

The relationship between cumulative credits and student learning outcomes: A cross-sectional assessment of information literacy and communication skills

Teresa Thonney¹ and Joe C. Montgomery²

Abstract: This article relates the efforts of faculty at one community college to define standards for achievement of two SLOs (critical thinking and effective communication) and to gather and analyze evidence of how well students meet those standards. Faculty from 13 disciplines assessed writing samples from 265 students. We found that, in general, students with more credits outperformed those with fewer credits. However, many students at every level demonstrated poor information literacy skills, indicating an area for curriculum improvement.

Keywords: critical thinking, information literacy, student learning outcomes, writing skills assessment

Introduction

Today, more than ever, accreditation commissions are asking colleges and universities to “provide substantive data-driven evidence of quality and effectiveness” (Baker, 2002, p. 5), and regional accreditors are finding an increasing number of institutions “out of compliance” with standards that require them to assess student achievement and improve programs based on their assessment findings (Nunley, Bers, & Manning, 2011).

At Columbia Basin College, a community college in Washington State, our accrediting body has called for a cycle of outcomes-driven assessment. In fact, our funding is increasingly based on demonstrating that students are achieving the institution’s student learning outcomes (SLOs). Those outcomes include:

- Apply information tools and resources
- Develop cultural awareness
- Think critically
- Communicate effectively
- Reason quantitatively and symbolically

These core competencies, listed among the student learning outcomes of most community colleges, cut across disciplines and are addressed in many courses; yet they are not mastered in any single course. They are instead developed over time.

According to the American Association for Higher Education 1999 Assessment Forum, assessment involves “setting appropriate criteria and high standards for learning quality; systematically gathering, analyzing, and interpreting evidence to determine how well performance matches those expectations and standards; and using the resulting information to document, explain, and improve performance” (cited in Ohlemacher & Davis, 2012, p. 11). In this article, we describe the efforts of faculty at Columbia Basin College to define standards for

¹ Columbia Basin College, TThonney@columbiabasin.edu

² Columbia Basin College, jmontgomery@columbiabasin.edu

student achievement of two SLOs—critical thinking and effective communication—and to gather and analyze evidence of how well our students meet those standards.

Critical Thinking and Information Literacy

Communication and critical thinking skills are, of course, essential for success in college and careers. In our digital age, one especially important kind of critical thinking is information literacy, specifically the ability to critically analyze and assess source information for reliability and relevance. It's a skill that is undeveloped in most first-year students, according to the findings of Project Information Literacy, an ongoing series of student surveys and interviews. In one of the project's recent surveys, for example, almost 50% of the 8,353 respondents expressed uncertainty about how to evaluate their research efforts (Head & Eisenberg, 2010).

Researchers involved with The Citation Project, another ongoing national study, are also evaluating how students select and use sources. They have found—by reviewing student papers from colleges of all types—that few students share faculty's definition of reliable information. In fact, say the project directors, many of the sources students cite are “stunningly cheesy” (Jamieson & Howard, 2011). Local studies routinely confirm that many undergraduates uncritically accept sources and source information (e.g., Daniels, 2010; Twait, 2005; Choinski & Emanuel, 2006; Swoger, 2011). Choinski, Mark, and Murphey (2003) found that even after taking an information resources class, many students are unable to recognize signs of a questionable website.

Given the importance of information literacy and given the evidence indicating it is a skill first-year students need to develop, our college's Teaching and Learning Committee designed a study to assess students' information literacy and written communication skills. These questions guided our research: 1) Would students recognize signs of bias or question the reliability of information in sources? and 2) Would students demonstrate what we considered to be college-level writing skills? Finally, in light of recent research indicating that little growth occurs in students' thinking and writing skills during the first two years of college (e.g., Arum & Roksa, 2011), we asked a third question: Do students in their second year of course work demonstrate higher critical thinking and written communication skills than students just beginning college? In other words, is there a correlation between the number of credits earned and students' information literacy and written communication skills?

Measuring Student Learning Outcomes

There are many tools for measuring critical thinking and communication outcomes at the institutional level, each with benefits and drawbacks. **Student satisfaction or experience surveys**, such as the National Survey of Student Engagement (NSSE), provide insight into students' satisfaction and attitudes (Zoellner, Samson, & Hines, 2008), as well as self-perceived gains in cognitive ability and writing skill (Cheng, 2001); but they may not be accurate measures of learning or skills (Coupe, 1993; Maughan, 2001; Maguire, Evans, & Dyas, 2001; Thompson, Pilgrim, & Oliver, 2005). In fact, the literature on self-perception indicates that novices tend to over-estimate their skill level (Schilling & Applegate, 2012; Kruger & Dunning, 1999).

Portfolios, common assignments, or other **course work** can provide more objective evidence of student achievement level. At the Community College of Baltimore County, for example, teams of faculty develop Common Graded Assignments. Students in designated

courses complete the assignment during a given term, and then normed faculty score random samples of the assignments using a six-item rubric that corresponds with the college's general education outcomes. (See <https://ccbcmd.edu/loa/great.html>.) Faculty at Miami Dade College have also designed writing tasks that allow them to measure several learning outcomes by scoring one artifact. Each term, willing faculty assign a common writing project, and teams of faculty use rubrics to score a random sample of completed projects from students who are about to graduate (Nunley, Bers, & Manning, 2011).

