# Scaling up Curriculum for Achievement, Learning, and Equity Project

## Year 4 Annual Report: 2005 - 2006

July 1, 2006

<table>
<thead>
<tr>
<th>Section</th>
<th>TABLE OF CONTENTS</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>BACKGROUND</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>PROJECT OVERVIEW</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>RESEARCH GOALS</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>PARTICIPANTS</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>What people have worked on your project?</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>What other organizations have been involved as partners?</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>Have you had other collaborators or contacts?</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>ACTIVITIES AND FINDINGS</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Q1. Experimental research activities and findings of the project</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>Q2. Ethnographic research activities and findings of the project</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Q3. Experience research activities and findings of the project</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>Q5. Fidelity of implementation activities and findings of the project</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>SQ. Supplemental grant research activities and findings of the project</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>SQ1. Hierarchical linear modeling</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>SQ2. Assessment validation</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>SQ3. Video data in professional development</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>SQ4. Characteristics of Implementation</td>
<td>47</td>
<td></td>
</tr>
<tr>
<td>Opportunities for training and development provided by the project</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Outreach activities</td>
<td>56</td>
<td></td>
</tr>
<tr>
<td>PUBLICATIONS AND PRODUCTS</td>
<td>60</td>
<td></td>
</tr>
<tr>
<td>Journal publications</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Books or other non-periodical, one-time publications</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>Web site</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>Other specific products</td>
<td>62</td>
<td></td>
</tr>
<tr>
<td>CONTRIBUTIONS</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>The principal discipline(s) of the project</td>
<td>63</td>
<td></td>
</tr>
<tr>
<td>Other disciplines of science or engineering</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Development of human resources</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Resources that form the infrastructure for research and education</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>Aspects of public welfare beyond science and engineering</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Special Requirements, Change in Objectives of Scope</td>
<td>65</td>
<td></td>
</tr>
<tr>
<td>Un-obligated Funds; Animal, Human Subjects, Biohazards</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>Changes in objectives</td>
<td>66</td>
<td></td>
</tr>
<tr>
<td>References</td>
<td>67</td>
<td></td>
</tr>
</tbody>
</table>
BACKGROUND

Despite the best intentions to increase equity and to close achievement gaps, the science education reform movement has failed to adequately respond to the diversity of the U.S. student population. For instance, eighth grade science data from the 2000 National Assessment of Educational Progress (NAEP) disaggregated by a proxy for SES--the “educational level of parents”--showed that for 13-year-old students whose parents had low levels of education, NAEP science scores decreased over the previous nine years (National Center for Educational Statistics, 2000). One possible reason for this situation is the dearth of good curriculum materials that improve student outcomes, as well as concerns that curriculum materials may not reach students currently underserved by science education. Moreover, reform-based curriculum materials designed to increase deep understanding of science concepts have failed to scale-up; traditional textbooks still dominate instruction in U.S. science classrooms. In response, the Scaling up Curriculum for Achievement, Learning, and Equity Project (SCALE-uP) is using both quantitative and qualitative research methods to explore the effectiveness and scale-up of three middle school science curriculum units with different American Association for the Advancement of Science (AAAS) Project 2061 Curriculum Analysis (AAAS, 1999, 2003) rating profiles.

PROJECT OVERVIEW

The purpose of SCALE-uP is to identify the conditions under which three middle school science curriculum units--one each in 6th, 7th, and 8th grade--with different AAAS Curriculum Analysis profiles improve student understanding of benchmark concepts. If the units prove to be successful in the classroom relative to the comparison condition, SCALE-uP then studies their "scale-up" (or the transition from idiosyncratic adoption of curriculum units to broad, effective implementation across a large and diverse school system). Scale-up can be demonstrated by showing a plan for the gradual and systematic implementation of the interventions (three curriculum units) to the teachers and students in 35 of Montgomery County Public Schools' (MCPS) middle schools (affecting about 67,000 students and 210 teachers). This Annual Report serves to summarize the fourth year (2005-2006, which is referred to as "Year 4") of this six-year scale-up study, with Year 0 (2001 - 2002) serving as the planning grant year. Table 1 outlines evidence of our scale-up efforts.

<table>
<thead>
<tr>
<th>Grant Year</th>
<th>Academic Year</th>
<th>Description of Scale-up Efforts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 0</td>
<td>2001 -2002</td>
<td>5 demographically matched pairs of Treatment and Comparison schools with Chemistry That Applies (CTA) in the Treatment ~ 3000 8th graders.</td>
</tr>
<tr>
<td>(Planning Grant)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Year 1</td>
<td>2002 - 2003</td>
<td>Replicate CTA with same schools ~ 3000 8th graders.</td>
</tr>
</tbody>
</table>
Year 2 2003 - 2004  
*CTA* in 20 Treatment schools ~ 6000 8th graders.  
The Real Reasons for Seasons (*RRS*) in a matched sample of 5 schools ~ 3000 7th graders.  
Exploring Motion and Forces (*M&F*) in a matched sample of 5 schools ~ 3000 6th graders.

Year 3 2004 - 2005  
*CTA* in 35 schools (~ 10,000 8th graders).  
Replicate study for *RRS* and *M&F* (~ 3000 7th; ~ 3000 6th graders).

Year 4 2005 - 2006  
*CTA* in 35 schools (~ 10,000 8th graders)  
*M&F* in 5 Treatment schools (~ 3000 6th graders).  

Year 5 2006 - 2007  
*CTA* in 35 schools (~10,000 8th graders)  
*M&F* in 35 schools (~10,000 6th graders)(dependant on results from Year 4)

**RESEARCH GOALS**

An interdisciplinary team including GWU researchers from the Graduate School of Education and Human Development, Department of Teacher Preparation and Special Education, and the Columbian College of Arts of Sciences, Department of Anthropology is conducting the study in partnership with Montgomery County Public Schools (MCPS), Maryland. During Year 4, the project team continued to identify the conditions under which effective evidence-based interventions improve student learning and achievement when applied on a large scale. In addition, the project continued to focus on the implementation effects of fidelity, scale, and school science departments’ increasing experience with highly rated science curriculum units.

This research compares mean scores on student outcome measures in the Treatment and Comparison conditions, disaggregates outcome data and tests for interactions between demographic groupings and curriculum conditions, uses ethnographic methods to analyze a complete enactment of a unit for a group of four middle school students per intervention to determine how the intervention functions in a diverse middle school classroom; and, examines the relationship between student outcomes and factors such as experience, scale, and fidelity of implementation.

To this end, during Year 4, the SCALE-uP study continued to examine the five original research questions. The five questions are:

Q1.  *Experimental*: Does the implementation of a science curriculum unit result in higher mean scores (on student outcome measures) than the mean scores of students in the Comparison condition? Does disaggregating outcome data and testing for interactions between demographic groupings and curriculum condition reveal important patterns not captured in the reporting of mean scores of Treatment and Comparison groups? (Years 0, 1, 2, 3, 4)

Q2.  *Ethnographic*: Using ethnographic methods to analyze a complete enactment of a unit for a group of four middle school students, how does the unit function in a diverse middle school classroom? (Years 0, 1, 2, 3, 4)
Q3. *Experience:* Does the effectiveness of curriculum units increase as schools and teachers become more experienced implementing it from year to year? (Years 0, 1, 2)

Q4. *Scale:* Is there a difference in the effectiveness of curriculum units when comparing Small Scale Treatments (i.e., five Treatment schools) with Large Scale Treatments (i.e., 20 or 35 Treatment schools)? (Years 1, 5)

Q5. *Fidelity of Implementation:* Is there a relationship between the degree to which curriculum units are taught as intended by their authors and student outcomes? (Year 4)

In addition, the SCALE-uP team received a Supplemental grant during Year 3 to expand upon four of the above Questions, in areas such as the following:

SQ1. *Hierarchical linear modeling:* How much of the variation in 6th and 7th grade students’ gain score is between classrooms? Which classroom-level variables are statistically significant predictors of gain score, after controlling for all individual and classroom level variables? Which individual level variables are statistically significant predictors of gain score, after controlling for all individual and classroom level variables? (Years 2, 3, 4)

SQ2. *Assessment Validation:* Is performance on the post-test significantly higher after experiencing the pre-test as opposed to not experiencing the pre-test? Is understanding of the motion and forces target ideas equivalent on all items judged to be aligned to the ideas (SCALE-uP and multiple-choice items)? Is understanding, as demonstrated on the written form of the MFA, consistent with understanding, as demonstrated on MFA items during a video-taped, transcribed, and coded interview?

SQ3. *Professional Development:* In what ways can video data be used to enhance teachers' understanding of student motivation, engagement, and argumentation? (Year 5)

SQ4. *Characteristics of Implementation (COI):* How are the Comparison and Treatment conditions different? Do such differences account for variances in student outcomes? (Years 3, 4)

Please see the Activities and Findings section of this report for more information collected during Year 4 on each of these research questions.

**PARTICIPANTS**

- What people have worked on your project?
- What other organizations have been involved as partners?
- Have you had other collaborators or contacts?

**What People Have Worked on the Project?**

**The George Washington University Research Staff: Year 3**

Senior Personnel
Sharon Lynch, Ph. D.
Principal Investigator
Professor of Secondary Education
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctor of Philosophy: Education, Wayne State University

Joel Kuipers, Ph. D.
Co-Principal Investigator
Professor of Anthropology & International Affairs and Human Sciences
Columbian College of Arts and Sciences
Department of Anthropology
Doctor of Philosophy: Anthropology, Yale University

Curtis Pyke, Ph. D.
Co-Principal Investigator
Associate Professor of Secondary Education
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctor of Philosophy: Curriculum and Instruction, University at Albany

Carol O'Donnell, M.S.
Research Scientist
Project Director, SCALE-uP
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctoral Candidate: Curriculum and Instruction, The George Washington University

Vasuki Rethinam, Ph.D.
Research Scientist
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctor of Philosophy: Educational Psychology, Texas Tech University

Gail Brendel Viechnicki, Ph. D.
Post-Doctoral Scientist
Columbian College of Arts and Sciences
Department of Anthropology
Doctor of Philosophy: Linguistics, University of Chicago

Graduate Students: Research Assistants
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

Emily Fugate-Brunino, M.S.
Research Assistant
Columbian College of Arts and Sciences
Department of Anthropology
Master of Science: Anthropology, The George Washington University

Annie Knight Hanson, B.S.
Research Assistant
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Master of Education Candidate: Curriculum and Instruction, The George Washington University

Elizabeth Edmonds Hatchuel, M.A.
Research Assistant
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctoral Candidate: Community Counseling, The George Washington University

Jesse Kimball, B.S.
Research Assistant
Columbian College of Arts and Sciences
Department of Anthropology
Master of Science Candidate: Anthropology, The George Washington University

Joelle Lastica, M.S.
Research Assistant
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctoral Candidate: Curriculum and Instruction, The George Washington University

Lindsey Jones Massoud, B.S.
Research Assistant
Columbian College of Arts and Sciences
Department of Anthropology
Master of Science Student: Anthropology, The George Washington University

Rob Ochsendorf, M.Ed.
Research Assistant
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Doctoral Candidate: Curriculum and Instruction, The George Washington University
William Watson, M.A.T.
Research Assistant
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
*Doctoral Candidate*: Curriculum and Instruction, The George Washington University

Laura Wright, M.S.
Research Assistant
Columbian College of Arts and Sciences
Department of Anthropology
*Doctoral Student*: Anthropology, Georgetown University

**Undergraduate and Graduate Assessment Raters**

Sebastian Andrews
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Forensic Science Graduate Student
Ethnicity: White
Total Number of Hours Worked on Project: 120

Brian Angolia
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Forensic Science Graduate Student
Ethnicity: White
Total Number of Hours Worked on Project: 120

Brandi Palmore
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Chemistry Doctoral Student
Ethnicity: African American
Total Number of Hours Worked on Project: 120

Rebecca Raese
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Forensic Science Graduate Student
Ethnicity: White
Total Number of Hours Worked on Project: 120
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

Raluca Teodorescu
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Physics Doctoral Student
Ethnicity: White
Total Number of Hours Worked on Project: 120

Julie Woodroffe
Assessment Rater
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
Forensic Science Graduate Student
Ethnicity: White
Total Number of Hours Worked on Project: 45

Undergraduate Students: Transcribers

Asa Boy
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

John Curan
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Maia Gil' Adi
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Vanessa Icolari
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Helen Lapedes
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Kara McGinnis
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Sara Mykietyn
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Ashley Rosen
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Julia Rosen
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Brooke Sadler
Transcriber
Columbian College of Arts and Sciences
Department of Anthropology
*Bachelor of Science Candidate*: Department of Anthropology

Technicians, Programmers

Tejas Bhise
Network Specialist
Columbian College of Arts and Sciences
Department of Anthropology
*Master of Science Candidate*: Computer Science Department
Hours on Project: 160

Sachin Patil, B.S.
Database and Streaming Video Specialist
Columbian College of Arts and Sciences
Department of Anthropology
*Master of Science Candidate:* Computer Science Department
Hours on Project: 160

Eswar Vommina, B.S.
Network Specialist
Columbian College of Arts and Sciences
Department of Anthropology
*Master of Science Candidate:* Computer Science Department
Hours on Project: 360

**Work Study**

Sarah Pelkey
Scanning Specialist
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
*Bachelor of Arts Candidate:* Communications
Hours on Project: 240

Olga Tsyganova
Research Aide
Graduate School of Education and Human Development
Department of Teacher Preparation and Special Education
*Bachelor of Arts Candidate:* Communications
Hours on Project: 240

**GWU SCALE-uP Project Staff Roles and Responsibilities**

*Principal Investigator*
- Hiring PD
- Professional Development - working with MCPS
- NSF Report
- Internal Budget
- NSF Budget
- Fidelity of Implementation

*Co-Principal Investigator: Ethnographic*
- Permissions
- IRB
- Post doctoral fellow position
- Hiring and overseeing CSAS RAs and transcribers
- Video data
Year 4 Annual Report: 2005 - 2006
July 1, 2006

- Year 0 budget CCAS

**Co-Principal Investigator: Experimental**
- Curriculum Analysis
- Assessment Development
- Scoring
- Quantitative data analysis
- Hiring, overseeing, supervising GSEHD RAs

**Project Director**
- Manage all aspects of project, including timelines.
- Work collaboratively with faculty, RAs, and MCPS teachers and staff.
- Maintain clear written and oral communications with project staff (budget, research, etc.).
- Write proposals, reports and publications related to the grant. Communicate with media.
- Review research materials related to the grant.
- Contribute to activities related to curriculum evaluation, assessment development, data analysis, fidelity of implementation, and the comparison condition.
- Work with MCPS personnel to create professional development programs for teachers.
- Develop evaluation instruments in conjunction with MCPS.
- Oversee the budget and other expenditures.
- Use SPSS, Atlas-ti, word processing, database and spreadsheet, Internet communication skills.
- Act as liaison between key personnel at GWU and within MCPS, including Advisory Board.
- Use excellent interpersonal skills to coordinate aspects of the project, especially collaboration between GWU and MCPS.

