AGT Evaluation Analysis (1)

Noboru Matsuda
University of Pittsburgh

June 14, 2004

1. Review of the AGT Evaluation

Procedure:

Participants were randomly assigned to one of two tutor conditions prior to the sessions. They first read a geometry booklet, which shows how to write a proof only for the assigned condition. They then take a pre-test (40 min, open book). After the pre-test they started using the tutor and worked on 11 problems. Finally, they took a post-test (40 min, open book). Table 1 shows the problems and postulates used in the evaluation study.

Participants:

60 people participated in the study. 2 dropped and 6 were too good (i.e., they scored 100% correct on the pre-test). As a consequence, there were 26 subjects in each tutor condition.

Independent variables:

Tutor: (FC) Forward chaining tutor
(BC) Backward chaining tutor

Test Item: (A) and (B). They are identical in terms of a set of postulates and their order need to apply to make a valid proof. A half of the subjects used Test-A as a pre-test and Test-B as a post test. Another half went the other way around.

Time spent to complete each problem.

Time spent on each message: measured as a difference between the time a message was displayed on the screen and the time the subject clicked [OK] button (which is the only available GUI activity at that time).

Dependent variable:

Pre- and Post-test scores

Table 1: Problems and Postulates used in the Evaluation.

		CPCTC	Identity	SAS	SSS	VerAng	Z	Mtri	ASA	Trans	Coll-para	TriM
Tutoring	N1	o	o	o								
	C1	X	o		o							
	N2	o	o	o								
	C0	X	o		o							
	N3	o		o		o						
	C5	X	o		o							
	C11	X		o		o	o					
	N6	o				o	o	o	o			
	N11	o				o	o	o	o	o	o	
	C12							X		o		o
	C13							X		o	o	o
Test A	N12	?	?						o			
	C14	o				?	?		?	o	o	
	N7	o		?								
	N4	o				o	o		o			
	C2	X	o	o			o					
	C8	X				o	o	o	o	o	o	
Test B	N13	?	?	o								
	C10	o				?	?	X	?	o	o	
	N10	o		?								
	N5	o				o	o		o			
	C4	X	o	o	o							
	C9	X				o	o	o	o	o	o	
		15	7	6	3	7	7	5	6	5	4	2

For Tutoring:

"o" shows that the corresponding postulate must be applied to compose a proof.

"X" shows that a construction is necessary to apply.

For Test A and Test B:

N12 through N7 are the fill-in-blank problems. "?" shows the postulates to be filled in.

N4 through C8 are write-a-proof problems. "X" shows that construction is necessary to apply corresponding postulate.

2. Scores on Pre- and Post-tests

Table 2 shows the pre- and post-test scored in each condition. Overall, there is no significant difference between BC and FC on the pre-test scores, but there is significant difference between BC and FC on the post-test scores.

Table 2: Pre- and Post-test scores

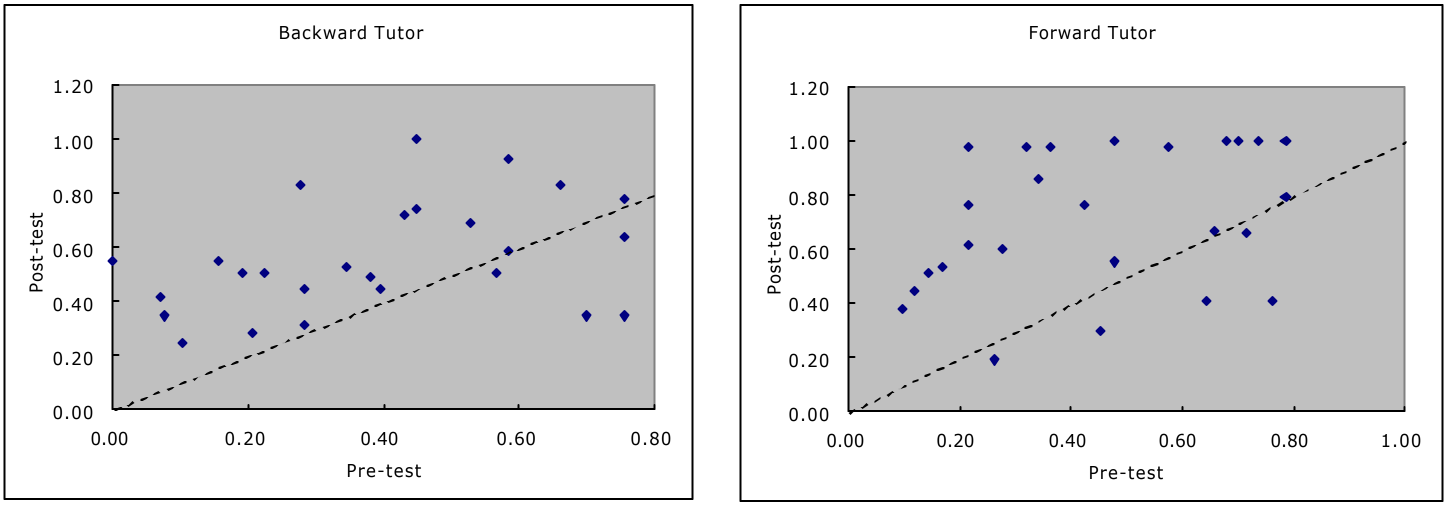
Group Statistics

	TUTOR	N	Mean	Std. Deviation	Std. Error Mean
PRETEST	BC	26	.392323	.2308381	.0452711
	FC	26	.446127	.2343356	.0459570
POSTTEST	BC	26	.559544	.2039039	.0399888
	FC	26	.705265	.2576926	.0505377