Common assignments, scored by normed faculty, provide a snapshot of the achievement level of a college's graduates, but the results do not necessarily tell faculty anything about how much students have learned or gained from their coursework. **Pre- and post-testing** is needed to demonstrate learning or skill development in students. A pre-test administered at the beginning of a course or program can gauge a student's initial skill or knowledge level, which can later be compared to the student's performance at the end of the course or program (Caspers & Bernhisel, 2005; Lombardo & Miree, 2003). However, post-tests administered immediately after a single course may reflect short-term retention rather than real learning (Cmor, Chan, & Kong, 2010); and pre- and post-tests at entrance and graduation are a challenge to administer at the institutional level (Nunley, Bers, & Manning, 2011).

Standardized aptitude tests are another direct measurement of student learning outcomes. Cisneros (2009), for instance, measured improvement in critical thinking skills by having graduate pharmacy students take the California Critical Thinking Skills Test (CCTST) at the beginning of the school year and the California Critical Thinking Disposition Inventory (CCTDI) at the end of the same school year. The CAAP Critical Thinking Test is another nationally normed exam for measuring students' ability to analyze and evaluate. These multiple-choice instruments have been validated and they are easy to administer, but they provide no opportunity to measure students' written communication skills.

The Collegiate Learning Assessment (CLA), on the other hand, is an open-ended test of analytic reasoning, critical thinking, problem solving, and writing skills. Its goal is "to provide a summative assessment of the *value-added* by the school's instructional and other programs (taken as a whole) with respect to certain important learning outcomes" (Klein, Benjamin, Shavelson, & Bolus, 2007, p. 418, original emphasis). Students analyze case studies and use information from sources provided to them to justify their decisions in real-world documents, such as memos or policy recommendations. Students don't necessarily take the CLA more than once, as a pre- and post-test, but schools can assess how their students perform in relation to students with the same mean SAT (or ACT) scores at other similar schools. Administering a standardized test like the CLA is, however, expensive (Swing & Coogan, 2010), and, like most community colleges, our institution does not require SAT or ACT scores, making a comparison of our students' CLA scores to those of students at comparable schools difficult.

To assess students' critical thinking and writing skills, we needed a means of assessment that, like the CLA, was open-ended, and that, like the common assignments used at the Community College of Baltimore County and Miami Dade College, would allow us to measure more than one outcome. We decided to have a sample of students write essays in response to a common prompt.

Methods

The Writing Prompt

In order to assess our students' ability to critically evaluate information from sources, we presented an argument and asked a sample of students to respond to that argument using data provided to them. We explicitly directed students to identify weaknesses in the evidence provided. Figure 1 includes the first page of the packet students received.

The Assignment: In a recent book, Mark Bauerlein, an English professor, argues that “the digital age stupefies young Americans and jeopardizes our future” by producing hyper-networked kids who rarely read books and who know more about the latest pop idol than they know about history, politics, economics, or culture. In fact, Bauerlein has dubbed today’s youth “The Dumbest Generation.”

Read and review the attached documents (Documents A-E), and then answer the following question:

Is the so-called intellectual decline in America as serious as critics like Bauerlein believe?

Respond in a well-organized, multi-paragraph essay (2 pages minimum, double-spaced). Include specific details from **all** of the attached documents in your essay. *In addition, explain the limitations of the attached data.* (For example, what additional information would help you write a more complete response to the question or help you assess the reliability of the attached data?)

Your essay will be evaluated according to how well you do the following:

- Respond appropriately to the assigned question (in bold type, above)
- Interpret the evidence provided
- Recognize the limitations or flaws of the evidence provided
- Respond in an organized and focused essay
- Use correct grammar, word choice, and spelling

Figure 1. The Writing Prompt

By directing students to “explain the limitations of the attached data,” highlighting this requirement in italics, and providing examples of questions they might ask about the data, we hoped to get an accurate picture of students' ability to assess evidence.

Five documents followed the prompt. Document A was a summary of Bauerlein's argument, written by Sharon Begley for *Newsweek Magazine*. The following sentences introduced Begley's two paragraph summary:

The following passage is taken from an article titled “The Dumbest Generation? Don't Be Dumb,” by Sharon Begley. It was originally published in *Newsweek Magazine*. In these paragraphs, Begley summarizes some of the evidence *Mark Bauerlein, a professor at Emory University, includes in his book.* (original emphasis)

A reader could not determine from Begley's summary whether or not Begley agrees with Bauerlein's argument, although her article title suggests she takes issue with Bauerlein's position.

Document B was a line graph showing U.S. high school graduation rates between 1870 and 2005. The graph indicates that the graduation rate rose steadily until 1970, when it peaked at

75% and then leveled off at roughly 70%. Document C was also a line graph, showing that college graduation rates have risen steadily since 1900. Document D included two tables, showing scores for the National Association of Educational Progress (NAEP) test for Math and Reading for two years (1975 and 2008), for three ethnicities (white, black, and Hispanic), and for three age groups (9-year-olds, 13-year-olds, and 17-year-olds). For all age groups and ethnicities, average math and reading scores have risen since 1975, although the rate of improvement has varied by ethnicity. The following citation appeared at the bottom of Documents B, C, and D:

Source: Murray, Charles. *Coming Apart*. Random House: New York, 2012.

The final document (E) was introduced in this way:

From *The Oprah Winfrey Show* titled *Waiting for "Superman": The Movie That Could Revolutionize Schools*. Published September 20, 2012. The screen shot below is taken from <http://www.oprah.com/relationships/Shocking-Education-Statistics>

The screen shot, titled "Shocking Education Statistics," with the subtitle "Get the truth about America's school system," included several figures. One figure claims that "approximately 7,000 kids drop out of school **every** day" (original emphasis). Other figures contradict information found in Documents B, C, and D. For example, one chart suggests that students' reading and math scores remained unchanged between 1971 and 2010 (no student ages or actual scores were identified). No source is identified for any of the statistics in Document E, and ads for McDonald's and birth control appear in the margins of the screen shot.