**Research Scientists**
- Work with the senior researchers to plan, conduct, analyze, interpret, and report on quantitative data focusing on middle school science education.
- Work with the research assistants in management, analysis, and maintenance of the database.
- Work with senior researchers to co-author papers.
- Attend professional conferences and workshops to keep up-to-date on scale-up research design and analysis, and present at conferences.
- Conduct analysis of Comparison and Treatment classrooms.
- Develop protocol for studying the Year 2 and 3 data at the individual-, classroom-, and school-level using hierarchical linear modeling.
- Attend planning meetings related to the study.

**Post-Doctoral Scientist**
- Develop and carry out advanced ethnographic research on topics related to schools, video, corpus linguistics, and/or science.
- Propose new research related to the existing project.
- Manage and supervise a research team analyzing digital data.
Facilitate database software.
Oversee the collection, transcription, and analysis of classroom video data. Work with and supervise undergraduate transcribers and graduate research assistants.
Use advanced video data analysis software (Primeview, Adobe, Atlas.ti, etc.) for digitizing, coding, and analyzing videotape data.
Assist PIs in developing data for the testing of hypotheses and the development of research publications.
Assist project personnel in other aspects of project activities, as required.

_GSEHD Research Assistants_
- Coordinate activities related to curriculum evaluation, assessment development, characteristics of implementation, and fidelity of implementation.
- Coordinate data transfer, assessment rating, instrument development, classroom observations, and teacher interviews.
- Assist in developing and maintaining data files in SPSS.
- Assist in data analyses and writing reports.
- Assist project personnel in other aspects of project activities, as required.

_CCAS Research Assistants_
- Code and analyze classroom video data.
- Work with and supervise classroom transcribers.
- Use advanced video data analysis software (Primeview, Adobe, Atlas.ti, etc.) for digitizing, coding and analyzing videotape data.
- Assist PIs in developing data for testing of hypotheses and development of research publications.
- Assist PIs in database management.

**What Other Organizations Have Been Involved as Partners?**

**Montgomery County Public Schools Research Staff: Year 4**

_Senior Personnel_

Michael Szesze, M. S.
Co-Principal Investigator, SCALE-uP (through April, 2006)
Program Supervisor, MCPS Science, K-12
_Master of Science: Curriculum and Instruction, Western Maryland College_

Bonnie Hansen-Grafton, B. S.
Instructional Specialist, Acting Co-Principal Investigator (as of April, 2006), SCALE-uP
Supervises the MCPS side of the NSF-IERI SCALE-uP Project
_Master of Education: Middle School Science_

_Technicians, Programmers_
HuldahLatchansingh
FiscalAssistant,MCPS,SCALE-uP
MCPSScience,K-12

GretchenBlankenship
FiscalAssistant,MCPS,SCALE-uP
MCPSScience,K-12

ElizabethCooper-Martin
ClassroomObserver
DepartmentofSharedAccountability,MCPS

TrishaMcGaughey
ClassroomObserver
DepartmentofSharedAccountability,MCPS

AngelaO.McKiney
Secretary,MCPS,SCALE-uP
MCPSScience,K-12

SuzanneMerchlinsky
Evaluationspecialist
DepartmentofSharedAccountability,MCPS

VinceParada
ClassroomObserver
DepartmentofSharedAccountability,MCPS

DonnaShipley
ClassroomObserver
DepartmentofSharedAccountability,MCPS

DebbieSymons
ClassroomObserver
DepartmentofSharedAccountability,MCPS

StephanieThliveris
ClassroomObserver
DepartmentofSharedAccountability,MCPS

NatalieWolanin
ClassroomObserver
DepartmentofSharedAccountability,MCPS
Research Aides (total hours worked = 75 per person)

Angela Clement  
Middle School Teacher, Student Journal Rater

Jacquelyn Geer  
Middle School Teacher, Student Journal Rater

Jill Maisch  
Middle School Teacher, Student Journal Rater

Samantha Phillips  
Middle School Teacher, Student Journal Rater

Melvin Smith  
Middle School Teacher, Student Journal Rater

Sheena Wright  
Middle School Teacher, Student Journal Rater

MCPS SCALE-uP Project Staff Roles and Responsibilities

**Instructional Specialist**
- Oversee implementation of the project within the MCPS school system.
- Perform administrative tasks.
- Play a critical role as the communication liaison between the MCPS administration, school principals, and participating science teachers and the GWU Principal Investigators (PI), Project Director, and GWU staff.
- Responsible for leading professional development.
- Select curriculum units.
- Coordinate with middle school MCPS science specialists.
- Arrange the distribution of equipment and other supplies.
- Manage budget for SCALE-uP subcontract.
- Assist in research and writing of reports.
- Assist in writing the Annual Report.
- 12-month position.

**Administrative Secretary**
- Aid the Instructional Specialist with the previously mentioned responsibilities.
- Perform clerical and secretarial duties.
- 12-month position.

Department of Shared Accountability
Disseminate, administer, collect, and clean pre- and post-unit assessments in each of the 6th grade science classes of 10 participating middle schools in the 2005-06 academic year.

Provide data exchange to GWU from MCPS.

Train observers for classroom observations.

Test reliability of classroom observation instrument.

Develop and refine the observation instruments.

Conduct Grade 6 Treatment and Comparison classroom observations.

Report Grade 6 Treatment and Comparison classroom observation data to GWU.

Participate in NSTA presentation of SCALE-uP implementation and outcomes.

Assist in writing Annual Report.

Participating Treatment Science Teachers (approximately 15 6th grade teachers and 112 8th)

- Attend professional development meetings.
- Administer pre- and post-unit assessments.
- Implement Treatment curriculum units.

Participating Comparison Science Teachers (approximately 15 6th grade teachers)

- Attend content-based professional development meeting.
- Administer pre- and post-unit assessments.
- Implement standard curriculum.

Implementation Teachers (ITs) (6 per grade level)

- Lead the professional development in Fall-Winter 2005-06 with the assistance of the MCPS Instructional Specialist and GWU team.
- Plan and lead the professional development for M&F and the scale-up professional development of CTA during the summer of 2005, in preparation for implementation, with the assistance of the curriculum unit developers.
- Participate on the curriculum and assessment teams meeting at GWU when needed.
- Help GWU in setting the standards for the pre- and post-unit assessments.

Have You Had Other Collaborators or Contacts?

Q1. Experimental Research Collaborations

Dr. Sharon Lynch collaborated with Dr. Juliana Taymans, Professor, Special Education, Department of Teacher Preparation and Special Education, The George Washington University on a paper submitted and accepted for publication in Exceptional Children. The title of the paper is, “Effectiveness of a Highly Rated Science Curriculum Unit for Students with Disabilities in General Education Classrooms.” Dr. Taymans served as second author. Other co-authors include William Watson, Robert Ochsendorf, Dr. Curtis Pyke, and Michael Szesze, all of SCALE-uP. The focus of the paper is the impact of the 8th grade Treatment unit, Chemistry That Applies (CTA), on students with disabilities, which is Dr. Tayman's specialty. The paper is currently in press.

Q2. Ethnographic Research Collaborations
Through the GWU Anthropology Department Discourse Lab, the SCALE-uP ethnographic team (Dr. Joel Kuipers, Dr. Gail Veietchnicki, Laura Wright, and Lindsey Jones) collaborated extensively with other like-minded scholars on research and publication. For a panel organized for the Annual Meetings of the American Anthropological Association (AAA) in November 2005, SCALE-uP ethnographers collaborated with Dr. Brian V. Street, Chair of the Department of Education at the University College, London; Dr. Stanton Wortham of the University of Pennsylvania; and, Dr. James P. Collins, Chair of the Department of Anthropology at State University of New York in Albany. Together, these ethnographers have prepared a volume of essays to be submitted as a special issue in the journal *Linguistics and Education*, to which Dr. Webb Keane of the University of Michigan has agreed to provide a preface.

In addition, Dr. Kuipers and his staff hosted Dr. Ray McDermott of Stanford University, Dr. Shelley Goldman of Stanford University, Dr. Janet Coffey of University of Maryland, Dr. Stanton Wortham of University of Pennsylvania, and Dr. Brian Street of University College London in a special meeting devoted to methods in the study of discourse and education. In the early spring of 2006, SCALE-uP ethnographers also hosted an afternoon advisory visit by Dr. Richard Duschl of Rutgers University, who gave a presentation on argumentation in science education.

**Q4. Scale up Collaborations**

Dr. Sharon Lynch collaborated with several other experts in the field of scale-up research during the American Educational Research Association Annual Meeting in April 2006. Participants in the symposium, titled *Identifying interventions that improve student learning on the large scale: Scale up in principle and scale up in practice*, included Chair Barbara Schneider (Michigan State University), Eva L. Baker, (UCLA), Geoffrey Borman (University of Wisconsin-Madison), James Dignam (University of Chicago), Larry Hedges (Northwestern University), Stephen Raudenbush (University of Chicago), and Robert Slavin (Johns Hopkins University).

**Q5. Fidelity of Implementation Collaborations**

Dr. Lynch and Carol O'Donnell were asked to serve as advisors to an NSF Instructional Materials Development (IMD) Proposal that focuses on the study of fidelity of implementation of Science and Technology for Children (STC) and Full Option Science System (FOSS) K-6 units implemented in Chicago Public Schools. This proposal, "*Applying Research on Science Materials Implementation: Bringing Measurement of Fidelity of Implementation to Scale,*" was submitted to NSF by the University of Chicago in the Spring 2006, under the direction of Principal Investigator, Dr. Andy Isaacs, Co-Director of the Center for Elementary Mathematics and Science Education (CEMSE) in the Physical Sciences Division at the University of Chicago; and, Co-PI and Project Director, Dr. Jeanne Century, Director of Science Education Research for CEMSE.

Dr. Lynch and Carol O'Donnell invited Dr. Douglas Clements, Principal Investigator of the IERI TRIAD grant, to GWU in the summer, 2006, to share ideas on how fidelity of
implementation can be observed in the classroom and to discuss the validation and reliability of fidelity of implementation instruments.

Dr. Lynch was the Chair of the Plenary Session on Fidelity of Implementation at the annual NSF/IERI PI meeting in collaboration with Dr. Jo Ellen Roseman, AAAS Project 2061; Nancy Romance, Florida Atlantic University; and, David Mostow, Carnegie Mellon University. The plenary session convened in Washington, DC in August 2005.

Collaborations with Curriculum Developers

Throughout Year 4, SCALE-uP staff collaborated with the developers of the curriculum units being implemented in the study. The developers are:

1. Dr. Theron Blakeslee, Director, Jackson County Mathematics and Science Center, is the developer of the unit: Chemistry That Applies (State of Michigan, 1993), hereafter referred to as CTA.

2. Dr. Bruce Ward, Project Manager of ARIES, Harvard Smithsonian Astrophysical Observatory, Science Education Department, is the developer of the unit: Exploring Motion and Forces (Harvard-Smithsonian Astrophysical Observatory, 2001), hereafter referred to as M&F.

3. Carolyn Willard, Network Director, Great Explorations in Math and Science (GEMS) Program, Lawrence Hall of Science, University of California, Berkeley, is the developer of the unit: The Real Reasons for Seasons (LHS, 2000), hereafter referred to as RRS.

Dr. Bruce Ward was consulted when SCALE-uP developed the Adherence Classroom Observation Protocol (ACOP) for M&F.

Bonnie Hansen-Grafton met with Carolyn Willard at the National Science Teachers Association Annual Meeting on April 7, 2006 to discuss the revisions to RRS.

Dr. Sharon Lynch, Dr. Curtis Pyke, Bonnie Hansen-Graton, and Carol O'Donnell hosted a meeting at GWU with the Dr. Elizabeth Stage, Director of the Lawrence Hall of Science, University of California at Berkeley, and Carolyn Willard, Director of GEMS, who came to GWU to discuss the SCALE-uP research and its impact on the development of GEMS curriculum materials.

Curriculum Analysis: Project 2061 Collaboration

Project 2061 launched a major effort to find curriculum materials aligned with benchmarks that meet a rigorous set of criteria consistent with current theories of learning and content specific instructional strategies that support learning (Kesidou and Roseman, 2002). The Project 2061 Curriculum Analysis consists of categories of questions for analyzing the instructional qualities of a unit. Rob Ochsendorf, Dr. Sharon Lynch, and Dr. Curtis Pyke collaborated on a paper summarizing the curriculum analysis of M&F, with consultation from
ARIES director, Dr. Bruce Ward, and Dr. Jo Ellen Roseman, AAAS, Project 2061. This paper is currently in preparation.

In June 2006, Carol O'Donnell, William Watson, Dr. Curtis Pyke, and Dr. Sharon Lynch co-authored a paper integrating the results from the curriculum analysis of RRS and outcomes from the implementation of RRS during Year 3. This paper includes responses from GEMS' developer, Carolyn Willard, who was given the opportunity to comment on and provide feedback to the Technical Report that summarizes the results from the RRS Curriculum Analysis. The paper is currently in review with the Journal of Research in Science Teaching. (See the Publications section for titles.)

SQ2. Assessment Validation Collaborations

During Year 4 of SCALE-uP, Dr. Curtis Pyke, Robert Ochsendorf, William Watson, and Carol O'Donnell of the SCALE-uP research staff, collaborated with Project 2061 to develop an interview protocol framework for exploring the validity of several Motion and Forces Assessment (MFA) items being used in the implementation study of M&F. Dr. Pyke and Rob Ochsendorf also collaborated with the Project 2061 team to develop a long-form of the MFA, which incorporated several assessment items developed by Project 2061. This work was undertaken with the idea that assessment developers, in general, could use groundwork from implementation research to explore questions about the content validity of particular standards-based assessment items. The goal in combining items from both SCALE-uP and Project 2061 in one assessment is to begin to strengthen the claims made from assessment items that comprise standards-based assessment instruments. See the Activities and Findings section of this report for additional information on the validation interviews’ goals and activities.

