Project M² Evaluation Report for The National Science Foundation (NSF)
The 2011 Executive Summary

Words & Numbers Research, Inc. compiled several program evaluation components that target Kindergarten and 1st grade students. Attachments A through E provide the full reports. The executive summary of the key findings document a program that is both remarkable and successful.

Kindergarten - Demographic and Pretest Statistical Profile - Attachment A (p. 3)

There were 223 (51%) students in the Intervention group and 212 students (49%) in the Comparison group. Elementary schools were located in Connecticut (5), Texas (2), South Carolina (2) and Kentucky (2).

- There were approximately 53% male and 47% female students.
- The ethnicities were Asian (6%), African American (21%), Hispanic (15%), and Caucasian (56%).
- Almost half (49%) of the students were eligible for a meal subsidy.
- Eight percent were identified as receiving Special Education services.
- Eight percent of the students were in ESL or ELL programs.

The pretest scores for the Intervention and Comparison groups on the ITBS were equivalent. There were no statistically significant differences. The Open Response scales and Total scores had no significant differences between the Intervention and Comparison groups. This is the ideal research finding at pretesting. There was group equivalence at the beginning of the intervention for Kindergarten students.

Unit Pre and Post Test Data Analysis - Attachment B (p. 8-11)

On the Grade 1 Measurement Unit, 1st graders achieved statistically significant ($p<.001$) gains from pre to posttesting on Different Weights, Same Weights, Weight Transitivity, Area, Area Transitivity, Measuring Length, Drawing Length and Total Scores. Almost every 1st grader (98%) made gains from pretesting to posttesting.

On the Kindergarten Measurement Unit, Kindergarteners achieved statistically significant ($p<.001$) gains from pre to posttesting on Inverse Length, Ordering Lengths, Capacity, Area, Written Volume, Written Length, and Total Scores. Every kindergartener (100%) made gains on Total Scores from pre to posttesting.

Kindergarten Teacher Professional Development Assessment - Attachment C (p. 12)

Professional development was measured for both pre and post training. Both quantitative and qualitative data reflected the professional development as excellent, according to Kindergarten teachers during the summer 2010. One hundred percent felt the quality of professional development was very satisfactory. There were statistically significant gains in confidence levels with mathematical content, covered in the training.

TEL (860) 489-5639 · 25 Maiden Lane · P.O. Box 1373 · Torrington, CT 06790-1373
E-mail: wordsnum@snet.net · www.wordsandnumbers.org

Pre and Post Mathematical Content Acquisition for 1st Grade Teachers - Attachment D (p. 19)

The mastery of content from pre to posttesting was very satisfactory for the 1st grade teachers and there were statistically significant gains. The content was mastered to a great extent with 11.25 out of a total of 16.5 points

Pre and Post Mathematical Performance by Intervention and Comparison Groups - Attachment E (p. 20)

There were two research questions empirically addressed in the evaluation research during PY4.

> **Research Question #1**: Is there an *increase in mathematics achievement* for the Intervention group of 1st grade students across all socioeconomic and ethnic backgrounds after exposure to a mathematical curriculum model that provides challenging standards-based curriculum and encourages high-level discourse?

> **Research Question #2**: Is there a *difference in mathematics achievement* between the Intervention group of 1st grade students who are exposed to the mathematics curriculum model and the Comparison group of 1st grade students of similar socioeconomic and ethnic backgrounds who participate in the regular math curriculum?

The first question affirmed a statistically significant increase in mathematical achievement for Intervention students on all mathematics performance measures - the Iowa Test of Basic Skills Mathematics scale, the Open Response Total test and its two subscales in Geometry and Measurement, respectively.

The second question affirmed statistically significant differences in mathematics performance for the Intervention group when compared to the Comparison group. A series of Hierarchical Linear Models were constructed to evaluate the performance on the ITBS and the Open Response Total and the two scale scores. Although there were no differences between the two groups on the ITBS, the Open Response Total and the Geometry and Measurement subscale scores were significantly higher for the group receiving the M² curriculum intervention.

The results affirm the research hypotheses. The Intervention group of 1st grade students made significant gains in mathematics performance from pre to posttesting due to the M² curriculum. Furthermore, there were statistically significant differences in favor of the Intervention group on mathematics performance. They outperformed the peer Comparison group on the Open Response Total and the Geometry and Measurement subscale scores.

From an evaluation perspective, this is a remarkable program given the outcomes achieved with students who are very young and with highly challenging content.

Respectfully submitted,
Susan Carroll, Ph.D.
Evaluation Consultant

Words & Numbers Research, Inc.
March 9, 2011

Attachment A

Kindergarten Students
Demographic and Pretest Statistical Profile

Submitted to: Dr. M. Katherine Gavin, Project Director
Submitted by: Dr. Susan Carroll, Evaluation Consultant
Submitted on: March 2011

Background on the Student Data Collected in Program Year 4 (PY4)

An **Individual Student Data Form** was designed by Words & Numbers Research, Inc. in order to ensure the uniform reporting of student data.

On the form, **archival or demographic variables** were requested. These included:

1. Public school
2. Grade level teacher
3. State of origin
4. Gender
5. Ethnicity
6. Eligibility for free / reduced lunch
7. Participation in Special Education
8. Participation in an ESL /ELL/ EL program
9. Participation in another Math program

Performance data were also requested. The standardized test utilized was the **Iowa Tests of Basic Skills** (ITBS). These data were reported as both Standard Scores and Raw Scores. Raw scores were converted into percentage of correct responses. There were 29 items on the ITBS. So the percentage was calculated based on the number of items out of 29 that the student responded to correctly.

Additionally, there were Open Response questions which covered two content areas: Measurement and Geometry. For Measurement the range of possible scores was 0 to 24. For Geometry the range of possible scores was 0 to 22. A Total Score was calculated as well. The range of scores was 0 to 46.

Demographic Data

- Approximately 435 students in Kindergarten were part of the **M²** project during the 2010-2011 school year. Of those, 212 (49%) were in the Comparison group and 223 (51%) were in the Intervention group. Students were drawn from the following schools:

Schools and State	Total	Comparison N	Comparison %	Intervention N	Intervention %
Corcoran, SC	45	21	47%	24	53%
Carleston, TX	40	20	50%	20	50%
Charter Oak, CT	31	15	48%	16	52%
Goodwin, CT	27	13	48%	14	52%
Lawhon, TX	34	17	50%	17	50%
Lincoln, KY	43	20	47%	23	53%
Midland Park, SC	41	21	51%	20	49%
Noah Webster, CT	32	16	50%	16	50%
Southeast, CT	32	17	53%	15	47%
Southern, KY	78	36	46%	42	54%
Vinton, CT	32	16	50%	16	50%
Total	**435**	**212**	**49%**	**223**	**51%**

Descriptive data were generated for the entire **M²** population and by Intervention and Comparison groups. Statistical comparisons were executed to determine if the two groups were similar on the demographic variables. The equivalence at pretesting was ensured. There were no differences on gender, ethnic background, family income, participation in Special Education, ELL programs or other math programs.

