Assessing Student Learning Outcomes in a Large, General Education, World Regional Geography Course

Roy Cole, Ph.D.
Department of Geography and Planning
B-4-203 Mackinac Hall
Grand Valley State University
Allendale, Michigan, 49401 USA

Abstract

In this experiment learning was assessed for 257 undergraduates in four sections of World Regional Geography using a pre-test posttest method. Content knowledge, critical thinking, and spatial cognition were measured as dependent variables. A variety of possible Independent variables were examined in the study. Data were analyzed with stepwise regression and ANOVA. Results indicate that the most important determinants of student learning are motivation and ability. There were some differences by sex, age and transfer status. Overall, those who were more motivated (higher GPA) or had greater ability (ACT) did better on the pretest and posttests, demonstrated higher critical thinking, and experienced the highest learning outcomes in the course.

Key words: assessment, student learning outcomes, world geography

1. Introduction

The purpose of the present study was to assess student learning outcomes of a general education world regional geography course of wide subscription at a mid-sized university in Michigan, Grand Valley State University. According to a well-known definition, "Assessment is the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development" (Palomba & Banta, 1999: 4).

Although the empirical assessment of learning in the United States began in 1900 according to Shavelson (2007), assessment of student learning outcomes (SLO) in higher education began in the United States in the mid-1980s on a small number of college campuses (El-Khawas 1995). Questions were raised at the time whether our students were learning what they should be learning, whether they were able to apply specialized knowledge and skills they had learned at university in the workplace, and whether they were able to communicate effectively and solve problems (Palomba & Banta 1999). In 1988, the U.S. Department of Education ordered all federally-approved accreditation bodies to include institutional outcomes in their assessment criteria (U.S. Department of Education 1988).

Not surprisingly, a spate of assessment literature has appeared in the last decade. Several collections or articles have been published by the Association for American Colleges and Universities and The Higher Learning Commission of the North Central Association of Colleges and Schools (Table 1).

The call for academic accountability that has permeated the university community has been echoed in the discipline of geography as well. For example, The National Council for Geographic Education published Geography for Life: National Geography Standards in 1994. The eighteen standards, the first national standards in education in the United States, provide a solid framework for the assessment of student learning for K-12 and higher education. Estaville, Brown and Caldwell (2006) demonstrate the important role of assessment for an entire geography undergraduate program (Figure 1) while Klein (2003) provides a clear example of a standards-based approach to assessing student learning outcomes in a university World Regional Geography course.

According to a dedicated issue of the *Journal of Geography* devoted to identifying issues of utmost importance for the discipline in the twenty-first century, Bettis (2001) opined that assessment must be a priority "... in order to ensure that we are graduating geographically literate students who are prepared to deal effectively with the spatial concerns, local to global, facing our society" (Bettis 2001: 171).

Grand Valley State University began systematically assessing individual student learning in 2007, although it had collected indirect measures of student learning many years prior to that. The effort followed a period of exploration which began in 2003 when the Provost called for the formation of a task force comprised of faculty with experience in assessment and charged them with the task of exploring methods to assess GVSU's General Education program. Prior to 2007, assessment of General Education courses was done rather indirectly through a form sent to each department chair for each General Education course taught by the department.

Although the present study has its roots in GVSU's General Education assessment of student learning outcomes, it is, in some ways, much more than the GVSU assessment program but, in other ways, much less. My purpose in conducting the research described in this paper will be evident if the reader considers a well-known definition of assessment. According to Suskie's (2004) micro-level definition (i.e. at the level of the instructor rather than the program), assessment is a process in which the instructor does the following.

1. Establishes clear and measurable expectations of student learning outcomes.
2. Ensures that students have sufficient opportunities to achieve expectations.
3. Systematically gathers, analyzes, and interprets empirical evidence to measure how well student learning matches expectations.
4. Uses the results to understand and improve student learning (Suskie 2004: 3).

The assessment undertaken in the present research may not even be considered assessment by some. I would argue that Paradis and Dexter's (2007) "learner-centered" assessment of a field geography course fits well with Suskie's assessment model. But although I "established," "ensured," "systematically gathered, analyzed, and interpreted" my data, and am beginning to "understand," I am not quite ready to "use" the results to improve student learning—and I may never be. Although I am interested in what students learned, at this point I am more interested in knowing what makes my highest achieving students different from the others.

The scholarly literature indicates that student learning is influenced by self-efficacy-beliefs (Pajares 1996), self-regulation (Zimmerman 2002), attendance (Newman-Ford, Fitzgibbon, Lloyd & Thomas 2008), socioeconomic status (Betts, Zau, & Rice 2003), class size, aptitude, and attitude (Harrington, Kulasekera, Bates, & Bredahl 2006), GPA (a proxy for effort, perhaps motivation), sex of the professor when it matches the student, sex, and personality type as measured by the Keirsey Temperament Sorter (McCarty, Padgham, & Bennett 2006). Pajares (1996) study of undergraduates found support for Bandura's research (1986) that showed "efficacy beliefs mediate the effect of skills or other self-beliefs on subsequent performance by influencing effort, persistence, and perseverance" Pajares (1996: 552)– i.e. internal motivation is the most important variable in student learning. This important behavioral finding was echoed by Fredricks, Blumenfeld, & Paris (2004) as "school engagement."

The ACT score as an independent variable was included in the present study based on recent, comprehensive research by Sackett, Borneman, and Connelly (2008) who reviewed and analyzed criticisms against the ACT score as a predictor for short-term and long-term academic performance. Their major findings indicate that 1) the ACT is a valid and robust measure of academic performance, 2) validity is not an artifact of socioeconomic status, 3) coaching is not a major determinant of test scores, and 4) motivational mechanisms (i.e. stereotype threat) are not major determinants of ACT test performance.

The present study is rather limited in scope and has no large agenda such as that described by Bettis (2001) or some of the literature briefly reviewed above. In the academic setting of Grand Valley State University, this study may be considered exploratory: its purpose was to give the author empirical data with which to understand his students and provide a baseline and direction for future research that *will* improve student learning outcomes.

2. Method

2.1. Population and sampling

Geography 235 (GPY235) World Regional Geography was probably among the first classes taught at Grand Valley State University when it first opened in 1963.

The course description in the GVSU course catalog says, "GPY 235 World Regional Geography. A survey of geography followed by an examination of specific geographic concepts. Physical, cultural, economic, and related factors will be given more emphasis than place-name geography. Three credits. Offered every semester. Fulfills General Education Social and Behavioral Sciences Foundations and World Perspectives requirements" (GVSU Course Catalog, 2007-2008: 471).

The GVSU General Education program distributes the courses into a Liberal Arts framework which students use to satisfy General Education requirements. As the course catalog indicates, GPY235 falls into two General Education categories: the Social and Behavioral Sciences Foundations and World Perspectives. In theory, instructors who teach World Regional Geography must demonstrate that their course meets the content and skills goals of both General Education categories (Table 2). In practice, only two instructors (this author is one of them) have developed course assessment plans (CAP) and conducted assessment of student learning outcomes. Two other faculty in the department have developed CAPs for GPY235 and it is expected that they will assess their sections of the course in Fall 2009.

Every year over 1000 students take World Regional Geography at Grand Valley State University. Depending on the classroom capacity, the maximum number of students in a single class ranges from 75 to 50. Summer classes generally enroll fewer students regardless of the size of the classroom (Table 3).

Students who enroll in World Regional Geography at Grand Valley State University come from a wide variety of majors or, particularly if the student is a freshman or sophomore, come to the class "undeclared." World Regional Geography is a lower-level general education course that students find attractive because it is included in two general education categories, Social and Behavioral Sciences and World Perspectives. On successful completion of the course students have satisfied two general education requirements.