ACTIVITIES AND FINDINGS

- Describe the major research and education activities of the project.
- Describe the major findings resulting from these activities.
- Describe the opportunities for training and development provided by your project.
- Describe outreach activities your project has undertaken.

Q1. Experimental Research Activities and Findings

SCALE-uP data is collected in one year and analyzed the subsequent year. During Year 4 of the SCALE-uP grant, SCALE-uP analyzed the Year 3 experimental data and collected the Year 4 experimental data. Therefore, this part of the Annual Report is divided into two sections: Year 3 (which summarizes the experimental research student and teacher participation data and findings from Year 3) and Year 4 (which outlines the demographic population studied in Year 4 and lists the data collection and distribution procedures employed in Year 4).
Participation Data

Student and teacher participation data for Year 3 are listed in Tables 2 and 3. (Note that these tables contained “Pending” data in the Year 3 Annual Report.) The Year 3 Experimental Research Findings in this Year 4 Annual Report are based on the data presented in Tables 2 and 3.
### Table 2: Year 3 (2004-05) Student and Teacher Participation

#### Grade 6 Motion and Forces Treatment and Comparison Schools

<table>
<thead>
<tr>
<th>School Profile Category (SPC)</th>
<th># Participating 6th Grade Teachers*</th>
<th># of Pre-unit Assessments (2-part)</th>
<th># of Post-unit Assessments (2-part)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td># of students sent assessments</td>
<td># of complete data sets rec’d from school</td>
</tr>
<tr>
<td>TREATMENT SCHOOLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>320</td>
<td>299</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>128</td>
<td>124</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>234</td>
<td>204</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>278</td>
<td>238</td>
</tr>
<tr>
<td>E1</td>
<td>5</td>
<td>404</td>
<td>375</td>
</tr>
<tr>
<td>E2</td>
<td>3</td>
<td>202</td>
<td>190</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>18</strong></td>
<td><strong>1566</strong></td>
<td><strong>1430</strong></td>
</tr>
<tr>
<td>COMPARISON SCHOOLS</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>328</td>
<td>315</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>211</td>
<td>192</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>244</td>
<td>226</td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>259</td>
<td>221</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>175</td>
<td>153</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>14</strong></td>
<td><strong>1217</strong></td>
<td><strong>1107</strong></td>
</tr>
</tbody>
</table>

* Non-participating 6th Gr. Teachers

- School E Treatment: 2 SpEd Teachers, 1 ESOL
- School D Treatment: 2 SpEd Teachers, 1 ESOL, 1 Long-Term Sub (4 classes) replaced Reg. teacher at the beginning of unit instruction
- School C Comparison: 1 SpEd Teacher
- School E1 Treatment: Data from this school was not used in Year 3 analysis.
Table 3: Year 3 (2004-05) Student and Teacher Participation
Grade 7 Seasons Treatment and Comparison Schools

<table>
<thead>
<tr>
<th>School Profile Category (SPC)</th>
<th># of Participating 7th Grade Teachers*</th>
<th># of students sent assessments</th>
<th># of complete data sets rec’d from school</th>
<th># of students sent assessments</th>
<th># of complete data sets rec’d from school</th>
<th># of students sent assessments</th>
<th># of complete data sets rec’d from school</th>
<th>Student Responsiveness Questionnaires*</th>
<th># of SRQ’s rec’d from school</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TREATMENT SCHOOLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>355</td>
<td>333</td>
<td>355</td>
<td>334</td>
<td>99</td>
<td>84</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>376</td>
<td>325</td>
<td>376</td>
<td>344</td>
<td>111</td>
<td>103</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>229</td>
<td>208</td>
<td>229</td>
<td>182</td>
<td>127</td>
<td>125</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>300</td>
<td>220</td>
<td>300</td>
<td>201</td>
<td>99</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>3</td>
<td>222</td>
<td>171</td>
<td>222</td>
<td>187</td>
<td>94</td>
<td>68</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>15</td>
<td>1482</td>
<td>1257</td>
<td>1482</td>
<td>1248</td>
<td>530</td>
<td>460</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>COMPARISON SCHOOLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>4</td>
<td>355</td>
<td>330</td>
<td>355</td>
<td>279</td>
<td>79</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>142</td>
<td>128</td>
<td>142</td>
<td>128</td>
<td>81</td>
<td>53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>285</td>
<td>204</td>
<td>285</td>
<td>225</td>
<td>108</td>
<td>78</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>3</td>
<td>372</td>
<td>315</td>
<td>372</td>
<td>287</td>
<td>101</td>
<td>72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>222</td>
<td>201</td>
<td>222</td>
<td>152</td>
<td>107</td>
<td>95</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>13</td>
<td>1376</td>
<td>1178</td>
<td>1376</td>
<td>1071</td>
<td>476</td>
<td>301</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Non-participating 7th Grade Teachers:
School E Comparison: 2 SpEd Teachers, 1 ESOL Teacher
School D Comparison: 2 SpEd Teachers School C Treatment: 1 SpEd Teacher
** Student Responsiveness Questionnaires were distributed only to students in the observed classrooms.
Motions and Forces Assessment (MFA) Results from Year 3

Analysis of the experimental data for MFA for Year 3 showed higher mean scores in both conditions at the post-test than at the pre-test. There was no statistically significant difference in MFA scores between the Treatment and Comparison conditions at post-test (see Figure 1). However, tests for interactions showed evidence that the student characteristics Gender, Ethnicity, and FARMS, influence the degree to which the Treatment condition affects MFA outcomes. The interactions indicate that some subgroups of students have lower scores in the Treatment condition than in the Comparison condition, suggesting that differences among subgroups might increase in the Treatment condition. The male, African American, Hispanic, and Now FARMS subgroups all scored lower in the Treatment condition than in the Comparison condition (see Table 4). However, these results appear to result in lower understanding in the Treatment condition relative to predetermined cut scores only for the African American subgroup.

For more discussion of the Motion and Forces results in Year 3, see Ochsendorf, Pyke, Lynch, & Watson (2006), an AERA conference paper.

For more discussion of the Motion and Forces results in Year 3, see Ochsendorf, Pyke, Lynch, & Watson (2006), an AERA conference paper.
Table 4  
Year 3 (2004-05) Curriculum Effect Sizes for all Levels of Independent Variables for the Motions and Forces Assessment

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>d</th>
<th>95% CI</th>
<th>$F_B$</th>
<th>$F_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>8.77*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1165</td>
<td>-0.12</td>
<td>-0.23 to -0.01</td>
<td>4.78*</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1191</td>
<td>0.07</td>
<td>0.00 to 0.18</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td>28.32*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>470</td>
<td>-0.34</td>
<td>-0.51 to -0.16</td>
<td>16.52*</td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>378</td>
<td>0.21</td>
<td>0.01 to 0.42</td>
<td>5.36*</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>455</td>
<td>-0.18</td>
<td>-0.36 to -0.01</td>
<td>4.37*</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1053</td>
<td>0.03</td>
<td>-0.09 to 0.15</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>FARMS</td>
<td>43.94*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1492</td>
<td>0.09</td>
<td>-0.02 to 0.19</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>Prior</td>
<td>259</td>
<td>-0.20</td>
<td>-0.44 to 0.04</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>605</td>
<td>-0.21</td>
<td>-0.36 to -0.05</td>
<td>8.07*</td>
<td></td>
</tr>
<tr>
<td>ESOL</td>
<td>16.89*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1836</td>
<td>-0.02</td>
<td>-0.11 to 0.07</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Prior</td>
<td>369</td>
<td>-0.02</td>
<td>-0.22 to 0.19</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>151</td>
<td>-0.15</td>
<td>-0.46 to 0.16</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>SPED</td>
<td>33.75*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2070</td>
<td>-0.03</td>
<td>-0.11 to 0.06</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>286</td>
<td>-0.02</td>
<td>-0.25 to 0.21</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Note.  $d$ = Cohen’s $d$ effect size; CI = Confidence Interval; $F_B$ = test of between group differences; $F_W$ = test of within group differences (Treatment vs. Comparison).

*p < .05
Causes of Seasons Assessment (CSA) Results from Year 3

Analysis of the experimental data for CSA in Year 3 (2004-2005) indicated that mean scores were higher on average at the post-test than at the pre-test in both conditions. Mean scores were statistically significantly higher in the Comparison condition than in the Treatment condition (see Figure 2). Tests for interactions showed that outcomes were influenced by only one student characteristic, SPED. However, the lack of interactions in other cases suggests that the performance of all subgroups of students, except Now SPED, on the CSA was lower when the Treatment curriculum materials were used than when other curriculum materials were used (see Table 5). Further, except for SPED, differences among subgroups were the same in both conditions. The overall low performance relative to predetermined cut scores is important to consider, as is the observation that all subgroups of students in both conditions performed within the same level of understanding at the post-test. Therefore, there were no differences between subgroups in terms of practical significance.

For more information on the results of the RRS unit, see the Internal Report of 2004-2005 Implementation Study for Exploring RRS (Year 3), which can be found on the SCALE-uP website. In addition, see the paper by O'Donnell, Pyke, Lynch, and Watson (in review).

### Adjusted Means for RSA Posttest Score

![Figure 2](image-url)

**Figure 2.** Adjusted mean scores for CSA between the Comparison and Treatment conditions Year 3 (2004-05). Means adjusted by the pre-test covariate, pre-test mean = 18.84.

**Note:** Levels of understanding: 0 - 20 = No understanding of concepts; 21-50 = Some understanding of concepts; 51-70 = Some fluency with concepts; 71 – 100 = Flexible understanding of concepts.
### Table 5

Year 3 (2004-05) Curriculum Effect Sizes for all Levels of Independent Variables for the Causes of Seasons Assessment (CSA)

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>d</th>
<th>95% CI</th>
<th>$F_B$</th>
<th>$F_W$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1165</td>
<td>-0.12</td>
<td>-0.23 to -0.01</td>
<td>4.78*</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1191</td>
<td>0.07</td>
<td>0.00 to 0.18</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>470</td>
<td>-0.34</td>
<td>-0.51 to -0.16</td>
<td>16.52*</td>
<td></td>
</tr>
<tr>
<td>Asian American</td>
<td>378</td>
<td>0.21</td>
<td>0.01 to 0.42</td>
<td>5.36*</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td>455</td>
<td>-0.18</td>
<td>-0.36 to -0.01</td>
<td>4.37*</td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>1053</td>
<td>0.03</td>
<td>-0.09 to 0.15</td>
<td>0.25</td>
<td></td>
</tr>
<tr>
<td>FARMS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1492</td>
<td>0.09</td>
<td>-0.02 to 0.19</td>
<td>3.54</td>
<td></td>
</tr>
<tr>
<td>Prior</td>
<td>259</td>
<td>-0.20</td>
<td>-0.44 to 0.04</td>
<td>3.19</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>605</td>
<td>-0.21</td>
<td>-0.36 to -0.05</td>
<td>8.07*</td>
<td></td>
</tr>
<tr>
<td>ESOL</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1836</td>
<td>-0.02</td>
<td>-0.11 to 0.07</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>Prior</td>
<td>369</td>
<td>-0.02</td>
<td>-0.22 to 0.19</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>151</td>
<td>-0.15</td>
<td>-0.46 to 0.16</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>SPED</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2070</td>
<td>-0.03</td>
<td>-0.11 to 0.06</td>
<td>0.44</td>
<td></td>
</tr>
<tr>
<td>Now</td>
<td>286</td>
<td>-0.02</td>
<td>-0.25 to 0.21</td>
<td>0.04</td>
<td></td>
</tr>
</tbody>
</table>

Note. $d$ = Cohen’s $d$ effect size; CI = Confidence Interval; $F_B$ = test of between group differences; $F_W$ = test of within group differences (Treatment vs. Comparison).

* $p < .05$
Year 4: Experimental Research Activities and Findings

Sampling and Demographic Characteristics

The data in Table 6 provide the demographic characteristics of the 10 MCPS middle schools that participated as Treatment and Comparison schools for the study of motion and forces (6th grade) during Year 4. To obtain these 10 schools, SCALE-uP first divided 35 of the middle schools in MCPS into five School Profile Categories (SPC’s) primarily indexed by socioeconomic status, with approximately seven schools in each category. One pair of middle schools that had not previously implemented the M&F unit was randomly selected from each SPC. One school from the pair was randomly assigned to implement the M&F curriculum unit, and the other school in the pair was assigned to use the menu of curriculum materials normally available to district teachers that addressed the motion and forces target benchmark. This sampling technique was expected to produce two equivalent samples, representative of the grade 6 population, with school level effects distributed evenly in each sample, and with enough students to provide power for significance tests on disaggregated subgroups. Although schools were used as the unit of randomization, students were the unit of analysis.

Assessment Distribution and Collection

During January 2006, pre-unit assessments were distributed and administered to a sample of classrooms randomly selected from the schools listed in Table 6. Consultants (former middle school teachers) were hired by the MCPS Department of Shared Accountability (DSA) to administer the pre-unit assessments to students in the randomly selected classrooms. To control for pre-test effects, classroom teachers did not have access to the pre-unit assessments.

Staff from DSA also packaged and sent all post-unit assessments to each Treatment and Comparison school’s designated coordinator, who is typically an Interdisciplinary Resource Teacher (IRT) or Resource Teacher (RT). These 6th grade Motion and Forces post-unit assessments were sent to schools before the teachers finished teaching the unit. Shipping dates varied from late February to early May. Assessments were packaged for each participating teacher and classroom. Each school coordinator received written distribution and collection instructions and each teacher received assessment administration instructions. All assessments had a pre-printed identifying student label. Each teacher received several additional unlabeled assessments for new students that may have entered the class. DSA staff verified IDs for all of the newly enrolled students.

Teachers completed a SCALE-uP Assessment Cover Sheet for each classroom. The cover sheet was used to identify special education or ESOL students who received accommodations and identified the nature of the accommodations. The cover sheet also noted any student who did not take the assessment and why. At the summer training (June 2005), teachers were reminded of the guidelines for determining which students should be exempt from the assessments for Year 4 of the study (2005–06 school year). It was suggested that self-contained ESOL and special education classes that do not teach the targeted science unit (e.g., special education students who are not diploma-bound) be exempt from the assessments. ESOL and special education students who are
mainstreamed into regular science classes were to take the assessments with their classmates, with accommodations as needed.

