Project M² Evaluation Reports for The National Science Foundation (NSF)
-Executive Summary-

There were several program evaluation components implemented by Words & Numbers Research, Inc. Attachments will provide the full reports. A summary of the key findings documents a highly effective Project M².

Grade 1 - Demographic and Pretest Statistical Profile - Attachment A (p. 3)

There were 202 (51%) students in the Intervention group and 193 students (49%) in the Comparison group. Elementary schools were located in Connecticut (5), Texas (2), South Carolina (2) and Kentucky (2).

- There were approximately 57% male and 43% female students.
- The ethnicities were Asian (10%), African American (18%), Hispanic (18%), and Caucasian (53%).
- Almost half (48%) of the students were eligible for a meal subsidy.
- Seven percent were identified as receiving Special Education services.
- Fifteen 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 Grade 1 students.

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

On the Measurement Unit, 2nd graders achieved statistically significant ($p<.001$) gains from pre to posttesting on Symmetry, Transformations, 3-D shapes, 3-D faces, Perspective, and Composing items. Ninety-nine percent made gains on Total scores from pre to posttesting.

On the Geometry Unit, 1st graders achieved statistically significant ($p<.001$) gains from pre to posttesting on Pentagon, Hexagon, Carroll Diagram , Right Angle, Venn Diagram, Decomposition, Symmetry, Congruence and Attributes. Every 1st grade student (100%) made gains on Total Scores from pre to posttesting.

Grade 1 Teacher Professional Development Assessment - Attachment C (p. 12)

Professional development was measured both pre and post training. Both quantitative and qualitative data reflected the professional development as excellent, according to 1st grade teachers during the summer 2009. 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.

Pre and Post Mathematical Content Acquisition for Grade 1 Teacher - Attachment D (p. 18)

The mastery of content from pre to posttesting was very satisfactory for the 1st grade teachers and there were statistically significant gains. The content in Measurement and Geometry was mastered to a great extent with 9.50 out of a total of 13 points.

Pre and Post Mathematical Performance by Intervention and Comparison Groups - Attachment E (p. 19)

There were two research questions empirically addressed in the evaluation research during PY3.

> **Research Question #1**: Is there an *increase in mathematics achievement* for the Intervention group of 2nd grade students across all socioeconomic and ethnic backgrounds after exposure to the advanced mathematics 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 2nd grade students who are exposed to the advanced mathematics model and the Comparison group of 2nd grade students who undertake the traditional mathematics 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 2nd 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, the return on the NSF investment was impressive.

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

Words & Numbers Research, Inc.
March 18, 2010

Attachment A

Grade 1 Students
Demographic and Pretest Statistical Profile

Submitted to: Dr. M. Katherine Gavin, Project Director
Submitted by: Dr. Susan Carroll, Evaluation Consultant
Submitted on: February 4, 2010

Background on the Student Data Collected in Program Year 3 (PY3)

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 Test 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 35 items on the ITBS. So the percentage was calculated based on the number of items out of 35 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 8. For Geometry the range of possible scores was 0 to 13. A Total Score was calculated as well. The range of scores was 0 to 21.

Demographic Data

- Approximately 395 students in the 1st grade were part of the **M²** project during the 2009-2010 school year. Of those, 193 (49%) were in the Comparison group and 202 (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	30	14	47%	16	53%
Carleston, TX	38	18	47%	20	53%
Charter Oak, CT	35	19	54%	16	46%
Goodwin, CT	34	17	50%	17	50%
Lawhon, TX	29	13	45%	16	55%
Lincoln, KY	41	19	46%	22	54%
Midland Park, SC	32	16	50%	16	50%
Noah Webster, CT	36	15	42%	21	58%
Southeast, CT	27	14	52%	13	48%
Southern, KY	62	33	53%	29	47%
Vinton, CT	31	15	48%	16	52%
Total	**395**	**193**	**49%**	**202**	**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 224 (57%) boys and 171 (43%) girls. There were no statistically significant differences in gender between the Intervention and Comparison groups.

