**Blended Learning:**

**Face-2-Face + Intelligent Tutor System = Differentiated Math Instruction**

Dr. Karen K Lucas  
Catawba College  
Salisbury, North Carolina  
United States  
klucas13@catawba.edu

**Abstract**

A blended program integrating an Intelligent Tutor System (ITS) with face-to-face instruction provided differentiated instruction resulting in increased student achievement in mathematics, with the greatest gains made by students who initially scored at the below-basic level. Seventy-five ninth-grade students who experienced the blended learning program were the intervention group, while 99 ninth-grade students who experienced face-to-face instruction alone served as the comparison group. Flow Theory was used to explain how artificial intelligence used in ITS programs contributed to the improvement in student achievement, particularly for those students who began the program with learning gaps in mathematics. Teacher facilitators overcame challenges, developed protocols, and employed best practices to successfully implement the program.

*Keywords:* Aptitude-Treatment Interaction (ATI), Blended Learning, Flow Theory, Intelligent Tutor System (ITS), Learning Gaps, Mathematics Achievement, Online Learning, Technological Pedagogical Content Knowledge (TPACK)
Blended Learning:

Face-2-Face + Intelligent Tutor System => Differentiated Math Instruction

In our knowledge-based economy, algebra is seen as the gatekeeper to high-level mathematics and many high-paying career opportunities (Capraro & Joffrion, 2006; Ladson-Billings, 1997; Stein, Kaufman, Sherman, & Hillen, 2011; Wu, 2001), and achievement in math has been identified as a concern in America (National Mathematics Advisory Panel, 2008). Therefore, it is important to identify what students need to become successful in learning algebra. The understanding of many mathematical concepts is prerequisite to learning algebra. This notion is supported by the National Mathematics Advisory Panel (2008) in *Foundations for success: The final report of the National Mathematics Advisory Panel*, which states that “The coherence and sequential nature of mathematics dictate the foundational skills that are necessary for the learning of algebra” (p. 18). This indicates that by the very nature of the discipline of mathematics, filling students’ learning gaps is essential prior to them learning higher level mathematical concepts. Thus, educators are tasked with the problem of finding efficient ways to identify individual student’s learning gaps and providing differentiated lessons to address each student’s instructional needs.

**Literature Review**

Online learning has entered K-12 classrooms for many reasons, including shrinking budgets, teacher shortages, and pressure for results. Integrating online learning with face-to-face instruction has resulted in *blended learning*, which has been defined as “any time a student learns at least in part at a supervised brick-and-mortar location away from home and at least in part through online delivery with some element of student control over time, place, path, and/or pace” (Horn & Staker, 2011, p. 3). Among the many computer assisted instruction programs that are
available to support learning, there are *Intelligent Tutor System* (ITS) programs, which are unique because they utilize artificial intelligence and cognitive psychology to assess students’ responses, provide appropriate hints or other immediate feedback, and direct students to exercises that meet their individual needs (Johnson, 2005).

Studies have been done that provide evidence that ITS programs can positively affect student achievement in mathematics (Barrus, Sabo, Joseph, & Atkinson, 2011; Chien, Md.Yunus, Ali, & Baker, 2008; Koedinger, Anderson, Hadley, & Mark, 1997). The ITS called the Practical Algebra Tutor (PAT) was implemented and studied in 1993-94 as part of the Pittsburg Urban Mathematics Project. The students who used PAT were online for 25 out of 180 days, and they scored 1.2 sigma higher on standardized tests than the control group (Koedinger, et al., 1997).

In another study, student outcomes after using computer assisted instruction (CAI) along with an ITS program were compared to outcomes of students who used CAI alone (Chien, et al., 2008). Both groups of students were online 1 hour per day for 8 days and were administered pre and post-tests, which revealed that the ‘CAI plus ITS’ group learned significantly more than the ‘CAI alone’ group. The difference in those student outcomes were attributed to the personalized feedback provided by the ITS (Chien, et al., 2008).

