Inquiry-Based and Didactic Instruction in a Computer-Assisted Context

John C. Mayer[#], Rachel D. Cochran*, Jason S. Fulmore*, Thomas O. Ingram*, Laura R. Stansell[#], and William O. Bond[#]

University of Alabama at Birmingham (UAB)
[#]Department of Mathematics
*Center for Educational Accountability (CEA)

Abstract

We compare the effect of incorporating inquiry-based sessions versus traditional lecture sessions, and a blend of the two approaches, in an elementary algebra course in which the pedagogy consistent among treatments is computer-assisted instruction. Our research hypothesis is that inquiry-based sessions benefit students significantly in terms of mathematical content knowledge, problem-solving, and communications. All students receive the same computer-assisted instruction component. Students are randomly assigned for the semester to one of three treatments (two inquiry-based meetings, two lecture meeting, or one of each, weekly). Measures, including pre- and post-tests with both open-ended and objective items, are described. Statistically significant differences have previously been observed in similar quasi-experimental studies of multiple sections of finite mathematics (Fall, 2008) and elementary algebra (Fall, 2009) with two treatments. Undergraduates, including many pre-service elementary teachers, who do not place into a credit-bearing mathematics course take this developmental algebra course.

Keywords. Elementary algebra, teaching experiment, computer-assisted instruction, inquiry-based instruction, didactic instruction.

The Greater Birmingham Mathematics Partnership is funded by NSF awards DUE-0632522 and DUE-0928665. The opinions expressed herein are those of the authors, and not necessarily those of the National Science Foundation.

Preliminary Research Report

One direction taken by course reform over the past few years has been the development of sophisticated computer-assisted instruction. This approach has been applied to large-enrollment service courses in mathematics, including algebra. Elementary algebra is typically taken by under-graduate students who do not place into a credit-bearing course. Traditionally, the goal of such a developmental algebra course has been to enhance students' "algebra skills," for example, dealing procedurally with rational numbers and expressions. Higher-order thinking may be largely absent. Alternately, one might focus on developing quantitative reasoning and communications skills, rather than, or in addition to, training to acquire a set of specific algebraic skills (Wiggins, 1989; Blais, 1988). Our position is that incorporating an inquiry-based component, either together with, or in place of, a didactic component, into a computer-assisted instructional environment may enhance student learning. Two previous studies in the literature bear this out (Mayer, 2009, 2010).

Fundamental Question. We compare three treatments in a quasi-experimental design: (GG) two weekly inquiry-based class meetings, (LL) two weekly lecture meetings, and (GL) one of each meeting weekly. The computer-assisted component is the same for all treatments. Our main hypothesis is that, of the three treatments, the one affording the most inquiry-based involvement to the students will differentially benefit the students in terms of mathematical content knowledge, reasoning and problem-solving ability, and communications. Secondarily, we expected no difference in student course grades among treatments. Typically, about 70% of students in this course earn an A, B, or C. Our hypothesis was supported in the areas of reasoning and problem-solving ability, communications, and course grades, but not supported in the area of mathematical content knowledge identified on the objective component of our pre/post-test.

Prior Research. Prior to the two most recent studies (Mayer, 2009, 2010), the methodology of simultaneously comparing different pedagogies within one semester, had few direct comparisons in the literature (Doorn, 2007). Some studies have compared different pedagogies over a longer time frame (Gautreau, 1997; Hoellwarth, 2005). The results of the quasi-experimental studies in (Mayer, 2009) of a finite mathematic course, and in (Mayer, 2010) of an elementary algebra course showed in both cases that students in the inquiry-based treatment did significantly better ($p<0.05$) comparing pre-test and post-test performance in the areas of problem identification, problem-solving, and explanation. Moreover, students, regardless of treatment, performed similarly (no statistically significant differences) when compared on the basis of course test scores. Outcomes of the two studies differed in gain in accuracy, pre- to post-test: in the finite mathematics study, there was no significant difference between treatments; in the elementary algebra study there was a significant difference between treatments in favor of the inquiry-based treatment. A limitation of both studies by Mayer was that accuracy was assessed on a small set of open-ended problems. The previous studies also did not test a blend of inquiry-based and traditional class meetings in a single treatment (Marrongelle, 2008).