Gender	Total	Intervention	Comparison
Males	224 (57%)	121 (60%)	103 (53%)
Females	171 (43%)	81 (40%)	90 (47%)

(Chi sq=1.72, df=1, p=.19)

- The ethnicities represented were Asian American (10%), African American (18%), Hispanic (18%), Caucasian (53%) and others (1%) including three Native American students. There were no statistically significant differences on ethnicity/ race between the Intervention and Comparison groups.

Gender	Total	Intervention	Comparison
Native American	3 (.5%)	3(1%)	--
Asian	39 (10%)	20 (10%)	19 (10%)
Black	70 (18%)	38 (19%)	32 (16%)
Hispanic	70 (18%)	34 (17%)	36 (19%)
White	209 (53%)	104 (52%)	105 (54%)
Other	4 (.5%)	3 (1%)	1 (.5%)

(Chi sq=4.40, df=5, p=.49)

- Almost half (48%) 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	174 (49%)	80 (44%)	94 (53%)
No	185 (51%)	101(56%)	84 (47%)

(Chi sq=2.66, df=1, p=.10)

- Seven 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	27 (7%)	10 (5%)	17 (9%)
No	368 (93%)	192 (95%)	176 (91%)

(Chi sq=2.31, df=1, p=.13)

- Fifteen percent of the students were in ESL or ELL programs. There were statistically significant differences between the Intervention and Comparison groups on proportions of ELL or EL students.

There were significantly more ESL or ELL students in the Comparison group. On school, Lawton had 100% of its Comparison group in the ESL/ELL program. [This did not appear to influence the group equivalence at pretesting as the ITBS and Open Response results will document.]

ELL or EL Participant	Total	Intervention	Comparison
Yes	58 (15%)	21 (10%)	37 (19%)
No	336 (85%)	181 (90%)	155 (81%)

(Chi sq=6.18, df=1, p=.013)

School Participation in ESL/ ELL	Comparison N	Comparison %
Corcoran, SC	0	0%
Carleston, TX	8	44%
Charter Oak, CT	6	32%
Goodwin, CT	0	0%
Lawhon, TX	13	100%
Lincoln, KY	0	0%
Midland Park, SC	8	50%
Noah Webster, CT	1	7%
Southeast, CT	1	7%
Southern, KY	0	0%
Vinton, CT	0	0%

- Two percent reported participating in another math program besides the current one. For 98% **M²** was the only math program that they were participating in. There were no statistically significant differences between the Intervention and Comparison groups.

ELL or EL Participant	Total	Intervention	Comparison
Yes	8 (2%)	6 (3%)	2 (1%)
No	387 (98%)	196 (97%)	191 (99%)

(Chi sq=1.86, df=1, p=.17)

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 56% 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 1st Graders

PRETEST SCORES	Intervention	Comparison			
	Mean (SD)	Mean (SD)	t	df	p
ITBS Standard Score (Highest possible score-188)	133.84 (9.56)	135.33 (10.99)	1.43	386	.16
ITBS Percent correct	56% (.15)	57% (.16)	1.10	386	.27
Measurement OR Scale (Highest possible score-8)	1.91 (1.40)	1.95 (1.45)	.23	389	.82
Geometry OR Scale (Highest possible score-13)	2.92 (1.55)	3.14 (1.58)	1.32	389	.19
Total Score OR Scale (Highest possible score-21)	4.85 (2.39)	5.09 (2.45)	.99	389	.32

Attachment B
Measurement Unit Pre and Post Test Data Analysis

The Measurement Unit: Grade 2

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

Submitted by: Susan R. Carroll, Evaluator
Words & Numbers Research, Inc.
July 14, 2009

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

- There were statistically significant gains from pre to posttesting on **Length**.
- There were statistically significant gains from pre to posttesting on **Measuring Tool**.
- There were statistically significant gains from pre to posttesting on **Finding Area**.
- There were statistically significant gains from pre to posttesting on **Ordering Area**.
- There were statistically significant gains from pre to posttesting on **Capacity**.
- There were statistically significant gains from pre to posttesting on the **Total Scores**.

These findings were true for the 187 students with complete pre and posttest sets.