Generally, a student will choose a section of World Regional Geography that fits into his or her schedule of courses for that semester. More specifically, early morning sections have lower enrolment, late morning and afternoon sections are generally full. It is likely that students do not plan their schedules around World Regional Geography; instead they plan around their major area of study and World Regional Geography as a double-dipping General Education requirement is fit in where convenient.

My students in two sections of World Regional Geography in Fall semester, 2007 and two sections in Winter Semester, 2008, comprised the subjects in this study. Both the Fall 2007 and Winter 2008 sections were held from 1:00 to 2:15 and from 4:00 to 5:15 pm. Although my subjects were not selected by a random process from the entire pool of students who enrolled in World Regional Geography for Fall 2007 and Winter 2008, on the face of it a reasonable person could conclude that my subjects represent the population of student who took World Regional Geography during those two semesters. Nevertheless, it was never my intention to represent World Regional Geography students in general.

Two hundred and eighty six students began the study: 138 in Fall, 2007, and 148 in Winter, 2008. The actual study population came to 257 because 29 students did not complete the course. Six of these 29 students never attended class at all, 15 dropped after the first exam, and 7 students had personal problems that prevented them from participating fully in the class. All of the latter students received a grade of "F" and, since their record in the course was extremely thin, were dropped from the study. One student was excluded from the study because he was in a graduate program. Of the 257 students who comprise the dataset, 127 were enrolled in Fall 2007 and 130 in Winter 2008. For the four sections of the course that were taught over the two semester period, there were 57 students in section 2, 70 in section 8, 68 in section 3, and 62 in section 6.

2.2. Measurement

At both the pretest and the posttest the students were asked to answer several questions on a cover sheet (Table 4, independent variables). Two pretests were given: the first on content knowledge in World Regional Geography and the second on spatial ability. there were 18 students (7% of the students) who added the course after the first day of class and consequently received no pretest.

The pretest 1 and posttest 1 (n=121, or 47.1% of the students)) measured content knowledge that would be covered in the class while the pretest 2 and posttest 2 (n=118, or 45.9% of the students) measured spatial ability. Critical thinking was measured with the GIS essay score (see below).

2.3. Computed variables

I computed four additional variables from the above. The computed variables, ABSQDIFF and ABSMapDiff represent the difference in the pre and post tests PRETQ and POSTQ and PREMAP and POSTMAP respectively. These will be discussed in the results section. I categorized the ACT score for ANOVA into terciles: High, Medium, and Low. I used an ACT score of 27 and above to indicate the "high" scoring group. The actual scores ranged from 27-33. I chose 27 because that is the cutoff for acceptance into the Honors College at Grand Valley State University as well as 89% on the national ranks for the ACT composite score. An ACT score of 24 to 26 was considered "medium" (75 to 85% on the national ranks), and any score less than <24 was considered "low" (less than 75% on the national ranks). The mean ACT score for the entire pool of students (n=241; 16 students did not report their scores), was 23.48 with a standard deviation of 3.28. The distribution of ACT scores was negatively skewed (0.39 with a standard error of 0.16) Since I felt that the 27 cutoff used by the GVSU Honors college was important to maintain I chose a distribution-free method of categorizing the ACT scores and sought to have roughly equal numbers in both the high and low groups (Table 5). I also categorized GPA for ANOVA from the numerical values into three groups: A, B, and C. There was one D student among the 257 subjects (Table 6). Both categorical variables, ACTCAT and GPACAT, are similar in that they isolate the top and the bottom ends of the performance ranges.

2.4. Conceptualizing and operationalizing critical thinking

Critical thinking is "skilled, active, interpretation and evaluation of observations, communications, information, and argumentation" (Fisher and Scriven 1997: 20). I defined critical thinking following Facione (2007) as "a process of purposeful, self-regulatory judgment that drives problem solving and decision making. Furthermore, Facione (2007) identifies a set of six cognitive skills (Table 7) and ten dispositions as being at the heart of critical thinking. For my purposes in the present research, the six cognitive skills were more important than the twelve behavioral dispositions.

I operationalized Facione's six cognitive skills in an assignment focused on the use of GIS. Students are required to produce two well-made maps with ArcView 3.3 depicting the percent of the regional vote taken by the candidates Muhammadu Buhari and Olusegun Obasanjo in the 2003 Nigerian presidential election and answer three essay questions. There is a distinct geographic pattern to the vote that students are asked to interpret. Each essay question is required to be exactly one page in length. The essay questions are:

1. What is the geographic pattern? What does the geographic distribution of the vote for the candidates tell you about the electorate in Nigeria?
2. How did this pattern develop? What is the deep or "ultimate" causes" (à la Jared Diamond) of this pattern?
3. What are the implications of this pattern for the future of Nigeria as a state, particularly with reference to the development of its "civil society," often touted as necessary for real democracy?

The questions are targeted toward the progressive use of critical thinking skills and it made sense to me to use the average score in percent for the three essays to assess critical thinking. To describe the geographic patterns on the map the student must use Interpretation and Explanation. To assign proximate causes to the patterns the student must use Interpretation, Analysis, Evaluation, Inference, Explanation, and Self-Regulation. Self-regulation will help the student avoid explaining with reference to ultimate causes at this point providing that the student has thoughtfully engaged classroom discussion about proximate and ultimate causality. All of the skills are necessary for the student to develop an argument on the ultimate causes of the pattern (Question 2). Question three is the most important question of all. To answer question three correctly, the student must understand the geographic pattern of the vote depicted on the two maps, the major proximate causes of the distribution, the ultimate causes (deeper causes in time and space) that set the pattern in place, and a profound understanding of the meaning of the term, "civil society." Lastly, the student must link all of the preceding to make an explanation about the future of Nigeria as a state.

Throughout the semester proximate and ultimate causality are discussed as are such concepts as "civil society" and I leave a breadcrumb trail of hints for students to follow. For example, on my World Regional Geography course syllabus the caption for a map of the language families and population of ethnolinguistic groups of Uganda states." Figure 4. Uganda. Language groups, major spoken languages, and approximate number of speakers.

What implications might Uganda's ethnolinguistic diversity have for the development of the idea of the state, the concept of citizenship, and the development of voluntary private organizations and interest groups ("civil society") that balance the raw coercive power of the state?" In my lecture on influences on development I discuss population and use Nigeria as an example of a state with three cultural cores, and two different and opposing legal systems: a mix of British common law and Nigerian traditional law in the south and Sharia in the north. Nevertheless, students are required to consult the scholarly literature on Nigeria to build their arguments about the geographic distribution of the vote in the 2003 Nigerian presidential election.

The successful student must use all of the cognitive skills listed in Table 7. Skill number six is particularly important because clear thinkers, and those who have done the appropriate research in the scholarly literature on Nigeria, will not be distracted by superficial explanations for the Nigerian vote.

2.5. Data collection

Data collection began in Fall Semester, 2007, on the first day of class and finished at the end of Winter Semester, 2008. On the first day of class students were given one of two pretests. The first pretest consisted of a blank page with instructions to draw a world map including the continents, the names of the continents, the names of the major oceans, and the equator. The second pretest was a multiple-choice test comprised of 25 questions (Appendix A). The pretest questions were focused on factual knowledge of the world regions, knowledge about the discipline of geography, and knowledge about theoretical frameworks used in geography. Students had several opportunities to demonstrate critical thinking in the course. For the present paper the total points from three essays from the required GIS exercise was used to assess critical thinking.

Map and question pretests were given to every other student such that each map maker had a multiple-choice test taker on either side and that each multiple-choice test taker had a map maker on either side. Data on the characteristics of the student, major, GPA, ACT score, grade student expected to get in the class along with some distractors (student's favorite food, book recommendation, etc.) were collected at both the pre and post tests. GPA and ACT were crosschecked with Banner, Grand Valley State University's student information system.