Research Methodology. Our methodology is quasi-experimental in that it seeks to remove from consideration as many confounding factors as possible, and to assign treatment on as random a basis as possible, constrained only by students being able to choose the time slot in which they take the course. All students involved in the courses received identical computer-assisted instruction provided in a mathematics learning laboratory. 86% of the grade in the course was determined by evaluation in the computer-assisted context (online homework and supervised online quizzes and tests). The remaining 14% of the grade was determined by one of three pedagogical treatments, described below. Students registered for one of three time periods in the Fall 2010 semester schedule, a 9:00 AM, 10:00 AM or noon time slot, for three days a week (MWF), for their 50 minute class meetings and 50 minute required lab meeting. Students in each time slot were randomly assigned to one of the three treatments for the semester. Three instructors agreed to participate in the experiment. Each instructor taught in three time slots. In one slot the instructor administered the twice-weekly inquiry-based treatment, in another time slot, the twice weekly lecture treatment, and in a third time slot, the blended treatment. The three instructors consisted of a full professor, a regular full-time instructor, and a graduate student with prior teaching experience. All instructors had previous experience in both didactic and inquiry-based teaching, and in computer-assisted instruction. A graduate teaching assistant worked with each instructor in the inquiry-based meetings, and in evaluating written student work product from such meetings. Each instructor also met with each class in the mathematics computer lab. The computer lab meeting for all treatments occurred on Wednesday, and the class meetings on Monday and Friday. In the GL treatment, the lecture meeting was on Friday.

The three pedagogies to be compared are:

GG: two sessions weekly of inquiry-based group work (random, weekly changing, groups of four) without prior instruction, on problems intended to motivate the topics to be covered in computer-assisted instruction;

LL: two sessions weekly of traditional summary lecture with teacher-presented examples on the topics to be covered in computer-assisted instruction, and

GL: a blend of treatments GG and LL, with one weekly meeting traditional lecture, and one weekly meeting inquiry-based group work.

In the inquiry-based treatments, each student turned in each class meeting a written report on his/her investigation and solution of the problem(s) posed in that class period. This report is evaluated based upon the same rubric as the open-ended items on the pre/post-test (see Appendix 1 for a copy of the rubric). Students were aware of the rubric and received written feedback consistent with the rubric. In the lecture treatment, the instructor gave a traditional lecture on the upcoming material. All instructors operated from the same outline of topics for each lecture. The 14% (140 of 1000 points) of the final grade determined by the class meetings differed among the three treatments as follows:

GG: 5 points are earned for each of the two weekly reports on the group work;

LL: 5 points are earned for attendance at each class meeting;

GL: 5 points are earned for the one weekly report on the group work meeting, and 5 points are earned for attendance at the lecture meeting.

This research was carried out in Fall, 2010. Data gathered included (1) course grades and test scores, (2) pre-test and post-test of content knowledge based upon a test which incorporates three open-ended problems, evaluated on rubric dimensions of conceptual understanding, evidence of problem-solving, and adequacy of explanation (3) pre-test and post-test of content knowledge based upon a test consisting of 25 objective questions, and (4) student course evaluations using the online IDEA system (IDEA, 2010), and (5) RTOP observations of the instructors in each of the nine class sections (RTOP, 2010; Sawada, 2002).