Table 1 documents the statistical outcomes by item and for Total Scores. The effect size was calculated for Cohen *d* at 2.90. Additionally, Table 2 shows the percentage of students whose scores increased or made gains from pre to posttesting. It is the majority in all cases.

Overall, there are impressive gains achieved with this unit, *Measurement*. Almost every student (99%) made gains on Total Scores from pre to posttesting. The content for the five items was very challenging. This unit did what it was supposed to do - offer advanced mathematical challenge to the young mathematicians it targeted.

Table 1: Correlated t-test Results
All Participants (N=187)

Measurement	Pre Mean	Post Mean	Mean Difference	t val-ue	df	*p*
Length	.69	2.38	1.69	18.38	186	***
Measuring tool	.57	2.90	2.34	22.79	186	***
Finding area	.24	2.60	2.36	20.99	186	***
Ordering area	1.13	3.40	2.27	21.65	186	***
Capacity	.38	1.30	.92	10.22	186	***
Total	**3.01**	**12.58**	**9.58**	**33.69**	186	*******

*** p <.001

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

All Participants (N=187)

Measurement	N	Gains %	Loss %	No change %
Length	187	83%	5%	12%
Measuring tool	187	86%	2%	12%
Finding area	187	84%	2%	14%
Ordering area	187	87%	2%	11%
Capacity	187	58%	11%	31%
Total	187	99%	0%	1%

Geometry Unit Pre and Post Test Data Analysis

The Geometry Unit: Grade 1

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

Submitted by: Susan R. Carroll, Evaluator
Words & Numbers Research, Inc.
February 1, 2010

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

- There were statistically significant gains from pre to posttesting on **Pentagon** (1 point).
- There were statistically significant gains from pre to posttesting on **Hexagon** (4 points)
- There were statistically significant gains from pre to posttesting on **Carroll Diagram** (3 points)
- There were statistically significant gains from pre to posttesting on **Right Angle** (2 points)
- There were statistically significant gains from pre to posttesting on **Venn Diagram** (3 points)
- There were statistically significant gains from pre to posttesting on **Decomposition (**3 points)
- There were statistically significant gains from pre to posttesting on **Symmetry** (3 points)
- There were statistically significant gains from pre to posttesting on **Congruence** (1 points)
- There were statistically significant gains from pre to posttesting on **Attributes** (4 points)
- There were statistically significant gains from pre to posttesting on the **Total Scores** (24 points).

These findings were true for the 202 students with complete pre and posttest sets for the nine item test.

Table 1 documents the statistical outcomes by item and for Total Scores. The effect size was calculated for Cohen d at 4.08. Additionally, Table 2 shows the percentage of students whose scores increased or made gains from pre to posttesting. It is the majority in all cases.

Overall, there are impressive gains achieved with this unit, *Geometry*. Every 1st grade student (100%) made gains on Total Scores from pre to posttesting. The content for the nine items was very challenging and put the mathematical skills of the young students to the test. They performed with outstanding results.

Table 1: Correlated t-test Results
All Participants (N=202)

Geometry	Points	Pre Mean	Post Mean	Mean Difference	t val-ue	df	*p*
Pentagon	1	.09	.81	+.72	21.45	200	***
Hexagon	4	1.17	3.61	+2.44	41.46	200	***
Carroll Diagram	3	1.31	2.51	+1.20	17.81	200	***
Right Angle	2	.37	1.60	+1.23	19.20	200	***
Venn Diagram	3	.09	2.11	+2.02	30.33	200	***
Decomposition	3	.52	2.07	+1.56	23.50	200	***
Symmetry	3	.49	2.50	+2.01	26.34	200	***
Congruence	1	.29	.93	+.63	16.82	200	***
Attributes	4	.59	2.93	+2.34	23.90	200	***
Total	**24**	**4.91**	**19.06**	**+14.15**	**54.99**	**201**	*******

*** p <.000

Table 2: Gains, Losses and No Changes from Pre to Post Testing
All Participants (N=202)

Geometry	N	Gains %	Loss %	No change %
Pentagon	201	73%	1%	26%
Hexagon	201	98%	0%	2%
Carroll Diagram	201	81%	7%	12%
Right Angle	201	78%	3%	19%
Venn Diagram	201	93%	0%	7%
Decomposition	201	89%	4%	7%
Symmetry	201	93%	3%	4%
Congruence	201	66%	2%	32%
Attributes	201	89%	3%	8%
Total	**202**	**100%**	**0%**	**0%**