Both the 25 multiple-choice questions and the map were included in the posttest. The posttest was comprised of 100 multiple-choice questions within which the 25 pretest questions were embedded. In addition to the multiple-choice questions, there was an essay question and instructions to draw a map of the world including the continents, the names of the continents, the names of the oceans, and the equator.

2.6. Scoring of the pre and posttests

The 25-questions of the pretest and the embedded 25 questions in the 100 question posttest (the final exam) were machine scored and each student's score was converted to a percentage. Evaluation of the hand-drawn map of the world was done in three areas: 1) continents, 2) oceans, 3) the equator. Each student's score on the map (49 total possible points) was converted to a percentage. Continents and oceans were evaluated on completeness, correct spelling, and placement accuracy (Table 8). Expected names for continents were Africa, Antarctica, Asia, Australia, Europe, North America, and South America. One point was awarded for each correct name. Two points were awarded for Eurasia if Asia and Europe were not indicated on the map. Expected names for major oceans were Atlantic Ocean, Pacific Ocean, and Indian Ocean. No points were awarded for additional oceans or seas. The accuracy of a continent's shape and its placement accuracy were evaluated by holding up the student's work to a National Geographic world wall map in the Robinson projection. Shape and placement accuracy evaluation was made using a dichotomous measure: 1 meaning accurate and 0 meaning inaccurate for each "continent." The placement accuracy of the equator was given five points. To earn the total points the student's drawn equator had to cross South America and Africa in the appropriate places. Three points were awarded if the student had drawn a line representing the equator but the line did not cross in the two test places, Ecuador and Gabon. If the student's line strayed even farther from the correct equator, more points were deducted (Table 8).

3. Results

Of the 257 students in the study population, 39.3% were Freshmen, 37.7% Sophomores, 14.8% Juniors and only 8.2% Seniors. Men were 47.5% and women 52.5%. There were no statistically significant differences in class standing by sex (Table 9).Students had an average age of 19.9 years, an average score on the ACT test of 23.48, a mean GPA of 3.13. Just under 32% of the students reported never having taken a geography course in high school but 46.3% (114) reported having had one geography course.

Almost 18% reported having had two geography courses in high school while 4% reported having 3 or more. Since the students were not asked about the specifics of the "geography" courses that they had reported, and since geography as a stand-alone subject is rarely offered at the high school level in Michigan, it is likely that the course remembered was in Social Studies or perhaps Earth Science. This variable was removed from further analysis because of the uncertainty. Likewise, the variable GPYU, the number of courses the student reported having taken at university was removed from the analysis. Results show that almost 90% of the students (218) had not taken a geography course at university, 8.9% (22) had taken one course, 2% (5) had taken 2 courses, and one student reported having taken five geography courses (Table 10).

At the pretest, students were very optimistic about their grades: 75.6% expected to earn an A or A- while 24.4% expected a grade in the B range while at the posttest the expectations had diminished somewhat (Figure 2). Actual grades were rather different (Figure 3).

3.1. Pre and posttests, GIS essays

Not surprisingly, there was a significant difference between the pre and posttest scores for the 25 questions (PRETQ and POSTQ). The pretest score on the 25 questions had a mean score of 11.3% (n=121) while the mean for the posttest score was 52.3% for just the pretest group (n=121) . The mean difference between the two test scores was 41%, the t-value was 27.2, significant and well below the .05 level (3.63e-053). On the 25-question posttest for the entire group of students (n=257), there was no significant difference between the group that received the 25-question pretest and the group that was not pretested (t of 1.16, sig. 0.249). Therefore no testing effect.

The difference between the pre and posttest of the world map was highly significant as well. With a pretest mean score of 69.9% (n=112) and a posttest score of 77.5%(n=112), a mean difference of 7.7%, a t-value of 6.8, and a probability of 4.79e-010. On the map posttest, there was no significant difference between the group that received the pretest and the group that was not pretested (t of 0.785, sig. 0.433). One can conclude that there was no testing effect.

Oneway ANOVA revealed that there were no statistically significant differences between class standing and the pretests, postests or GIS essays (Table 11) or class standing and sex (table not presented). But there were significant differences both by ACT group and GPA group (Tables 13 and 14).

There were no differences between GPA groups on the pretest (PRETQ) but, similar to the ACT results, there were differences on the posttest (POSTQ). The A group scored significantly higher than either the B or C groups. The GPA group A was significantly different from groups B and C on the GISESSAYS: there was a 17.2% difference between groups A and C and a 10.9% difference between groups A and B while the difference between B and C was 6.3%. Attendance varied significantly by GPA group as well. GPA group A differed from C by 15.3% and group B by 7.5% while the attendance for group B was 7.8% higher than group C. The ACT groups did not differ significantly on attendance.

Regression analysis was used to test for a relationships with the scores on the pre and post tests and GIS essays. Of the independent variables used in this analysis, GPA, ACT, Sex, Age EXGRPRE none was significant in explaining the PRETQ scores. The case was different for the POSTQ. The ACT score, the GPA and age in years were significant variables (R^2=26.3 for the model, F=28.1, sig. <.0001; GPA: slope t=3.26, sig .001; ACT: slope t=6.80, sig. <.0001; AGE: slope t=2.68, sig. .008). The difference between the PRETQ and POSTQ (ABSQDIFF) scores was best explained by GPA (R2=14.3; F=18.3, sig. <.0001, slope t=4.27, sig. <0001).

The pretest score on the map (PREMAP) was best explained by ACT and SEX (R^2=15.7, F=6.29, sig. .001). The slope coefficients were significantly different from zero: ACT, slope t=2.66, sig. .009; SEX, slope t=-2.24, sig. .027. The ACT score had the strongest relationship with the posttest score on the map (POSTMAP): R^2 was 12.3; F=11.1, sig. <.0001; slope t=4.33, sig. < .0001. SEX was not significant. There were no significant relationships between any of the predictor variables and the change in scores on the map, ABSMapDiff. GPA and SEX were the most important predictors of the students' scores on the GISESSAYS (R^2= 21.5; F=32.45, sig. <.0001; GPA slope t=7.55, sig. <.0001; SEX t=-3.52, sig. .001). SEX was negatively correlated with the GISESSAY score with women scoring lower than men.

Although student attendance (AttendPCT) was not significant in any of the regression analyses as an independent variable, it was best predicted with AGE, and GPA.

For the entire model F was 19.55, sig. <.0001; coefficients for GPA and AGE were significant with a t value of 5.3 (sig. <.0001) and -3.5 (sig. <.0001) respectively. For every unit of increase in age there was a 1.8% decline in attendance but for every unit increase in GPA (scale of 0 to 4) there was an increase of 10.4% in attendance.

3.2. Student learning

Figure 4 illustrates learning from the pre to the posttest for the 25 questions (PRETQ and POSTQ) categorized by ACT score and GPA. The graphed data were taken from Tables 12 and 13.

The difference between the scores on the pre and post test for the 25 questions was significantly different for students in the High ACT condition (27-33). High ACT students increased their scores by 49%, Medium ACT students increased their scores by 40% and low ACT students increased their scores by 36%.

Although there were no significant differences on the 25 question pretest (PRETQ) between the three ACT groups, significant differences (sig. <.0001) were apparent at the posttest where there was a mean difference of 9.5% and 16.2% between the High and Medium and the High and Low ACT conditions respectively and a 6.7% difference between the Medium and Low ACT conditions (Figure 4).

GPA groups were not significantly different in their scores on the PRETQ while the A GPA group differed significantly from both the B and C groups (7.7% and 9.5% respectively; both sig. <.0001). In a similar fashion, the posttest (POSTQ) scores of the A group were 17.2% higher than the C group and 11.4% higher than the B group. The C group was 5.9% less than the B group.