Teachers were asked to administer the post-unit assessments to their individual classrooms within one week of completing their motion and forces unit. All teachers in both the Treatment and Comparison condition received a letter from the Principal Investigators outlining their responsibilities in completing the post-unit assessments. This letter was sent via email from Bonnie Hansen-Grafton; a hard copy of the letter was also sent to the teachers via US postal mail. Ms. Hansen-Grafton emailed the teachers to ensure that all schools had received the post-unit assessments. DSA sent follow-up emails and Ms. Hansen-Grafton phoned the schools that did not send in their post-unit assessments within the estimated time of completing the unit. The school in SPC D, however, did not respond to these communications. Visits to the school by the project directors indicated that the assessments were not administered. As a result, no assessments were collected from SPC School D.

All post-unit assessments, except for those from SPC D, were collected at the school level by the coordinator and sent to DSA’s Program Evaluation Unit. The assessments were “logged-in,” checked for missing data, and student names were removed for anonymity and security purposes. Blind data were then given to The George Washington University researchers for analysis.

The Student Questionnaires were distributed with the post-unit assessments to a sample of students in classrooms where observations took place. Teachers were instructed to give students the Student Questionnaire after they had completed the post-unit assessments. Teachers then sent the completed Student Questionnaires back to DSA for processing (i.e., removal of student names for anonymity).

M&F student journals were also collected from Treatment classrooms only. These journals were sent to the Treatment teachers at the beginning of the unit with the other instructional materials and supplies, and they were returned completed to DSA with their post-unit assessments. As with the other data sources, DSA removed student names from the journals before sending them to GWU for analysis.

When completed, the data in Table 7 will show the number of teachers and students who completed the Motion and Forces pre- and post-unit assessments, and the number of students who completed the Student Responsiveness Questionnaire (SRQ) for Year 4 (2005-06 school year). Data pending in Table 7 will be reported in the Year 5 (2006-07 school year) Annual Report. Note that Comparison School D is highlighted in gray in Table 7, because this data was reported as absent from the Year 4 data set.

**Student Demographic Data**

For use in disaggregating the assessment data, the data set for Year 4 contained the following variables, which were sent by DSA to GWU on June 30, 2006:

- Student’s pseudo-ID number (randomly assigned for this study)
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report:  2005 - 2006
July 1, 2006

- School name
- School number
- Course code
- Course description
- Teacher’s name
- Class grades (for 1st, 2nd, 3rd, and 4th marking periods)
- 2005 grade 5 MSA results for this year’s grade 6 students
- Class level (e.g., regular, Honors, Magnet)
- Class period
- Class section number
- Student date of birth
- Ethnicity
- Gender
- FARMS status (Free and Reduced-price Meal Services)
- Special education status
- ESOL status (English for Speakers of Other Languages)

The ESOL and FARMS data have three possible values:
- NEVER - The student has never, past or present, received the service.
- PRIOR - The student has received the service in past years, but did not receive the service this year.
- NOW - The student has received the service at some time during the current school year.

The special education variables indicate whether the student receives:
- No special education services;
- Less than 15 hours of special education services weekly; or
- More than 15 hours of special education services weekly.

All values are mutually exclusive.
Table 6
Year 4 (2005-06) Demographic Characteristics of Participating Schools
Grade 6 Motion and Forces Treatment and Comparison Schools

<table>
<thead>
<tr>
<th>School Profile Category (SPC) Code</th>
<th>Total Enrollment</th>
<th>% Mobility *</th>
<th>% ESOL</th>
<th>% FARMS</th>
<th>% SpEd</th>
<th>% Gender Female</th>
<th>% Gender Male</th>
<th>% Racial/Ethnic Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Af Am</td>
<td>Am Indian</td>
<td>Asian</td>
</tr>
<tr>
<td>TREATMENT SCHOOLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>695</td>
<td>9.3</td>
<td>3.0</td>
<td>11.8</td>
<td>14.0</td>
<td>45.9</td>
<td>54.1</td>
<td>9.5</td>
</tr>
<tr>
<td>B</td>
<td>803</td>
<td>12.3</td>
<td>3.5</td>
<td>30.4</td>
<td>9.2</td>
<td>48.4</td>
<td>51.6</td>
<td>20.7</td>
</tr>
<tr>
<td>C</td>
<td>918</td>
<td>12.8</td>
<td>5.0</td>
<td>27.3</td>
<td>15.4</td>
<td>46.7</td>
<td>53.3</td>
<td>19.4</td>
</tr>
<tr>
<td>D</td>
<td>873</td>
<td>16.4</td>
<td>6.3</td>
<td>39.5</td>
<td>17.2</td>
<td>49.9</td>
<td>50.1</td>
<td>39.3</td>
</tr>
<tr>
<td>E</td>
<td>588</td>
<td>20.4</td>
<td>6.8</td>
<td>44.2</td>
<td>16.8</td>
<td>48.0</td>
<td>52.0</td>
<td>27.2</td>
</tr>
<tr>
<td>COMPARISON SCHOOLS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>994</td>
<td>5.8</td>
<td>2.2</td>
<td>9.3</td>
<td>12.6</td>
<td>46.8</td>
<td>53.2</td>
<td>8.7</td>
</tr>
<tr>
<td>B</td>
<td>845</td>
<td>15.0</td>
<td>5.2</td>
<td>22.0</td>
<td>13.6</td>
<td>48.8</td>
<td>51.2</td>
<td>14.7</td>
</tr>
<tr>
<td>C</td>
<td>776</td>
<td>17.5</td>
<td>5.5</td>
<td>33.1</td>
<td>15.2</td>
<td>48.1</td>
<td>51.9</td>
<td>25.5</td>
</tr>
<tr>
<td>D</td>
<td>741</td>
<td>22.0</td>
<td>9.0</td>
<td>43.2</td>
<td>23.8</td>
<td>46.0</td>
<td>54.0</td>
<td>37.7</td>
</tr>
<tr>
<td>E</td>
<td>647</td>
<td>16.4</td>
<td>7.1</td>
<td>42.8</td>
<td>15.1</td>
<td>49.9</td>
<td>50.1</td>
<td>24.1</td>
</tr>
</tbody>
</table>

All data from Montgomery County Public Schools “2005-2006 Schools at a Glance”
*2004-2005 School Year Data
Note: Post-unit assessments from Comparison School D were not administered during Year 4 and therefore are absent from the data set.
### Year 4 (2005-06) Student and Teacher Participation

**Grade 6 Motion and Forces Treatment and Comparison Schools**

<table>
<thead>
<tr>
<th>School Profile Category (SPC)</th>
<th># of Participating 6th Grade Teachers</th>
<th># of selected students</th>
<th># of complete data sets</th>
<th># of students sent Form A</th>
<th># of complete data sets rec’d from school Form A</th>
<th># of students sent Form B</th>
<th># of complete data sets rec’d from school Form B</th>
<th># of students sent questionnaires</th>
<th># of complete data sets rec’d from school</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TREATMENT SCHOOLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>2</td>
<td>31</td>
<td>pending</td>
<td>51</td>
<td>pending</td>
<td>193</td>
<td>pending</td>
<td>162</td>
<td>pending</td>
</tr>
<tr>
<td>B</td>
<td>2</td>
<td>83</td>
<td>pending</td>
<td>27</td>
<td>pending</td>
<td>222</td>
<td>pending</td>
<td>167</td>
<td>pending</td>
</tr>
<tr>
<td>C</td>
<td>4</td>
<td>21</td>
<td>pending</td>
<td>67</td>
<td>pending</td>
<td>202</td>
<td>pending</td>
<td>194</td>
<td>pending</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>58</td>
<td>pending</td>
<td>29</td>
<td>pending</td>
<td>188</td>
<td>pending</td>
<td>158</td>
<td>pending</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>27</td>
<td>pending</td>
<td>26</td>
<td>pending</td>
<td>164</td>
<td>pending</td>
<td>137</td>
<td>pending</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td><strong>220</strong></td>
<td>pending</td>
<td><strong>200</strong></td>
<td>pending</td>
<td><strong>969</strong></td>
<td>pending</td>
<td><strong>818</strong></td>
<td>pending</td>
</tr>
<tr>
<td><strong>COMPARISON SCHOOLS</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>3</td>
<td>29</td>
<td>pending</td>
<td>52</td>
<td>pending</td>
<td>256</td>
<td>pending</td>
<td>171</td>
<td>pending</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>50</td>
<td>pending</td>
<td>37</td>
<td>pending</td>
<td>214</td>
<td>pending</td>
<td>123</td>
<td>pending</td>
</tr>
<tr>
<td>C</td>
<td>2</td>
<td>51</td>
<td>pending</td>
<td>22</td>
<td>pending</td>
<td>211</td>
<td>pending</td>
<td>134</td>
<td>pending</td>
</tr>
<tr>
<td>D</td>
<td>2</td>
<td>31</td>
<td>NOT GIVEN</td>
<td>0</td>
<td>NOT GIVEN</td>
<td>204</td>
<td>NOT GIVEN</td>
<td>173</td>
<td>NOT GIVEN</td>
</tr>
<tr>
<td>E</td>
<td>2</td>
<td>26</td>
<td>pending</td>
<td>46</td>
<td>pending</td>
<td>160</td>
<td>pending</td>
<td>133</td>
<td>pending</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>12</strong></td>
<td><strong>187</strong></td>
<td>pending</td>
<td><strong>157</strong></td>
<td>pending</td>
<td><strong>1045</strong></td>
<td>pending</td>
<td><strong>734</strong></td>
<td>pending</td>
</tr>
</tbody>
</table>

*Post-unit Assessment Form A = Extended Standard Post-unit assessment (Long Form)*

Post-unit Assessment Form B = Standard Post-unit assessment

**Student Questionnaires** were distributed only to students in the observed classrooms

**Note:** Post-unit assessments from Comparison School D were not administered during Year 4 and therefore are absent from the data set.
Year 4: M&F Implementation

In order to more fully understand the findings from the Motion and Forces Assessment in Year 3, SCALE-uP studied the implementation of M&F in 5 new Treatment schools and compared student outcomes in those schools to student outcomes in 5 new Comparison schools during Year 4 (2005-2006). (Demographic data for these 10 schools are shown in Table 6.) Data collected during Year 4 are currently being analyzed and will be available by August 31, 2006.

Year 4: RRS Implementation

Given the results of the Causes of Seasons Assessment during Year 3, it was decided by the Montgomery County Public Schools in conjunction with the SCALE-uP Principal Investigators, that RRS would not be scaled-up in Year 4. This decision was made on the basis of the experimental results from Year 3, which indicated that RRS is not more effective than the MCPS seasons curriculum materials or the combination of materials staff are presently using. It was not scaled up past the first 5 initial Treatment schools. During Year 4, 7th grade teachers in the remaining 33 schools received the surplus equipment for RRS and Professional Development for the unit was offered on how to instruct the activities in RRS. However, teachers were told they did not have to use the activities in RRS and could use them as they see fit. The data were presented to the Resource Teachers (RTs) and they received instructions on how RRS could be used as supplemental material.

Q2. Ethnographic Research Activities and Findings

As with the experimental data, ethnographic data is collected in one year and analyzed in subsequent years. Therefore, this part of the Annual Report is divided into three sections: Year 2 (which details the Year 2 ethnographic findings for M&F and compares those results to CTA Year 1); Year 3 (which discusses the preliminary analysis of Year 3 ethnographic findings for RRS); and, Year 4 (which outlines the ethnographic activities conducted in Year 4).

Year 2: Analysis of Ethnographic Data for M&F

Codes

Transcription of the video ethnographic data collected in Year 2 from the M&F curriculum unit was completed in Year 4. This data has been coded and analyzed for Class Period Segments and Scientific Term Use. Coding and preliminary analysis has begun for Literacy Events and Object Manipulation and will be completed by the end of Year 4. The codes are defined as the following:

1) **Classroom Period Segments:** Class period segments is a code that allows us to capture when during a class period certain events are likely to be found. Each day is divided into segments labeled “warmup,” “presentation,” “lab activity,” “reflection,” and “transition.”

2) **Scientific Term Use (STU):** According to Project 2061, in order to effectively learn science, children should receive instruction that supports
scientific term use in context. We asked MCPS teachers to list the key terms in each of the three units and coded the transcriptions for these terms and all the terms’ inflections.

3) **Literacy Events**: This code examines students' use of print media in the learning of scientific concepts and is operationalized in two ways ethnographically: “writing” (student puts pen to paper) and “reading” (student reads aloud from the text).

4) **Object Manipulation (OM)**: Since AAAS encourages instructional materials to promote direct hands-on experiences of scientific phenomena, we have developed a code that measures whether, within a sample one-minute period, a student actually comes into contact with a relevant object.

By using these codes as analytical dimensions for interpreting the video and transcription data, we are able to sort, classify and analyze the video data in ways that provide insights into how the unit functions from the standpoint of the students themselves.

**Comparison of CTA Year 1 and M&F Year 2**

*Classroom Period Segments.* Videoethnographic data from Year 1 (CTA) and Year 2 (M&F) have been compared in order to locate salient differences between the ways that these two curriculum units function in the classroom. We recognize that Year 1 is the second year of implementation for CTA, while Year 2 is the first year of implementation for M&F. We also recognize that these units are designed for different grades. However, we believe comparing the two units is a necessary step in understanding how highly rated units function in diverse classrooms.

The most notable distinctions between the implementations of CTA and M&F are the following. The video of M&F depicts a high portion of the class period on group exploration and transition, while the video of CTA demonstrates a larger percent on presentation and reflection than M&F. In other words, CTA devotes more class time to preparing for and reflecting on the explorations, while M&F spends more time on the actual explorations, as well as transitioning between activities (see Figure 3).