A number of red flags make the information in Document E at best limited and at worst suspect. The reported reading and math scores, for example, are dubious if for no other reason than they have stayed exactly the same for 40 years; and without knowing the dropout rates for earlier decades, the 7000-dropouts-a-day figure is of no use for responding to Bauerlein's argument.

Of course, we hoped that students would critically read the other documents as well, as some students did. For example, many students recognized that in order to respond to Bauerlein's argument, they must consider graduation rates and test scores in historical context. On the other hand, no student thought to ask who Charles Murray is, despite the fact that several figures were attributed to him. We do not assume that this was because students are familiar with the scholar. Some students missed other rhetorical clues, wrongly attributing claims in Begley's summary (of Bauerlein), for instance, to Begley. In addition, the provided documents addressed only one marker of intelligence (formal education), something only a few students noted in their essays.

Study Procedure

We recruited faculty volunteers willing to have their students write the essay. At the beginning of the class hour, each instructor read a script provided to them summarizing the instructions and grading criteria. Students received the writing instructions (Figure 1), followed by the collection of five sources (documents A-E) and paper on which to write their essays

In all, 265 students wrote essays in response to the writing prompt during a 60-minute class session. The students were enrolled in either a 100 or 200-level business, composition, history, nursing, psychology, or sociology course ($n = 176$) or a 1-credit first-year introduction (FYI) course ($n = 89$). The FYI course is required for all students and is taken before they earn any college credit. The sample was not random or representative, but it did include students who

had not yet taken any courses, students in their first year of course work, and students in their second year of course work.

Students enrolled in the FYI course wrote their essays during spring break of 2013; students enrolled in a college-level course wrote their essays during the first three weeks of spring quarter 2013. To ensure that we measured summative skill level rather than short-term learning acquired in a specific course, none of the instructors discussed the writing topic with students beforehand. In addition, with the exception of composition, the focus of the courses was not on written communication or information literacy.

The faculty volunteers told students that the writing was “required,” although it was left to the instructor to determine the weight of the assignment. In this way, we hoped to avoid the bias inherent in volunteer samples. However, making the essay a course requirement was likely more effective in the college-level courses than in the FYI course, where students receive a grade of pass or fail, a point we return to later.

Scoring Procedure

A team of 23 faculty members, representing 13 disciplines, met to score the essays using a rubric designed and previously revised by members of the college’s Teaching and Learning Committee. Used for decades to assess student writing, rubrics have also been used to evaluate other SLOs, including information literacy skills (Choinski, Mark, & Murphey, 2003; Knight, 2006; Oakleaf, 2009). Using a rubric allowed us to quantify and compare assessments of various graders.

The rubric included different performance levels for five criteria:

- responds appropriately to the prompt;
- interprets the evidence correctly;
- recognizes limitations or flaws in the evidence;
- responds in an organized and coherent essay; and
- uses correct grammar, word choice, and spelling.

The first three items in the rubric addressed the college’s critical thinking SLO; the last two items addressed communication.

Before rating essays individually, faculty scored sample essays and discussed their scores, repeating the process until sufficiently “normed.” Each criterion was rated on a four-point scale with “1” being the highest performance and “4” being the lowest. (See Figure 2.) Giving each criterion a separate score allowed us to measure several outcomes from a single writing sample.

Two scorers read each essay. When there was discrepancy of more than one point on any criterion, a third person read the essay. For example, if one rater gave an essay a score of “4” for “recognizes limitations or flaws in the evidence” and another rater gave the same element a score of “2,” a third rater read the essay, even if scores for the other rubric criteria were identical. In all, 40 of 265 essays (15%) received a third reading because scores awarded for one or more criteria by the first two readers differed by more than one point.

We averaged the readers’ scores to create a single rating for each element. We classified a criterion receiving a rating between 1.0 and 3.0 to be at “college level,” meaning that specific criterion within the paper would warrant a grade of “A,” “B,” “C,” or “D” if the paper were submitted in a college-level course. We classified ratings above 3.0 as below college level, meaning the specific criterion within the paper would warrant a grade of “F.” Because of the

different rubric criteria, some aspects of a paper could be deemed to be at college level (i.e., warranting a grade of A-D); other aspects could be deemed to be below college level (i.e., warranting a grade of F).

Inter-Rater Reliability

An earlier version of the rubric, used in a smaller pilot study in 2012, included six rating levels for each criterion, but that pilot resulted in low inter-rater reliability, attributed to both insufficient norming and too many gradations of evaluation. Raters using the revised rubric (Figure 2), however, had high levels of agreement.

Element	Superior (Always displays this element) 1	Skilled (Displays this element most of the time) 2	Minimal Competence (Displays this element occasionally or somewhat) 3	Inadequate/Below College Level (Displays this element rarely or not at all) 4
Responds appropriately to the prompt <i>Addresses the assigned question and all sections are related to the assigned question (i.e., are on topic).</i>				
Interprets the evidence correctly <i>Interprets evidence correctly and connects the evidence to the assigned purpose.</i>				
Recognizes limitations or flaws in the evidence <i>Recognizes flaws in the evidence or recognizes how the evidence is insufficient to address the problem.</i>				
Responds in an organized and coherent essay <i>Ideas and paragraphs are focused, sufficiently developed, logically connected, and related to the assigned purpose.</i>				
Uses correct grammar, word choice, and spelling				

Figure 2. Rubric Used to Assess Critical Thinking and Written Communication SLOs

We evaluated inter-rater reliability using the intra-class correlation coefficient (ICC) provided by the SPSS RELIABILITY procedure (SPSS, Version 19). High levels of the ICC indicate that raters gave similar ratings; low levels of agreement indicate that raters gave varying ratings. ICC values exceeding 0.7 were classified as “High” agreement, 0.5 to 0.7 as “Moderate” agreement, and 0.3 to 0.5 as “Fair” agreement. Values below 0.3 were labeled “Low” reliability.