For the map pre and post test, except for the middle ACT group, there were no such dramatic differences within groups. Between groups differences were large. The High ACT category scored high on the map pretest (79.3%) and remained high at the posttest (81.5%), while the Medium ACT group pretested at 70.1% and posttested at 77.5%, scoring the largest difference of any of the three groups. The Low ACT group didn't change (67.9% at the pretest and 67% at the posttest). The only significant difference within the three groups was the Medium ACT category: mean difference 9.2%, significant at <.0001 level (Figure 5). At the pretest, the High ACT group differed significantly from the other two groups (a 9.2% and 11.3% difference respectively) while the Medium and Low ACT groups were the same. At the posttest, however, both the High and Medium ACT groups were significantly different from the Low group (14.4% and 10.5% respectively). The posttest scores of the Medium group rose to the level of the High ACT group and there was no significant difference between the High and Medium groups.

GPA group scores did not differ significantly from one another on the map pretest (PREMAP) but the Group A (mean score 80.7%) posttest score (POSTMAP) differed significantly from the B group (mean score 74.3%). There were no differences with the C group (mean score 73. 6%).

With regard to the change from the map pretest to the map posttest (PREMAP to POSTMAP), there were no significant differences by GPA group, each group increased its score by about 5 to 7%: A group increased from 73.6% to 80.7%, B group from 68.7% to 74.4%, and C group from 66.2% to 73.7%.

4. Conclusion

The results of the present research support the general findings in the research literature on the important role of motivation (GPA) and ability (ACT) and age. Students entered my course knowing little about world geography and by the end of the semester certain groups had distinguished themselves from the others. Older students with high motivation and ability scored higher on the multiple-choice posttest. Motivation (GPA) provided the only significant explanation of change from pre to post tests. Those students with higher ability (ACT) performed the best on the spatial cognition (POSTMAP) exercise. The findings also suggest that differences in spatial cognition between men and women may be cultural rather than biological. Women differed significantly from men in spatial cognition on the first day of class but, at the time of the posttest on the last day of class, did not. It may be that the experience of taking World Regional Geography helped women, on average, overcome the difference in spatial cognition that existed on the pretest of spatial cognition (PREMAP). This result is supported by the findings of Stempke (2003) who, in a comparative study of undergraduate men and women from 2000 to 2003, found that experience closed the gender gap in measured in spatial ability. The evidence suggests that critical thinking also seems to be a product of motivation (GPA) rather than ability (ACT). A surprising result that deserves further investigation was the significant difference between men (+) and women (-) in critical thinking.

5. Bibliography

Association of American Colleges and Universities. (2004). *Taking responsibility for the quality of the baccalaureate degree. A report from the greater expectations project on accreditation and assessment.* Washington, D.C.: AACU.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*.

Englewood Cliffs, New Jersey: Prentice Hall.

Betts, J.R., A.C. Zau, & Rice, L.A. (2003). *Determinants of student achievement: New evidence from San Diego*. San Francisco: Public Policy Institute of California.

Bettis, N.C. (2001). Assessment issues in geographic education for the twenty-first century. *Journal of Geography*, 100(4), 171-174.

El-Khawas, E. (1995). *Campus trends, 1995*. Higher Education Panel Report 85. Washington, D.C.: American Council on Education. In C. A. Palomba & T. W. Banta. (Eds.) (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.

Estaville, L. E., B. J. Brown, & Caldwell, S. (2006). Geography undergraduate program assessment. *Journal of Geography*, 105(6), 239-248.

Facione, P. A. (2007). *Critical thinking: What it is and why it counts*. San Francisco: California Academic Press.

Fisher, A. & Scriven, M. (1997). *Critical thinking. Its definition and assessment*. Point Reyes, California: Edgepress.

Fredricks, J. A., P.C. Blumenfeld, & Paris, A.H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74(1), 59-109.

Harrington, D. R., K. Kulasekera, R. Bates, & Bredahl, M. (2006). *Determinants of student performance in an undergraduate financial accounting class*. Working Papers, Department of Agricultural Economics and Business. Guelph, Ontario: University of Guelph.

Klein, P. (2003). Active learning strategies and assessment in world geography classes. *Journal of Geography*, 102(4), 146-157.

Leskes, A. & Wright, B.D. (2005). *The art and science of assessing general education outcomes: A practical guide*. Washington, D.C.: Association of American Colleges and Universities.

McCarty, C., G. Padgham, & Bennett, D. (2006). Determinants of student achievement in Principles of Economics. *Journal for Economics Educators*, 6(2), 1-9.

Miller, R. (2007). *Assessment in cycles of improvement: Faculty designs for essential learning outcomes.* Washington, D.C.: Association of American Colleges and Universities.

Newman-Ford, L., K. Fitzgibbon, S. Lloyd, & Thomas, S. (2008). A large-scale investigation into the relationship between attendance and attainment: A study using an innovative, electronic attendance monitoring system. *Studies in Higher Education*, 33(6), 699-717.

Pajares. F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, 66 (4), 543-578.

Palomba, C. A. & Banta, T. W. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education*. San Francisco: Jossey-Bass.

Paradis, T. W. & Dexter, L. R. (2007). Teaching and assessment in an undergraduate field analysis course. *Journal of Geography*, 106(4),171-180.

Sackett, P. R., M. J. Borneman, & Connelly, B. S. (2008). High-stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist*, 63(4), 215:227.

Shavelson, R. J. (2007). *A brief history of student learning assessment: How we got where we are and a proposal for where to go next*. Washington, D.C.: Association of American Colleges and Universities.

Stempke, J. (2003). *Mapping the Middle East: Gender differences and the spatial cognition of students* (Unpublished senior thesis). Department of Geography and Planning, Grand Valley State University, Allendale, Michigan.

Suskie, L. (2004). *Assessing student learning: A common sense guide*. Bolton, Massachusetts: Anker.

U.S. Department of Education. (1988). Secretary's procedures and criteria for recognition of accrediting agencies. *Federal Register*, 53(127), 25088-25099.

Van Kollenburg, S. E. (ed.) (2003). The future focused organization: Focusing on the effectiveness of teaching and learning. Volume 3 of *A collection of papers on self-study and institutional improvement*. Chicago, Illinois: The Higher Learning Commission.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64-70.