The data in Figure 3 aligns not only with the nature of the benchmark ideas (exploring motion and forces versus conservation of matter), but it also aligns with the pedagogical goals of the curriculum developers. In M&F, the goal of the developer (Bruce Ward, ARIES) was to expose children to group exploration—the hands-on experience of model building and repeated trials from which students were expected to individually draw inferences about the more abstract benchmark ideas. In CTA, by contrast, where more time was devoted to reflection, students were expected to participate in the process of discussion and analysis of results as a group.
Figure 3. Total percentage of each coded class period segment for M&F (Year 2) and CTA (Year 1)

*Scientific Term Use (STU).* Consistent with these different allocations of time, the children at the lab table for M&F during Year 2, as shown on the left in Figure 4, used substantially fewer scientific terms (as defined by a sample of teachers of the unit) per turns at talk than did the children in CTA during Year 1, where more time was allocated to reflection and discussion. When more time is provided for discussion and debate, in the context of debating relevant phenomena, then it appears that the children in CTA used more scientific terms in their appropriate context (as shown on the right of Figure 4). Use of scientific terms in context is one of the criteria used by AAAS to evaluate successful forms of instructional support.
Figure 4. Use of scientific terms per turns at talk for students at the focus table of M&F Year 2 (shown on the left) and CTA Year 1 (shown on the right)

**Literacy Events.** In addition, in Year 4 we examined students’ use of print media in the learning of scientific concepts. To do this, we operationalized two types of “literacy events” for ethnographic coding: “writing” and “reading.” To define “writing,” we first divided a sample of the M&F and CTA videos that contained “explorations” (labs) into one-minute segments. Then we noted whether a child put pen to paper during that segment. We then compared the individual children’s writing within and across units (see Figure 5). Next, we defined “reading” as those moments when a child read aloud from the text during those same samples. We compared the children within and across the units (see Figure 6). We found that children in M&F (as shown on the left in Figure 5) engaged in substantially more writing than in CTA (shown in the right in Figure 5), while these same children read from the text less frequently. Our preliminary interpretation is that M&F places heavier literacy demands on children than CTA, with less textual support from the student guide.
Object Manipulation (OM). Because of its labor-intensive nature and because of the rapidly evolving methods and theory, our analysis of object manipulation (OM) is not complete for M&F. Preliminary analysis of Year 2 data, however, suggests that students in M&F engage in more OM overall, but those acts of OM are not linked sequentially with scientific term use (STU) or literacy events in the same way as with CTA.
Year 3: Preliminary Analysis of RRS Ethnographic Data

In Year 3, videoethnographic data were collected on RRS. This data set was transcribed during Year 4, and is currently being coded for class period segments and scientific term use. It will be completed by the end of Year 4 (September 2006). Preliminary analysis suggests that interactionally, RRS is a far more teacher-centered curriculum unit than the other two curriculum units that SCALE-uP has examined (CTA and M&F).

Year 4: Ethnographic Research Activities

Videotaping Process

One classroom from the M&F Treatment schools was selected for videotaping during Year 4. The students in this selected classroom and their parents signed an Internal Review Board (IRB) agreement and video release form prior to filming and were aware of the conditions of the study.

A total of 25 students in a single classroom were filmed for 35 days during Year 4, in a rotating block schedule (a combination of 90- and 45-minute class periods). The unit of study, M&F, lasted 31 days; 4 days of filming took place prior to the beginning of M&F. One camera was used to film the same four students everyday, while another camera rotated around the room filming one of the other tables of three or four students each day.

These video data have been digitized and are currently being transcribed (Summer 2006). Analysis of the video data will begin in Fall 2006 and will continue throughout Year 5.

Interviews

The overall goal of performing student interviews during Year 4 was to gain a better understanding of how students were interacting with the M&F curriculum. These interviews will serve as a tool for analyzing how highly rated units function in the classroom and will give additional context to the ethnographic data. Specifically, we wanted to gain insight on:

- what the students did and did not like about the M&F curriculum unit
- how the students interpreted the various activities of the unit
- how the students interact socially within the classroom
- previous experiences students have had with curriculum-related objects or activities
- interactions the students have had with key concepts outside of class
- misconceptions the students may have had about motion and forces
- how the students work through conceptual trouble both during class and at home
Q3. Experience Research Activities and Findings

SCALE-uP Research Question Q3 asks, "Does the effectiveness of the curriculum unit increase as schools and teachers become more experienced implementing it from the first year to the second year?" A summary of the findings from this research question, which was analyzed during Year 4 of the SCALE-uP grant, is outlined below.

Motion and Forces: Looking Across Two Years of Data (Years 2 and 3)

One of the important SCALE-uP research questions to be explored in Year 4 of the study derives directly from analysis of the previous years’ implementation data. When looking across two years of data for M&F, the experimental results from the Year 2 implementation suggested that students, given the opportunity for concept learning with M&F, outperform their peers who learn with existing curriculum materials. However, there was little practical significance between these groups. Disaggregated data revealed main effects, interaction effects, and greater effect sizes for certain subgroups in the Treatment condition (White, Asian American, Never FARMS). The typically underserved subgroups (Hispanic, African American, Now FARMS) did not seem to gain as much as their peers in the Comparison condition when learning with the M&F curriculum materials in Year 2; but, again, there was no practical significance between groups. In Year 3 however, as noted earlier in this report, there was no statistical or practical significance overall between Treatment and Comparison groups on the concept learning measure. Therefore, based on these two years of data, it appears that the effectiveness of M&F does not increase as teachers and schools become more experienced with the unit (see Figure 7).

In addition, lower overall post-test scores were observed in the Treatment condition in Year 3 as compared to Year 2. In Year 2, Treatment group performance on the post-test improved 15.65 points over the pre-test score, while in Year 3 Treatment group performance improved only 12.98 points from pre-test to post-test. This is further supported by effect size data from both years (Table 8). The overall effect size declined from .10 in Year 2 to -.06 in Year 3. In addition, patterns of effect sizes for subgroups suggest that the effectiveness of the curriculum unit for particular subgroups does not increase with experience. For example, the negative effect sizes observed for African American and Hispanic subgroups in Year 2 are persistent in Year 3 and in fact become more negative. The same is true of the Year 2 and Year 3 effect sizes for the Prior and Now FARMS subgroups as well.
Figure 7. Adjusted MFA scores for Treatment and Comparison groups in Years 2 and 3.

Table 8
Curriculum effect sizes for all levels of Independent Variables for MFA in Years 2 and 3

<table>
<thead>
<tr>
<th>Variable</th>
<th>Year 2</th>
<th>Year 3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>d</td>
</tr>
<tr>
<td>Overall</td>
<td>2169</td>
<td>0.10</td>
</tr>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1108</td>
<td>0.12</td>
</tr>
<tr>
<td>Female</td>
<td>1064</td>
<td>0.08</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>African American</td>
<td>486</td>
<td>-0.04</td>
</tr>
<tr>
<td>Asian American</td>
<td>354</td>
<td>0.17</td>
</tr>
<tr>
<td>Hispanic</td>
<td>394</td>
<td>-0.10</td>
</tr>
<tr>
<td>White</td>
<td>938</td>
<td>0.21</td>
</tr>
<tr>
<td>FARMS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1349</td>
<td>0.23</td>
</tr>
<tr>
<td>Prior</td>
<td>248</td>
<td>-0.08</td>
</tr>
<tr>
<td>Now</td>
<td>575</td>
<td>-0.10</td>
</tr>
<tr>
<td>ESOL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>1717</td>
<td>0.12</td>
</tr>
<tr>
<td>Prior</td>
<td>309</td>
<td>-0.16</td>
</tr>
<tr>
<td>Now</td>
<td>146</td>
<td>0.26</td>
</tr>
<tr>
<td>SPED</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1977</td>
<td>0.13</td>
</tr>
<tr>
<td>Now</td>
<td>195</td>
<td>-0.20</td>
</tr>
</tbody>
</table>

Note. $d =$ Cohen’s $d$ effect size.
For more discussion of the *Motion and Forces* results in Years 2 and 3, see Ochsendorf, Pyke, Lynch, & Watson (2006), an AERA conference paper.

**Seasons: Looking Across Two Years of Data (Years 2 and 3)**

Year 2 experimental data for *RRS* suggests that mean scores are higher on average at the post-test than at the pre-test in both the Treatment and Comparison conditions. However, mean scores were statistically significantly lower for students receiving the *RRS* Treatment curriculum unit when compared to the mean scores for those using other curricula. This means that in Year 2 of the *RRS* implementation, the patterns were highly consistent with the Year 3 patterns reported earlier (see Figure 8). Tests for interactions in Year 2 showed that outcomes can be influenced by student characteristics (SPED). However, the lack of interactions in other cases revealed no change in gaps for Ethnicity, FARMS, and ESOL.

![Figure 8. Adjusted mean scores for RSA between the Comparison and Treatment conditions Year 2 (Trial 1) and Year 3 (Replication). Means adjusted by the pre-test covariate.](image)

While overall post-test scores for *RRS* did increase slightly from Year 2 to Year 3, experimental data from Years 2 and 3 of *RRS* implementation suggest that the effectiveness of the *RRS* curriculum unit, measured relative to the Comparison group does not increase as teachers and schools gain in experience from the first year of implementation to the second year of implementation.

**Q5. Fidelity of Implementation Research Activities and Findings**

Fidelity of Implementation (FOI) is defined as the determination of how well an intervention is being implemented in comparison with the original program design (Mihalic, 2002). During Year 4, the SCALE-uP project collected FOI data during the implementation of the *M&F* unit in order to better understand the relationship between FOI and student outcomes. The SCALE-uP project considers FOI in two ways: *Fidelity to Structure* and *Fidelity to Process.*
Unlike other fidelity scales, which may only measure fidelity to the structure of the program or intervention, the SCALE-uP conceptualization of FOI is compatible with how practitioners think about treatment implementation. This measurement strategy may be especially helpful in solving some of the problems that have cropped up in developing fidelity scales for more complex evidence-based practices, such as teaching. This multiple measure approach avoids shortcomings of program-level fidelity scales that have been criticized as being too focused on structural elements of a curriculum unit and neglect the instructional competence of the teacher and classroom environment, as well as the response that students' have to that implementation. Ultimately, a multiple method approach to FOI is essential for both research and practical purposes. That is, it is important to measure the teacher's fidelity to the structure of the curriculum program and to measure his/her fidelity to its processes (we call this Teacher FOI). It is also important to measure the students' fidelity to the structure of the program (i.e. how much of the program they complete) and to measure the students' perceptions of the level of fidelity to process (we call this Student FOI).

**Fidelity to Structure**

The first way of measuring FOI is to examine Treatment teachers' and students' adherence to the structural components of the curriculum unit. Mobray refers to this as fidelity to structure (Mobray et al, 2003); Sarama and Clements refer to it as "Near Fidelity" or fidelity to the actual program materials (Sarama & Clements, in review). This measure of FOI is independent of Project 2061 criteria and instructional processes. It explores to what extent teachers in Treatment classrooms are implementing the curriculum unit as it was designed by the curriculum developer and to what frequency students are completing the work intended by the curriculum guide.

**Study sample**

During Year 4, 30 Treatment classrooms were randomly selected from a total number of 44 Treatment classrooms (refer back to Table 6). Six classrooms from each Treatment school were selected in order to obtain teacher and student variability.

**Data collection**

Data collection for FOI to Structure included three phases. The first phase involved a series of classroom observations using the Adherence Classroom Observation Protocol (ACOP), which was used to observe each teacher's adherence to the following six structural components of a M&F lesson: Materials (1 item), Recording Ideas (2 items), Procedures in the Student Journal (4 items), Interpreting Ideas (2 items), Procedures in the Teacher's Guide (1 item), and Sequencing (1 item).

Because the ACOP only measures adherence to one lesson, it was important to also capture each teacher's fidelity to the entire unit. To do this, the second phase for understanding teachers' Fidelity to Structure involved interviewing the Treatment teachers about their adherence to the entire M&F curriculum unit. The Adherence Interview Protocol (AIP) was used to guide
our understanding of the extent to which teachers adhered to the following six structural components of the M\&F unit: Materials (1 item), Recording Ideas (2 items), Procedures in the Student Journal (3 items), Interpreting Ideas (2 items), Procedures in the Teacher's Guide (1 item), and Sequencing (4 items). Interviews also included demographic and background data to determine the Treatment teachers' experience, education, formal science knowledge, experience teaching motion and forces.

The final phase for assessing Fidelity to Structure involved counting the number of completed questions in each student's M\&F Student Journal. Journal raters recorded the total number of questions filled in by the student in each of the 18 lessons. Raw data was recorded in Excel (then transferred to SPSS) for the following three components of each lesson of the Student Journal: Recording Ideas, Procedures, and Interpreting Ideas. Over 750 journals were collected and rated.

Research Questions and Planned Analysis

The analysis of the Year 4 Fidelity to Structure will be guided by the following research questions:
1. Is there a relationship between Teacher FOI (as measured by the AIP and ACOP) and Student FOI (as measured by the frequency of completed questions in the Student Journal) in the ARIES: Motion and Forces Treatment classrooms? (bivariate analysis)
2. Is there a relationship between Teacher FOI (as measured by the ACOP and AIP) and student learning in the ARIES: Motion and Forces Treatment classroom? (bivariate analysis).
3. Is there a relationship between Student FOI (as measured by the frequency of completed questions the Student Journal) and student learning in the ARIES: Motion and Forces Treatment classrooms? (bivariate analysis)
4. If statistically significant relationships exist, how much of the variation in student learning is predicted by each variable (Teacher FOI and Student FOI)? (multiple regression--note that only items that are statistically significantly correlated with student learning will be put into the regression model)

Fidelity to Process

The second way to measure FOI, using the same 30 Treatment classrooms described above, is to determine whether there is fidelity to the Project 2061 instructional analysis criteria as enacted in sample Treatment classrooms. Mobra称 refers to this as fidelity to process (Mobra称 et al, 2003); Sarama and Clements refers to it as "Far Fidelity", or fidelity to good teaching practices (Sarama & Clements, in review). Two criteria are used to determine Fidelity to Process- -quality of delivery using the Instructional Strategies Classroom Observation Protocol (ISCOP) and student responsiveness using the Student Responsiveness Questionnaire (SRQ).

Data Collection
ISCOP. The ISCOP measures a teacher’s “quality of delivery” through the observation of classroom instructional strategies. Quality of delivery is defined by the SCALE-uP project as “the extent to which the Project 2061 instructional characteristics are evident during the enactment of a target idea.” The ISCOP consists of 20 items that represent Categories I through V from the Project 2061 Instructional Analysis: (I) Identifying sense of purpose (6 items); (II) Taking account of student ideas (5 items); (III) Engaging students with relevant phenomena (2 items); (IV) Developing and using scientific ideas (5 items); and, (V) Promoting student thinking about phenomena, experiences, and knowledge (2 items). Category I is based on a dichotomous scoring rubric (0 = Not Evident; 1 = Evident). Categories II through V during Year 3 were based on a three-point Likert-like scale (0 = Not Evident; 1 = Evident; 2 = Evident With Emphasis). The ISCOP was revised in Year 4 to include a fourth level for Categories II through V (0 = Not Evident; 1 = Evident; 2 = Evident More than One Time; 3 = Evident With Emphasis).