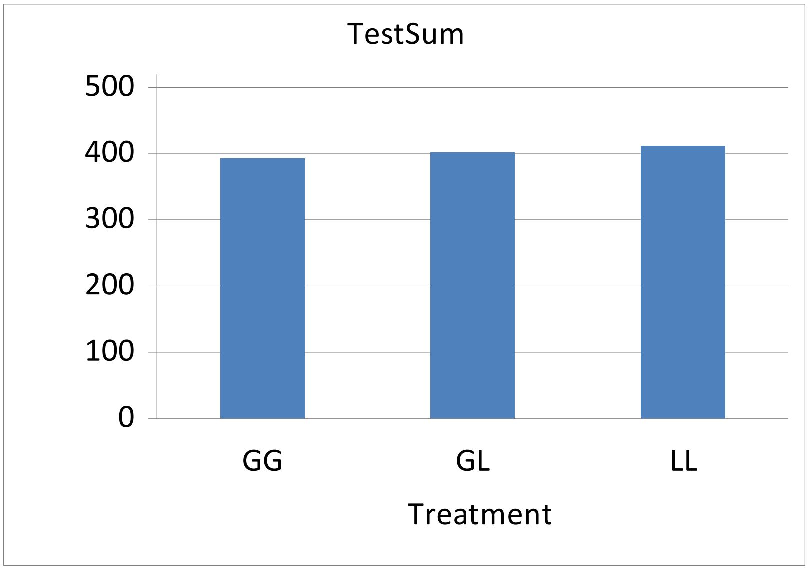
A limitation of the studies by Mayer (2009, 2010) is that the pre/post-test consisted of only three or four open-ended problems which made a reliable evaluation of accuracy gains, if any, problematic. The pre/post-test in the study described herein consists of two parts:

Part 1: three open-ended problems, evaluated by a rubric as described above, and

Part 2: 25 objective questions which had been validated for testing algebraic content knowledge in previous studies.

A battery of the previously validated (for content) objective questions was piloted in Summer 2010 on students in the same course, and item analysis was used to select the items for the pre/post-test in this study. As a result of the more careful test design, we expected that differential gains in accuracy between treatments, if present, would be more detectable than in the two earlier studies cited.

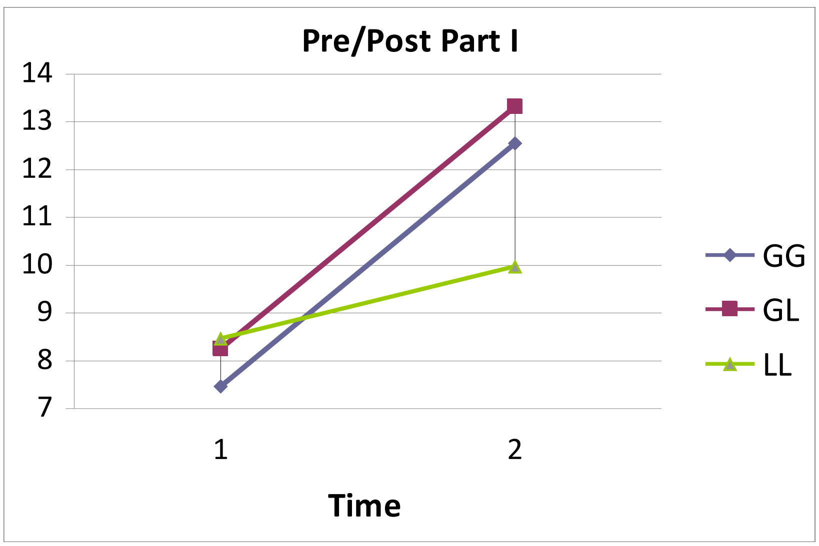
Results. Students in all three treatments take the same five tests during the term. The tests are administered in the mathematics learning lab and are graded by computer. Tests are short answer rather than multiple choice. Each test is worth 130 points. Since the tests are the most significant determiner of student grades in the course, we used the sum of the first four of the five tests as our measure of the impact of the treatments on student grades. The maximum possible score is thus 520. The following graph shows the average test sum by treatment.



(N=315; GG=100; GL=106; LL=109.)

There was no significant difference ($p<0.05$) in the sum of the four test grades between any pair of treatments. Nor were there significant differences on any single test. The first four tests were used because some students who are satisfied with their accumulated course grade prior to test 5 elect not to put forth much effort on the fifth test. (A test sum of 400 points on the first four tests is well on the way to a "B" in the course, other grade factors being typical.)