- There were approximately 230 (53%) boys and 205 (47%) girls. There were no statistically significant differences in gender between the Intervention and Comparison groups.

Gender	Total	Intervention	Comparison
Males	230 (53%)	120 (53%)	110 (52%)
Females	205 (47%)	103 (47%)	102 (48%)

(Chi sq=.16, df=1, p=.69)

- The ethnicities represented were Asian American (6%), African American (21%), Hispanic (15%), Caucasian (56%) and others (1%). There were no statistically significant differences on ethnicity/ race between the Intervention and Comparison groups.

Gender	Total	Intervention	Comparison
Asian	28 (6%)	14 (6%)	14 (7%)
Black	93 (21%)	51 (23%)	42 (20%)
Hispanic	67 (15%)	33 (15%)	34 (16%)
White	242 (56%)	122 (55%)	120 (56%)
Other	5 (1%)	3 (1%)	2 (1%)

(Chi sq=.83, df=4, p=.94)

- Almost half (49%) of the students were eligible for a meal subsidy. There were no statistically significant differences between the Intervention and Comparison groups on proportions of students with subsidies.

Subsidy	Total	Intervention	Comparison
Yes	205 (49%)	97 (47%)	108 (51%)
No	214 (51%)	110 (53%)	104 (49%)

(Chi sq=.70, df=1, p=.40)

- Eight percent were identified as receiving Special Education services. There were no statistically significant differences between the Intervention and Comparison groups on proportions of students in Special Education.

Participates in Special Ed.	Total	Intervention	Comparison
Yes	33 (8%)	19 (9%)	14 (7%)
No	402 (92%)	204 (92%)	198 (93%)

(Chi sq=.57, df=1, p=.45)

- Eight percent of the students were in ESL or ELL programs. There were no statistically significant differences between the Intervention and Comparison groups on proportions of ELL or EL students.

ELL or EL Participant	Total	Intervention	Comparison
Yes	35 (8%)	19 (8%)	16 (8%)
No	400 (92%)	204 (92%)	196 (92%)

(Chi sq=.14, df=1, p=.71)

- Four percent reported participating in another math program besides the current one. For 96% **M²** was the only math program that they were participating in. There were no statistically significant differences between the Intervention and Comparison groups.

Other Math Program	Total	Intervention	Comparison
Yes	17 (4%)	9 (4%)	8 (4%)
No	418 (96%)	214 (96%)	204 (96%)

(Chi sq=.02, df=1, p=.89)

Pretest Performance Data

The pretest data for the Intervention and Comparison groups on the ITBS were equivalent. There were no statistically significant differences. At pretesting students in both groups were able to answer 45% of the ITBS items correctly. The Open Response scales and Total scores had no significant differences between the Intervention and Comparison groups either. These pretest findings are assurance that the two groups were equivalent prior to the Intervention. This is the ideal research finding.

ITBS and Open Response Pretest Scores for Kindergarten Students

PRETEST SCORES	Intervention	Comparison			
	Mean (SD)	Mean (SD)	t	df	p
ITBS Standard Score	120.65 (8.30)	121.42 (8.14)	.98	430	.33
ITBS Percent correct	45% (.12)	46% (.12)	.95	430	.34
Measurement OR Scale (Highest possible score-22)	5.50 (2.77)	5.70 (2.74)	.77	430	.45
Geometry OR Scale (Highest possible score-24)	3.64 (2.39)	4.15 (2.56)	2.14	430	.03
Total Score OR Scale (Highest possible score-46)	9.14 (4.17)	9.85 (4.36)	1.74	430	.08

(Bonferroni adjustment applied (p<.05/5=.01)

Attachment B
Grade 1 Measurement Unit Pre and Post Test Data Analysis

The Measurement Unit: Grade 1

Summary of Findings:
Pre-Post Data Analysis for Measurement Unit

Submitted by: Susan R. Carroll, Evaluator
Words & Numbers Research, Inc.
June 28, 2010

Using SPSS for data analysis, correlated t-tests were executed.

- There were statistically significant gains from pre to posttesting on **Different Weights** (3 points).
- There were statistically significant gains from pre to posttesting on **Same Weights** (1 point)
- There were statistically significant gains from pre to posttesting on **Weight Transitivity** (2.5 points)
- There were statistically significant gains from pre to posttesting on **Area** (2.5 points)
- There were statistically significant gains from pre to posttesting on **Area Transitivity**(4 points)
- There were statistically significant gains from pre to posttesting on **Measuring Length (**3 points)
- There were statistically significant gains from pre to posttesting on **Drawing Length** (2 points)
- There were statistically significant gains from pre to posttesting on the **Total Scores** (18 points).

These findings were true for the 193 students with complete pre and posttest sets for the seven item test.

Table 1 documents the impressive gains on Measurement. **Total Scores** increase 6.6 points from pre to posttesting. The effect size was calculated for Cohen d at 2.32. Yet, the content on Area Transitivity and Measuring Length were challenging for the 1st grade mathematicians. Given the possible point spread, their posttest scores on those areas reflect the demanding nature of the advanced curriculum unit.

Table 2 shows the percentage of students whose scores increased or made gains from pre to posttesting. Overall, there are impressive gains achieved with this Measurement unit. Almost every 1st grade student (98%) made gains on Total Scores from pre to posttesting.

Table 1: Correlated t-test Results
All Participants (N=193)

Geometry	Points	Pre Mean	Post Mean	Mean Difference	t val-ue	df	*p*
Different Weights	3	.61	1.68	+1.07	13.63	192	***
Same Weights	1	.62	.88	+.26	6.67	192	***
Weight Transitivity	2.5	.63	1.95	+1.32	20.26	192	***
Area	2.5	.78	2.23	+1.45	22.92	192	***
Area Transitivity	4	.63	1.59	+.95	10.90	192	***
Measuring Length	3	.16	1.03	+.87	12.10	192	***
Drawing Length	2	.34	1.01	+.67	8.78	192	***
Total	**18**	**3.77**	**10.37**	**+6.60**	**30.42**	192	*******

*** p <.000

Table 2: Gains, Losses and No Changes from Pre to Post Testing

All Participants (N=193)

Measurement	N	Gains %	Loss %	No change %
Different Weights	193	72%	8%	20%
Same Weights	193	35%	6%	59%
Weight Transitivity	193	89%	5%	6%
Area	193	92%	4%	4%
Area Transitivity	193	64%	6%	30%
Measuring Length	193	59%	3%	38%
Drawing Length	193	49%	8%	43%
Total	193	98%	0%	2%

Measurement Unit Pre and Post Test Data Analysis

The Measurement Unit: Kindergarten

Summary of Findings:
Pre-Post Data Analysis for Measurement Unit

Submitted by: Susan R. Carroll, Evaluator
Words & Numbers Research, Inc.
April 2011

Using SPSS for data analysis, correlated t-tests were executed for the six items and total scores.

