# New Mexico State University – Las Cruces
# Assessment of General Education Student Learning Outcomes
# 2013-14 Final Report

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

Julie Fitzsimmons – Art
Greg Armfield – Communication Studies
Ryan Goss – Plant & Environmental Sciences
David Smith – Chemistry and Biochemistry
Kassia Wosick – Sociology
Amal Mostafa – Math
Cynthia Pierard – Library

**October 2nd, 2014**

## Overview

The Committee for the Assessment of Student Learning in General Education (CASL-GE) is responsible for assessing student achievement of general education (GE) learning outcomes and for supporting improvement of student learning in lower-division general education courses. Annual assessments, performed at a program level rather than course by course, measure student achievement of general education (GE) learning outcomes related to a subset of New Mexico's mandated general education competencies. During the 2013-2014 academic year, the committee performed a broad assessment of STEM (science, technology, engineering, and mathematics) competencies among undergraduate students primarily ranked as juniors or seniors, students who had presumably completed the majority of their required lower-division GE courses. A pilot assessment of writing competencies was also implemented in anticipation of supporting NMSU's Quality Initiative during the 2014-2015 academic year.

## Assessment Process

The STEM assignment and scoring rubric were developed by a working subcommittee of the CASL-GE and implemented on a pilot scale during the 2012-2013 academic year. In the assignment, students were placed in the role of a city water manager and were told they were responsible for responding to a fictional, impending water supply crisis. They were presented with contextual information, data and charts. They were asked to calculate and graph projected water demand from a quadratic water demand model, predict water shortage time scales from projected supply and demand data, and provide a recommended course of action with a brief rationale. As shown in Table 1, instrument questions align with the following state competencies for mathematics and laboratory sciences:

- Mathematics (Area II):
  - II(a) – Construct and analyze graphs and/or data sets
  - II(b) – Use and solve various kinds of equations
  - II(d) – Demonstrate problem solving skills within the context of mathematical applications
- Laboratory sciences (Area III):
  - III(c) – Communicate scientific information
  - III(d) – Apply quantitative analysis to scientific problems
  - III(e) – Apply scientific thinking to real world problems

Table 1: Alignment of Instrument questions with state common core competencies

| | | **Question #** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| **Area** | **Item** | **1-2** | **3(a), 3(b)** | **4** | **5** | **6-7** | **8** | **9(a), 9(b)** |
| Mathematics | II(a) | X | | X | | X | | |
| | II(b) | | X | | X | | | |
| | II(d) | | | | | | X | X |
| Laboratory Sciences | III(c) | | | | | X | X | |
| | III(d) | | X | | X | | | |
| | III(e) | | | | | | | X |

Questions 6-9 also align with Baccalaureate Experience "Effectiveness of Communication" and "Critical Thinking" outcomes.

Assessment data were collected by CASL-GE members or their representatives from randomly selected Viewing a Wider World (VWW) courses. 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. Following an IRB-approved procedure, the assignment administrator read a script to students regarding the purpose of the assessment and the voluntary nature of their participation. Students desiring to participate were provided with an informed consent letter (requiring their signature) and a scientific calculator and were given 30 minutes to complete the assignment. Data were collected from a total of 291 students in 10 different VWW courses over a time period from January 30th to March 18th.

In April 2014, Assessment Liaisons and interested faculty participated in a scoring session held at the NMSU Teaching Academy. A total of thirty-six faculty members attended the scoring session. Participants were trained in use of the assignment rubric by the chair of the subcommittee that developed the STEM instrument. Scoring of student work then proceeded according to the following process:

- Four to six graders were seated at each of several tables in the scoring room, with each assigned a grader number.
- A single copy of each student paper was printed and identified only by the student's Banner ID number. Stacks of student papers and rubrics were placed on each of the tables.
- Graders scored individual student papers according to the rubric, recorded scores and their grader number on the rubric sheet, and stapled the sheet to the back of the student paper. Each question was graded according to the following scheme: 0 = no evidence of competence, 1 = emerging, 2 = competent and 3 = skillful.
- Each student paper was scored independently by two different graders. The second grader compared scores and, when differing by more than 1 on any item, initiated discussion with the first grader to improve agreement.
- Once a student paper was scored by two graders, it was submitted to an administrative assistant who recorded scores and the students Banner ID in an Excel spreadsheet.