Attachment C

Pre and Post M² Teacher Professional Development: Assessment of the Summer Institute 2009

In order to prepare the twelve 1st grade teachers for the actual implementation of the **M²** project, teacher training was a key component. The training session was conducted during July 2009 on the campus of the University of Connecticut. Because of its importance, the training was evaluated from three vantage points:

- Satisfaction with the training implementation
- Content and skill acquisition from pre to post training
- Background of participating teachers

With helpful input from project staff, Words & Numbers Research, Inc. revised the two instruments used in previous teacher evaluations.

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 2009 training period. Each was rated on a three point rating scale:

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

The results were exemplary, 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 teach-

ers. Furthermore, 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. There was not one single rating of <u>not</u> *satisfactory*. The training was a job well done in the eyes of participants. Please see Table 1.

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 trainers- execution of objectives	100%	---	---
3. Ability of presenters to provide concrete examples to illustrate mathematics	100%	---	---
4. The quantity and depth of content covered in the training	100%	---	---
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 (8:30-3.30)	92%	8%	---
8. The logistics - comfort of rooms, location, equip, refreshments	100%	---	---
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 comments were offered on the aspects of training.

Verbatim Comments

- ✓ *I was very happy to have been made to feel so welcome. Everyone listens to what you contribute to discussion and responses make you feel that they really think of you as a contributing partner. Everyone has been friendly and supportive. I love that they did not make you feel inferior for not already doing these strategies in your room. I am very happy to have been here.*

- ✓ *I loved the concrete examples. I liked how we got to do activities that we are going to teach and how you demonstrate the Talk Moves and Think Alouds.*

- ✓ *Great. Nothing was disappointing!*

- ✓ *Really excited to implement the program in addition -very excited materials will be prepared.*

- ✓ *The professors were great! It was a very pleasant conference. Thank you.*

- ✓ *I am excited to implement the units this school year. The students are going to absolutely have a blast with math. I love the idea of the Think Deeply questions being tied to writing. What made the training this week interesting and exciting were the hands-on activities and the different speakers/presenters for the chapters. It kept our attention.*

- ✓ *I especially appreciated having different speakers handling different skills, topics, etc. Also, having a teacher come in who has actually taught the material along with the video was great. I think it would be helpful to see more examples like the video so we could watch children performing and teachers teaching in the classroom. I was concerned that the days would be long, but the way you organized them made the time pass quickly. I found it helpful to be able to use the materials we will use with our students.*

- ✓ *The video helped me a lot in seeing how to implement these ideas. Overall, I enjoyed the classes and learned a lot while I was here.*

- ✓ *This was an exceptional workshop. Kudos to the presenters as well as the content they have developed. The presentation of the material was very clear. And concise. I feel highly confident about presenting these units. Thankfully, additional training will be done in the early part of the school year in regard to the writing component - that is sure to boost my confidence with that component of the program.*

- ✓ *I really enjoyed the friendly atmosphere and the use of tools for demonstrations of the lessons.*

A couple of teachers voiced suggestions.

- ✓ *More video or deeper review of each chapter would be helpful.*

- ✓ *I would like to see email set up for students to use to email Zani, Obbo and Imi. I think a little bio sheet on each member of the group (email address, phone numbers, etc.) would be great.*