SRQ. The SRQ was designed to describe students’ direct experience with curriculum materials. SCALE-uP researchers developed 18 items using the language of Categories I through V of the Project 2061 Instructional Analysis. The SRQ attempts to capture whether or not students experience the instructional strategies during the curriculum implementation. Items 1-3 attempted to assess the sense of purpose perceived by students; items 4-6, how well the implementation took account of student ideas; items 7-9, the level to which students were engaged with relevant phenomena; items 10-13, whether or not students used scientific ideas; and items 14-18, if students thought about phenomena during implementation. The items are scored using a 5-point Likert-type scale. The scale ranges from 1-5, one being “Not True” and 5 being “Very True.” All the items are constructed in a positive direction, because previous experience with the study population suggested that questions stated in a negative direction might be confusing.

We hypothesized the presence of five factors in the instrument that corresponded to each of the five categories upon which items were based. To test for these factors, factor analysis, inter-item reliability (Cronbach’s alpha), and face validity were employed. For more information on the specifics of these tests and their results, see Watson, Lynch, Rethinam, and O’Donnell (2006), an NARST conference paper.

Research Questions and Planned Analysis

The analysis of the Year 4 Fidelity to Process data will be guided by the following research questions:

1. Is there a relationship between Teacher FOI (as measured by the ISCOP) and Student FOI (as measured by the SRQ) in the ARIES: Motion and Forces Treatment classrooms? (bivariate analysis)

2. Is there a relationship between Teacher FOI (as measured by the ISCOP) and student learning in the ARIES: Motion and Forces classrooms? (bivariate analysis)

3. Is there a relationship between Student FOI (as measured by the SRQ) and student learning in the ARIES: Motion and Forces classrooms? (bivariate analysis)
4. If relationships exist, how much of the variation in student learning is predicted by each variable? (multiple regression–note that only items that are statistically significantly correlated with student learning will be put into the regression model)

**SQ: Supplemental Grant Research Activities and Findings**

A supplemental grant was received during Year 3 of the SCALE-uP grant, and funding for this supplemental research was received October 1, 2004. The supplemental research grant was acquired to further investigate four of the original research questions during Year 4, in areas such as Hierarchical Linear Modeling (SQ1), Assessment Validation (SQ2), Video Data and Professional Development (SQ3), and Characteristics of Implementation (SQ4). This section of the Annual Report summarizes the activities and findings from each of these four supplemental research questions.

**SQ1. Hierarchical Linear Modeling Activities and Findings**

During Year 4, a more sophisticated exploratory analysis (hierarchical linear modeling [HLM]) of individual, classroom, and school effects was retroactively conducted across samples involved with the RRS and M&F curriculum units during Year 2. In conducting the additional analyses, Dr. Pyke took on additional oversight in the development of the analytic models of the study and collaborated with the new Research Scientist, Dr. Vasuki Rethinam, in carrying out the analyses and communicating the results.

**Year 2 HLM: Motion and Forces Assessment (MSA)**

Because students are nested within classrooms, the Year 2 data of M&F was analyzed using a hierarchical linear modeling (HLM) technique. The null model indicates that the proportion of variance in the individual gain scores attributed to systematic classroom effects is 16%. The proportion of variance in the gain score that exists at the individual level accounted for 84% of the variance. Analysis of the experimental data for the Motion and Forces Assessment (MFA) for Year 2 shows higher gain scores in the Treatment condition. MFA scores are higher when the Treatment curriculum is used than when the Comparison curriculum is used. There was no interaction found between the student demographic variables and the experimental condition. However, HLM analyses yielded a significant main effect for individual and classroom level variables. Students’ FARMS status had a significant effect on gain scores indicating that high SES students gain more than low SES students. Students who were and are eligible for ESOL status had a higher gain score than students in the never ESOL group. Students’ prior science GPA had a negative effect on gain scores. The classroom level that was significant was percent of African American students. Classrooms that had higher percentage of African Americans had greater gains than other classrooms. However, further descriptive analysis indicated that the classrooms with a higher percentage of African American students had low pre-test scores (see Table 9).
Table 9  
Year 2 (2003-2004) Significant Classroom and Individual Level Predictors for the Motions and Forces Unit

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-Value</th>
<th>Effect Size (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Level Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>-.025</td>
<td>.04</td>
<td>.55</td>
<td>…</td>
</tr>
<tr>
<td>Experimental Condition</td>
<td>.24</td>
<td>.09</td>
<td>.01</td>
<td>.60</td>
</tr>
<tr>
<td>% African American</td>
<td>.22</td>
<td>.11</td>
<td>.05</td>
<td>.55</td>
</tr>
<tr>
<td>Individual Level Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FARMS</td>
<td>-.11</td>
<td>.06</td>
<td>.06</td>
<td>-.28</td>
</tr>
<tr>
<td>ESOL</td>
<td>.10</td>
<td>.05</td>
<td>.06</td>
<td>.25</td>
</tr>
<tr>
<td>Prior Science GPA</td>
<td>-.11</td>
<td>.04</td>
<td>.01</td>
<td>-.28</td>
</tr>
</tbody>
</table>

Year 2 HLM: Causes of Seasons Assessment (CSA)

Year 2 data of the *RRS* unit was analyzed using HLM, since students are nested within classrooms. The null model indicates that the proportion of variance in the individual gain scores attributed to systematic classroom effects is 14%. The proportion of variance in the gain score that exists at the individual level accounted for 86% of the variance. Analysis of the experimental data for CSA for Year 2 shows higher gain scores in the Comparison condition (see Table 10). Students in the Comparison condition gained more than the students in the Treatment condition. There was a significant interaction effect for ethnic/race group and experimental condition. Asian, African American, Hispanic students and the reference group (White) in the experimental condition had a differential effect on learning favoring students in the Comparison condition. HLM analyses yielded a significant main effect for individual and classroom level variables. Students’ gender had a significant effect on gain scores indicating that male students had higher gain scores than female students. Students who are eligible for SPED had lower gain scores than students in the never SPED group. Students’ prior science GPA had a positive effect on gain scores. The classroom level that was significant was percent SPED. Classrooms with a higher percentage of students eligible for SPED status had lower gains than other classrooms. However, the effect was very small. Further descriptive analysis indicated that there were classrooms that contained 100% SPED students.

For more discussion on SCALE-uP's HLM analysis of the Year 2 data, see Rethinam, Pyke, and Lynch (2006), an AERA conference paper currently in preparation for publication.
Table 10
Year 2 (2003-2004) Significant Classroom and Individual Level Predictors for the Seasons Unit

<table>
<thead>
<tr>
<th>Independent Variables</th>
<th>Coefficient</th>
<th>Standard Error</th>
<th>P-Value</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classroom Level Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Constant</td>
<td>.014</td>
<td>.03</td>
<td>.67</td>
<td>…</td>
</tr>
<tr>
<td>Experimental Condition (Expt.)</td>
<td>-.44</td>
<td>.07</td>
<td>.0001</td>
<td>-1.16</td>
</tr>
<tr>
<td>Percent SPED</td>
<td>-.06</td>
<td>.03</td>
<td>.04</td>
<td>-.16</td>
</tr>
<tr>
<td>Individual Level Variables</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.13</td>
<td>.04</td>
<td>.002</td>
<td>-.34</td>
</tr>
<tr>
<td>SPED</td>
<td>-.23</td>
<td>.07</td>
<td>.001</td>
<td>-.61</td>
</tr>
<tr>
<td>Prior Science GPA</td>
<td>.21</td>
<td>.03</td>
<td>.0001</td>
<td>.55</td>
</tr>
<tr>
<td>Asian</td>
<td>.06</td>
<td>.06</td>
<td>.322</td>
<td>…</td>
</tr>
<tr>
<td>African American</td>
<td>-.03</td>
<td>.06</td>
<td>.66</td>
<td>…</td>
</tr>
<tr>
<td>Hispanic</td>
<td>.01</td>
<td>.059</td>
<td>.87</td>
<td>…</td>
</tr>
<tr>
<td>Interaction effect a</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Asian x Expt.</td>
<td>-.34</td>
<td>.13</td>
<td>.008</td>
<td>-.57</td>
</tr>
<tr>
<td>African American x Expt.</td>
<td>-.27</td>
<td>.12</td>
<td>.03</td>
<td>-.46</td>
</tr>
<tr>
<td>Hispanic x Expt.</td>
<td>-.34</td>
<td>.12</td>
<td>.005</td>
<td>-.57</td>
</tr>
</tbody>
</table>

* Ethnicity gap is fixed as it did not vary systematically between classrooms. Effect size is computed by dividing each interaction coefficient by the standard deviation of the slope, .59 (for all three ethnic groups), calculated by multiplying standard error for that coefficient shown in the main effects HLM model by the square root of the sample size (98). For example, .06 x √98 = .59 (Lee, Loeb, & Lubeck, 1998).

SQ2. Assessment Validation Research Activities and Findings

One of the threats to internal validity of the SCALE-uP study design has to do with a testing effect. That is, a student's pre-test experience could influence performance on the subsequent post-test. Similarly, there were concerns that a teacher's experience with and knowledge of the pre-test could influence his/her subsequent instruction of the target ideas in a way that might affect student performance in the Treatment or Comparison condition. Consequently, pre-testing was eliminated in Year 4 for all but a small sample of randomly selected classrooms in both the Treatment and Comparison conditions. Data from this small sample of classrooms will be used to explore the presence or absence of a testing effect in the study.

In addition, a further threat to validity is related to the MFA instrument itself. To provide further information about the content validity of the MFA instrument, an additional form of the assessment was created in Year 4. This form (Form A) contains both the original SCALE-uP items in addition to a small sample of multiple choice items judged to be aligned to the same target ideas. These items were developed, analyzed, and revised by an outside organization.
(Project 2061). The other form (Form B) consists of the original SCALE-uP assessment items that have been used in Years 2 and 3 of the study (the MFA). SCALE-uP also eliminated several survey instruments (engagement and motivation and epistemology) that were given in the past.

**Research Questions**

The following research questions related to the collection of validity evidence for MFA will be answered with data collected in Year 4 of the study:

1. Is performance on the post-test significantly higher after experiencing the pre-test as opposed to not experiencing the pre-test?
2. Is understanding of the motion and forces target idea equivalent on all items judged to be aligned to the idea (SCALE-uP and multiple-choice items)?
3. Is understanding, as demonstrated on the written form of the MFA, consistent with understanding, as demonstrated on MFA items during a videotaped, transcribed, and coded interview?

**Methodology**

For the first research question, 14 of the 98 total classrooms were randomly selected from both Treatment and Comparison groups to administer pre-tests (n=379 students). In this analysis, post-test scores for the two groups will be compared (the 14 classrooms that took the pre-test and the 84 classrooms that did not take the pre-test during Year 4).

For the second research question, an additional 14 classrooms were randomly selected during Year 4 from Treatment and Comparison conditions to administer Form A of the post-test (n=328 students). The remaining 84 classrooms were administered the MFA (Form B). In the analysis, student performance on Form A will be judged. Specifically, we will compare how students perform on both sets of items contained in Form A that are all judged to be aligned to the motion and forces target ideas of interest.

For the third research question, a structured interview protocol with several items from Form A was developed and used with a sample of 13 students from one classroom. These interviews were videotaped and transcribed. Analyses will focus on the similarities and differences between student understanding as demonstrated on the written assessment and student understanding as demonstrated during the interview. This will be discussed in more detail below.

**Validation Interviews**

During Year 4 of SCALE-uP, Dr. Curtis Pyke, Rob Ochsendorf, Carol O’Donnell, and William Watson of the SCALE-uP research staff, developed an interview protocol framework for collecting evidence of the content and construct validity of the assessment items being used in the implementation study of M&F. The validation interviews and interview protocol were developed to answer the third research question.
The interviews were conducted with 13 students from the Treatment classroom in May 2006 after the students had the opportunity to complete the M&F unit and take the MFA in their typical classroom setting. The interview protocol was designed so that subjects would first experience the item stimulus, have an opportunity to describe the stimulus, and then provide an answer to the item in writing. After the student responded in writing, the interview protocol was designed to: a) elicit confirmatory evidence of the students’ understanding of the idea of interest, b) elicit confirmatory evidence of the students’ lack of understanding of the target idea of interest c) assess students’ comprehension of the tasks/diagrams, and d) provide other opportunities for students to express their understanding of the target ideas in other contexts (i.e., by using 3-dimensional models or other representations of the phenomena). Each interview was videotaped and will eventually be transcribed and coded in a manner consistent with the goals of the interviews.

For more information on the assessment validation, see Ochsendorf and Pyke (2006), an AERA conference paper.

**SQ3. Video Data for Professional Development and Argumentation**

Under the guidance of Co-principal investigator, Dr. Joel Kuipers, the Discourse Lab staff, over the course of Years 3 and 4 of the SCALE-uP grant, has been researching the processes of argumentation and reasoning as they appear in science classrooms. As a result of this work, Laura Wright, Lindsey Massoud, Joelle Lastica, and Carol O'Donnell will be working on creating a professional development video based on an ethnographic account of argumentation as it appears in the science curriculum units being studied. The staff will choose a few short video clips of argumentation in the classroom and will explain how these relate to other processes in the classroom, such as literacy events, object manipulation, and scientific term use. This video will be created during Year 5 in coordination with Bonnie Hansen-Grafton at MCPS in order to make this product useful to MCPS science teachers and administration.

**SQ4. Characteristics of Implementation Research Activities and Findings**

During Year 3, SCALE-uP researchers acquired a supplemental grant to study the Comparison classroom conditions. At the same time the examination of the Comparison curriculum was taking place, SCALE-uP researchers were seeking to understand surprising results from the Year 2 Treatment implementation of RRS. The unexpected results of RRS raised even more questions about what was happening in the Comparison condition. More specifically, in what ways were the Comparison and Treatment conditions similar and/or alike? Answering this question became part of the focus of the supplemental research for Years 3 and 4.

The investigation of Characteristics of Implementation (COI) during Years 3 and 4 was guided by two research questions:

1. What instructional materials did Comparison teachers use to teach the seasons target idea?
2. Is there a relationship between the characteristics of implementation and the Treatment and Comparison classrooms and student outcomes?