As shown in Table 1, there were high levels of inter-rater agreement for four of the five elements: “responds appropriately to the prompt” (0.78); “interprets the evidence correctly”

(0.81); “recognizes limitations or flaws in the evidence” (0.85); and “responds in an organized and coherent essay” (0.80). The remaining element, “uses correct grammar, word choice, and spelling,” showed moderate to fair levels of agreement (0.59). By definition, papers read by three raters had lower inter-rater reliability than papers read by two raters. These values ranged from a high of 0.72 for “recognizes limitations or flaws in the evidence” to a low of 0.36 for “uses correct grammar, word choice, and spelling.”

Table 1

Inter-Rater Reliability

Number of Raters	Number of Cases	Q1. Responds appropriately to the prompt		Q2. Interprets the evidence correctly		Q3. Recognizes limitations or flaws in the evidence		Q4. Responds in an organized, coherent essay		Q5. Uses correct grammar, word choice, and spelling	
		ICC	Rater agreement	ICC	Rater agreement	ICC	Rater agreement	ICC	Rater agreement	ICC	Rater agreement
2 raters	225	0.78	High	0.81	High	0.85	High	0.8	High	0.59	Moderate
3 raters	40	0.47	Fair	0.61	Moderate	0.72	High	0.6	Moderate	0.36	Fair

Results

Overall Findings

Of the five rubric elements, as shown in Table 2, “uses effective grammar, word choice, and spelling” had the highest percentage of papers rated at “passing” level performance (85%), followed closely by “responds appropriately to the prompt” (82%). “Interprets the evidence correctly” and “responds in an organized and coherent essay” had 70% at passing level. “Recognizes limitations of the evidence” had the lowest rate of college-level performance (54%). While many students could correctly interpret the data and write a coherent paper, few challenged the data or questioned its sufficiency to answer the question in the prompt.

Table 2

Percentage of students scoring at college level for each criterion

Rubric Criteria	Percent of students who scored at college level
Uses effective grammar, word choice, spelling	85%
Responds appropriately to the problem	82%
Interprets the evidence correctly	70%
Responds in an organized and coherent essay	70%
Recognizes limitations of evidence	54%

Demographics, Ability, and Performance Variables

Additional analyses were performed to identify demographic or ability variables that might be associated with, and predictive of, rubric ratings. For example, one might anticipate that students with a higher cumulative grade point average, students with a greater number of credits, or students who performed at college level on the college's reading, writing, or math placement test (COMPASS) would earn higher ratings on the rubric elements. Consequently, multiple regression analyses were performed using the five rubric elements as the dependent variables and age, cumulative GPA, cumulative credits, and COMPASS outcomes as independent variables. Multiple regression analyses have the benefits of showing the total predictability of the dependent variable and the unique significance of each independent variable in predicting the dependent variable. The analysis holds constant the remaining independent variables while assessing the contribution of a single independent variable.

Results of the multiple regression analyses are shown in Tables 3 - 7. Table 3 shows the results for prediction of the first rubric element ("responds appropriately to the prompt") from age, cumulative GPA, and placement at college- or pre-college levels for writing, reading, and mathematics. The overall R-squared value of 0.15 was significant ($p < .01$), although the only significant predictor was cumulative credits ($\beta = -0.261$, $p < .0001$). That is, those with more credits received significantly higher ratings on this criterion.

Table 3

Multiple regression: Predicting the element "responds appropriately to the prompt"

Responds Appropriately to the Prompt Regression Table					
Variables	Unstandardized Coefficients		Standardized Coefficients	T	Sig.
	B	Std. Error	Beta		
Age	.013	.007	.115	1.710	.089
Cumulative GPA	-.025	.050	-.033	-.504	.615
Cumulative Credits	-.003	.001	-.261	-3.824	.000
College-level Writing	-.221	.116	-.143	-1.908	.058
College-level Reading	-.202	.118	-.125	-1.719	.087
College-level Math	.018	.131	.009	.136	.892

Notes: $R^2 = .15$ ($ps < .01$)

Table 4 shows a similar analysis for the second rubric element ("interprets evidence correctly") as the dependent variable. The R-squared value was again 0.15 ($p < .01$) and cumulative credits was again a significant predictor ($\beta = -0.228$, $p < .001$). However, placement in writing was also a significant predictor ($\beta = -0.167$, $p < 0.026$), with those placing at college-level in writing scoring significantly higher for this rubric element.

Table 5 shows the multiple regression results for predicting the element "recognizes limitations of evidence." The R-squared value for this analysis was 0.22 ($p < 0.01$), and there were three significant predictors. The strongest predictor was cumulative credits ($\beta = -0.302$,

$p < .0001$), followed by placement in writing ($\beta = -0.183$, $p < .011$) and placement in reading ($\beta = -0.271$, $p < .032$). Those with a higher number of accumulated credits and college-level placement in writing and reading received significantly higher ratings.