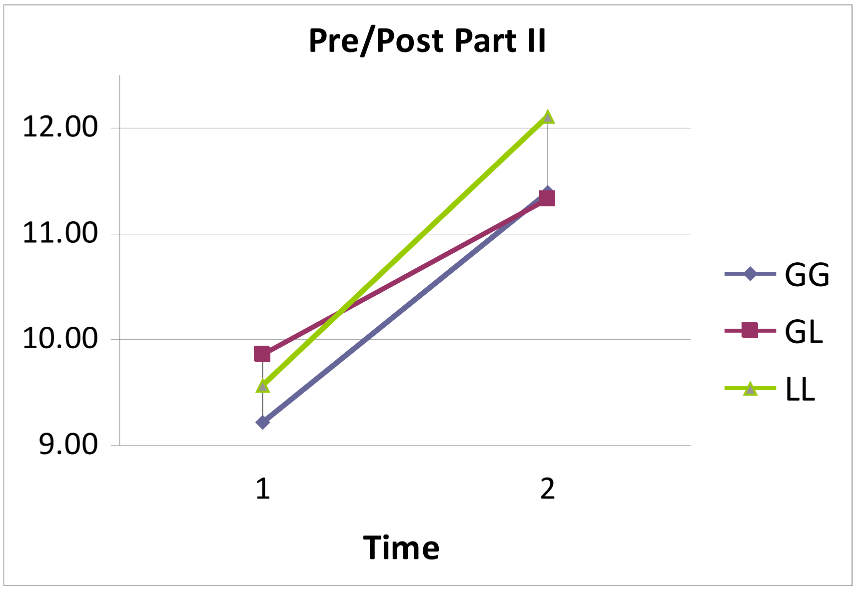
Part 1 of the pre/post-test consisted of three open-ended problems graded by a four-part rubric. On each part of the rubric (see Appendix 1 for a copy of the rubric), a student could score up to two points, for a total of 8 points per question, and a total of 24 points on Part 1 of the test. The graph below shows the results by treatment on Part 1 of the pre/post-test.



(N=272; GG=85; G =93; LL=94.)

One way ANOVA on the pre-test scores showed that there was no significant difference ($p<0.05$) between any pair of treatments. The GG and GL treatments differed significantly from the LL treatment. Repeated measures ANOVA showed that both the Time effect and the Time*Treatment interaction effect were statistically significant ($p<0.05$). Wilkes Lambda for the Time effect was $\lambda=0.690$ and for the Time*Treatment interaction effect was $\lambda=0.921$. One way ANOVA on the post-test scores confirmed this finding: there was no significant difference between the GG and GL treatments, but there were significant differences between the GG and LL and between the GL and LL treatments.

Part 2 of the pre/post-test consisted of 25 objective questions, some yes/no, some multiple choice, and some always/sometimes/never. The expected value of the test (answering at random) was 10.38. The following graph shows the results by treatment on Part 2 of the pre/post-test.



(N=273; GG =88; GL =91; LL =94.)

One way ANOVA on the pre-test scores showed that there was no significant difference ($p<0.05$) between any pair of treatments. The Time effect was statistically significant ($p<0.05$), and Wilkes lambda for the Time effect was $\lambda=0.759$. There was no significant Time*Treatment interaction effect. One way ANOVA on the post-test scores confirmed this finding: there was no significant difference between any pair of treatments.

We also computed the effect size (difference of means divided by the pre-test standard deviation) for each of the treatments. The effect sizes ranged from medium (>0.40) to large (>0.80). However, in the absence of any statistically significant treatment effect, it is difficult to know how to interpret the results concerning effect size.