In recent years, ITS programs have become more readily available, which led to a study comparing the effects of two different off-the-shelf ITS programs, Carnegie Learning’s Cognitive Tutor and ALEKS (Barrus, et al., 2011). In this study, both groups used the ITS programs 4 hours per day for 14 days of summer school. Both groups’ scores improved significantly over time based on assessments administered on day 1, day 7, and day 13; however, gains did not differ significantly based on which ITS program was used.
Purpose

This study contributes to the growing body of knowledge about ITS programs by analyzing the amount of growth made by students based on their initial aptitude levels. After student achievement data collected from the intervention group were compared to that of the comparison group as a whole, the data were further analyzed to determine if there was an Aptitude-Treatment Interaction (ATI). If the effect of the instructional treatment differed based on the students’ aptitude levels, then an ATI was present (Snow, 1991). Another focus of this study was the documentation of challenges teachers encountered and best practices that they used while implementing the blended learning program. The purpose of the analyses done in this study was to inform educators’ decisions regarding the implementation of ITS programs. Therefore, the research questions addressed in this study were the following: Does student achievement differ based on instructional program, face-to-face plus an ITS program or face-to-face alone? Do achievement gains differ based on students’ initial aptitude levels? What challenges do teachers encounter, and what practices do they use when implementing the ITS program?

Theoretical Framework

A case can be made that student achievement would be higher for the students in the intervention group, who experienced the ITS program, based on Flow Theory. Mihaly Csikszentmihalyi proposed Flow Theory in 1975 to explain motivation. Flow is described as a state of deep concentration that is evident when a person is intensely immersed in performing a task such as painting, acting, or performing in athletics (Liao, 2006; Scherer, 2002). When in the flow of learning, students simultaneously display concentration, interest, and enjoyment (Shernoff, Csikszentmihalyi, Schneider, & Shernoff, 2003). According to Flow Theory, for a
student to experience flow there must be a balance between the challenge of tasks and the student’s skills. Because ITS programs use artificial intelligence, they adapt lessons to individualize instruction ensuring that the challenge of the tasks presented closely matches each student’s skill level and all students have opportunities to experience success (Fleisher, 2006; Huffmyer, 2008; Johnson, 2005). Therefore, implementing the ITS program to support students in learning mathematics was expected to improve learning flow and increase achievement.

**Methodology**

This study took place at Haywood High School (pseudonym), an urban Title I school located in the Southeastern region of the United States. All of the ninth-grade students who were taking Algebra 1 at Haywood HS in the spring of 2012 received the intervention. This was done in order to avoid the potential problem of resentful demoralization, which is a negative social interaction threat that could have arisen if any students who would have been denied the intervention gave-up on learning out of anger about not getting access to the online program. Therefore, the ninth-grade students who took Algebra 1 the previous spring at Haywood HS served as the comparison group. The intervention group consisted of 75 ninth-grade students who were 85% Black, 12% White, 1% Hispanic, and 1% other and took Algebra 1 in the spring semester of 2012. The comparison group consisted of 99 ninth-grade students who were 89% Black, 10% White, 1% Hispanic, and 1% other and took Algebra 1 in the spring semester of the previous year. In both groups, approximately 90% of the students were economically disadvantaged based on free or reduced lunch data.

The ITS program implemented in this study, called Apangea Math © [subsequently renamed Think Through Math © (“Think Through Math,” 2014)], was developed based on research by the U.S. Air Force Research Laboratory and the National Science Foundation to
assist students who struggle with learning mathematics ("Apangea Learning," 2011). The design of the study was quasi-experimental research that took place in a natural school setting with intact classes. Apangea Math was delivered online to the intervention group using a rotation model of blended learning, meaning the students rotated between face-to-face and online instruction facilitated by the same teacher on a fixed schedule (Horn & Staker, 2011). The Algebra 1 courses in this study were one semester long and met 5 days per week for 90 minutes each day. During each week, the intervention group students were scheduled to learn via online Apangea lessons 4 days per week for 40 minutes each day and via face-to-face instruction the remainder of the time. The comparison group students learned via face-to-face instruction alone.