Table 4

Multiple regression: Predicting the element “interprets evidence correctly”

Interprets the Evidence Correctly					
Regression Table					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Age	.008	.007	.077	1.153	.250
Cumulative GPA	-.048	.050	-.062	-.955	.340
Cumulative Credits	-.003	.001	-.228	-3.344	.001
College-level Writing	-.259	.116	-.167	-2.239	.026
College-level Reading	-.123	.117	-.076	-1.046	.297
College-level Math	-.119	.131	-.060	-.908	.365

Notes: $R^2 = .15$ ($ps < .01$)

Table 5

Multiple regression: Predicting the element “recognizes limitations of evidence”

Recognizes Limitations of Evidence					
Regression Table					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Age	.009	.008	.071	1.109	.268
Cumulative GPA	.016	.054	.019	.308	.759
Cumulative Credits	-.004	.001	-.302	-4.629	.000
College-level Writing	-.318	.124	-.183	-2.570	.011
College-level Reading	-.271	.125	-.150	-2.161	.032
College-level Math	-.177	.140	-.080	-1.264	.207

Notes: $R^2 = .22$ ($ps < .01$)

Table 6 shows the multiple regression analysis for the fourth rubric element, “responds in an organized and coherent essay.” The R-squared value of 0.24 was significant ($p < .01$) and cumulative credits was the strongest predictor ($\beta = -0.296$, $p < .0001$), followed by placement levels in writing ($\beta = -0.187$, $p < .009$) and reading ($\beta = -0.172$). Those with more credits and higher placement scores in writing and reading received significantly higher ratings.

Table 7 provides the results of the final multiple regression analysis, with the rubric element “uses effective grammar, word choice, and spelling” as the dependent variable. The R-squared value of 0.26 was significant ($p < .01$). Cumulative credits was the strongest predictor (beta = - 0.293, $p < .0001$), followed by placement levels in writing (beta = -0.242, $p < .001$) and reading (beta = -0.175, $p < .011$). Again, those with a higher number of college credits and college-level placement in writing and reading received significantly higher ratings.

Table 6

Multiple regression: Predicting the element “responds in an organized and coherent essay”

Responds in an Organized and Coherent Essay					
Regression Table					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Age	.001	.007	.014	.212	.832
Cumulative GPA	-.052	.047	-.069	-1.120	.264
Cumulative Credits	-.004	.001	-.296	-4.580	.000
College-level Writing	-.284	.107	-.187	-2.647	.009
College-level Reading	-.272	.109	-.172	-2.496	.013
College-level Math	-.021	.122	-.011	-.174	.862

Notes: $R^2 = .24$ ($ps < .01$)

Table 7

Multiple regression: Predicting the element “uses effective grammar, word choice, and spelling”

Uses Effective Grammar, Word Choice, and Spelling					
Regression Table					
Variables	Unstandardized Coefficients		Standardized Coefficients	t	Sig.
	B	Std. Error	Beta		
Age	-.010	.006	-.102	-1.631	.104
Cumulative GPA	-.013	.042	-.018	-.305	.761
Cumulative Credits	-.003	.001	-.293	-4.601	.000
College-level Writing	-.335	.096	-.242	-3.472	.001
College-level Reading	-.252	.098	-.175	-2.574	.011
College-level Math	.084	.109	.048	.769	.443

Notes: $R^2 = .26$ ($ps < .01$)

For all five rubric elements, “cumulative credits obtained” was the most highly significant predictor of ratings. For each analysis, higher levels of credits were associated with

higher ratings on the elements, with the significance levels at $p < .001$ or higher. The standardized regression coefficient, indicating the relative predictor importance, was considerably higher for cumulative credits than for any other predictor. COMPASS writing placement was a significant predictor for four of the elements, and COMPASS reading placement was a significant predictor for three of the elements. Note that three predictors did not achieve statistical significance for any of the five analyses: age, cumulative GPA, and COMPASS math placement. That is, being older, having higher overall grades, and having higher math placement level were not associated with higher ratings on the rubric.

These analyses suggest the importance of accumulating college credits to receiving higher ratings on the critical thinking and communications rubrics. The analyses also downplay the likelihood that the results can be explained by maturation effects, ability effects, or student attrition (as discussed further below).

Analysis by Credit Levels

We performed a follow up analysis to obtain a better subjective sense of the importance of credit levels on rubric ratings. Subjects were categorized as having 0 college-level credits (entering students, $n = 89$), having 1-44 college-level quarter credits (first-year students, $n = 90$), or having 45+ college-level quarter credits (second year, if full time, $n = 86$). Students with 45+ credits outperformed the other two student groups on all five elements. (See Figure 3.) For the element showing the weakest performance, “recognizes limitations of evidence,” over 70% of those with 45+ quarter credits achieved college-level performance, compared to 34% of entering students. For all five elements, students with 45+ quarter credits were rated at college-level performance by at least a 20 percentage-point margin over entering students.