2.00 Pre and Post Content /Skill Acquisition

There were nineteen 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 19 items on a five-point rating scale. This was undertaken before training began (the pretest) and after training ended (the post-test). 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. Teaching students to describe shapes using their attributes and properties	4.33	4.92	.58	3.92	11	.**
2. Teaching students how to read a scale to compare weight	4.00	4.67	.67	2.60	11	*
3. Modeling mathematically valid writing with students	3.00	4.25	1.25	4.49	11	**
4. Teaching students to repeat what another student has stated	3.67	4.67	1.00	5.75	11	**
5. Teaching students how to use a Venn Diagram to sort shapes	4.17	5.00	.83	4.02	11	**
6. Integrating verbal discourse into math classes	3.42	4.17	.75	2.69	11	*
7. Helping students compose and decompose geometric shapes	3.27	4.64	1.36	4.40	11	**
8. Understanding how to order objects by weight using transitivity	2.58	4.50	1.91	9.93	10	**
9. Encouraging students to build on what others say by adding on ideas beyond the initial solution to a problem	3.33	4.50	1.17	3.92	11	**
10. Understand how to integrate writing into math classes	2.92	4.17	1.25	4.49	11	**
11. Teaching students to identify right, acute, and obtuse angles	3.00	4.67	1.67	5.38	11	**
12. Supporting students in explaining their mathematical reasoning in writing	2.67	4.00	1.33	4.30	11	**
13. Teaching students to find lines of symmetry using paper folding and mirrors	3.50	4.67	1.17	3.63	11	**
14. Using tree diagrams to sort shapes	2.42	4.25	1.83	5.01	11	**
15. Teaching students to identify and describe congruent figures	2.92	4.83	1.92	6.13	11	**
16. Setting up an environment for learning that promotes listening and sharing ideas with one another	3.58	4.58	1.00	3.63	11	**
17. Teaching students how to agree or disagree with others' reasoning verbally and tell why	3.33	4.42	1.08	4.17	11	**
18. Recording students' ideas to help them write about mathematics	3.25	4.42	1.17	3.92	11	**
19. Using Carroll diagrams to sort shapes	1.92	4.50	2.58	7.22	11	**

** $p < .01$

* $p < .05$

The gains from pre test to posttest means were impressive and ranged from 2.58 to .58. 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 12. The range was a low of 3 years to a high of 27 years. In terms of years in their current position, the average was six years although the range was from 2 to 20 years.

- The majority (84%) 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. Eleven (91%) reported having some **professional development related to mathematics** during the school year. Six reported having one PD session a year in the field of mathematics; three reported two to three times. Two teachers had PD related to math several times a year.

- Given those findings, it was important to determine how they felt about teaching mathematics in their classrooms. 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 and median scores were seven, signifying a definite preference for teaching math. They were asked to assess their **background knowledge in math** and the mean and median were 6.5, moderately strong. Finally, the teachers were asked to evaluate their **comfort level** with teaching mathematics. The mean and median scores were 7.0, suggesting a comfort level with the content.

- Next, teachers were asked how often they used **math manipulatives** in their classrooms. All (100%) claimed that they used them *often*. Some of those listed by teachers included: pattern blocks, unifix cubes, 3D blocks, coins, clocks, geoboards, macaroni, buttons, beads, tangrams, thermometers, polydrons, rules, scales, counters, links, dice, playing cards, tape measures, calendars, attribute blocks, sea creatures, straws, value mats and others.

- Teachers were also asked how often they **differentiated instruction** when students needed **more support**. The majority reported either *often* (50%) or *sometimes* (42%). Examples included: small group instruction, computer programs, self made games, one on one activities, work stations, paraprofessional support, peer support or support from parents, volunteers or pupil personnel.

- 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* (33%) or *sometimes* (50%). Examples included: problem solving cards, strategic games, math puzzles, mind benders, computer games, small group work, individualized instruction, flash cards, and collaboration with 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 (4)
Bridges (3)
Trailblazers (2)
Harcourt
HSB
Houghton Mifflin

4.00 Summary

The evaluation results for the **M²** professional development training during July 2009 at the University of Connecticut were outstanding. The data document a successful training experience for the participating 1st grade teachers. Not only were logistics and training sessions implemented well, but also the targeted content was delivered effectively. The training significantly increased the confidence levels of teachers regarding the knowledge and skills that are critical to successful Project **M²** implementation.

Attachment D
M2 Pre and Post Content Acquisition Results for Teachers

October 15, 2009

Approximately, twelve complete sets of pre and post questions were analyzed for teachers in the M² project. This was a retrospective analysis. Pretest data were collected during the summer training in 2009. Posttest data were collected after each unit was completed. Although the sample size was small, correlated t-tests were applied to the data to determine if there were statistically significant gains on the three units of content.