Both the intervention group and the comparison group were administered the same online benchmark tests, Discovery Assessments (DA) by Discovery Education © (2008), in January and April which were used as pre-test and post-test to measure the gains students made in achievement over the course of each semester. Data regarding the amount of time the intervention group students spent using Apangea and the number of Apangea lessons they passed were statistics that were available through the Apangea program. At the end of each course, all the Algebra 1 students were administered the same standardized high-stakes Algebra 1 end-of-course (EOC) exam to measure their final achievement levels. The teachers who implemented the blended learning program responded to open-ended questions weekly regarding challenges encountered and practices used, and the researcher conducted observations and facilitated meetings with teachers.

To compare the effects of the instructional treatments, an independent t-test was conducted on the students’ Algebra 1 EOC exam scores to determine if the final proficiency levels of the intervention group and the comparison group differed significantly. Furthermore,
because the design of this study was quasi-experimental, the gains that students showed in learning mathematics over the course of each semester was determined based on pre and post DA test scores, and those gains were compared to determine if student growth differed based on instructional program, face-to-face plus the ITS program or face-to-face alone. The pre and post-test scores were analyzed by conducting a mixed ANOVA with the between-subjects independent variable being treatment (2 levels: intervention and comparison), the within-subjects independent variable being time (2 levels: pre and post), and the dependent variable being the percent correct on the DA tests. This was done to show if there was an interaction between time and treatment on achievement.

Additional analyses were conducted to determine if achievement gains differed based on students’ initial aptitude levels in order to reveal if there was an aptitude-treatment interaction. Initial aptitude levels were determined by students’ pre-test DA scores, which aligned with the Algebra 1 EOC high-stakes exams’ cut-scores: less than 28% correct indicated below basic, 28 to 42% was basic, 43 to 58% was proficient, and 59% or more correct was advanced. An independent t-test was conducted for each initial aptitude level in order to determine if the gains made by the intervention group differed significantly from the gains made by the comparison group based on initial aptitude.

The intervention group’s data alone were analyzed further by conducting an ANOVA on pre and post DA test scores to determine if gains made by the students who experienced the blended learning program differed significantly based on students’ initial aptitude levels. The intervention group’s data was also examined to arrive at a way to predict students’ EOC scores. This was accomplished by conducting a standard multiple regression on predictor variables,
which were found to be significantly correlated with EOC scores, resulting in an equation that
could be used to predict EOC scores.

Finally, in order to determine what challenges teachers encountered and what practices
they used when implementing the ITS program, the qualitative data, which were collected from
teachers’ weekly forms, classroom observations, and meeting notes, were analyzed according to
Snoeyink and Ertmer’s (2001) categorization of change as first-order (external) or second-order
(internal). The data was segmented, coded, reviewed, and adjusted iteratively until clear
consistent themes emerged.

Results

Comparing the Intervention Group to the Comparison Group

On the Algebra 1 EOC exams, the intervention group students’ scores \((M = 82.65, SD =
10.95)\) were significantly higher than the comparison group students’ scores \((M = 71.21, SD =
14.45)\), \(t(170.96) = 5.92, p < .001, d = .88\). See Figure 1.

<table>
<thead>
<tr>
<th>Intervention Group</th>
<th>Comparison Group</th>
</tr>
</thead>
<tbody>
<tr>
<td>(M = 82.7, SD = 10.95)</td>
<td>(M = 71.2, SD = 14.45)</td>
</tr>
</tbody>
</table>

![Pie chart](image)

**Figure 1.** Means and standard deviations of Algebra 1 EOC exam scores and achievement levels by treatment group.

53% scored Proficient or Advanced   |   16% scored Proficient or Advanced
As previously described, this study was conducted in a natural school setting with intact classes. Thus, it was important to examine the outcomes of the non-equivalent groups by analyzing the gains students made in achievement over time rather than focusing simply on final EOC exam scores. The number of students who were present for the administrations of both the pre and post DA tests were 67 in the intervention group and 92 in the comparison group. Table 1 shows the means and standard deviations for the Discovery Assessments by time and group.