	Mean	*Mean*	*Standard Deviation*	*Standard Deviation*	*Effect Size*
Treatment	Pre	Post	Pre	Post	
GG	9.22	11.39	3.02	2.98	0.72
GL	9.86	11.33	3.44	3.38	0.43
LL	9.57	12.11	3.00	3.32	0.84

The IDEA student evaluations produce both numerical data that can be compared to the IDEA database of courses nationwide and within the discipline (mathematics), as well as written comments from students. Most students who complete the survey do not write comments. The course instructors and two additional readers were given student comments identified only by a code letter blindly corresponding to treatment and asked to identify "themes" in the comments. Following are the themes identified by the two additional readers (which are consistent with those identified by the instructors).

Reader 1: (General comment: I didn't count those who just said "good teacher" among the like. I was more looking for comments on the format.)

GG: Never taught/no help very frequent (13); other negative (unspecific) (7); only two positive.

GL: Like/learned a lot very frequent (14); some think online not helpful (3).

LL: Helped/understood better (6); lab good/class bad (4); tests should be closer to when material is learned (2).

Reader 2: (General comment: These are my observations concerning the major themes in the three sets of student comments.)

GG: The lack of lectures was harmful (the teacher did not teach).

GL: The instructor was the key to learning and made a great difference in the value received.

LL: The math learning lab is very helpful and was a key element in learning.

Student comments on the GG treatment appeared to be overwhelmingly negative in the view of all the readers, including the three course instructors. The comments on the GL treatment were more "middle of the road" and the comments on the LL treatment were mixed, but generally positive.

The table below represents the IDEA survey overall ratings of student evaluation of instruction. The Raw Average in each category is on a 5-point Likert scale, where 5 represents "strongly agree" and 1 "strongly disagree."

	Raw Average		Converted
GG	Excellent Teacher	Excellent Course	Score
Instructor 1	2.4	2.2	25
Instructor 2	4.2	4.0	52
Instructor 3	2.5	2.6	30

	Raw Average		Converted
GL	Excellent Teacher	Excellent Course	Score
Instructor 1	3.7	3.6	45
Instructor 2	4.3	4.0	52
Instructor 3	4.8	4.1	56

	Raw Average		Converted
LL	Excellent Teacher	Excellent Course	Score
Instructor 1	4.3	3.9	51
Instructor 2	4.3	4.0	50
Instructor 3	4.3	4.0	49

The form in which we receive the IDEA ratings does not permit aggregating student responses by treatment. The Converted Score (an adjusted average of the two raw scores) allows one compare student evaluations of instruction to the IDEA database of all courses: scores in the 45-55 range place the course/instructor evaluation in the middle 40% of all student ratings. Scores 37 or lower are in the lowest 10% of all student ratings.

Conclusions. We summarize our conclusions based upon the above results as follows.

- The inclusion of group work class meetings in lieu of lecture does not appear to affect adversely student success as measured by grades.
- Inquiry-based group work does have a positive effect on problem-solving and communications abilities as measured by the rubric score (Part 1) of the pre/post-test.
- Inquiry-based group work does not appear to affect accuracy as measured by the objective part (Part 2) of the pre/post-test.
- Two group work sessions do not appear to be significantly better than one per week, as indicated by the statistical indistinguishabilty of the GG and GL treatments.
- Student satisfaction is moderate with both the LL and GL treatments, but students are unsatisfied with the GG treatment.

Implications. This research will inform our teaching of elementary algebra. We do not know if the results of this study are generalizable to other mathematics courses with differing content. However, we expect to extend this study in subsequent years to credit courses such as intermediate algebra, college algebra, and pre-calculus algebra and trigonometry (Oerhtman, 2008). The aforementioned courses all are computer-assisted. This further study will allow us to examine the cumulative effect on students of multiple classes incorporating inquiry-based sessions in a computer-assisted context.