All data from the scoring session were submitted to the Office of Institutional Analysis for processing to determine information on range, central tendency, and inter-rater reliability, and for correlation of those items with student class standing (junior/senior or freshman/sophomore).

On September 12, 2014, an Open Forum was held at the NMSU Teaching Academy to report statistical findings of the assessment and engage the campus community in discussion of its significance.

The writing instrument under development at this time was administered in a similar fashion to a small audience of online VWW students in the spring. Following scoring and revision by CASL-GE members, a revised writing instrument was administered to a group of COMM 265 students in the summer with scoring and analysis to follow in the fall of 2014. Results of the pilot assessment will not be discussed further in this document.

## STEM Assessment Findings

### <u>Analysis Pools</u>

Statistical findings were determined for four different student pools:

1. All students completing the assessment (N = 291)
2. Students with valid a NMSU ID (N = 186)
3. Students with junior or senior standing and with a valid NMSU ID. (N = 152)
4. Students with freshman or sophomore standing and with a valid NMSU ID. (N = 34)

Results presented and discussed below focus primarily on the third pool since juniors and seniors were the target audience for the assessment. A comparison of the performance of the junior/senior pool and the freshman/sophomore pool is also discussed briefly.

Tabulated Findings

Results for students with junior or senior standing are listed in Table 2 and displayed graphically in Figures 1-5. In each figure, the left panel contains a description or graphic of the questions represented in that figure and the right panel gives the distribution of results and lists percentages of students achieving "competent" scores of 2 or higher.

Table 2: Findings for junior and senior students only

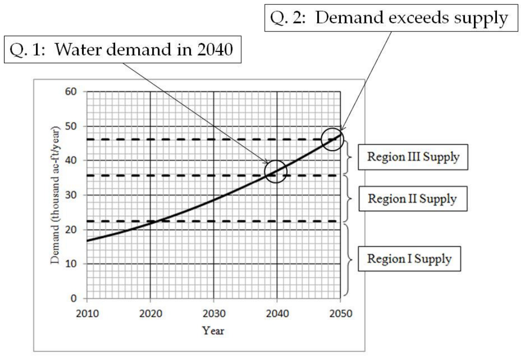
| Statistics | Q 1 | Q 2 | Q 3a | Q 3b | Q 4 | Q 5 | Q 6 | Q 7 | Q 8 | Q 9* | Overall |
|---|---|---|---|---|---|---|---|---|---|---|---|
| *Range* (%) | | | | | | | | | | | |
| <0.5 | 8.6 | 9.9 | 21.1 | 25.7 | 32.2 | 15.8 | 29.6 | 19.1 | 30.3 | 17.1 | 3.9 |
| 0.5-0.9 | 0.0 | 0.0 | 2.6 | 1.3 | 0.0 | 6.6 | 2.6 | 5.9 | 7.9 | 9.9 | 5.3 |
| 1.0-1.4 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.2 | 13.8 |
| 1.5-1.9 | 1.3 | 0.0 | 1.3 | 1.3 | 1.3 | 0.7 | 1.3 | 4.6 | 3.9 | 8.6 | 21.1 |
| 2.0-2.4 | 10.5 | 2.6 | 2.6 | 3.3 | 10.5 | 7.2 | 1.3 | 7.2 | 5.9 | 11.2 | 16.4 |
| 2.5-2.9 | 5.9 | 0.7 | 4.6 | 5.3 | 2.0 | 2.0 | 1.3 | 8.6 | 4.6 | 13.2 | 28.9 |
| 3.0 | 73.0 | 86.2 | 62.5 | 62.5 | 52.6 | 64.5 | 59.9 | 50.7 | 38.8 | 25.0 | 10.5 |
| *Central Tendency* | | | | | | | | | | | |
| Mean | 2.58 | 2.66 | 2.13 | 2.11 | 1.87 | 2.20 | 1.93 | 2.02 | 1.58 | 1.63 | 2.07 |
| Median | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 1.50 | 1.50 | 2.15 |
| *Condensed Range* (%) | | | | | | | | | | | |
| < 2.0 | 10.5 | 10.5 | 30.3 | 28.9 | 34.9 | 26.3 | 37.5 | 33.6 | 50.7 | 50.7 | 44.1 |
| >= 2.0 | 89.5 | 89.5 | 69.7 | 71.1 | 65.1 | 73.7 | 62.5 | 66.4 | 49.3 | 49.3 | 55.9 |
| *Krippendorff's Alpha* | | | | | | | | | | | |
| Alpha | 0.870 | 0.978 | 0.930 | 0.882 | 0.968 | 0.949 | 0.945 | 0.845 | 0.921 | 0.819 | 0.974 |