Table 1. Means and Standard Deviations for Discovery Assessments by Time and Group

<table>
<thead>
<tr>
<th>DA Scores</th>
<th>n</th>
<th>Pre-test</th>
<th>Post-test</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>M  (SD)</td>
<td>M  (SD)</td>
<td>M  (SD)</td>
</tr>
<tr>
<td>Intervention Group</td>
<td>67</td>
<td>42.54 (13.50)</td>
<td>51.99 (15.74)</td>
<td>47.26 (15.36)</td>
</tr>
<tr>
<td>Comparison Group</td>
<td>92</td>
<td>33.67 (10.55)</td>
<td>38.32 (13.21)</td>
<td>35.99 (12.15)</td>
</tr>
<tr>
<td>Total</td>
<td>159</td>
<td>37.41 (12.63)</td>
<td>44.08 (15.81)</td>
<td>40.74 (14.67)</td>
</tr>
</tbody>
</table>

The students’ scores from pre to post-test were analyzed by conducting a mixed ANOVA, which revealed a significant interaction effect between time and treatment group, $F(1,157) = 5.25, p < .05$, partial eta$^2 = .032$, observed power = .63, which indicated that from pre-test to post-test the intervention group’s gains (9.45 points) were significantly greater than the comparison group’s gains (4.65 points). The eta for this interaction effect was approximately .18, which is a small to medium effect size (Leech, Barrett, & Morgan, 2008).

Additional analyses were conducted to determine if the intervention group made significantly greater gains than the comparison group for students at each initial aptitude level. For each level, based on students’ pre-test scores (below basic, basic, proficient, and advanced), an independent t-test was conducted. The independent t-test for students who initially scored at
the below basic level showed that gains made by the intervention group’s below-basic students (24.50 points) from pre-test ($M = 20.50, SD = 3.59$) to post-test ($M = 45.00, SD = 11.16$) were significantly greater than gains made by the comparison group’s below-basic students (10.91 points) from pre-test ($M = 22.78, SD = 5.18$) to post-test ($M = 33.69, SD = 10.59$), $t(38) = 3.15, p < .01$. See Figure 2 comparing the gains made by students whose initial aptitudes were below basic based on treatment group. Among students whose initial aptitude levels were basic or proficient, the intervention and comparison groups’ gains did not differ significantly. No $t$-test was conducted to compare the gains made by intervention and comparison students with advanced aptitudes because of an inadequate sample size.

![Figure 2](image)

*Figure 2.* DA gains among students with below basic aptitude by treatment group.
Analyzing the Intervention-Group Data Alone

Prior to conducting analyses on the intervention-group data alone, statistics gathered from the Apangea program were analyzed to confirm that the amount of time students spent online and the number of Apangea lessons that they passed did not differ significantly based on initial aptitude level. Then an analysis was conducted on the intervention group students’ pre and post-test scores to determine if their gains differed significantly based on initial aptitude levels. Figure 3 shows the means of the intervention group’s pre-test scores (vertical blue bars), gains (horizontal red bars), and post-test scores (complete bars) by initial aptitude levels.

The ANOVA conducted to analyze the pre and post-test data of the intervention group students showed that their gains differed significantly by aptitude level, $F(3,62) = 5.11, p < .01$, and a Bonferroni post hoc test showed that below-basic students’ gains ($M = 24.50, SD = 11.46$) were statistically greater than basic students’ gains ($M = 7.46, SD = 13.67$), proficient students’ gains ($M = 7.46, SD = 9.74$), and advanced students’ gains ($M = 5.25, SD = 11.41$).
In an effort to analyze how variables in this study might explain students’ EOC scores, a Pearson $r$ correlation was conducted to identify predictor variables. The amount of time that students spent on Apangea and the number of Apangea lessons that they passed were highly inter-correlated so a combined average of those two variables was computed as one new variable, called *Apangea combined*. See Table 2 for the means, standard deviations, and inter-correlations for EOC scores and predictor variables.

