Ideas, 61% met or exceeded the desired level of performance. For the other outcomes, only 9% scored competent or above on Cultural Awareness, and 11% on Historical Perspective. Just over 15% of students (14 total), achieved an overall composite score that indicated a performance of competent or above on the desired outcomes.

Of specific concern are student scores on Cultural Awareness. Cultural Awareness is an identified learning objective from Area IV courses of the New Mexico State Common Core competencies. It falls under NMSU’s broad GE outcomes of Cultural Awareness, and the NMSU Baccalaureate Experience learning objectives of Diversity and Self-Awareness. This same weakness was identified in our Spring 2012 Baccalaureate Experience assessment of Self-Awareness:

[http://assessment.nmsu.edu/files/2013/04/BE-Assessment-Report-2011-12.Final\\_.pdf](http://assessment.nmsu.edu/files/2013/04/BE-Assessment-Report-2011-12.Final_.pdf)



While discussion at the Open Forum acknowledged both improvement in and continued challenges of the assessment process and tools, there was also agreement that students are not, as a whole, performing at the desired level to indicate success in achieving GE learning outcomes. It was concluded that it is in the best interest of the campus community to take what we can learn from this assessment, what we can continue to learn in the future through similar processes, to collectively and effectively increase student success in GE learning.

The ongoing discussion needs to be about how we, as a campus, address identified weaknesses in student learning. Cultural awareness and historical perspective/knowledge, for example, have been identified through this assessment as problematic areas. The institution's challenge is to generate a comprehensive response to these findings.