- There were statistically significant gains from pre to posttesting on **Inverse Length** (8 points).
- There were statistically significant gains from pre to posttesting on **Ordering Length** (3 points)
- There were statistically significant gains from pre to posttesting on **Capacity** (2 points)
- There were statistically significant gains from pre to posttesting on **Area** (7 points)
- There were statistically significant gains from pre to posttesting on **Written Volume** (7 points)
- There were statistically significant gains from pre to posttesting on **Written Length (**5 points)
- There were statistically significant gains from pre to posttesting on the **Total Scores** (32 points).

These findings were true for the 216 students with complete pre and posttest sets for the six item test.

Table 1 documents the extraordinary gains on Measurement. **Total Scores** increased 14.33 points from pre to posttesting. Each of the six individual items also reflects gains.

Table 1: Correlated t-test Results
All Participants (N=216)

Measurement	Points	Pre Mean	Post Mean	Mean Difference	t value	df	*p*
Inverse Length	8	4.01	6.83	+2.82	13.26	215	***
Ordering Length	3	1.74	2.33	+.59	.47	215	***
Capacity	2	.56	1.67	+1.11	.95	215	***
Area	7	2.33	6.07	+3.74	3.46	215	***
Written Volume	7	.93	4.14	+3.21	2.96	215	***
Written Length	5	.73	3.59	+2.86	2.64	215	***
Total Score	**32**	**10.31**	**24.64**	**+14.33**	**13.60**	**215**	***

*** p <.000

Table 2 shows the percentage of students whose scores increased or made gains from pre to posttesting. Overall, there are impressive gains achieved with this Measurement unit. Every Kindergarten student (100%) made gains on Total Scores from pre to posttesting. Large percentages of students made gains on the six items covered in the measurement test. These data document a very successful achievement of content acquisition. The project has accomplished remarkable results with very young student population.

Table 2: Gains, Losses and No Changes from Pre to Post Testing
All Participants (N=216)

Measurement	N	Gains %	Loss %	No change %
Inverse Length	216	**70%**	11%	19%
Ordering Length	216	**51%**	8%	41%
Capacity	216	**67%**	9%	24%
Area	216	**92%**	4%	4%
Written Volume	216	**91%**	1%	8%
Written Length	216	**89%**	3%	8%
Total Score	216	**100%**	0%	0%

Attachment C

Pre and Post M² Kindergarten Teacher Professional Development: Assessment of the Summer Institute 2010

Submitted by: Susan R. Carroll, Evaluator
Words & Numbers Research, Inc.
August 2, 2010

1.00 Background

Professional development during the summer was offered to nineteen Kindergarten educators in order to prepare them for the actual implementation of the **M²** project. The training session was conducted during July 2010 on the campus of the University of Connecticut. Because of its importance, the training was evaluated from two vantage points: *satisfaction with the training* and *content and skill acquisition*.

2.00 Satisfaction with the Training

The teachers were asked to assess their **level of satisfaction** with several aspects of the training. There were eleven discrete areas evaluated at the end of the July 2010 training period. Each was rated on a three point rating scale:

- **Very satisfactory**
- **Satisfactory**
- **Not satisfactory**

The results reflect a superior professional development experience, as Table 1 indicates. There were very high levels of satisfaction across the areas evaluated. The overall quality of the training was rated as "very satisfactory" by 100% of the teachers. Additionally, the training was perceived as useful, delivered by skilled presenters, just the right amount of content, supported by pertinent handouts, placed in appropriate facility and implemented with opportunity to interact. The training was a job well done in the eyes of participants and the quantitative and qualitative data reflect.

Table 1: Satisfaction with the Training

Aspects of Training	Very Satisfactory	Satisfactory	Not Satisfactory
1. The level of expertise/ knowledge base of presenters	100%	---	---
2. The delivery of the content by presenters	90%	10%	---
3. Ability of presenters to provide concrete examples to illustrate mathematics	95%	5%	---
4. The quantity and depth of content covered in the training	85%	10%	5%
5. The quality of the materials to support the content	100%	---	---
6. The opportunity for discussion, questions and interchange	100%	---	---
7. The length of the training (8:30-3.00)	100%	---	---
8. The logistics - comfort of rooms, location, equipment, refreshments	90%	10%	---
9. The usefulness of the content and skills presented	100%	---	---
10. The organization of the summer institute	100%	---	---
11. Overall quality of the professional development training	100%	---	---

The following verbatim comments were offered on the aspects of training.

- ✓ *I loved having a variety of different presenters each with their own area of expertise and each with a slightly different perspective. Everything was explained very well and the videos really helped crystallize my thinking. I wish there had been more videos, especially of the K pilot classrooms from last year. All the staff associated with the project was welcoming, warm and knowledgeable. I especially appreciated having Nita Copley who is an early childhood expert. The special learning and developmental needs of 4 and 5 year olds are met in this curriculum and this is not always the case for new programs and curricula I am asked to teach.*

- ✓ *This is one of the best training sessions I have had on math and teaching math to Kindergarten students. The concepts were well presented with research to back it up and the materials fit the activities and objectives exactly. I loved the presenters... wonderful, warm and supportive. I felt welcome and accepted and know that my teaching is valued. The facility was excellent and food, great. I felt that the workshop was encouraging and promises support if needed. Thank you!*

- ✓ *The training has been great. I think this program is going to be very successful. All of the labs and centers were very well presented. The hands on will be very useful for the students and fun. I think the students will enjoy these units and learn a lot more of the higher level thinking than our regular curriculum.*

- ✓ *This training was a wake up call for me as a teacher. My love is Language Arts and I put a lot of emphasis in that area. Math was not strong for me in school and I did not enjoy teaching it. This training has helped me to realize how much deeper my math instruction should go and how much I lack with terminology and problem solving. Thank you for the hands on activities that will drive my instruction. This is a wonderful math program and a new language and understanding I have in this area.*

- ✓ *The organization of the entire session has been super. I have been sheltered, fed, and entertained far better than I could have imagined. I have been treated as an equal in the project and made to feel that I am important. I have learned so much. The presenters are energetic, patient and supportive. I have loved every minute of this week. Thank you!*

- ✓ *Exemplary professional, responsive conference leaders. Hands on experiences. Opportunity to dialogue with colleagues. Research based lessons and materials. I do have concerns about the ability to discourse meaningfully with 4 and 5 year olds. Because of this, I would have liked to see more video samples of young children engaged in meaningful dialogue so I could know better how this will look in my classroom.*

- ✓ *All the presenters were so knowledgeable and helpful and supportive of learning about this program. They worked exceptionally well, an outstanding model of teamwork. They each brought their own style, strength and viewpoints. I feel very excited and confident about implementing M2 in the classroom. I also feel more confident in general about how student discourse should and could look. The set up of the training was well thought out with a combination of different activities, videos, etc. I am thrilled to be part of this program and can't wait to use it this year.*