Table 1. *Means, Standard Deviations, and Intercorrelations for EOC Scores and Predictor Variables (N = 66)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>$M$</th>
<th>$SD$</th>
<th>DA pre-test</th>
<th>Apangea combined</th>
<th>DA gains</th>
</tr>
</thead>
<tbody>
<tr>
<td>EOC Scores</td>
<td>84.17</td>
<td>(9.60)</td>
<td>.57**</td>
<td>.53**</td>
<td>.32**</td>
</tr>
<tr>
<td>Predictor Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DA Pre-test</td>
<td>42.47</td>
<td>(13.59)</td>
<td>_</td>
<td>.35**</td>
<td>-.30**</td>
</tr>
<tr>
<td>Apangea Combined</td>
<td>14.31</td>
<td>(3.94)</td>
<td>_</td>
<td></td>
<td>.34**</td>
</tr>
<tr>
<td>DA Gains</td>
<td>9.26</td>
<td>(12.91)</td>
<td>_</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*p < .05; **p < .01

A standard regression was then run with the predictor variables that were significantly correlated with EOC scores; namely, DA pre-test, Apangea combined, and DA gains. Those three predictor variables significantly predicted EOC scores, $F(3,62) = 31.30, p < .001$, with a large effect. See Table 3 for the standard multiple regression analysis. The multiple regression analysis indicated that all three variables contributed positively to predicting the EOC scores. The following equation predicts the EOC scores for intervention group students where $Y$ represents EOC score, $X_1$ is DA pre-test, $X_2$ is Apangea combined, $X_3$ is Gains, and $e$ is error:

$$Y = 58.01 + .47(X_1) + .15(X_2) + .35(X_3) + e$$
Examining the Challenges Encountered and the Practices used during Implementation

The qualitative data collected regarding the challenges encountered and the practices used by teachers during the implementation of the blended learning intervention program were coded and recoded in an iterative process to reveal themes that emerged. The data was also categorized as first-order (external) or second-order (internal) according to the Snoeyink and Ertmer (2001) framework. See Table 4 for a summary of the challenges encountered and the practices used to support the successful implementing of the blended learning program.

Table 4. Challenges encountered and Practices used by Teachers implementing the ITS program

<table>
<thead>
<tr>
<th>Challenges</th>
<th>Practices</th>
</tr>
</thead>
<tbody>
<tr>
<td>First Order Challenges</td>
<td>First Order Practices</td>
</tr>
<tr>
<td>• Network issues</td>
<td>• Establish Protocols</td>
</tr>
<tr>
<td>• Software Limitations</td>
<td>• Provide Incentives</td>
</tr>
<tr>
<td>• Limited Adherence</td>
<td>• Provide Personal Instruction</td>
</tr>
<tr>
<td>• Help Abuse</td>
<td>• Create Learning Pathways</td>
</tr>
<tr>
<td>Second Order Challenges</td>
<td>Second Order Practices</td>
</tr>
<tr>
<td>• Lack of time</td>
<td>• Self-Directed Learning</td>
</tr>
<tr>
<td>• Disbelief in the program</td>
<td>• Developing Technological Pedagogical Content Knowledge (TPACK)</td>
</tr>
<tr>
<td>• Off-task behavior</td>
<td>• Student Burn-out</td>
</tr>
</tbody>
</table>
The first-order (external) challenges that teachers reported were network issues, software limitations, some students not adhering to implementation protocols, and abuse of online hints. Network issues that were encountered included slow and unreliable internet connections, which were problems that were solved by administration and technology support staff district during the first week or two of the implementation of the intervention. The theme, software limitations, emerged because occasionally mathematical expressions and equations entered by students were not accepted as correct by the ITS program if they were entered by the student in an unconventional way. For example, if the correct answer to a problem was $a^2 + b^2 = c^2$, the ITS might not accept an alternate form of the equation, such as $c = \sqrt{a^2 + b^2}$. These types of software limitations caused some frustration occasionally for both students and teachers, but personnel from the ITS company were readily available via live chat or phone to address difficulties with individual problems. Limited adherence and help abuse were themes used to describe times when students did not adequately follow protocols for using the online program. For example, some students would sometimes fail to utilize the headphones, calculators, and paper and pencil as prescribed, and some students would click to access online hints excessively. To reduce these undesirable behaviors, teachers explained proper procedures for engaging in the online lessons and motivated students by assigning grades based on how many online lessons they passed, which was not the same as how many lessons they clicked through.