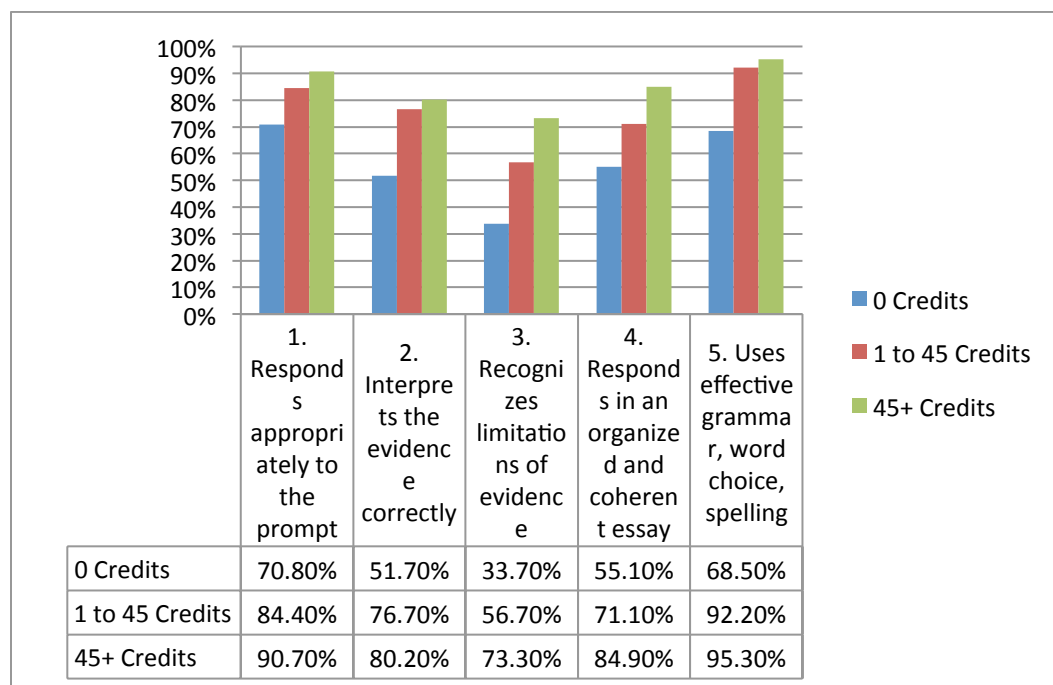


Figure 3. Percent of Students Performing at College Level by Number of Quarter Credits Earned

Discussion of Results

When faculty see inferior sources cited in student papers, they often assume that students have used the first sources they encountered in order to complete an assignment with minimal effort. In other words, inferior evidence is thought to be a symptom of lazy research. But our results suggest the explanation is not that simple. Students in this study were provided with sources and asked to identify the limitations of the evidence. Yet despite being provided with some “stunningly cheesy” sources (e.g., the *Oprah Winfrey Show* website), 46% of the students failed to recognize any limitations. Like the students Choinski, Mark, and Murphey (2003) observed, many of the students in the present study did not recognize indicators of an unreliable website. This finding is of concern, given that students rely on the Web more than the library when conducting research (Head, 2013).

Many of today’s digital-savvy students enter college aware that the Web is rife with misinformation (Manuel, 2005); so why did so many in our sample fail to recognize the evidence in the *Oprah Winfrey Show* site as flawed or incomplete? One possible explanation is that while students may “know” that they should not believe everything they read online, they may not know how to determine what not to believe. That is, students may cite inferior sources because they believe they are citing quality sources (Grimes & Boening, 2001). Certainly, in the case of the *Oprah Winfrey Show* website, the ethos of Winfrey herself may have influenced students. Students might assume that information from her show’s website would be credible. Another possible reason so many students failed to critically assess the *Oprah Winfrey Show* website, in particular, is the medium of presentation (print vs. web). We provided students with a screen shot in hopes that the ads in the margins would signal to students that the site was not an academic one. However, seeing a screen shot reproduced on a page is not the same as encountering the website online. Presentation affects how viewers rate the reliability of Internet sources (Wathen & Burkell, 2002). Did students assume that sources distributed by their professors would be credible, despite being directed to critique the sources? Would students have been more skeptical of the information if they had discovered it for themselves online? Results from the Citation Project suggest the answer to the latter question is no, and, indeed, many of the students who cited information from the *Oprah Winfrey Show* website made no mention of the suspect nature of the information.

Many others students, perhaps recognizing that the data they received conflicted, opted to write one-sided arguments, picking and choosing the evidence that helped make their case but ignoring the rest. This is not surprising, as it is common for novice writers to draw from only those sources who agree with each other or with whom they agree (Penrose & Geisler, 1994). In fact, Jonathan Baron’s (1995) research into “myside” bias in student arguments indicates that many students regard a one-sided argument to be *more* convincing than one that recognizes other viewpoints. Further research is needed to determine the most effective methods for training students to consider multiple viewpoints in their writing.

Closer analysis of our rating data revealed another interesting finding: For all five rubric criteria, students with 0 credits had the lowest percentage performing at college level, those with 1-44 quarter credits had a higher percentage at college level, and those with 45 or more quarter credits had the highest percentage at college level. This finding was somewhat unexpected, given what other research has shown about the rate of student development. Arum and Roksa (2011), for example, found that for 45% of the 2,322 four-year college or university students they studied there were “no statistically significant gains in critical thinking, complex reasoning, and

writing skills” during their first two years of college (p. 36). Similarly, when Cisneros (2009) measured the critical thinking skills of graduate pharmacy students at the start and end of a given school year, he found no significant improvement in students’ total scores. Miller (2004) documented critical thinking gains in pharmacy students between *admission* and *graduation*, but he also found no significant difference *from year to year*, suggesting that measurable and significant gains in thinking are cumulative, occurring over a time span longer than one year.

Ultimately, the study procedures do not allow us to determine why students with the most credits earned the highest critical thinking and communication ratings. Some of the difference between student groups might be attributed to attrition. Completion rates at community colleges are routinely around 50% (Nunley, Bers, & Manning, 2011). It seems plausible that those who enter college with the lowest critical thinking and writing skills would be among those most likely to drop out, and those who enter college with higher than average critical thinking and writing skills would be more likely to persist. In addition, as already noted, students with 0 credits (enrolled in FYI) had the least motivation to do their best. However, it bears repeating that the multiple regression analyses served to control for age, GPA, and ability level (through placement test results), that these three factors had *no* significant relationship to the ratings, and that cumulative credits showed strong, significant relationships to the ratings for all five rubric criteria. These findings suggest that improvements in critical thinking and communication skills were not due simply to maturation effects, academic success, or attrition of students with lesser ability levels. Even after removing FYI student results from the data, significant differences remained between students with 45 or more quarter credits and students with 1-44 credits. It is difficult to discount the finding that across all rubric items students with more credits generally displayed better critical thinking and writing skills than did students with fewer credits.