All treatments had statistically indistinguishable test scores and course final grades. Any valued-added from including inquiry-based sessions is not being captured in course grades. This is not surprising given that 86% of the course grade is determined by the computer-assisted component. The rubric-scored Part 1 of the pre/post-test appears to capture at least part of the value-added: students are better able to express their thinking about mathematics problems. However, it is depressing that the objective Part 2 of the pre/post-test does not show any difference among treatments. It is possible that cumulative exposure to learning constructively will have a detectable objective effect. We are also continuing to study the pre/post-test itself, the beginnings of which are presented in Appendix 2.

Many of our students in elementary algebra did not appear to find value in the inquiry-based components. But they did, as witnessed by the reactions to the GL treatment, tolerate inquiry-based sessions when blended with traditional instruction. As instructors, we value the deeper thinking that can arise in learning constructively. We provide an illustrative story of this in Appendix 3. While we do not discount student satisfaction, as teachers, we ask ourselves if pedagogical decisions ought to be based on whether or not students perceive (in the short term) any value in the inquiry-based components of the learning process? If blended treatments are more accepted by students then this perhaps should be our direction in the short term.

References

Blais, D. M. (1988), Constructivism -- a Theoretical Revolution for Algebra. Mathematics Teacher, Nov. 1988: 49--56.

Doorn, D. and O'Brien, M. (2007), Assessing the Gains from Concept Mapping in Introductory Statistics. International Journal for the Scholarship of Teaching and Learning 1(2); online at \url{http://academics.georgiasouthern.edu/ijsotl/v1n2/articles/doorn/index.htm}.

Gautreau, R. and Novemsky, L. (1997), Concepts First-a Small Group Approach to Physics Learning. Am. J. Phys. 65(5): 418-429.

Hoellwarth, C., Moelter, M. J., and Knight, R.D. (2005), A direct comparison of conceptual learning and problem solving ability in traditional and studio style classrooms. Am. J. Phys. 73(5): 459-463.

IDEA (2010), The IDEA Center. \url{http://www.theideacenter.org/our-services/student-ratings-instruction/0030-idea-online}.

Marrongelle, K. and Rasmussen, C. (2008), Meeting New Teaching Challenges: Teaching Strategies that Mediate Between All Lecture and All Student Discovery. In M. Carlson and C. Rasmussen (Eds.), Making the Connection: Research to Practice in Undergraduate Mathematics, pp. 167-178. Mathematical Association of America, Washington, DC.

Mayer, J.C., et al. (2009), Incorporating Inquiry-Based Class Sessions with Computer Assisted Instruction. Electronic Proceedings of 2009 SIGMAA Conference on Research in Undergraduate Mathematics Education, Raleigh, NC, February, 2009.

Mayer, J.C., et al. (2010), Blending Inquiry-Based and Computer Assisted Instruction in a Basic Algebra Course: a Quasi-Experimental Study, Electronic Proceedings of Conference on Research in Undergraduate Mathematics Education, Raleigh, NC, February, 2010.

Oerhtman, M., Carlson, M., and Thompson, P. (2008), Foundational Reasoning Abilities that Promote Coherence in Students' Function Understanding. In M. Carlson and C. Rasmussen (Eds.), Making the Connection: Research to Practice in Undergraduate Mathematics, pp. 65-80. Mathematical Association of America, Washington, DC.

Pyrczak, F. (1973). Validity of the discrimination index. Journal of Educational Measurement 10(3): 227-231.

RTOP (2010), Reformed Teaching Observation Protocol. Developed by ACEPT Project (Arizona State University). \url{http://physicsed.buffalostate.edu/AZTEC/RTOP/RTOP_full/}

Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R. & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. School Science and Mathematics, 102(6), 245-253.

Wiggins, G. (1989), The Futility of Trying to Teach Everything of Importance. Educational Leadership, Nov. 1989: 71-78.