Second-order (internal) challenges that teachers discussed were perceived lack of time, disbelief in the program, concern about off-task behaviors, and student’s experiencing burn-out. Some teachers initially expressed dissatisfaction with the ITS program when it would direct some students to work on mathematical content that did not match the face-to-face lessons for the week. Teachers perceived this as a waste of time until they recognized that the program was
adapting to student input and helping students fill their learning gaps. Off-task behaviors and student burn-out were themes that emerged to categorize teachers’ descriptions of a few students who would sometimes access the drawing tools on their computers instead of interacting with the ITS program and would become tired of the online lessons after the novelty of working online had worn off.

Most of the challenges that were encountered were met and overcome by practices that the teachers and administrators enacted. The most commonly described first-order (external) practice that contributed to the success of the implementation of the blended learning program was establishing protocols. Those protocols included agreeing on and defining the schedule, procedures, and expectations for the program. Technology staff configured the tablet computers that were used in the math classrooms so the only accessible website was Apangea, which reduced off-task behaviors. Recognitions were developed to reward the classes that passed the most Apangea lessons each week. Teachers utilized Apangea’s reporting features to assess students’ progress and provided individual face-to-face tutoring.

A helpful second-order (internal) practice resulted from teachers explicitly coaching students about thinking through problems and accessing online hints and help only when necessary; thus, students developed self-directed learning skills. Another noteworthy second-order practice that was recognized as important was the development of teachers’ TPACK which happened progressively as they received support and learned to facilitate student learning through the blended program.

**Discussion**

The results of this study showed that the increased gains displayed by the intervention group where driven primarily by the gains made by the students who demonstrated initial
aptitude levels that were below-basic. This aptitude-treatment interaction may be attributed to the fact that through the use of artificial intelligence, the ITS program adapted to meet those students’ needs. It provided appropriate instruction on concepts that were more basic than the mathematical content that teachers routinely teach in Algebra 1 courses. The ITS contributed to the closing of learning gaps, which substantiated the notion that many mathematical concepts are prerequisites to learning algebra (National Mathematics Advisory Panel, 2008).

Teachers adjusted to the new learning model and indicated that well established network connectivity and easy access to technical support were critical to successfully implement blended learning. Teachers also emphasized that it was important to have clear communication between administration and teachers and to define program protocols, meaning it was a good practice to explicitly spell out expectations for when and how the ITS was to be used. During the study it became apparent that providing time for ongoing professional development and reflection increased teachers’ TPACK, which increased their acceptance of and belief in the blended learning program. By the end of the study, teachers described the ITS program as the best differentiation tool they had available to them.

Recommendations for future research include conducting similar studies with groups that are randomly selected in order to implement a true experimental design, which would improve internal validity. It would also be desirable to conduct similar studies with larger, more diverse groups of students in order to improve generalizability and generate more data regarding possible aptitude treatment interactions. Furthermore, conducting research on blended learning models of instruction that provide face-to-face and the ITS program in different proportions of time should be studied because altering those variables may result in different outcomes.
References