Without both formative and summative assessment, measuring gains in any individual student’s abilities isn’t possible. Unfortunately, at community colleges, assessing at entrance and graduation would capture only a small percentage of students, thanks to an ever-shifting student population (Nunley, Bers, & Manning, 2011). In addition, most community college students work, raise families, have attended other colleges, and/or temporarily leave and then return, making it difficult for any college to “assert the student attained the [general education] knowledge and skills as a result of the courses and activities in which they engaged at the college” (Nunley, Bers, & Manning, 2011, p. 15). Further, the likelihood of obtaining pre- and/or post-test scores for a *non-college* attending control group is virtually nil. In short, as much as we would like to see the feature film version of each student’s development, we must settle for the occasional snapshot. Nonetheless, the snapshot of our students led us to two important conclusions: Students with the most accumulated credits demonstrate the highest critical thinking and writing skills, yet many students at every credit level fail to recognize the difference between quality and inferior source information.

The college’s research team will continue to assess students’ critical thinking and communication skills, as well as other general education outcomes. It will be important in these further studies to introduce a more theoretical basis into the research. At this point, it is unclear why so many students (70%) could correctly interpret the data yet so few (54%) could evaluate the quality of data. According to Bloom’s Taxonomy (1956) and the Revised Taxonomy (Anderson, Krathwohl, & Bloom 2001), challenging assumptions and questioning data reflect higher level thinking skills. In addition, the willpower literature (e.g., Baumeister & Tierney, 2011) suggests that cognitive resources are limited and that thinking and reasoning skills decline as these resources are expended. Perhaps the effort required to read and interpret the data, and to

write an analytic essay, depleted the cognitive resources needed for questioning and challenging the data. At any rate, future research should incorporate more of a theory base and develop *a priori*, theory-based hypotheses for testing.

Conclusion

For faculty, the reasons for measuring student achievement at the institutional level go beyond fulfilling the requirements of administrators and accrediting bodies. Without institutional-level research, faculty cannot know whether or not student learning lasts beyond the span of a single course. Significant gains in thinking and writing skills, in particular, don't occur in 12-15 weeks, making it difficult for faculty to determine whether or not "the environments they create are having their intended effects on student outcomes" (Astin & Antonio, 2012, p. 141).

This article describes how faculty at our community college assessed students' critical thinking and written communication skills. It is the first step in a cycle of assessment that we hope leads to data-driven curriculum changes. In our case, we found evidence that students are not getting sufficient practice in evaluating source information. The Teaching and Learning Committee has shared these findings with all college faculty, and we hope that information literacy will now be more frequently addressed across the curriculum—in assignment prompts, in course requirements, and in course lectures. As Angelo (1999) notes, "Assessment should be first and foremost about improving student learning and secondarily about determining accountability for the quality of learning produced. In short: Though accountability matters, learning still matters most."

References

- Anderson, L. W., Krathwohl, D. R., & Bloom, B. S. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Angelo, T. A. (1999). Doing assessment as if learning matters most. *AHHE Bulletin*, 51, 1-10. Retrieved from http://assessment.uconn.edu/docs/resources/ARTICLES_and_REPORTS/Thomas_Angelo_Doing_Assessment_As_If_Learning_Matters_Most.pdf
- Arum, R., & Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago: University of Chicago Press.
- Astin, A. W., & Antonio, A. L. (2012). *Assessment for excellence: The philosophy and practice of assessment and evaluation in higher education* (2nd ed.). Blue Ridge Summit, PA: Rowman & Littlefield Publishers.
- Baker, R. L. (2002). Evaluating quality and effectiveness: Regional accreditation principles and practices. *Journal of Academic Librarianship*, 28(1), 3-7. doi: 10.1016/S0099-1333(01)00279-8
- Baron, J. (1995). "Myside bias in thinking about abortion." *Thinking and Reasoning*, 1, 221-235.

Thonney, T. & Montgomery, J.C.

Baumeister, R. F., & Tierney, J. (2011). *Willpower*. New York: Penguin Books. doi: 10.1080/13546789508256909

Bloom, B. S. (Ed.) (1956). *Taxonomy of educational objectives, handbook I: Cognitive domain*. New York: Longman.

Caspers, J., & Bernhisel, S. (2005). What do freshmen really know about research? Assess before you teach. *Research Strategies*, 20(4), 458-468. doi: 10.1016/j.resstr.2006.12.016

Cheng, D. X. (2001). Assessing student collegiate experience: Where do we begin? *Assessment & Evaluation in Higher Education*, 26(6), 525-538. doi: 10.1080/02602930120093869

Choinski, E., & Emanuel, M. (2006). The one-minute paper and the one-hour class: Outcomes assessment for one-shot library instruction. *Reference Services Review*, 34(1), 148-155. doi: 10.1108/00907320610648824

Choinski, E., Mark, A. E., & Murphey, M. (2003). Assessment with rubrics: An efficient and objective means of assessing student outcomes in an information resources class. *Portal: Libraries & the Academy*, 3(4), 563-575. doi: 10.1353/pla.2003.0078

Cisneros, R. M. (2009). Assessment of critical thinking in pharmacy students. *American Journal of Pharmaceutical Education*, 73(4), 1-7. doi: 10.5688/aj730466

Cmor, D., Chan, A., & Kong, T. (2010). Course-integrated learning outcomes for library database searching: Three assessment points on the path to evidence. *Evidence Based Library and Information Practice*, 5(1), 64-81.