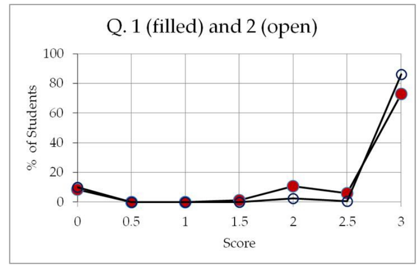
**Combination of Questions 9a and 9b*

Inter-Rater Reliability

Inter-rater reliability (IRR) was measured for each question on the assignment using Krippendorf's alpha, 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).

## Graphical Representation of Findings




Scores greater than or equal to 2
90% (Q. 1) and 90% (Q. 2)

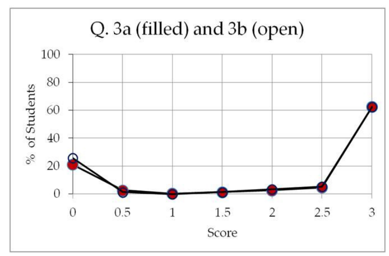
Figure 1: In questions 1 and 2, students are asked to read information from the graph shown in the left panel.

Q. 3(a): calculate water use *rate*

$$\frac{19{,}500 \text{ ac-ft}}{114{,}000 \text{ people}} = 0.171 \text{ ac-ft/person}$$

Q. 3(b): *convert* to gallons per person

$$\frac{0.171 \text{ ac-ft}}{\text{person}} \times \frac{326{,}000 \text{ gallons}}{1 \text{ ac-ft}} = 55{,}800 \text{ gallons}$$




Scores greater than or equal to 2
70% (Q. 3a) and 71% (Q. 3b)

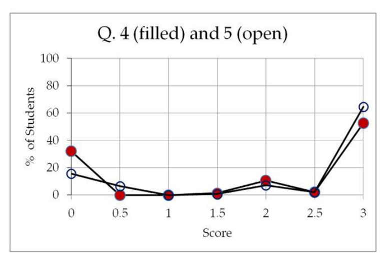
Figure 2: In questions 3(a) and 3(b), students are asked to calculate a water usage rate and convert their calculated rate from acre-feet per person into gallons per person.

Q. 4: Gathering and interpreting tabulated data

$22.5 + 10.5 = 33$ thousand ac-ft/year

Q. 5: Calculations using a quadratic water demand model:

$$D_{revised} = 0.0060(t-2005)^2 + 0.14(t-2005) + 17$$




Scores greater than or equal to 2
65% (Q. 4) and 74% (Q. 5)

Figure 3: In questions 4 and 5, students gather and interpret data from a table and do calculations by substituting into a quadratic formula.