- ✓ *The presenters were very intelligent and taught me some things that I was unaware of. I can't wait to use this with my kids. They will love the frogs. Thank you for ideas and prizes. I really enjoyed this opportunity.*

- ✓ *The trainers and staff went out of their way to make sure everyone was comfortable, well cared for and relaxed about undertaking this new venture. The atmosphere was always warm and friendly, never threatening which makes me eager to try something new. I feel that my students and I are in good hands!*

- ✓ *This has been the best professional development. I have learned so much about math and can't wait to get started with it in my classroom. All the presenters have so much expertise and are so supportive of the classroom teacher. Thank you for all the fun materials that we will be using to teach the M2 units. I had so much fun! It was also great meeting and being able to share ideas with teachers from other states.*

- ✓ *Enjoyed the training. Friendly, informative people to learn from and with. The first days of Talk Moves was confusing but appreciated the clarification the next day.*

- ✓ *Appreciated going through lessons/activities/talk frame when they were done as models for how units would be completed. Sometimes too much talking and lecturing and not sure it was helpful.*

- ✓ *The detail and quality of the lessons, strategies and materials was exemplary. The warm and supportive instructors made everyone feel at ease with and willing to try a new curriculum.*

- ✓ *The authors are knowledgeable and personable. They use humor in marvelous ways. Their knowledge of content and children is always impressive. It is an honor to work with them!*

- ✓ *Having the opportunity to actually meet the authors and tap into their experiences and expertise was enlightening. It did increase my understanding of the unit by helping to build on background and content knowledge. Thank you!*

- ✓ *After this training, I feel that I have a better understanding of measurement and geometry. I am very excited to implement this in my classroom. The presenters did a wonderful job!*

- ✓ *Presenters are extremely knowledgeable and approachable. These four days have gone by very quickly- a couple of days I was surprised at how fast 3:00 came. Food was great; need to diet for two weeks. Looking forward to watching these two units develop. Excellent professional development opportunity. Thank you!*

- ✓ *Materials provided (just the fact that the teachers receive them) is fabulous. Books are written to provide background knowledge for teachers, include higher level questions, teacher tips, materials list, etc. Nothing*

is left out. Also, loved the vignettes to know what should be expected. Absolutely, everything about this program is exemplary- five star rating.

A few of teachers voiced suggestions.

- ✓ *Share with us the 1st grade and 2nd grade objectives so that we know how the program builds.*
- ✓ *Go over one talk frame at the beginning (all the way through).*
- ✓ *I would have liked a complete lesson to have been taught and modeled by the presenters and then had it taught by one of the K teachers with support of the presenters. This would have been especially important with the Talk Moves. Teachers need to be practicing those- rather than just listening.*

2.00 Pre and Post Content /Skill Acquisition

There were 16 discrete item stems that reflected the targeted content to be delivered in the training. To ascertain the success of the training, teachers were asked to rate each of the 16 items on a five-pint rating scale. This was undertaken before training began (the pretest) and after training ended (the posttest). The rating scale is below.

5 = Very high confidence level / Very well-informed
4 = High confidence level / Well-informed
3 = Moderate confidence level/ Adequately informed
2 = Low confidence level / Partially informed
1 = Negligible confidence level / Not informed

In order to determine whether there were statistically significant gains on knowledge/skill acquisition, correlated t-tests were applied to the items. Pre and posttest scores were compared to determine if any change occurred. The data were statistically analyzed using SPSS, a statistical software package.

The findings were very favorable. There were statistically significant gains on 100% of the items, as Table 2 indicates. This means that the teachers perceived an increase in their knowledge base /skills as a result of the training intervention.

Table 2: Correlated t-test Results on 19 Items

Content and Skills Targeted in the Training	Pre Mean	Post Mean	Gain	t	df	p
1. Encouraging your students to describe 3-dimensional shapes using their properties and the names of the shapes	3.68	4.74	1.05	4.73	18	**
2. Helping students understand the meaning of capacity and volume	3.21	4.63	1.42	6.87	18	**
3. Encouraging students to express their math ideas in writing (i.e., any representation on paper)	3.37	4.47	1.11	4.85	18	**
4. Having students repeat what another student has stated and checking with the student to make sure they heard correctly	3.95	4.63	.68	2.82	18	*
5. Helping students understand the meaning of area and how to measure area	3.42	4.63	1.21	5.75	18	**
6. Integrating verbal discourse into math classes	3.84	4.47	.63	2.47	18	*
7. Helping students understand the inverse relationship between the size of a unit and the number of units needed to measure an object	3.42	4.53	1.11	4.85	18	**
8. Teaching students to measure, compare, and order lengths	4.11	4.95	.84	6.10	18	**
9. Encouraging students to build on what others have said by adding on new ideas	4.11	4.63	.53	2.54	18	*
10. Deciding how to integrate the unit centers and /or management with those that I use in my classroom	3.37	4.53	1.16	4.73	18	**
11. Supporting students in explaining their mathematical reasoning verbally.	3.68	4.47	.79	3.03	18	**
12. Having students realize the need for equal-sized units to measure length, area and volume	3.79	4.84	1.05	5.41	18	**
13. Teaching students to measure and compare the capacity/volume of containers	3.16	4.74	1.58	7.63	18	**
14. Setting up an environment for learning that promotes listening and sharing ideas with one another	4.21	4.84	.63	2.88	18	*
15. Teaching students how to agree or disagree with others' reasoning verbally and tell why	3.63	4.63	1.00	4.14	18	**
16. Recording students' ideas on the talk frame to help with them write about mathematics	2.84	4.05	1.21	6.17	18	**

** $p < .01$
* $p < .05$

The gains from pre test to posttest means were impressive and ranged from 1.58 to .53. An increased level of confidence was acquired in the targeted content/ skills from the beginning of the training to its conclusion. Clearly, the quantitative data support the positive impact of the training.

3.00 Background of Participating Teachers

- Most teachers involved in the project were in the field of education for many years. The average **number of years in the field of education** was 20. The range was a low of 4 years to a high of 39 years. In terms of years in their current position, the average was 7 years, although the range was from 1 to 21 years.

- Most of the teachers in the M² project came to it without an **undergraduate major** or minor in the field of math, nor a **graduate degree** in mathematics. Their focus was in the field of elementary or early childhood education. However, three teachers had math as an area of concentration in their undergraduate program, while two reported that math was a concentration at the graduate level.

- Almost all (94%) reported having some **professional development** related to mathematics during the school year. Six teachers reported having one session and six reported having two to three. Three teachers had four to five sessions of training in math. Two teachers had PD related to math more than five times per year. Only one teacher had none available during the school year.

- Teachers were asked to rate on a scale of 1(low) to 10 (high) where math fell in the range of **preferred subjects** to teach. The mean score was 8.16, signifying a definite preference for teaching math. They were asked to assess their **background knowledge in math** and the mean was 7.63, moderately strong. Finally, the teachers were asked to evaluate their **comfort level** with teaching mathematics. The mean was 8.11, suggesting a high comfort level with the content.