Figures

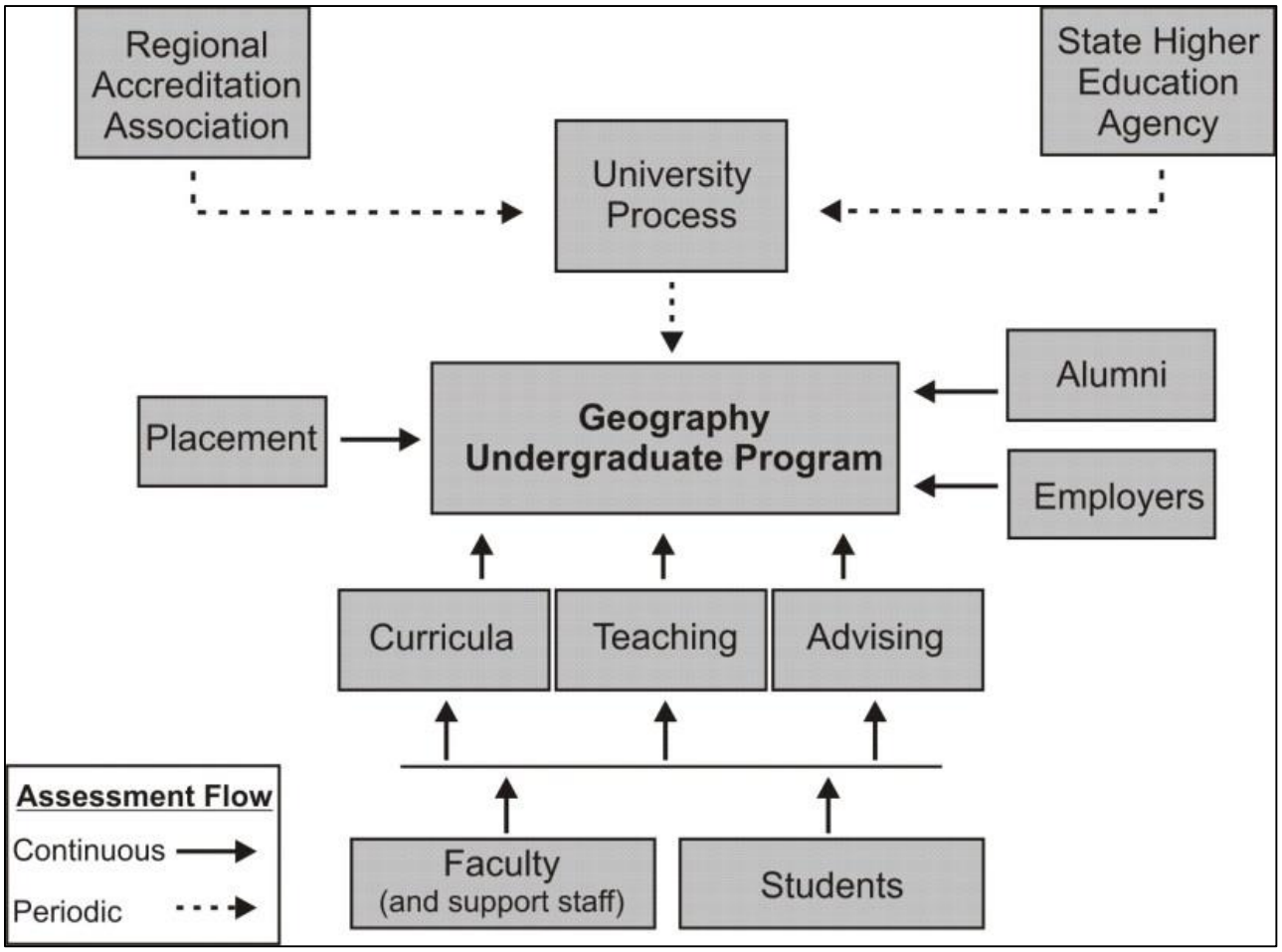

Figure 1. Integrative model of continuous and periodic assessment processes of an undergraduate geography program. Adapted from Estaville, Brown, and Caldwell (2006).

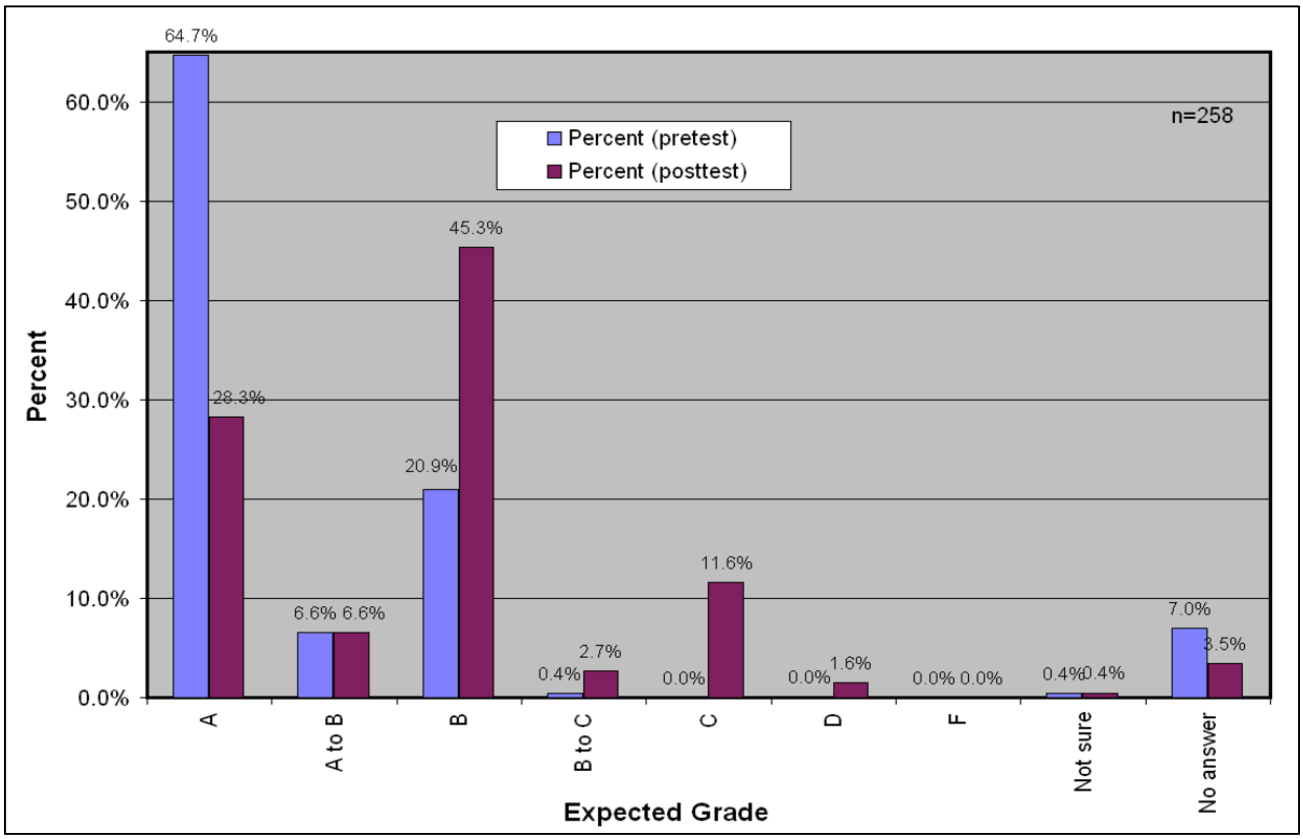

Figure 2. Students expected grade at the pre and posttests.