There were statistically significant gains from pre to post testing on two units, the **Geometry** and **Measurement** units. The gains in content acquisition from pre to posttesting were impressive for both Geometry (+4.12) and Measurement (+1.54).

The mastery of content at posttesting was very satisfactory for Measurement (90%). The median was 4.50 out of a total of 5 points. Geometry (63%) showed room for additional content acquisition for the teachers; the median was 5.00 out of 8 possible points. Overall, the mastery at posttest was 73% of the total content with 9.50 out of a total of 13 points. The content clearly was challenging for the teachers who participated in the project, M². Please 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	.88 (.86)	5.00 (.71)	4.12	13.94	11	.000
Measurement	5	2.67 (.58)	4.21 (1.29)	1.54	5.17	11	.000
Total	13	3.54 (.99)	9.21 (1.50)	5.67	17.00	11	.000

Table 2: Mastery of Targeted Content

Test Content and Points	Pretest Median	Posttest Median	Average Mastery at Posttest
Geometry - 8 points	.50	5.00	63%
Measurement - 5 points	2.50	4.50	90%
Total - 13 points	3.25	9.50	73%

TEL (860) 489-5639 · FAX (860) 496-8939 · 25 Maiden Lane · P.O. Box 1373 · Torrington, CT 06790-1373
E-mail: wordsnum@snet.net · www.wordsandnumbers.org

Attachment E

M² Student Mathematics Performance: 2nd Grade

Summative Evaluation

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 2nd grade students across all socioeconomic and ethnic backgrounds after exposure to the advanced mathematics 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 2nd grade students who are exposed to the advanced mathematics model and the Comparison group of 2nd grade students who undertake the traditional mathematics curriculum?

Data Collection Methods and Instrumentation

The 2nd 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 Test 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 2nd 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 20.9 points from a pretest score of 152.72 to a posttest score of 173.58.
- ✓ The Open Response Total score rose 9.12 points from a pretest score of 3.04 to a posttest score of 12.16.
- ✓ 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. Graphic portrayals of the performance are found on page 28-29.

Table 1

Pretest to Posttest Gains for 2nd Grade Intervention Group on Mathematics Achievement Indicators

Mathematics Measures *n=191*	*Pre M (SD)*	*Post M (SD)*	*Gain*	*t value*	*df*	*p*
Iowa Test of Basic Skills (ITBS)	152.72 (14.59)	173.58 (17.62)	+20.86	23.51	190	***
Open Response Total (ORT)	3.09 (2.23)	12.16 (5.00)	+9.12	28.88	190	***
Open Response Geometry (ORG)	1.42 (1.10)	6.62 (2.80)	+5.20	25.98	190	***
Open Response Measurement (ORM)	1.63 (1.54)	5.54 (2.82)	+3.92	22.16	190	***

*** $p<.001$ Bonferroni adjustment (.0125)

Research Question #2:

Is there a *difference in mathematics achievement* between the Intervention group of 2nd grade students, who are exposed to the mathematics curriculum model, and a Comparison group of students of similar socioeconomic and ethnic backgrounds?

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 Outcome Variables

Mathematics Measures	Intervention (n=191) *Post M (SD)*	Comparison (n=190) *Post (SD*
Iowa Test of Basic Skills (ITBS)	173.58 (17.62)	172.48 (15.74)
Open Response Total (ORT)	12.16 (5.00)	8.39 (3.92)
Open Response Geometry (ORG)	6.62 (2.80)	4.16 (2.17)
Open Response Measurement (ORM)	5.54 (2.82)	4.24 (2.23)

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 achieved homogeneous Level 1 variances but the three 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	.14	.45	.37	.43
Kurtosis	-.36	-.55	-.69	-.58
Homogeneity of Variance				
χ^2 (df)	27.28 (23)	71.11 (23)	56.96 (23)	50.69 (23)
p	.24	<.001	<.001	<.001

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 .16 indicating that 16% of the variance in the ITBS posttest scores lay between classes and 84%, within classes. The ICC for the Open Response assessment was somewhat more variable. The ICC for the Total Score was .35, indicating that 35% of the variance in the Open Response posttest scores lay between classes and 65%, within classes. The Geometry subscale score ICC was .37 and the Measurement subscale score ICC was .25, 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 decision 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 .78 with a SE of .05 (p<.001). For the Open Response Total Scores pretest (γ_{10}), the parameter estimate was .90 with an SE of .09 (p<.001). The Open Response Geometry subscale pretest (γ_{10}) parameter estimate was .81 with an SE of .06 (p<.001) and the Open Response Measurement subscale pretest (γ_{10}) parameter estimate was .67 with a SE of .11 (p<.001).