- Next, teachers were asked how often they used **math manipulatives** in their classrooms. All (89%) claimed that they used them *often,* and 11%, *sometimes*. Examples included: dice, legos, unifix cubes, square tiles, teddy bear counters, dominoes, pattern blocks, clocks, money, geoboards, tangrams, links and others.

- Teachers were also asked how often they **differentiated instruction** when students needed **more support**. The majority reported either *often* (56%) or *sometimes* (44%). Examples in-

cluded: small group instruction, games, one on one instruction, restating, modifications, peer tutoring, targeted homework assignments, supplemental practice, cluster grouping and others.

- The frequency of differentiating instruction was presented to teachers but with groups of students who needed **more challenge**. The majority reported that they did this either *often* (47%) or *sometimes* (47%). One teacher (6%) said *rarely*. Examples included: extra homework, higher level questioning, small group work, oral dialogue, independent work activity, extended practice, enrichment, acceleration, cluster grouping, tiered assignments, computer challenges and others.

- Teachers were asked to identify which math textbook, if any, they used in their classroom. These are the ones that were identified on the evaluation form.

Everyday Mathematics (n=7)
Bridges (n=4)
Trailblazers (n=2)
HSP Math (n=2)
M2
M3
Harcourt
Saxon Math

4.00 Summary

The evaluation results for the **M²** professional development training during July 2010 at the University of Connecticut were superior. The training significantly increased the confidence levels of teachers regarding the knowledge and skills that are critical to successful Project **M²** implementation. Additionally, a successful training experience for the participating Kindergarten grade teachers was documented. There was a high level of teacher satisfaction both in quantitative and qualitative results. The entire summer experience was excellent.

Attachment D
M2 Pre and Post Content Acquisition Results for 1st Grade Teachers

Submitted by:
Susan Carroll, Ph.D.
Evaluation Consultant
October 16, 2010

Content acquisition was assessed for 1st grade teachers in the M² project. Although the sample size was small, correlated t-tests were applied to the data. There were statistically significant gains from pre to post testing on the **Geometry** and **Measurement** units. There was no statistical gain in Math Writing, but there were increased scores. There was a statistically significant gain overall (+3.54) in total scores.

The mastery of content at posttesting was very satisfactory for Geometry (88%). The median was 7.00 out of a total of 8 points. Geometry (50%) showed room for additional content acquisition for the teachers; the median was 3.50 out of 7 possible points. Math Writing (67%) was 1.00 out of 1.50 points. Overall, the mastery at posttest was 67% of the total content with 11.00 out of a total of 16.5 points. The content clearly was challenging for the teachers who participated in the project, M². Refer to Tables 1 and 2.

Table 1: Correlated t test results

Units	Possible Points	Pre Mean (SD)	Post Mean (SD)	Mean Difference	t value	df	p
Geometry	8	5.00 (1.09)	7.36 (.67)	2.36	7.63	10	.000
Measurement	7	2.17 (1.19)	3.58 (1.16)	1.41	2.38	11	.037
Math Writing	1.5	.83 (.39)	1.96 (1.45)	.13	.61	11	.555
Total	16.5	7.71 (1.86)	11.25 (2.38)	3.54	4.68	11	.001

Table 2: Mastery of Targeted Content

Units	Possible Points	Pretest Median	Posttest Median	Average Mastery at Posttest
Geometry	8	5.00	7.00	88%
Measurement	7	2.50	3.50	50%
Math Writing	1.5	1.00	1.00	67%
Total	16.5	7.75	11.00	67%

Attachment E

M² Student Mathematics Performance: 1st Grade

Summative Evaluation

Submitted to: Dr. M. Katherine Gavin, Project Director
Submitted by: Dr. Susan Carroll, Evaluation Consultant

Submitted: July 2010

Methodology

The evaluation design consisted of multiple levels of performance testing to establish project efficacy. To address this design feature, there were two major research questions.

Empirical Questions - Research Hypotheses

Research Question #1:

Is there an *increase in mathematics achievement* for the Intervention group of 1st grade students across all socioeconomic and ethnic backgrounds, after exposure to an intervention model that provides challenging standards-based curriculum and encourages high level discourse?

Research Question #2:

Is there a *difference in mathematics achievement* between the Intervention group of 1st grade students, who are exposed to the mathematics curriculum model, and a Comparison group of students of similar socioeconomic and ethnic backgrounds who participated in the regular math curriculum?

Data Collection Methods and Instrumentation

The 1st grade students in both the Intervention and Comparison groups were tested *before* **Project M²** began [PRE] and at the *conclusion* of the intervention [POST]. This corresponded to the beginning of the school year and the ending of the school year. There were corroborative methods of instrumentation to respond to the previously stated empirical research questions related to mathematics achievement.

- The first tool was a standardized measure, the *Iowa Tests of Basic Skills (ITBS).* Concepts and Estimation was the subtest /scale related to the mathematics achievement targeted in the project.

- There were performance-based measurement tools developed by **Project M²** staff in order to address content appropriate to the age of the students targeted for the intervention. These were *Open Response Assessments Total Score* with subscales in *Geometry* and *Measurement*.

Statistical Results

Research Question #1:

Is there an *increase in mathematics achievement* for the Intervention group of 1st grade students across all socioeconomic and ethnic backgrounds, after exposure to an intervention model that provides challenging standards-based curriculum and encourages high level discourse?

For **Research Question #1,** a pre and post statistical analysis was undertaken using paired or correlated t-tests. The results documented project success. There were statistically significant gains for the Intervention group from pre to posttesting on each of the four mathematics performance indicators.

- ✓ The ITBS scores rose 17.18 points from a pretest score of 133.90 to a posttest score of 151.08.
- ✓ The Open Response Total score rose 10.09 points from a pretest score of 4.87 to a posttest score of 14.96
- ✓ The Geometry subscale and Measurement scales scores, which compose the Open Response Total Score, also made statistically significant gains in an upward direction.

Please refer to Table 1 for the results of the Intervention Group on the pre to post mathematics achievement indicators for the Intervention group.

Table 1

Pretest to Posttest Gains for 1st Grade Intervention Group on Mathematics Achievement Indicators

Mathematics Measures *n=186*	*Pre M (SD)*	*Post M (SD)*	*Gain*	*t value*	*df*	*p*
Iowa Test of Basic Skills (ITBS)	133.90 (9.66)	151.08 (15.15)	+17.18	20.61	185	***
Open Response Total (ORT)	4.87 (2.41)	14.96 (3.71)	+10.09	40.22	185	***
Open Response Geometry (ORG)	2.93 (1.56)	9.55 (2.55)	+6.62	36.83	185	***
Open Response Measurement (ORM)	1.94 (1.44)	5.41 (1.74)	+3.47	23.44	185	***

*** $p<.001$

Research Question #2:

Is there a *difference in mathematics achievement* between the Intervention group of 1st grade students, who are exposed to the mathematics curriculum model, and a Comparison group of students of similar socioeconomic and ethnic backgrounds who participated in the regular math curriculum?