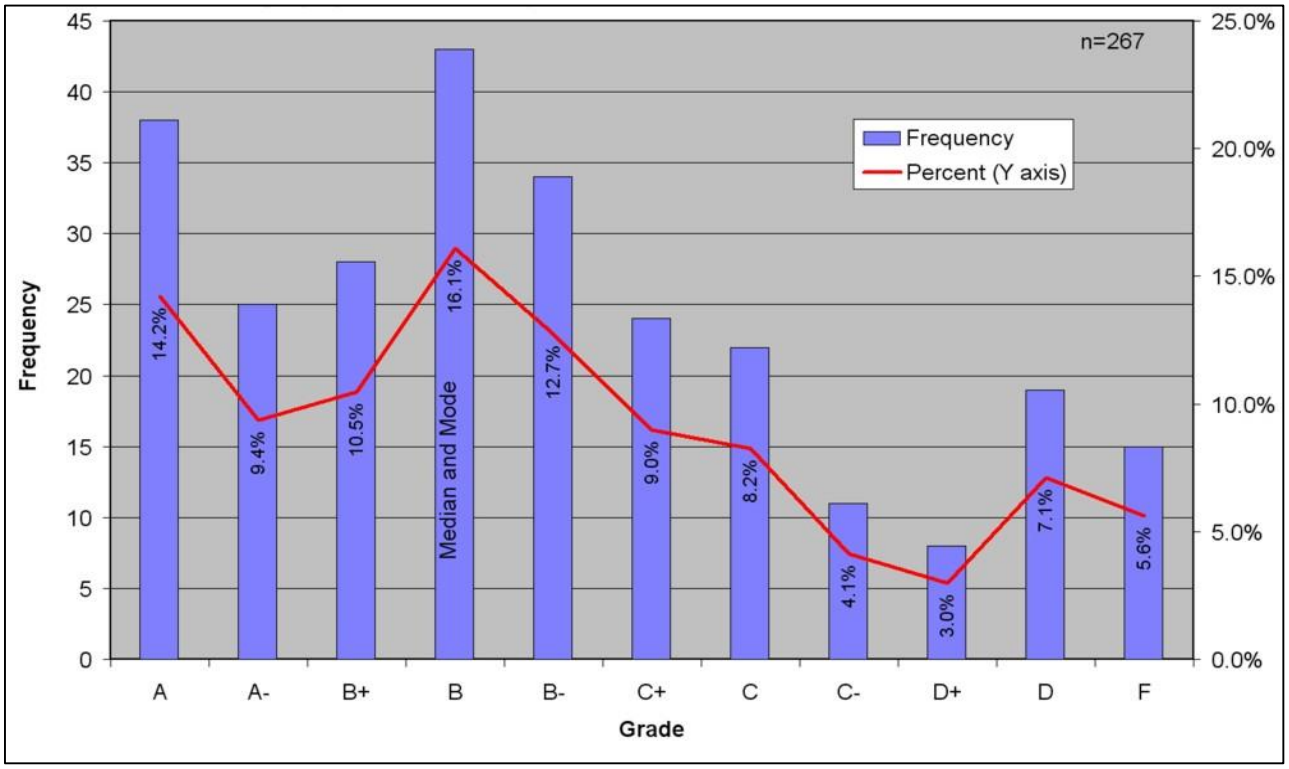

Figure 3. Students' actual grades at the end of the semester.

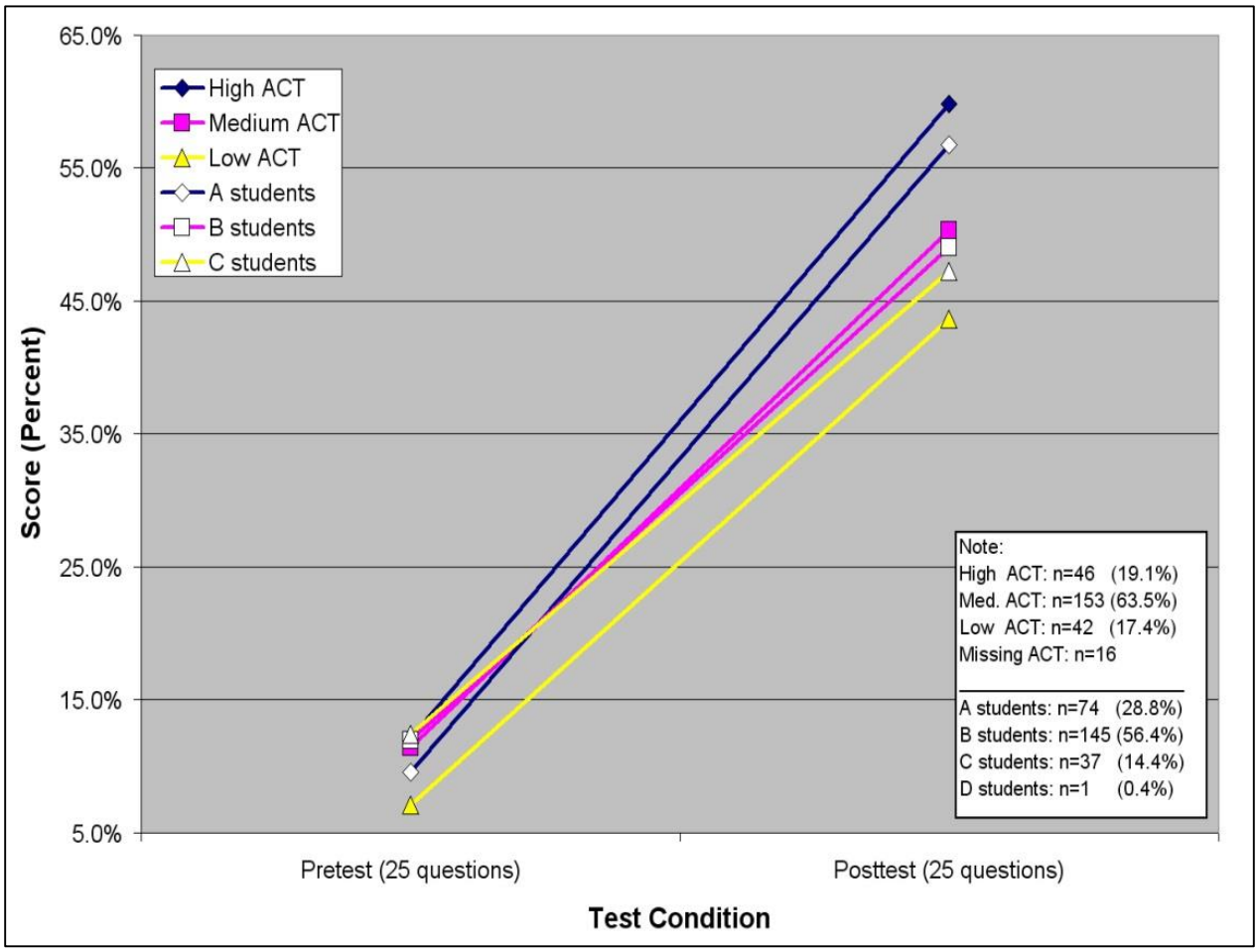

Figure 4. Change in score from pre to posttests, 25 questions, ACT and GPA groups.