This section is divided into two parts: Year 3 (which outlines the COI data collected from seasons classrooms during Year 3 and analyzed during Year 4), and Year 4 (which describes the sample, methodology, and proposed analysis of the COI data collected from motion and forces classrooms during Year 4).

**Year 3: Characteristics of Implementation of Seasons Classrooms**

**Sample**

During Year 4, SCALE-uP researchers analyzed the COI data that was collected from seasons classrooms during Year 3. In Year 3, 20 Treatment classrooms and 19 Comparison classrooms from 57 Treatment and 55 Comparison classrooms respectively were randomly selected from SCALE-uP schools to participate in the COI study.

**Data Collection and Findings**

It was the contention of SCALE-uP researchers that a combination of interviews, classroom observations, and student questionnaires would be the most comprehensive method of determining whether or not differences exist between the Treatment and Comparison conditions at the classroom level. SCALE-uP researchers developed four instruments for the COI investigation: (1) a Comparison Teacher Interview Protocol (CTIP), (2) an Instructional Strategies Classroom Observation Protocol (ISCOP), (3) a Lesson Flow Classroom Observation Protocol (LFCOP), and (4) a Student Responsiveness Questionnaire (SRQ). Two of these instruments were also used to study Fidelity of Implementation (FOI).

**CTIP.** Questions on the CTIP were organized into five sections: (1) instructional materials; (2) delivery conditions; (3) professional development; (4) experience; and (5) teacher’s knowledge about the SCALE-uP project. The CTIP data from the Comparison teachers who taught the seasons benchmark indicated that they were using a wide variety of materials to teach the seasons target idea. Material selection varied from school to school although teachers within each school tended to follow similar lesson plans for the same amount of time, using the same materials. Some of the most commonly used materials included Prentice Hall Astronomy textbook (readings on the seasons), Bill Nye videos, and MCPS labs drawn from the curriculum guide “Interactions with the Solar System.”

The majority of the teachers draw their labs from the MCPS curriculum guide on astronomy. Although no one reported 100% adherence to the lesson plans in the MCPS county curriculum guide, most had used at least 1-2 labs from the guide. Comparison teachers indicated that the labs offered hands-on learning opportunities for the students. Many of the teachers described their frustrations with the lack of resources to teach about the seasons. Few had a background in astronomy and as a result felt like additional resources would have helped them to better conceptualize the unit and in turn, choose better materials for their students. Several also
vocalized a desire for more trainings and/or opportunities to collaborate with other teachers in the county regarding their material selection.

One major point of variation in the data is the length of time teachers used to teach the seasons target ideas. Since teachers within each school tend to use the same types of lessons, differences were only observable between schools. For example, teachers on a “normal” bell schedule took anywhere from five to nine class sessions to teach the seasons target ideas. This represents a range of 225-405 minutes. Teachers on a block schedule took between two to nine sessions to teach the target ideas. The range here is even greater at 180-900 minutes.

ISCOP. Using the same ISCOP instrument described earlier in this report (see Q5 Fidelity of Implementation), trained observers during Year 3 sampled instruction in both RRS and Comparison classrooms by observing one lesson in each classroom in its entirety. All eight lessons of RRS were observed. Observers recorded whether or not the strategies associated with the Project 2061 instructional criteria were in use during the teaching of the seasons target benchmark. The higher the rating on the ISCOP, the more evidence that the strategy was in use for a given classroom. Inter-rater reliability scores ranged from .63 to .95.

Classroom observations revealed 8 of the instructional strategies listed on the ISCOP were observed overall in the RRS classrooms; 9 were observed overall in the Comparison classrooms. The modes for each instructional strategy observed in the RRS and Comparison classrooms are shown in Table 11. We hypothesized that RRS would support teachers in using the instructional strategies listed on the Project 2061 Instructional Analysis, and this seemed to play out in the classroom. For example, in most cases where RRS scored Satisfactory or Excellent on a criterion, RRS teachers were observed implementing the instructional strategy to some extent—that is, the observation of that instructional strategy was rated as either 1= Evident or 2= Evident with Emphasis. In some cases where RRS scored Poor on a criterion, RRS teachers were not observed implementing the instructional strategy. However, in two of the remaining criteria, the instructional strategy was observed as 1 = Evident overall during RRS classroom implementation, even though the evaluators of RRS rated the criteria Poor. These results were opposite of what was expected.

Where Seasons was rated as Excellent on the Project 2061 Instructional Analysis, the RRS teachers were observed using the instructional strategies supported by the unit to the fullest extent. In two of these cases, the implementation of these instructional strategies was more evident in RRS classrooms than in Comparison classrooms. However, while we hypothesized that there would be distinct differences between instructional strategies in the RRS and Comparison classrooms, few differences were evident. Thus, it appeared that RRS did not provide enough instructional support to distinguish the RRS teachers' use of these strategies from those in the Comparison condition.
Table 11
Mode of each observed instructional strategy in the RRS and Comparison classrooms compared with the RRS Instructional Analysis ratings

<table>
<thead>
<tr>
<th>Instructional Analysis Ratings</th>
<th>Instructional Analysis</th>
<th>Classroom Observation Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excellent=●; Very Good= []; Satisfactory= []; Fair= []; Poor= ○</td>
<td>RRS (rating)</td>
<td>RRS (mode)</td>
</tr>
<tr>
<td>N/R = No Rating, n/a = Not analyzed</td>
<td>0 = Not Evident</td>
<td>1 = Evident</td>
</tr>
</tbody>
</table>

Note: Items 1a-1c are dichotomous (0 and 1).

I. Identifying a Sense of Purpose

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>RRS (rating)</th>
<th>RRS (mode)</th>
<th>Comparison (mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Conveying unit purpose</td>
<td>N/R</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>b. Conveying lesson/activity purpose</td>
<td>●</td>
<td>1</td>
<td>1*</td>
</tr>
<tr>
<td>c. Justifying lesson/activity sequence</td>
<td>◒</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

II. Taking Account of Student Ideas

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>RRS (rating)</th>
<th>RRS (mode)</th>
<th>Comparison (mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Attending to prerequisite knowledge and skills</td>
<td>○</td>
<td>0</td>
<td>1*</td>
</tr>
<tr>
<td>b. Alerting teacher to commonly held student ideas</td>
<td>N/R</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>c. Assisting teacher in identifying own students’ ideas</td>
<td>○</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>d. Addressing commonly held ideas</td>
<td>◒</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

III. Engaging Students with Relevant Phenomena

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>RRS (rating)</th>
<th>RRS (mode)</th>
<th>Comparison (mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Providing variety of phenomena</td>
<td>○</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>b. Providing vivid experiences</td>
<td>◒</td>
<td>2*</td>
<td>2</td>
</tr>
</tbody>
</table>

IV. Developing and Using Scientific Ideas

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>RRS (rating)</th>
<th>RRS (mode)</th>
<th>Comparison (mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Introducing terms meaningfully</td>
<td>◒</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>b. Representing ideas effectively</td>
<td>●</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>c. Demonstrating use of knowledge</td>
<td>◒</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>d. Providing practice</td>
<td>○</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

V. Promoting Student Thinking about Phenomena, Experiences, and Knowledge.

<table>
<thead>
<tr>
<th>Instructional Categories</th>
<th>RRS (rating)</th>
<th>RRS (mode)</th>
<th>Comparison (mode)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Encouraging students to explain their ideas</td>
<td>○</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>b. Guiding student interpretation and reasoning</td>
<td>●</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>c. Encouraging students to think about what they’ve learned</td>
<td>○</td>
<td>n/a</td>
<td>n/a</td>
</tr>
</tbody>
</table>

Note: Means, and not modes, were used to test for statistically significant differences. Asterisk indicates statistically significant difference at \( p < .05 \).
The following items on ISCOP were found to be significantly correlated with class gain score for the Treatment classrooms:

1. Ia. The teacher communicates the purpose of the lesson to students by displaying it in writing (positively correlated with class gain score $r = .47$).
2. IIb. The teacher asks students to make predictions or questions them about their prior knowledge individually though a writing activity (negatively correlated with class gain $r = -.53$).

There was no significant correlation found between items on ISCOP and students leaning in comparison classrooms.

**LFCOP.** During Year 3, SCALE-uP researchers were only able to conduct Lesson Flow observations in five RRS Treatment and seven Comparison classes in the sample. The LFCOP is designed to capture how the critical features of the Treatment implementation differ from the standard or more traditional curriculum implementation. These critical features include topic (on or off), centeredness (teacher, student, or group) and type of activity (lab-based or non lab-based). Descriptive results from the LFCOP showed that, on average, 58% of the lesson time was spent on teacher-centered activities in Comparison classes. In Treatment classrooms, about 71% of the time was teacher-centered. One item on the LFCOP was found to be significantly correlated with class gain score for the RRS Treatment classrooms. Individual student centeredness was found to be highly correlated with class mean gain scores ($r = .90$). There was no significant relationship found between LFCOP and student learning in the Comparison classrooms.

**SRQ.** Using the same SRQ described earlier in this report (see Q5. Fidelity of Implementation) responses were collected for each individual student and were aggregated to the classroom level. An independent samples t-test was conducted for each subscale of the SRQ to test for the differences in class mean experience with curriculum materials between the Treatment and Comparison classrooms. Only one of the five subscales was found to be significantly different between the Treatment and Comparison classrooms. “Developing and using scientific idea” was found to be significant, $t(32) = 2.03$, $p < .05$. The mean was higher for Comparison condition ($M = 3.45$) than the Treatment ($M = 3.19$). Correlation analysis was conducted using the 5 SRQ subscales and class mean gain. Table 12 outlines the three subscales that were significantly correlated with class mean gain scores for the Treatment classrooms:

| Table 12: SRQ subscales that correlated with class mean gain scores |
|-------------------------|-------------------------|
| Items                   | R                       |
| Identifying a sense of purpose | .53*                   |
| Engagement with relevant phenomena | .62**                  |
| Developing and using scientific ideas | .55*                   |

Note: * $p < .05$, ** $p < .01$
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

There was no significant correlation found between the 5 subscales and class mean gain for the Comparison classrooms.

The results from the Year 3 COI study provide valuable information regarding similarities and differences between the Treatment and Comparison Seasons classrooms. The investigation further aided in SCALE-up’s understanding into why the RRS unit did not appear to be more effective than the Comparison condition.

See the Characteristics of Implementation (COI) Internal Report on the SCALE-uP web site for complete information. See also the conference paper written by Lynch, O'Donnell, Merchlinsky, Hatchuel, Rethinam, & Watson (2006).

Year 4: Characteristics of Implementation of Motion and Forces Classrooms

Research Questions

For Year 4, SCALE-uP researchers decided to turn their focus towards the Motion and Forces unit. The COI research questions from Year 3 were expanded upon during Year 4 to include the following:

1. What curriculum materials are used in the Comparison condition to teach the motion and forces target ideas? What are the delivery conditions and types of professional development opportunities attended by teachers?
2. Are there differences in demographic and background characteristics between teachers in Treatment and Comparison groups?
3. Are there differences between Treatment and Comparison classrooms on the variables of characteristics of implementation?
4. Is there a relationship between the variables of characteristics of implementation and student learning?
5. If relationships exist, how much of the variation in student learning is predicted by each variable?

Study sample

Thirty Treatment and 30 Comparison classrooms were randomly selected from a total number of 93 classrooms (49 Comparison and 44 Treatment classrooms, respectively) in the 10 Year 4 schools (refer back to Table 6). Six classrooms from each school were selected in order to obtain teacher and student variability.

Data collection

Similar to Year 3, data collection for the Year 4 COI investigation was a two-stage process. The first stage involved the CTIP, which surveyed the Comparison teachers about their use of curriculum materials to teach the Motion and Forces essential indicators. The next stage of the COI investigation involved a series of classroom observations (ISCOP and LFCOP) during
the Motion and Forces unit and the distribution of student questionnaires (SRQ). An additional instrument was created to obtain “duration” data from the teachers, defined as the total number of minutes spent on teaching the target benchmark. Each teacher completed a Duration Log indicating the date he/she started and stopped teaching the motion and forces target ideas and the number of minutes he/she taught the motion and forces target idea on each date.

**Proposed data analysis**

Data has not yet been analyzed for Year 4, but will be analyzed using correlation and regression.

**Opportunities for Training and Development**

**Professional Development**

Montgomery County Public Schools (MCPS) science education staff in conjunction with The George Washington University (GWU) researchers provided opportunities for training and Professional Development (PD) to teachers throughout the SCALE-uP project.

The model of PD used in MCPS is a combination of direct PD and "train the trainers." For the purposes of the grant, 19 Instructional Teachers (IT)–approximately six per grade level–were selected specifically to facilitate curriculum implementation for the SCALE-uP project. Each IT agreed to serve as a PD specialist and, over the life of the five-year grant, has provided leadership and training for the curriculum unit being implemented in his/her grade level. Note that during Year 4, one IT did not complete the year and was replaced with an instructor in the Treatment group.

Throughout the 2005-2006 school year, the IT’s for each unit, in conjunction with the MCPS SCALE-uP Project Specialist, Bonnie Hansen-Grafton, met to plan the professional development for the curriculum modules. The ITs also communicated with the curriculum developers either through e-mail or in person to create PD opportunities for the Treatment schools selected for Year 4 implementation.