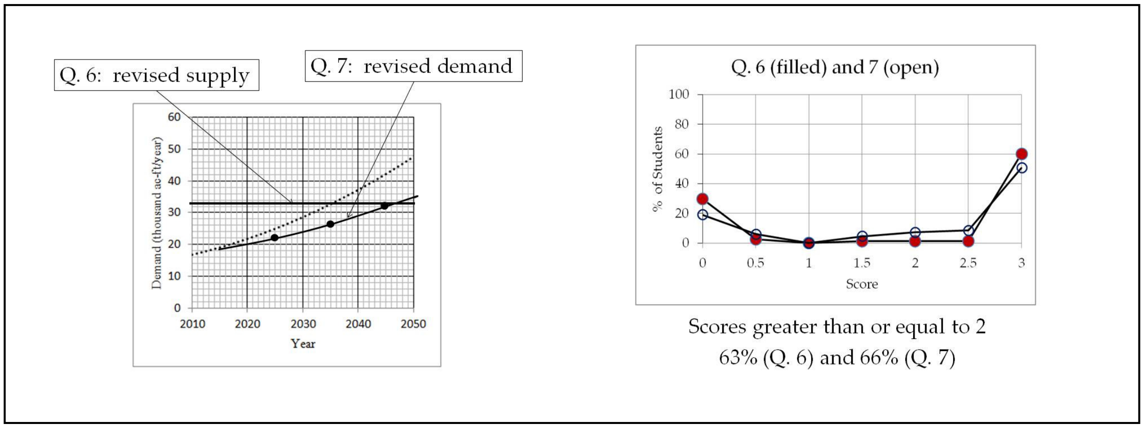



Figure 4: In questions 6 and 7, students are asked to draw a horizontal line to represent the revised water supply (calculated in Q. 4) and to graph the quadratic curve representing the revised water demand (calculate in Q. 5).

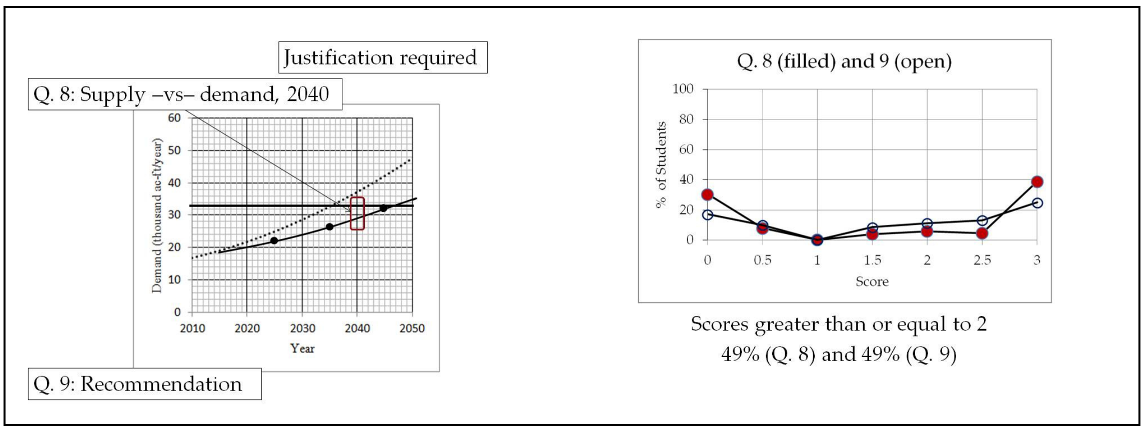



Figure 5: In questions 8 and 9, students are asked to predict whether the water supply will be adequate in 2040 (Q. 8, justification based on their graph required) and to make a recommendation to the City Council based on the information in the entire document.

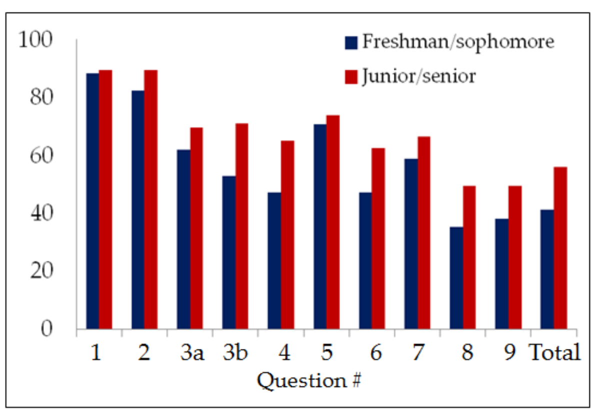



Figure 6: Comparison of performance of freshman/sophomore students with junior/senior students. Bars represent the percentage of students scoring in the competent range of 2 or greater.

## Representativeness of Sample

Demographic characteristics of students who took the assessment were compared against those who did not to evaluate whether the assessment sample is representative of the broader NMSU student population. Results are summarized in Table 3. In this measure, a smaller P-value indicates a greater difference between the junior and senior assessment takers and the broader NMSU population of juniors and seniors.