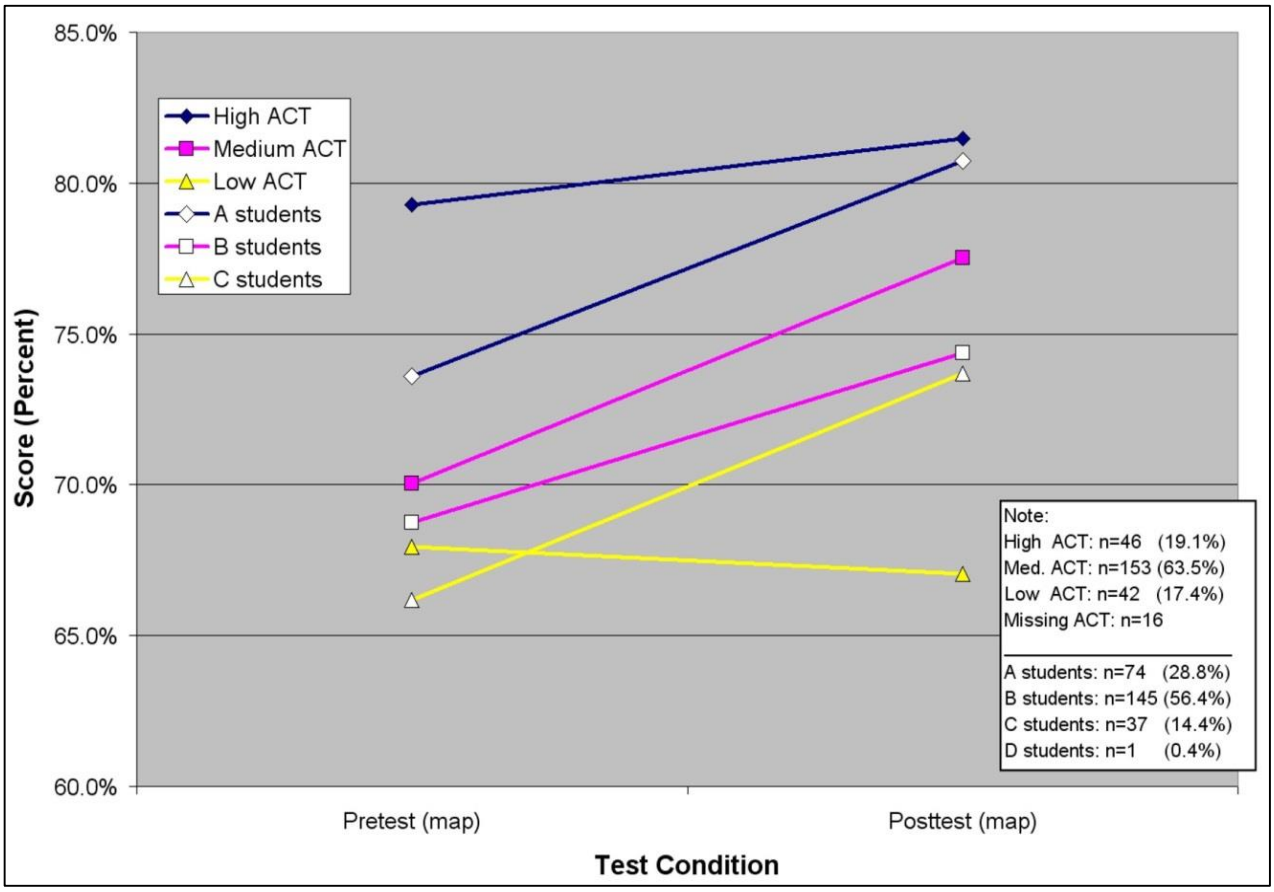

Figure 5. Change in scores from pre to posttest, map prompt, ACT and GPA groups.

Author	Date of Publication	Title	Publication details
Miller, R.	2007	Assessment in cycles of improvement: Faculty designs for essential learning outcomes.	AACU
Van Kollenburg (ed.)	2003	The future-focused organization: Focusing on the effectiveness of teaching and learning.	HLC of the NCA
Leskes and Wright	2004	*The art and science of assessment: Assessing general education outcomes, a practical guide.*	AACU
Greater Expectations Project	2004	*Taking responsibility of the quality of the baccalaureate degree.*	AACU
Suskie	2004	*Assessing student learning: A common sense guide.*	Anker
Palomba and Banta	1999	*Assessment essentials: Planning, implementing and improving assessment in higher education.*	Jossy-Bass

Table 1. Recent "how to" assessment literature.

Goal Category	Description
Content Goals: Social and Behavioral Sciences Foundations	The introduction of how knowledge in the discipline of Geography is created and applied.
	The major approaches, methods, theories, and substantive findings in the discipline of Geography.
	An informed critical stance that will allow students to weigh and apply ideas and claims from the discipline of Geography outside the classroom.
World Perspectives	To examine how culture affects people's efforts to understand, use, and survive in their environments, and how these efforts, in turn, affect culture.
	To examine within a cultural context the world views, language, or ways of life of societies, nations, regions, or peoples located outside of the United States.
Skills	To engage in articulate expression through effective writing and speaking.
	To think critically and creatively.
	To locate, evaluate, and use information effectively.
	To integrate different areas of knowledge and view ideas from multiple perspectives.

Table 2. General Education goals: Social and Behavioral Sciences Foundation, World Perspectives, and Skills Goals.

Semester	Year	Number of students	Number of sections	Class size		
				Maximum	Minimum	Average
Fall	2007	459	9	75	27	51
Winter	2008	586	10	75	47	58.6
Spring/Summer	2008	53	3	25	13	17.7
Total		1098	22			
Fall	2008	543	10	70	21	54.3
Winter	2009	582	9	70	46	64.7
Spring/Summer	2009	Data not available				

Table 3. Number of students, number of sections, and class size data for GPY235 World Regional Geography, Fall 2007 to Winter 2009.

Variable	Name	Description	Level of measurement
Independent	ACT	American College Testing score	Interval
	AGE	Age in years	Ratio
	CLASS	Freshman, Sophomore, Junior, Senior	Ordinal
	GPA	Student's Grade Point Average	Interval
	SEX	Biological sex	Nominal
	GPYHS	Number of geography courses previously taken in high school	Ratio
	GPYU	Number of geography courses taken at university	Ratio
	EXGRPRE	The grade the student expected to get on the first day of class	Ordinal
	EXGRPOST	The grade the student expected to get on the day of the final exam	Ordinal
Dependent	PRETQ	Percent correct on a 25-multiple choice question test	Ratio
	POSTQ	Percent correct on a 25-multiple choice question posttest	Ratio
	PREMAP	Percent correct on a hand drawn map of the world	Ratio
	POSTMAP	Percent correct on a hand drawn map of the world	Ratio
	GISESSAY	Average score in percent of three essays.	Ratio

Table 4. Variables used in the study.

ACT score group	Frequency	Percent
27-33	46	19.1
24-26	153	63.5
<24	42	17.4
Missing	16	
Total	241	100

Table 5. ACT score groups, frequency and percent.

GPA group	Frequency	Percent
A range (3.4-4.0)	74	28.8
B range (2.7-3.3)	145	56.4
C range (1.7-2.6)	37	14.4
D range (1.0-1.6)	1	0.4
Missing	0	0
Total	257	100

Table 6. GPA categories, frequency and percent.

Number	Cognitive Skill	Description
1	Interpretation	To understand and express the meaning of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures, or criteria.
2	Analysis	To identify the intended and actual inferential relationships among statements, questions, concepts, descriptions, or other forms of representation intended to express belief, judgment, experiences, reasons, information, or opinions.
3	Evaluation	To assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief, or opinion; and to assess the logical strength of the actual or intended inferential relationships among statements, descriptions, questions or other forms of representation
4	Inference	To identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation
5	Explanation	To be able to present in a cogent and coherent way the results of one's reasoning. This means to be able to give someone a full look at the big picture: both "to state and to justify that reasoning in terms of the evidential, conceptual, methodological, criteriological, and contextual considerations upon which one's results were based; and to present one's reasoning in the form of cogent arguments.
6	Self-regulation	To self-consciously monitor one's cognitive activities, the elements used in those activities, and the results educed, particularly by applying skills in analysis, and evaluation to one's own inferential judgments with a view toward questioning, confirming, validating, or correcting either one's reasoning or one's results.

Table 7. Cognitive skills that are at the heart of critical thinking. Source: Facione (2007).

Area	Sub area	Points
1. Continents	Names, completeness	7
	Names, correct spelling	7
	Names, placement accuracy	7
	Continents, placement accuracy	7
	Continents, shape accuracy	7
2. Oceans	Names, completeness	3
	Names, correct spelling	3
	Names, placement accuracy	3
3. Equator	Placement accuracy	5
Total		49

Table 8. Areas of evaluation of student maps.

Class standing	Sex				Total
	Male		Female		
	Frequency	Percent	Frequency	Percent	
Freshman	51	41.80	50	37.04	101
Sophomore	47	38.52	50	37.04	97
Junior	17	13.93	21	15.56	38
Senior	7	5.74	14	10.37	21
Total	122	100.00	135	100.00	257

Table. 9. Student class standing by sex.

Variable	N	Minimum	Maximum	Mean	Standard Deviation
AGE	256	17.00	41.00	19.91	2.70
ACT	241	17.00	33.00	23.48	3.28
GPA	257	1.06	4.80	3.13	0.53
GPYHS	246	0.00	6.00	0.96	0.89
GPYU	246	0.00	5.00	0.15	0.50
EXGRPRE	238	3.00	4.00	3.76	0.43
PRETQ	121	0.00	0.44	0.11	0.11
POSTQ	257	0.20	0.84	0.51	0.13
Difference PRETQ and POSTQ	121	-0.04	0.76	0.41	0.17
PREMAP	112	0.02	0.96	0.70	0.15
POSTMAP	257	0.00	1.00	0.76	0.16
Difference PREMAP, POSTMAP	111	-0.16	0.48	0.07	0.12
GISESSAY	257	0.00	0.99	0.71	0.15
ATTENDPCT	257	10.0	100.00	85.15	16.98

Table 10. Descriptive statistics for the entire group of students.