To investigate the differences in mathematics achievement between the Intervention and Comparison groups, a series of 2-level multilevel models using hierarchical linear modeling HLM version 6.06 (Raudenbush, Bryk, Cheong, Congdon, & du Toit, 2004) were employed. Specifically, four HLM procedures were executed. The respective dependent or outcome variables were the mathematics subscale score on the ITBS, the Total Score on the Open Response assessment, and the two subscale scores on the Open Response assessment that represented the Geometry and Measurement constructs. Table 2 reports the descriptive statistics for the dependent or outcome variables for the both the Intervention and Comparison groups.

Table 2
Descriptive Statistics for Intervention and Comparison Groups on the Dependent or Outcome Variables

Mathematics Measures	Intervention (n=186) *Post M (SD)*	Comparison (n=174) *Post M (SD*
Iowa Test of Basic Skills (ITBS)	151.08 (15.15)	149.41 (14.21)
Open Response Total (ORT)	14.96 (3.71)	8.74 (2.85)
Open Response Geometry (ORG)	9.55 (2.55)	5.67 (1.89)
Open Response Measurement (ORM)	5.41 (1.74)	3.06 (1.66)

Evaluation data were collected at the student level with concomitant goal of testing classroom-level effects. A major advantage of the HLM procedure is that it addresses the fact that students "nested" in the same classes with the same teacher are interdependent versus independent.

Level 1 of the HLM consisted of the mathematical outcome scores for the students along with the corresponding pretest scores to control for students' prior knowledge. The independent variable, and the focus of the null hypothesis to be tested, was group membership in Project M². The research hypothesis tested whether there was a differential effect on mathematics achievement for students in the Intervention group who were exposed to the Project M² curriculum as opposed to students in the Comparison group who experienced the traditional mathematics curriculum. At Level 2, dummy coding was employed with the

Intervention group coded as 1 and the Comparison group coded as 0. The restricted maximum likelihood estimation (REML) for HLM analyses was employed to address the Level 2 sample size.

Prior to conducting the HLM analyses for the four outcome variables, preliminary statistical tests to meet the requisite assumption of HLM were undertaken. A test for homogeneity of Level 1 variances was tested with the alpha level set at a conservative level of .02. Level 1 descriptive statistics on skewness and kurtosis were also generated. Table 3 reports these results. The ITBS scores and the Measurement scale of the Open Response achieved homogeneous Level 1 variances but the Geometry Scale of the Open Response and Total Open Response scores did not. Therefore, the assumption of homogeneity of variances was rejected. The robust standard errors for the statistical analyses were chosen for use in the HLM statistical procedure, as they are able to more fully sustain the violations of homogeneity than are the conventional standard errors.

Table 3
Level 1 Variables and Homogeneity of Variance Results

	ITBS	Total OR	Geometry OR	Measurement OR
Distribution of Outcome Scores				
Skewness	.26	.11	.10	.14
Kurtosis	-.49	-.79	-.91	-.94
Homogeneity of Variance				
χ^2 (df)	30.93 (23)	40.45 (23)	44.69 (23)	22.89 (23)
p	.12	<.01	<.01	>.50

Null or Unconditional Model for the Outcome Variables

The first step in the HLM procedures involved estimating the null or baseline models for each of the four outcome measures where there are no predictors at either Level 1 or Level 11. Each model's intraclass correlation (ICC) is estimated. The ICC is a measure of the proportion of variance between classes in relation to the total variance.

For the ITBS, the ICC was .18 indicating that 18% of the variance in the ITBS posttest scores lay between classes and 82%, within classes. The ICC for the Open Response assessment was somewhat more variable. The ICC for the Total Score was .56, indicating that 56% of the variance in the Open Response posttest scores lay between classes and 44%, within classes. The Geometry subscale score ICC was .54 and the Measurement subscale score ICC was .39, respectively.

The results of the null or baseline models for each of the four outcome scores are in Tables 4 through 7.

Random Coefficients Model

Random assignment of teachers and their classrooms was utilized in the design of the study. Because random assignment of students to groups was not possible, it was important to be sure that both the Intervention and Comparison groups began at the same starting position with respect to math achievement. To ensure that group equivalence existed, the students' pretest scores were included in the statistical analysis as a Level I covariates, and grand mean centered. This would account for any pretreatment differences in the Intervention and Comparison groups.

In fact, each of the grand mean-centered pretest scores were significant predictors of the posttest scores. Specifically, The ITBS pretest (γ_{10}) the parameter estimate was .91 with a SE of .05 (p<.001). For the Open Response Total Scores pretest (γ_{10}), the parameter estimate was .60 with an SE of .05 (p<.001). The Open Response Geometry subscale pretest (γ_{10}) parameter estimate was .42 with an SE of .06 (p<.001) and the Open Response Measurement subscale pretest (γ_{10}) parameter estimate was .28 with a SE of .07 (p<.001).

For the ITBS scores, the pretest score accounted for 42% of the variance within classes. There was no statistically significant variance in pretest slope (τ_{11})=.01, χ^2 = 21.23, p>.50). The Open Response Total pretest score explained 19% of the within class variance with no statistically significant variance in pretest slope (τ_{11})=.00, χ^2 = 13.70, p>.50). The Open Response Geometry subscale pretest score explained 7% of the within class variance with no statistically significant variance in pretest slope (τ_{11})=.00, χ^2 = 17.11, p>.50). The Open Response Measurement subscale pretest score explained 5% of the within class variance with no statistically significant variance in pretest slope (τ_{11})=.01, χ^2 = 28.68, p=.19). For each of these variables, the variance component was not allowed to randomly vary.

The results for the Random Coefficients Models for each of the four outcome scores are in Tables 4 through 7.

Full or Contextual Model

The final step in the HLM was estimating the full Level 2 models comparing the mathematics achievement of the Project M² Intervention and Comparison groups. The null hypothesis tested if there were dif-

ferences between the Intervention and Comparison groups in mathematics achievement after accounting for pretest scores. The coefficient for Project M² status in the Intervention group was γ_{01} while membership in the Comparison group was γ_{00}; the coefficients can be used to determine the predicted scores on each of the mathematics measurements in the study.

For the 1st grade students in project M², the predicted posttest ITBS for the Comparison group was 148.53 (γ_{00}) while that of the Intervention group (γ_{01}) was 151.69 (148.53+3.16). This was not a statistically significant difference (γ_{01})3.16, t=1.61, p=.12). The conclusion is that there were no differences in mathematics achievement on the ITBS between the Intervention and Comparison groups after controlling for pretest scores. However, the predicted score was slightly higher for the Intervention group. Please refer to Table 4.

A different finding was true on the Open Response assessments for the Total Score and both of the subscales, Geometry and Measurement. The students who were exposed to the M² mathematics curriculum outperformed their peers who had received the traditional mathematics curriculum.