For the ITBS scores, the pretest score accounted for 44% of the variance within classes. There was no statistically significant variance in pretest slope (τ_{11})=.01, χ^2 = 24.83, p=.36). The Open Response Total pretest score explained 24% of the within class variance with no statistically significant variance in pretest slope (τ_{11})=.07, χ^2 = 31.99, p=.10). The Open Response Geometry subscale pretest score explained 15% of the within class variance with no statistically significant variance in pretest slope (τ_{11})=.01, χ^2 = 7.69, p>.50). For these three variables, the variance component was not allowed to randomly vary.

The Open Response Measurement subscale pretest score explained 21% of the within class variance with statistically significant variance in pretest slope (τ_{11})=.17, χ^2 = 51.09, p> .001). The variance component was allowed to randomly vary in the Contextual model.

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 differences 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 2nd grade students in project M², the predicted posttest ITBS for the Comparison group was 172.85 (γ_{00}) while that of the Intervention group (γ_{01}) was 173.89 (172.85+1.04). This was not a statistically significant difference (γ_{01})1.04, t=.53, p=.60). 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 2nd grade students, the predicted Open Response Total Score for the Comparison group was 8.20 ($\gamma_{00)}$ while that of the Intervention group was 12.63 (8.20+4.43). This was a statistically significant difference (γ_{01} 4.43, t=5.46, p<.001). The Open Response Geometry subscale score was predicted to be 4.09 for the Comparison group while the Intervention group's was 6.83 (4.09+2.74). This was statistically significant (γ_{01} 2.74, t=5.12, p<.001). The Open Response Measurement subscale score was predicted to be 4.20 for the Comparison group while the Intervention group's was 5.89 (4.20+1.69). This was statistically significant (γ_{01} 1.69, t=4.79, p<.001). Please refer to Table 4 through 7.

The full model for the Open Response Measurement subscale explained the variance in the pretest slopes. The coefficient was statistically significant. (γ_{11}=.50, t=2.33, p<.03). This means that a one point increase in the pretest score results in a .96 point increase on the posttest score for the Intervention group and a .46 increase in the posttest score for the Comparison group. Whereas students with average pretest scores had predicted Open Response Measurement scores of 5.89 and 4.20, respective of the Intervention and Comparison groups, those with a pretest score 1 point above the average would have higher scores. Specifically, posttest scores would be 6.85 (5.89+.50 +.46) for the Intervention group and 4.66 (4.20+.46) for the Comparison group. The results for the Full Models for each of the four outcome scores are in Tables 4 through 7. Graphic portrayals of the performance are found on pages 28-29.

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

	Unconditional Model		Random Coefficients Model		Contextual Model	
Parameter	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	173.49***	1.54	173.36***	.98	172.85 ***	1.37
M2 Intervention (γ_{01})					1.04	1.95
Pretest (γ_{10})			.78***	.05	.78***	.04
Variance estimate						
Level 1 Variance (σ^2)	235.94		133.17		133.15	
Intercept variance (τ_{00})	43.93 ***		15.67***		16.55***	
Deviance	3191.92 (2)		2968.80 (2)		2967.13 (2)	
(Number of REML parameters)						

REML= Restricted Likelihood Estimation
ITBS 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

	Unconditional Model		Random Coefficients Model		Contextual Model	
Parameter	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	10.47***	.62	10.43***	.61	8.20 ***	.48
M2 Intervention (γ_{01})					4.43***	.81
Pretest (γ_{10})			.90***	.09	.91***	.09
Variance estimate						
Level 1 Variance (σ^2)	15.98		12.08		12.07	
Intercept variance (τ_{00})	8.45 ***		8.56***		3.65***	
Deviance &	2188.75 (2)		2088.71 (2)		2070.74 (2)	
(Number of REML parameters)						