p = .41

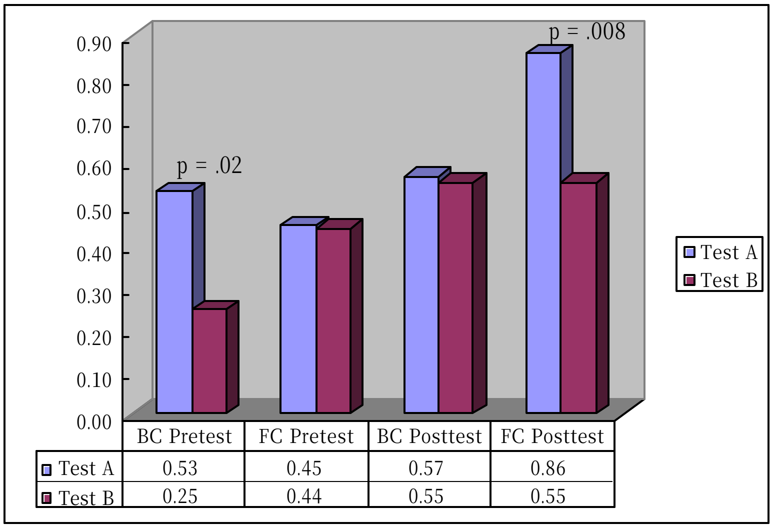
p = .028 FC >> BC

Table 3: Scatter plot of Pre-test against Post-test



3. Effect of Test-Item Difference on the Test Scores

There is a significant difference between the test items A and B on (1) the pre-test in the BC condition and (2) the post-test in the FC condition.



	BC Pretest	FC Pretest	BC Posttest	FC Posttest
Test A	0.53	0.45	0.57	0.86
Test B	0.25	0.44	0.55	0.55

Tests of Between-Subjects Effects

Dependent Variable: PRETEST

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.550[a]	3	.183	4.012	.013
Intercept	9.139	1	9.139	200.052	.000
TUTOR	.038	1	.038	.824	.369
ITEM1	.266	1	.266	5.813	.020
TUTOR * ITEM1	.247	1	.247	5.399	.024
Error	2.193	48	.046		
Total	11.882	52			
Corrected Total	2.743	51			

a. R Squared = .200 (Adjusted R Squared = .151)

Tests of Between-Subjects Effects

Dependent Variable: POSTTEST

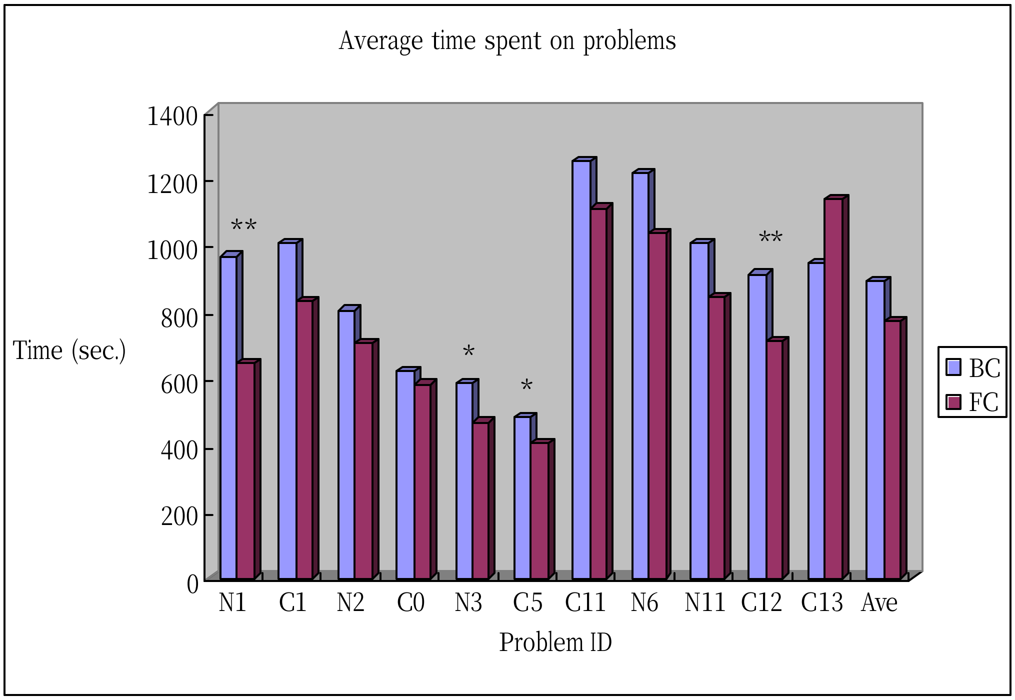
Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	.891[a]	3	.297	6.844	.001
Intercept	20.797	1	20.797	478.963	.000
TUTOR	.276	1	.276	6.358	.015
ITEM2	.334	1	.334	7.684	.008
TUTOR * ITEM2	.282	1	.282	6.489	.014
Error	2.084	48	.043		
Total	23.772	52			
Corrected Total	2.976	51			