Appendix 1

MA 098/110 Scoring Guide

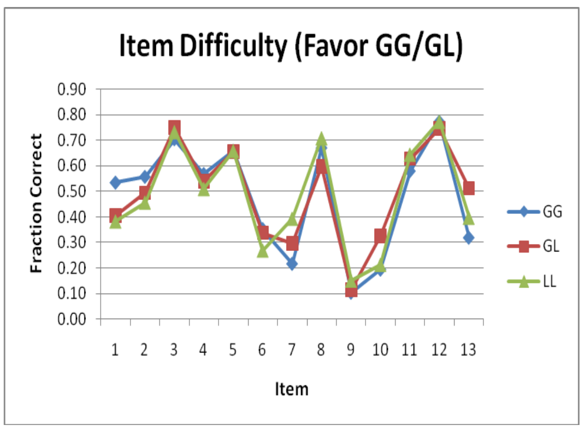
	Conceptual Understanding: *Interpreting the concepts of the task and translating them into mathematics*	**Evidence Of Problem Solving:** *Choosing strategies that can work, and then carrying out the strategies chosen.*	**Explanation:** *Using pictures, symbols, and/or vocabulary to convey the path to the identified solution*	**Accuracy:** *Providing a complete and accurate solution appropriate for the given problem*
2	The translation of the task into adequate mathematical concepts using relevant information is completed.	Pictures, models, diagrams, symbols, and/or words used to solve the task are complete.	Explanation is clear and complete.	Solution is correct and complete with no errors.
1	The translation of the major concepts of the task is partially completed and/or partially displayed.	Pictures, models, diagrams, symbols, and/or words used to solve the task may be only partially useful and/or partially recorded.	The explanation is partially complete and/or partially developed with gaps that have to be inferred.	Solution is appropriate and demonstrates understanding, but is either not quite complete or contains minor errors.
0	Does not achieve minimal requirements for 1 point.	Does not achieve minimal requirements for 1 point.	Does not achieve minimal requirements for 1 point.	Does not achieve minimal requirements for 1 point.

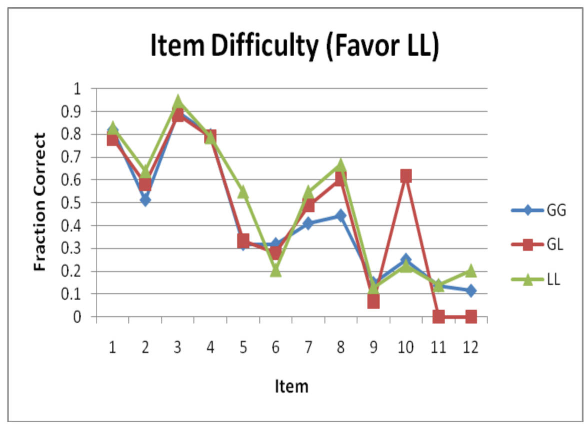
Loosely adapted from the Oregon Department of Education’s 1995-2003 statewide assessments (originally a 5-point rubric).

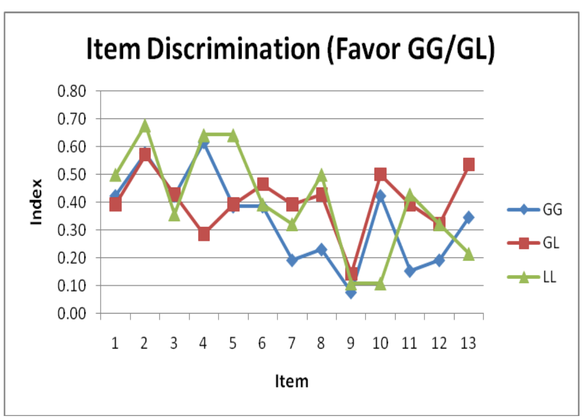
Appendix 2. Preliminary Analysis of Part 2 of Pre/Post-Test.