For the 1st grade students, the predicted Open Response Total Score for the Comparison group was 8.66 (γ_{00}) while that of the Intervention group was 14.97 (8.66+6.31). This was a statistically significant difference (γ_{01} 6.31, t=11.17, p<.001). The Open Response Geometry subscale score was predicted to be 5.69 for the Comparison group while the Intervention group's was 9.53 (5.69+3.84). This was statistically significant (γ_{01} 3.84, t=9.06, p<.001). The Open Response Measurement subscale score was predicted to be 3.03 for the Comparison group while the Intervention group's was 5.39 (3.03+2.36). This was statistically significant (γ_{01} 2.36, t=10.58, p<.001). Please refer to Table 4 through 7.

Table 4
Summary of the REML Parameter Estimates for the Two Level Model: ITBS

Parameter	Unconditional Model Parameter Estimate	SE	Random Coefficients Model Parameter Estimate	SE	Contextual Model Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	149.97***	1.45	150.12***	1.04	148.53 ***	1.18
M2 Intervention (γ_{01})					3.16	1.96
Pretest (γ_{10})			.91***	.05	.91***	.05
Variance estimate						
Level 1 Variance (σ^2)	177.55		103.75		103.78	
Intercept variance (τ_{00})	40.25 ***		19.89***		18.25***	
Deviance & (Number of REML parameters)	2916.07 (2)		2725.03 (2)		2717.60 (2)	

REML= Restricted Likelihood Estimation
Pretest scores were grand-mean centered.
*p<.05 ** p<.01 *** p<.001

Table 5
Summary of the REML Parameter Estimates for the Two Level Mode: OPEN RESPONSE TOTAL SCORE

Parameter	Unconditional Model Parameter Estimate	SE	Random Coefficients Model Parameter Estimate	SE	Contextual Model Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	11.79***	.70	11.82***	.70	8.66 ***	.27
M2 Intervention (γ_{01})					6.31***	.57
Pretest (γ_{10})			.60***	.05	.60***	.05
Variance estimate						
Level 1 Variance (σ^2)	9.21		7.45		7.45	
Intercept variance (τ_{00})	11.79 ***		11.86***		1.62***	
Deviance & Number of REML parameter	1888.86 (2)		1821.78 (2)		1777.69 (2)	

REML= Restricted Likelihood Estimation
Pretest scores were grand-mean centered.
*p<.05 ** p<.01 *** p<.001

Table 6
Summary of the REML Parameter Estimates for the Two Level Model: OPEN RESPONSE GEOMETRY SUBS-CALE

Parameter	Unconditional Model Parameter Estimate	SE	Random Coefficients Model Parameter Estimate	SE	Contextual Model Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	7.60***	.45	7.61***	.45	5.69 ***	.19
M2 Intervention (γ_{01})					3.84***	.42
Pretest (γ_{10})			.42***	.06	.42***	.06
Variance estimate						
Level 1 Variance (σ^2)	4.15		3.85		3.85	
Intercept variance (τ_{00})	4.78 ***		4.71***		.93***	
Deviance & Number of REML parameter	1600.45 (2)		1578.74 (2)		1542.93 (2)	

REML= Restricted Likelihood Estimation
Pretest scores were grand-mean centered.
*p<.05 ** p<.01 *** p<.001

Table 7
Summary of the REML Parameter Estimates for the Two Level Model: OPEN RESPONSE MEASUREMENT SUBSCALE

Parameter	Unconditional Model Parameter Estimate	SE	Random Coefficients Model Parameter Estimate	SE	Contextual Model Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	4.19***	.27	4.20***	.27	3.03 ***	.14
M2 Intervention (Y γ_{01})					2.36***	.22
Pretest (γ_{10})			.28***	.07	.31***	.06
Variance estimate						
Level 1 Variance (σ^2)	2.64		2.52		2.52	
Intercept variance (τ_{00})	1.68 ***		1.60***		.16**	
Deviance & Number of REML parameter	1425.84 (2)		1412.53 (2)		1372.45 (2)	

REML= Restricted Likelihood Estimation
Pretest scores were grand-mean centered.
*p<.05 ** p<.01 *** p<.001

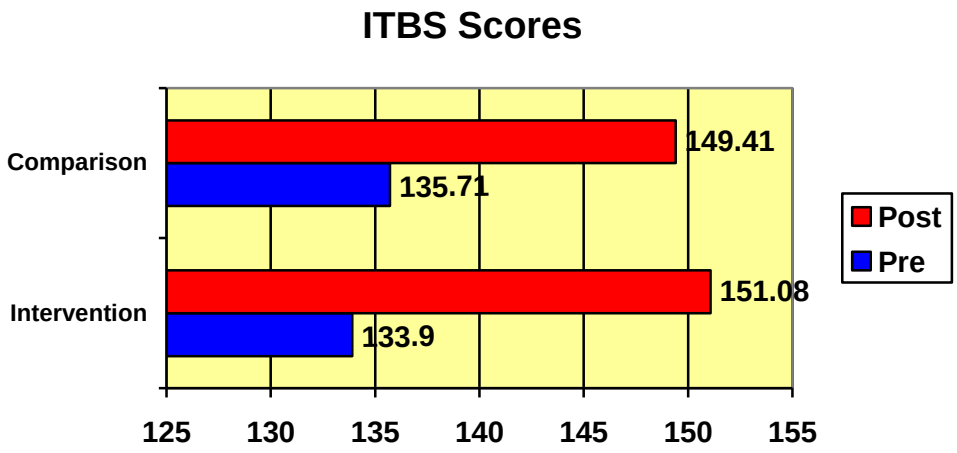
In summary, there were statistically significant differences between the Intervention and Comparison groups on the Total Open Response scores and both of its component scales of Geometry and Measurement. There differences were in favor of the Intervention group which outperformed the Comparison group, statistically equated by the inclusion of covariate pretest scores. The *Cohen d* statistics were very impressive ranging from 1.38 for Measurement, to 1.73 for Geometry, with the Total score at 1.88. The ITBS performance between the Intervention and Comparison groups was not statistically different - although the Intervention group had a slightly higher predicted score at posttesting.

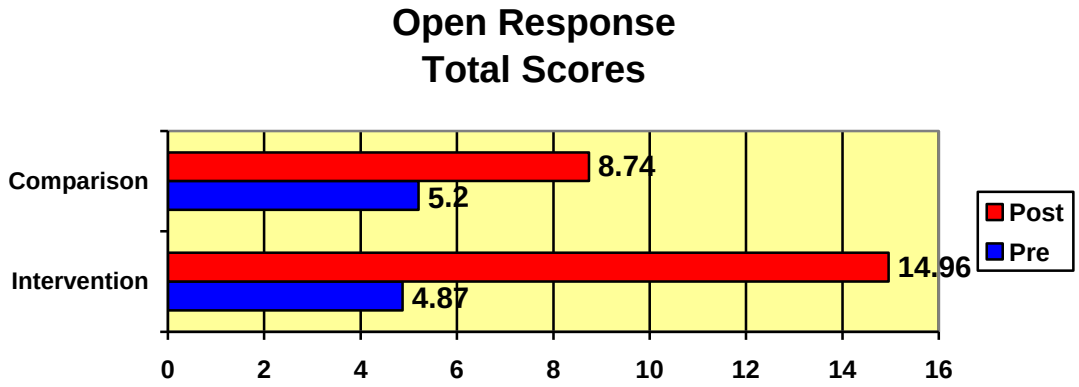
The HLM summary data, as well as the *Cohen d* effect sizes, are found in the Summary Table 8.