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

	Unconditional Model		Random Coefficients Model		Contextual Model	
Parameter	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	5.48***	. 36	5.47***	.39	4.09 ***	.31
M2 Intervention (γ_{01})					2.74***	.54
Pretest (γ_{10})			.81***	.06	.80***	.06
Variance estimate						
Level 1 Variance (σ^2)	5.04		4.29		4.29	
Intercept variance (τ_{00})	2.92 ***		3.49***		1.61***	
Deviance (Number of REML parameters)	1752.35 (2)		1697.91 (2)		1682.12 (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 MEASUREM.

	Unconditional Model		Random Coefficients Model		Contextual Model	
Parameter	Parameter Estimate	SE	Parameter Estimate	SE	Parameter Estimate	SE
Fixed Effect						
Intercept (γ_{00})	4.98***	.29	5.02***	.25	4.20 ***	.21
M2 Intervention (Y γ_{01})					1.69***	.37
Pretest (γ_{10})			.67***	.11	.46***	.09
Intervention(γ_{11})					.50	.22
Variance estimate						
Level 1 Variance (σ^2)	5.27		4.19		4.15	
Intercept variance (τ_{00})	1.74 ***		1.23***		.53***	
Pretest variance (τ_{11})			.17**		.16***	
Deviance & 1668.27 (4) (Number of REML parameters)	1758.44 (2)		1686.37 (4)			

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. The 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 impressive ranging from .51 for Measurement to .98 for Geometry with the Total score at .84. 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.

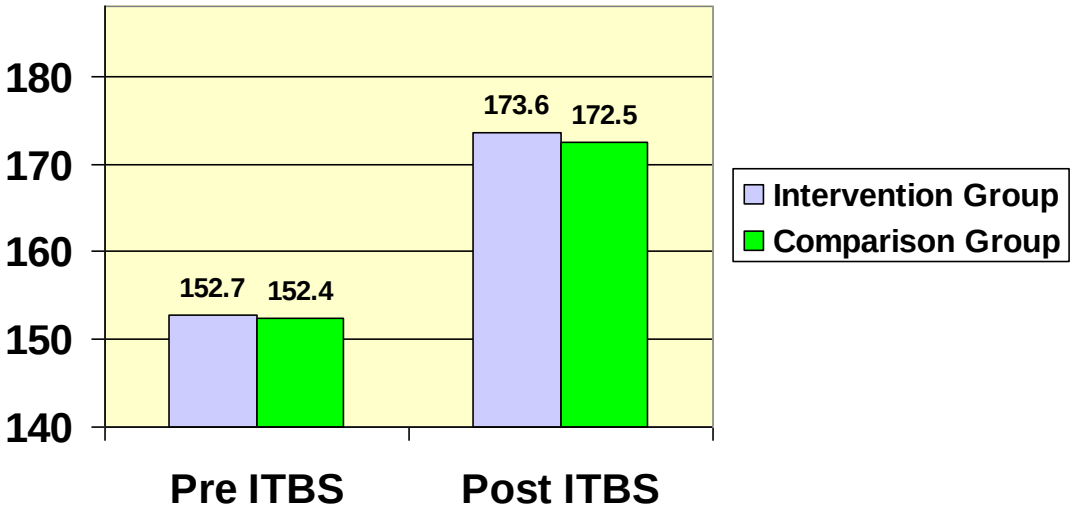
The HLM summary data, as well as the Cohen d effect sizes, are found in the Summary Table 8.

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

Outcome Variables	Coefficient[a] (SE)	t (df)	*p*	*d*
ITBS	1.04 (1.95)	.53 (22)	.60	.07
Open Response Total Score	4.43 (.81)	5.45 (22)	<.001***	.84
Geometry	2.74 (.54)	5.12 (22)	<.001***	.98
Measurement	1.69 (.37)	4.79 (22)	<.001***	.51

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

*p<.05 ** p<.01 *** p<.001


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