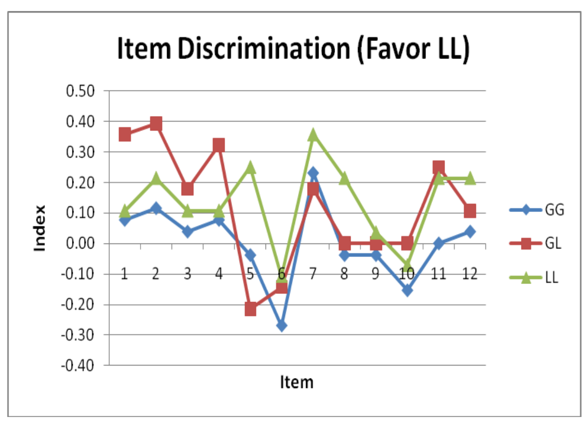
Because our expectation of improved performance on the objective part (Part 2) of the pre/post-test by those receiving more inquiry-based instruction was not realized, we have begun a more detailed analysis of Part 2. Here we present some preliminary results.

To begin, Instructors/TAs were asked to rate on a scale of 1 (lowest) to 5 (highest) the extent to which students received instruction that would help them in responding to each pre/post-test question, with each treatment being considered separately. Based upon the responses, we tentatively identified 13 questions on Part 2 as possibly favorable to students in the GG or GL treatments and 12 as indifferent or possibly favoring the LL treatment. We then performed item difficulty analysis and a simple item discrimination analysis (Pyrczak, 1973) on each so-identified group of questions separately, with respect to each treatment. The following graphs illustrate the results.







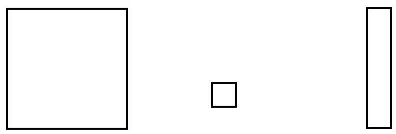


No clear pattern emerges and further study is required.

Appendix 3: An Example of a Group Learning Situation

The group learning experience is vastly different than the traditional class lecture experience. The following example of a group learning situation in one of the group work sessions illustrates this difference.

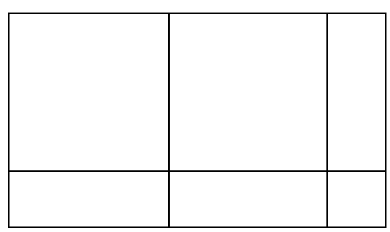
Students (working in groups of four) were given a set of algebra tiles at the beginning of a group work session. There are three different-sized pieces in the algebra tile sets. There are large square pieces, small square pieces, and (non-square) rectangular pieces. The rectangular piece has width equal to the smaller square's width and length equal to the larger square's width.



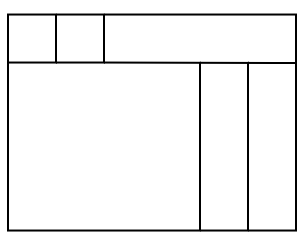
Students were given a warm-up problem using a set of these tiles. Students were asked to view the three tile pieces as a 10 x 10 square, a 1 x 10 square, and a 1x 1 square. Then students were asked to show a rectangular area model of 264 using the tiles.

After the warm-up problem, students were asked to construct the following figures: an x-by-x square, a 1-by-x square, and a 1-by-1 square and label the area of each of these figures. Next, students were directed to represent a trinomial such as $2x^2 + 3x + 1$ as a rectangular area model using the tiles. An overwhelming majority of students use the larger square from the algebra tiles as the x-by-x square and the smaller square as the 1-by-1 square.

Then they create a rectangular area model similar to the one below.



In contrast to the rest of the students, one student in the class used the smaller square as the x-by-x square and the larger square as the 1-by-1 square. This student's rectangular area model is shown below.



The student and his group struggled with the idea of why there were two different models representing the same algebraic expression. After about twenty-five minutes and some questions from the instructor, the student who chose the "unusual" depiction of $2x^2 + 3x + 1$ realized that he could put certain restrictions on the value of the variable (x >1 or x <1) to explain the two different representations. This student and his group initially expressed concern that his model was different and assumed that only one model could be correct.

This type of experience is unlikely to occur in a lecture setting. Students do not have the opportunity to explore different scenarios and models for a problem in a lecture. The example above shows how students can arrive at a deeper understanding of a concept through encountering unexpected outcomes and working through the reasoning behind these unexpected outcomes.