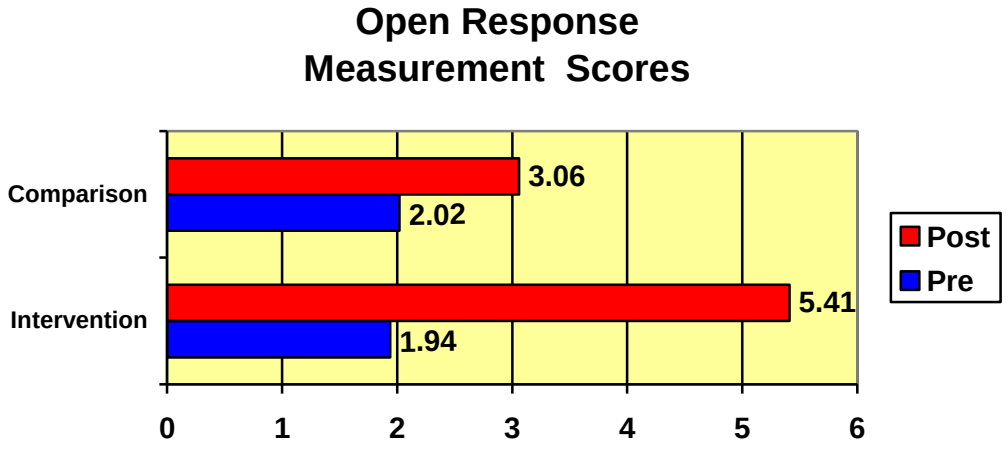
Table 8
Summary of Outcome Variables, Intervention Group Differential and Effect Size

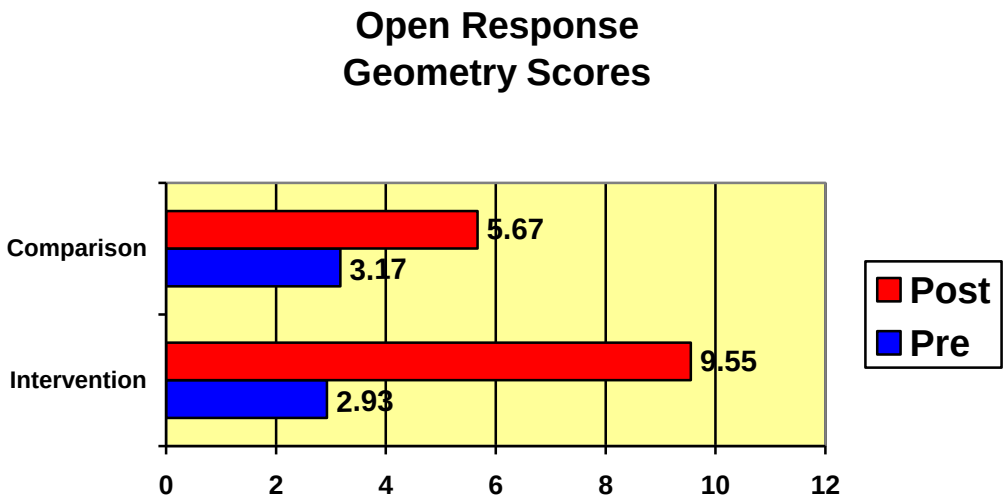
Outcome Variables	Coefficient[a] (SE)	t (df)	*p*	*d*
ITBS	3.16 (1.96)	1.61 (22)	.12	.11
Open Response Total Score	6.31 (.57)	11.17 (22)	<.001	1.88
Geometry	3.84 (.42)	9.06 (22)	<.001	1.73
Measurement	2.36 (.22)	10.58 (22)	<.001	1.38

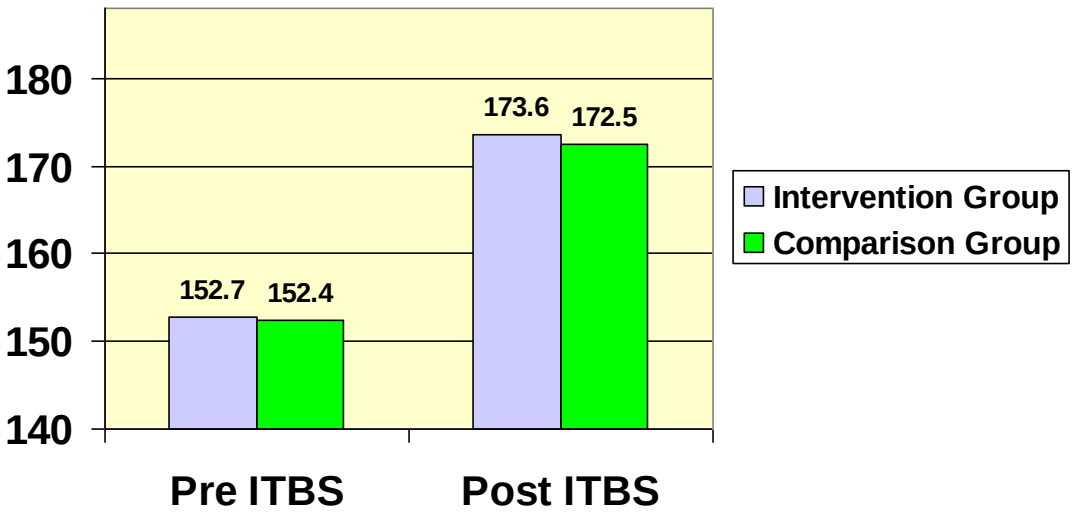
[a]This is the differential for the Intervention group's gain over the Comparison group after adjusting for the pretest scores.


ITBS Scores
Comparison
Intervention
149.41
135.71
151.08
133.9
Post
Pre
125
130
135
140
145
150
155


Open Response
Total Scores
Comparison
Intervention
8.74
5.2
14.96
4.87
Post
Pre
0
2
4
6
8
10
12
14
16


Open Response
Measurement Scores
Comparison
Intervention
3.06
2.02
5.41
1.94
Post
Pre
0
1
2
3
4
5
6


Open Response
Geometry Scores
Comparison
Intervention
5.67
3.17
9.55
2.93
0
2
4
6
8
10
12
Post
Pre


Pre to Post ITBS Math Scores
180
170
160
150
140
152.7
152.4
173.6
172.5
Intervention Group
Comparison Group
Pre ITBS
Post ITBS

Pre to Post Open Response Scores