Dependent variable	Student's sex	N	Mean	Standard Deviation
Grade Point Average	Male	122	3.07	0.55
	Female	135	3.18	0.52
Pretest score (25 questions) (pct)	Male	55	0.12	0.12
	Female	66	0.11	0.11
Posttest score (25 questions) (pct)	Male	122	0.52	0.12
	Female	135	0.50	0.13
Difference between pre and post tests (questions) (pct)	Male	55	0.41	0.16
	Female	66	0.41	0.17
Pretest score (map)	Male	58	0.73	0.14
	Female	54	0.67	0.16
Posttest score (map)	Male	122	0.78	0.17
	Female	135	0.75	0.14
Difference between pre and post tests (map)	Male	58	0.07	0.11
	Female	53	0.07	0.12
GIS essays average score (pct)	Male	122	0.74	0.13
	Female	135	0.69	0.16
Attendance (pct)	Male	122	0.85	0.17
	Female	135	0.86	0.17

Table 11. Descriptive statistics by sex.

Class standing	Statistic	PRETQ	POSTQ	Pre-Post difference	PREMAP	POSTMAP	Pre-Post difference	GIS Essays
Fresh.	N	49	101	49	40	101	40	101
	Mean	0.11	0.51	0.42	0.73	0.74	0.07	0.70
	sd	0.09	0.13	0.16	0.13	0.19	0.12	0.15
Soph.	N	41	97	41	47	97	46	97
	Mean	0.11	0.51	0.41	0.69	0.78	0.07	0.73
	sd	0.12	0.13	0.18	0.17	0.13	0.11	0.15
Junior	N	18	38	18	17	38	17	38
	Mean	0.11	0.51	0.41	0.71	0.81	0.07	0.72
	sd	0.13	0.12	0.17	0.11	0.10	0.12	0.17
Senior	N	13	21	13	8	21	8	21
	Mean	0.14	0.49	0.37	0.60	0.70	0.08	0.68
	sd	0.13	0.13	0.15	0.23	0.18	0.15	0.13
Total	N	121	257	121	112	257	111	257
	Mean	0.11	0.51	0.41	0.70	0.76	0.07	0.71
	sd	0.11	0.13	0.17	0.15	0.16	0.12	0.15

Table 12. Pretests and posttests by class standing.

Dependent Variable	ACT Group (I)	ACT Group (J)	Mean Difference (I-J)	Std. Error	Sig.
PRETQ	27 to 33	24 to 26	0.007	0.025	1.000
		Less than 24	0.050	0.033	0.386
	24 to 26	27 to 33	-0.007	0.025	1.000
		Lessthan24	0.044	0.029	0.397
	Less than 24	27 to 33	-0.050	0.033	0.386
		24 to 26	-0.044	0.029	0.397
POSTQ	27 to 33	24 to 26	.095(*)	0.020	0.000
		Less than 24	.162(*)	0.025	0.000
	24 to 26	27 to 33	-.095(*)	0.020	0.000
		Less than 24	.067(*)	0.021	0.004
	Less than 24	27 to 33	-.162(*)	0.025	0.000
		24 to 26	-.067(*)	0.021	0.004
ABSQDIFF	27 to 33	24 to 26	.090(*)	0.035	0.037
		Less than 24	.128(*)	0.047	0.022
	24 to 26	27 to 33	-.090(*)	0.035	0.037
		Less than 24	0.038	0.041	1.000
	Less than 24	27 to 33	-.128(*)	0.047	0.022
		24 to 26	-0.038	0.041	1.000
PREMAP	27 to 33	24 to 26	.092(*)	0.035	0.032
		Less than 24	.113(*)	0.043	0.031
	24 to 26	27 to 33	-.092(*)	0.035	0.032
		Less than 24	0.021	0.033	1.000
	Less than 24	27 to 33	-.113(*)	0.043	0.031
		24 to 26	-0.021	0.033	1.000
POSTMAP	27 to 33	24 to 26	0.039	0.024	0.288
		Less than 24	.144(*)	0.030	0.000
	24 to 26	27 to 33	-0.039	0.024	0.288
		Less than 24	.105(*)	0.024	0.000
	Less than 24	27 to 33	-.144(*)	0.030	0.000
		24 to 26	-.105(*)	0.024	0.000
ABSMapDIFF	27 to 33	24 to 26	-0.049	0.032	0.392
		Less than 24	0.009	0.039	1.000
	24 to 26	27 to 33	0.049	0.032	0.392
		Less than 24	0.058	0.029	0.154
	Less than 24	27 to 33	-0.009	0.039	1.000
		24 to 26	-0.058	0.029	0.154
GISESSAYS	27 to 33	24 to 26	.069(*)	0.025	0.020
		Less than 24	.084(*)	0.032	0.028
	24 to 26	27 to 33	-.069(*)	0.025	0.020
		Less than 24	0.015	0.026	1.000
	Less than 24	27 to 33	-.084(*)	0.032	0.028
		24 to 26	-0.015	0.026	1.000
AttendPCT	27 to 33	24 to 26	0.187	2.886	1.000
		Less than 24	5.278	3.663	0.453
	24 to 26	27 to 33	-0.187	2.886	1.000
		Less than 24	5.091	2.990	0.270
	Less than 24	27 to 33	-5.278	3.663	0.453
		24 to 26	-5.091	2.990	0.270

Table 13. Post hoc (Bonferroni) results of ANOVA with ACT group as the independent variable.

* The mean difference is significant at the .05 level.

Dependent Variable	GPA classes (I)	GPA classes (J)	Mean Difference (I-J)	Std. Error	Sig.
PRETQ	C	B	0.005	0.030	1.000
		A	0.028	0.032	1.000
	B	C	-0.005	0.030	1.000
		A	0.023	0.023	0.948
	A	C	-0.028	0.032	1.000
		B	-0.023	0.023	0.948
POSTQ	C	B	-0.018	0.023	1.000
		A	-.095(*)	0.025	0.000
	B	C	0.018	0.023	1.000
		A	-.077(*)	0.017	0.000
	A	C	.095(*)	0.025	0.000
		B	.077(*)	0.017	0.000
ABSQDIFF	C	B	-0.059	0.041	0.471
		A	-.173(*)	0.045	0.001
	B	C	0.059	0.041	0.471
		A	-.114(*)	0.032	0.002
	A	C	.173(*)	0.045	0.001
		B	.114(*)	0.032	0.002
PREMAP	C	B	-0.026	0.046	1.000
		A	-0.074	0.050	0.428
	B	C	0.026	0.046	1.000
		A	-0.049	0.033	0.427
	A	C	0.074	0.050	0.428
		B	0.049	0.033	0.427
POSTMAP	C	B	-0.007	0.029	1.000
		A	-0.070	0.032	0.081
	B	C	0.007	0.029	1.000
		A	-.064(*)	0.022	0.015
	A	C	0.070	0.032	0.081
		B	.064(*)	0.022	0.015
ABSMapDIFF	C	B	0.017	0.036	1.000
		A	0.025	0.039	1.000
	B	C	-0.017	0.036	1.000
		A	0.008	0.026	1.000
	A	C	-0.025	0.039	1.000
		B	-0.008	0.026	1.000
GISESSAYS	C	B	-.0629(*)	0.026	0.045
		A	-.172(*)	0.028	0.000
	B	C	.0629(*)	0.026	0.045
		A	-.109(*)	0.020	0.000
	A	C	.172(*)	0.028	0.000
		B	.109(*)	0.020	0.000
AttendPCT	C	B	-7.762(*)	3.012	0.032
		A	-15.285(*)	3.293	0.000
	B	C	7.762(*)	3.012	0.032
		A	-7.523(*)	2.336	0.004
	A	C	15.285(*)	3.293	0.000
		B	7.523(*)	2.336	0.004

Table 14. Post hoc (Bonferroni) results of ANOVA with GPA group as the independent variable.

* The mean difference is significant at the .05 level.