a. R Squared = .300 (Adjusted R Squared = .256)

4. Difference in Time Spent on Problems

Table 4 shows average time spent on each problem in each condition. Overall, there is no significant difference between BC and FC conditions, except on the test items N1 (p=0.001) and C12 (p=.045). There are marginal difference on test items N3 (p=.055), and C5 (p=.068).

Table 4: Time spent on Problems

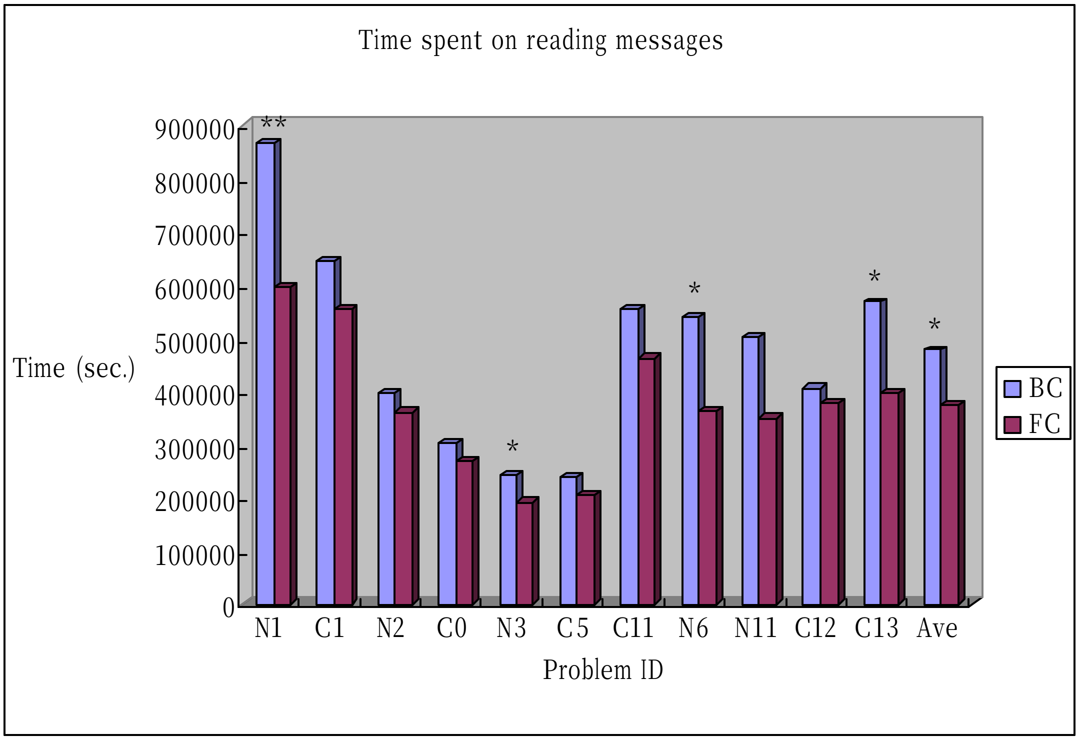


In the table above, the problems are listed in the order that they are set in the tutor. There is a big jump from C5 from C11. Most of the subjects (44 out of 52, or 85%) divided the entire session into two sub sessions and they solved the first 6 problems (N1 through C5) on the first day. As shown in Table 1, the problems N1 through C5 repeatedly used 4 postulates (CPCTC, Identity, SAS, SSS). The tutor started to use a new postulate (or two) for each new problem from C11. The number of postulates involved also started to increase from C11. These might account for the big jump between C5 and C11.

5. Difference in Time Spent on Reading Messages from Tutor

There is a marginal difference in the time spent on tutor's messages between BC and FC conditions (FC >> BC, p = .094). The problems N3, N6, and C13 shows marginal difference as well (p = 0.65, 0.90, and 0.89 respectively). Only the problem N1 showed a significant difference (p = .005).

Table 5: Time spent to read messages from the tutor



6. So, what makes FC learn more?

Something facilitated FC learning? Or, something hindered BC learning?

I will analyze relation between learning gain and (1) # of errors made, (2) # of bottom-out hint (or the tutor's solution) provided, (3) # of postulate applications made, (4) time spent on each postulate application, etc…