Table 3: Sample Representativeness (Assessment Takers vs. Non-Takers)

| Demographic Variable | P-Value | Possible Sample Biases |
|---|---|---|
| Citizenship | 0.4000 | None |
| Class | 0.9884 | None |
| College | <0.0001 | Business over-represented |
| Ethnicity/Race | 0.0025 | Whites over-represented |
| First generation | 0.4643 | None |
| Gender | 0.0035 | Males over-represented |
| Minority | 0.0007 | Minorities under-represented |
| Status (FT vs. PT) | <0.0001 | Full-time over-represented |
| Student population | 0.1144 | None |

# Discussion

## Reliability Issues

"Krippendorf's alpha" inter-rater reliability scores (Table 2) were in the "very good agreement" range (0.8 – 1.0) across all questions. This suggests suggest that the rubric and/or training used in this evaluation exercise encourage consistent evaluation between different judges.

Measures of the representativeness of sample (Table 3) showed significant differences between test takers and non test takers for several demographic categories including the over-representation of whites, males, and students from the Business College, and under-representation of minorities. This likely reflects similar representations of students enrolled in the particular VWW courses selected for data collection. While the significance of these findings is unclear, they do suggest that greater care may be needed in selection of courses for data collection to assure a more representative sample.

## Junior/Senior Student Performance

Most of the discussion of findings during the Sept. 12th Open Forum centered on the junior/senior performance data as shown in Figures 1-5 and in Table 2. The following items were near consensus opinions of that audience in response to the findings:

- Questions 1 and 2: Most students are able to accurately extract simple information from graphs.
- Questions 3(a) and 3(b): It is disappointing (but not surprising to STEM instructors) that a large percentage (~30 %) of students performed poorly on these questions involving fundamental skills related to rates, ratios, and unit conversions.

- Questions 4 and 5: It is somewhat surprising that around 1/3 of our students underperformed on the very simple data gathering and analysis task in question 4. The calculations involved in question 5 were much more complex, yet students did significantly better at this "plug and chug" activity than in the data gathering task. This likely reflects a weakness among students in higher-level thinking processes involved in gathering and organizing information.
- Questions 6 and 7: Students had more difficulty creating graphs than with reading them (questions 1-2), but it should be noted that the tasks in questions 6 and 7 were inherently more complex than those in questions 1-2. Open forum participants were more satisfied with student performance on these questions than on many others.
- Questions 8 and 9: It is unfortunate but not surprising that students performed poorly on these questions involving higher-level skills of analysis and justification/communication of information.

Overall, Open Forum participants were in general agreement that students were weakest on tasks involving higher-level thinking skills. Many agreed that students' weakness in reading comprehension may have negatively impacted their performance. While not unexpected, these data still suggest that focusing on things such as higher-level thinking and reading comprehension may the most productive approach to improvement of student outcomes in math and science competencies. All agreed that students certainly would not benefit from further training in "plug and chug" skills that have little relevance outside of a particular course.

<u>Development of competencies</u>

This STEM assessment was not designed to reveal how students develop GE competencies during their academic careers. Nevertheless, some information on this process may be found in the comparison of performance of students with different class standings (Figure 6). Students with junior/senior standing outperformed freshman/sophomore level students on every question, although the small sample size for the freshman/sophomore group makes the differences statistically significant only for a few items (Q. 3(b), 4, and overall score). It is interesting that the larger effects are seen in questions involving some higher-level thinking skills (Q. 4, 6, 8, and 9).

<u>Assessment Process</u>

There was significant discussion at the end of the Open Forum on ways to improve the quality of STEM assessment information. Some of the suggestions involved processing of existing data to correlate performance explicitly with STEM and math course completion, or with incoming ACT scores. Ongoing conversations with a representative of Institutional Assessment will clarify which of these suggestions may be feasible to implement and provide meaningful information. There was also significant discussion on how we might acquire baseline data on incoming students to better assess how students develop competencies while studying at NMSU. Some suggested that use of a nationally-normed instrument would be beneficial to assess the capabilities of our students.