Coupe, J. (1993). Undergraduate library skills: Two surveys at Johns Hopkins University. *Research Strategies*, 11(4), 188-201.

Daniels, E. (2010). Using a targeted rubric to deepen direct assessment of college students' abilities to evaluate the credibility of sources. *College & Undergraduate Libraries*, 17(1), 31-43. doi: 10.1080/10691310903584767

Grimes, D. J., & Boening, C. H. (2001). Worries about the Web: A look at student use of Web resources. *College & Research Libraries*, 62, 11-23. doi: 10.5860/crl.62.1.11

Head, A. J. (2013). *Learning the ropes: How freshmen conduct course research once they enter college*. Project Information Literacy Research Report. Retrieved from http://projectinfolit.org/images/pdfs/pil_2013_freshmenstudy_fullreport.pdf

Head, A. J., & Eisenberg, M. B. (2010). *Truth be told: How college students evaluate and use information in the digital age*. Project Information Literacy. Retrieved from http://projectinfolit.org/images/pdfs/pil_fall2010_survey_fullreport1.pdf

Jamieson, S., & Howard, R. M. (2011). Unraveling the citation trail. Project Information Literacy smart talk, no. 8. Retrieved from <http://projectinfolit.org/st/howard-jamieson.asp>

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. *Evaluation Review*, 31(5), 415-439.

Knight, L. A. (2006). Using rubrics to assess information literacy. *Reference Services Review*, 34(1), 43-55.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, 77(6), 1121-1134. doi: 10.1037/0022-3514.77.6.1121

Lombardo, S. M., & Miree, C. E. (2003). Caught in the web: The impact of library instruction on business students' perceptions and use of print and online resources. *College & Research Libraries*, 64, 6-22. doi: 10.5860/crl.64.1.6

Maguire, S., Evans, S. E., & Dyas, L. (2001). Approaches to learning: A study of first-year geography undergraduates. *Journal of Geography in Higher Education*, 25(1), 95-107. doi: 10.1080/03098260125539

Manuel, K. (2005). What do first-year students know about information research? And what can we teach them? In H. Thompson (Ed.), *Currents and convergence: Navigating the rivers of change, proceedings of the ACRL 12th National Conference* (401-417). Chicago: Association of College and Research Libraries.

Maughan, P. D. (2001). Assessing information literacy among undergraduates: A discussion of the literature and the university of California-Berkeley assessment experience. *College & Research Libraries*, 62(1), 71-85. doi: 10.5860/crl.62.1.71

Miller, D. (2004). An assessment of critical thinking: Can pharmacy students evaluate clinical studies like experts? *American Journal of Pharmaceutical Education*, 68(1), 1-6. Retrieved from <http://archive.ajpe.org/aj6801/aj680105/aj680105.pdf>

Nunley, C., Bers, T., & Manning, T. (2011). *Learning outcomes assessment in community colleges*. National Institute for Learning Outcomes Assessment (NILOA). Retrieved from <http://www.learningoutcomeassessment.org/documents/communitycollege.pdf>

Oakleaf, M. (2009). Using rubrics to assess information literacy: An examination of methodology and interrater reliability. *Journal of the American Society for Information Science and Technology*, 60(5), 969-983. doi: 10.1002/asi.21030

Ohlemacher, J., & Davis, A. (2012). Community college strategies. *Assessment Update*, 24(3), 11-13.

Thonney, T. & Montgomery, J.C.

Penrose, A. M., & Geisler, C. (1994). Reading and writing without authority. *College Composition and Communication*, 45(5), 505-520. doi: 10.2307/358762

Schilling, K., & Applegate, R. (2012). Best methods for evaluating educational impact: A comparison of the efficacy of commonly used measures of library instruction. *Journal of the Medical Library Association*, 100(4), 258–268. doi: 10.3163/1536-5050.100.4.007

Swing, R. L., & Coogan, C. S. (2010). *Valuing assessment: Cost-benefit considerations* (NILOA Occasional Paper No. 5). Urbana: University of Illinois and Indiana University, National Institute for Learning Outcomes Assessment. Retrieved from http://learningoutcomeassessment.org/documents/SwingCoogan_001.pdf

Swoger, B. J. M. (2011). Closing the assessment loop using pre- and post-assessment. *Reference Services Review*, 39(2), 244-259. doi: [10.1108/00907321111135475](https://doi.org/10.1108/00907321111135475)

Thompson, G., Pilgrim, A., & Oliver, K. (2005). Self-assessment and reflective learning for first-year university geography students: A simple guide or simply misguided? *Journal of Geography in Higher Education*, 29(3), 403-420. doi: 10.1080/03098260500290959

Twait, M. (2005). Undergraduate students' source selection criteria: A qualitative study. *Journal of Academic Librarianship*, 31(6), 567-573. doi: 10.1016/j.acalib.2005.08.008

Wathen, C. N., & Burkell, J. (2002). Believe it or not: Factors influencing credibility on the web. *Journal of the American Society for Information Science and Technology*, 53(2), 134-144. doi: 10.1002/asi.10016

Zoellner, K., Samson, S., & Hines, S. (2008). Continuing assessment of library instruction to undergraduates: A general education course survey research project. *College and Research Libraries*, 69(4), 370–383. doi: 10.5860/crl.69.4.370