To move teachers along the novice to expert continuum--from Initial Implementation to Replication to Middle (or Mid-) Implementation, to Full Scale--MCPS and GWU researchers standardized plans for a series of PD institutes and workshops. Each PD workshop was assigned a letter (A – G) to designate its plan of implementation. Table 13 outlines the SCALE-uP Professional Development attendance for Years 1 through 4 for each type of implementation. The seven PD Plans, established in Year 1, are outlined as follows (see the Year 2 Annual Report for full descriptions of each code):

**Initial Implementation Year:**

A = Professional Development Plan A: Summer Institute Session I Initial Implementation PD with unit Curriculum Developers (two-day)
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

B = Professional Development Plan B: Refresher - Intro PD (one day) (all Treatment teachers invited but geared toward new teachers)
C = Professional Development Plan C: Mid-unit Content Training (all teachers--Treatment and Comparison invited) (after school 3 hours)
D = Professional Development Plan D: Follow-up Unit Session I Initial Implementation PD (before or after school 3 hours)

Replication Year:
DD = Professional Development Plan DD: Follow-up Unit Session I Replication Year PD (before or after school 3 hours)

Mid-Implementation Year:
E = Professional Development Plan E: Summer Institute Session II Mid-Implementation PD with Unit Curriculum Developers - (two-day)
F = Professional Development Plan F: Follow-up Unit Session II Mid-Implementation PD (before or after school 3 hours)

Table 13
Years 1 - 5 SCALE-uP Professional Development Attendance

<table>
<thead>
<tr>
<th>Grade</th>
<th>Number of Teachers</th>
<th>Yr 1</th>
<th>Year 2</th>
<th>Year 3</th>
<th>Year 4</th>
<th>Year 5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>8</td>
<td>6</td>
<td>6</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>Summer Institute Invited</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>E</td>
<td>A</td>
</tr>
<tr>
<td>Attended</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>E</td>
<td>E</td>
<td>A</td>
</tr>
<tr>
<td>Refresher Invited</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Attended</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>Mid-unit Content Training Invited</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Attended</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>Follow-up Invited</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Attended</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
<td>D</td>
</tr>
<tr>
<td>Total Teachers Invited</td>
<td>45</td>
<td>90</td>
<td>162</td>
<td>249</td>
<td>41</td>
<td>47</td>
</tr>
<tr>
<td>Total Attended</td>
<td>29</td>
<td>90</td>
<td>125</td>
<td>148</td>
<td>39</td>
<td>31</td>
</tr>
</tbody>
</table>

+ Indicates that all MCPS teachers (Comparison and Treatment) of this grade level were invited. The number listed in the invited row reflects the number of Treatment teachers only. The number listed in the attended row reflects the number of Treatment and Comparison teachers.
Table 14 outlines a list of professional development institutes conducted for SCALE-uP throughout Year 4.

### Table 14

<table>
<thead>
<tr>
<th>Plan</th>
<th>Unit and Event</th>
<th>Date</th>
<th>Audience</th>
<th>Topic</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>CTA “refresher”</td>
<td>9/27 – 9/29 three sessions All Day</td>
<td>8th grade staff new to the 36 schools or teaching 8th grade</td>
<td>Curriculum training and fidelity of implementation</td>
<td>32 teachers B. Hansen</td>
</tr>
<tr>
<td></td>
<td>CTA scoring</td>
<td>This was handled by the IRT’s in the building with 1 IT going out to schools which needed help</td>
<td>8th grade staff new to the 36 schools</td>
<td>Scoring for CTA</td>
<td>32 teachers IRT’s 1 IT</td>
</tr>
<tr>
<td>E</td>
<td>CTA “during”</td>
<td>Did not occur this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>CTA “after”</td>
<td>Did not occur this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>GEMS “refresher”</td>
<td>4/25-27 3 sessions</td>
<td>Staff who wanted to use materials and were not ever invited to PD</td>
<td>Curriculum training and fidelity of implementation</td>
<td>27 staff 6 IT’s B Hansen-Grafton</td>
</tr>
<tr>
<td>C</td>
<td>GEMS “during”</td>
<td>Did not have this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>GEMS “after”</td>
<td>Did not have this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>ARIES IT planning</td>
<td>12/09/05 12/02/05</td>
<td>ARIES IT’s and IRT’s</td>
<td>Planning on PD for the 2005</td>
<td>ARIES IT’s B. Hansen GWU staff</td>
</tr>
<tr>
<td>B</td>
<td>ARIES “refresher”</td>
<td>9/10 – 12/06</td>
<td>Six Year 4 Treatment staff and 18 other</td>
<td>Curriculum training and fidelity of implementation</td>
<td>ARIES IT’s B. Hansen GWU staff</td>
</tr>
<tr>
<td>C</td>
<td>ARIES planning “during”</td>
<td>Did not occur this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>ARIES “after”</td>
<td>Did not occur this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>Summer Institute Curriculum training</td>
<td>Did not occur this year</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The Effect of Scale up on the Proposed PD

At the close of each year of the SCALE-uP grant, MCPS holds a PD summer institute for the following year's curriculum unit implementation. MCPS did not hold a summer institute in 2006 for the following reasons. First, the results from the Year 4 implementation of M&F, which should be complete by the end of August 2006, will determine whether M&F will be scaled up. If the findings show significant gains for students in the M&F Treatment group, PD for new M&F users will be held in the winter of Year 5. If findings do not show significant gains for M&F, it will not be scaled up in MCPS and no PD will be necessary. Second, CTA has been fully scaled-up in MCPS, therefore the PD has scaled back; and, per the PD plan, no summer institute is held during scale-up. Once a unit is institutionalized, only new staff are offered PD, which many times is completed in each school. Finally, RRS was not scaled–up during Year 4; therefore, there was no follow-up training except for teachers who had a desire to integrate the RRS lessons into a larger unit.

COMA Training

During the Fall of 2005, 25 MCPS 8th grade middle school science teachers, who were new to teaching CTA, attended a one-day PD. The goal of the PD was to familiarize the teachers with the CTA unit and with the goals of the Conservation of Matter Assessment (COMA). Specifically, teachers were trained on how to administer the COMA assessment to their students, how to understand the features of the assessment items, and how to use the rating guides to score the assessments. COMA Manuals (prepared by SCALE-uP) were distributed to each teacher for their use in recording pre and post-test scores for each of their students. These MCPS teachers, during Year 4 of CTA implementation, used the COMA manual and the COMA assessment actively in their classrooms as an integral part of the curriculum. Rob Ochsendorf, GWU Research Assistant, led this professional development. The training took place in MCPS on September 27-29, 2005.

Outreach Activities

MCPS Outreach Activities

In addition to professional development activities, GWU researchers and MCPS science education staff presented a summary of data and outcomes from the SCALE-uP Grant Years 0 - 3 to administrators, middle school (MS) and high school (HS) teachers, Department of Shared Accountability (DSA) staff, the Board of Education, and project staff of other MCPS grants. These presentations served as a form of outreach for the project. Table 15 outlines a list of outreach activities conducted for the SCALE-uP project during Year 4.
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

Table 15
Year 4 Outreach Activities for the SCALE-uP Project

<table>
<thead>
<tr>
<th>Event</th>
<th>Date</th>
<th>Audience</th>
<th>Topic</th>
<th>Number of Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>IRT/RT meetings</td>
<td>9/14, 10/19, 11/16, 12/14, 1/18, 2/8, 3/8, 4/19</td>
<td>HS RTs, MS IRTs, Select science teachers</td>
<td>Updates on SCALE-uP Project</td>
<td>40-100+ science teachers</td>
</tr>
<tr>
<td>Middle School Principals meetings</td>
<td>Quarterly meetings</td>
<td>37 Middle school principals</td>
<td>Middle school programs and issues</td>
<td>37 middle school principals</td>
</tr>
<tr>
<td>Board of Education (BOE) presentation</td>
<td>April, 2006</td>
<td>BOE</td>
<td>Present budget from the SCALE-uP project year 1</td>
<td>BOE</td>
</tr>
<tr>
<td>Maryland Department of Education Briefings (MSDE)</td>
<td>Fall 2005, Winter and Spring 2006</td>
<td>MSDE science leadership Maryland counties science supervisors</td>
<td>Participated in Maryland State Science Curriculum programs</td>
<td>30 Science Supervisors 3 MSDE staff</td>
</tr>
<tr>
<td>Maryland State Supervisors Association (MSSA)</td>
<td>5/10/06 All day Fall, Winter Spring</td>
<td>Science supervisors from each MD County</td>
<td>Discussions about the State Voluntary Curriculum</td>
<td>35 supervisors and staff</td>
</tr>
<tr>
<td>DCI/OCIP</td>
<td>Second Tuesday of month</td>
<td>All Central Office curriculum staff</td>
<td>Share the SCALE-uP Project with specialists in other disciplines</td>
<td>100</td>
</tr>
<tr>
<td>Science office meetings</td>
<td>Every Wednesday 9-11</td>
<td>Science office staff--elementary, middle, and high school</td>
<td>Share weekly updates about the SCALE-uP Project</td>
<td>6</td>
</tr>
<tr>
<td>GWU weekly/monthly Grant meetings</td>
<td>varies</td>
<td>GWU grant staff</td>
<td>Various presentations from grant staff</td>
<td>varies</td>
</tr>
<tr>
<td>MCPS Grant staff meetings</td>
<td>First Tuesday Of the month 1-2</td>
<td>Michael Szesze Bonnie Hansen Gary Hedges Angela McKinley Gretchen Blankenship</td>
<td>To share info about the 2 grants with secretarial staff; inform others on progress</td>
<td>6</td>
</tr>
<tr>
<td>Curriculum Advisory</td>
<td>4 times a year</td>
<td>Community</td>
<td>Share curriculum projects with the community</td>
<td>Varies 5-20</td>
</tr>
<tr>
<td>NSTA 2006</td>
<td>4/03/06</td>
<td>National science teachers</td>
<td>Using data to improve student achievement</td>
<td>science teachers nationally</td>
</tr>
</tbody>
</table>

One science Resource Teacher (RT) from each of the 38 middle schools--a position not dependent upon the SCALE-uP project--participated in many of the outreach activities. Each RT serves as a liaison between the teachers in his/her own school and the MCPS SCALE-uP Instructional Specialist, Bonnie Hansen-Grafton. In addition, RTs provide leadership in implementing the MCPS curricula, conduct classroom observations, and assist teachers in the implementation of professional development plans. On July 5, 2006, 75 high school and middle school RT's convened at Walt Whitman High School for a three-day workshop. Dr. Sharon Lynch, Dr. Gail Brendel Viechnicki, William Watson, and Carol O'Donnell of the SCALE-uP staff conducted a 90-minute presentation briefing the RT's about the results of the SCALE-uP
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

grant, Years 0-3, including describing how these units function ethnographically in actual classrooms. The goal of this presentation was to help MCPS RT’s to understand the effectiveness of the units both experimentally and ethnographically and to apply these results to their own classroom experiences.

GWU Outreach Activities

Sixteen masters students and one doctoral student participated in the graduate level course, TRED 217, Recent Developments in Secondary Science Education, taught by SCALE-uP Principal Investigator, Dr. Sharon Lynch, Summer 2006. These participants had access to the SCALE-uP ethnographic video data set as a part of their class project on student science discourse and reasoning from evidence.

Dr. Lynch served on the Planning Committee for the NSF-administered Interagency Education Research Initiative (IERI) Principal Investigators’ Conference on scale-up, which is scheduled for August 30-31, 2006.

Dr. Lynch has proposed and had accepted a four-paper science education policy section for the journal Science Education. She will write the lead framing article, as well as one of the policy research articles. The other two authors for the papers are Okhee Lee (University of Miami FL) and Nancy Songer (University of Michigan) as well as consulting with Sarah Kay MacDonald of the University of Chicago, DRDC/NORC, and Janet Earle, Program Officer for SCALE-uP, of the National Science Foundation. If the series is successful, then Science Education plans on using a similar format to explore other NSF-funded policy research.

Carol O’Donnell discussed the SCALE-uP work on Fidelity of Implementation at the National Science Foundation (NSF) Instructional Materials Development Conference in February, 2006, sponsored by the American Geological Institute (AGI) and NSF. A booklet outlining the proceedings and Carol’s invited keynote on Fidelity of Implementation at this same conference in February 2005 was published by NSF/AGI in October, 2005 and distributed to all participants at the 2006 conference.

In June 2006, Dr. Sharon Lynch was invited to speak at the Educational Development Center Curriculum Development Conference, sponsored by the National Science Resources Center in Baltimore, MD, on the role of curriculum evaluation in secondary science. Proceedings from the conference will be published in a monograph.

Dr. Curtis Pyke, Dr. Sharon Lynch, and Carol O’Donnell served as panel reviewers for NSF Research and Evaluation on Education in Science and Engineering (REESE) Proposals (June, 2006), which includes proposals for both empirical research and synthesis studies.
Conference Papers:

Other outreach efforts include the following conference paper presentations by SCALE-uP researchers at national meetings during Year 4:


Year 4 Annual Report: 2005 - 2006
July 1, 2006


PUBLICATIONS AND PRODUCTS

- Journal publications
- Books or other non-periodical, one-time publications
- What Web site or other Internet site have you created?
• What other specific products (databases, physical collections, educational aids, software, instruments, or the like) have you developed?

Journal Publications

Publications:


Kuipers, J., Ed. (in review). Objectification and the inscription of knowledge in science classrooms.


Books or other non-periodicals and one-time publications

conducted at the meeting annual meeting of the American Association for the Advancement of Science.


Citations


Web site

Two web sites have been developed to support the SCALE-uP study:

1. The George Washington University SCALE-uP Project Web Site:
   
   http://www.gwu.edu/~scale-up

2. The Montgomery County Public Schools Science, K-12 SCALE-uP Project Web Site:
   
   http://www.mcps.k12.md.us/curriculum/science/

Other Specific Products (databases, physical collections, educational aids)

The SCALE-uP grant has developed the following research tools:

Q1. Experimental Research Tools

1. Assessments
   a. Causes of Seasons Assessment (CSA)
   b. Motion and Forces Assessment (MFA)
   c. Conservation of Matter Assessment (COMA)
d. Science Learning Orientation and Engagement Questionnaire (SLOESQ), an instrument composed of previously validated scales (Marks, 2000; Midgely et al., 2000).

2. Validation interview protocol
   a. Conservation of Matter
   b. Motion and Forces
   c. Causes of Seasons

Q2. Ethnographic Research Tools

Q5. Fidelity of Implementation
   4. Adherence Classroom Observation Protocol (ACOP)
   5. Adherence Interview Protocol (AIP)
   6. Instructional Strategies Classroom Observation Protocol (ISCOP)

SQ. Characteristics of Implementation Tools
   7. Instructional Strategies Classroom Observation Protocol (ISCOP)
   8. Lesson Flow Classroom Observation Protocol (LFCOP)
   9. Student Responsiveness Questionnaire (SRQ)
   10. Comparison Interview Protocol (CIP)
   11. Duration Log

See the Supplemental Grant Research Activities and Findings for more information on each of these instruments.

CONTRIBUTIONS

Below are the unique contributions, major accomplishments, innovations and successes of SCALE-uP relative to:

- the principal discipline(s) of the project;
- other disciplines of science or engineering;
- the development of human resources;
- the physical, institutional, or information resources that form the infrastructure for research and education;
- or other aspects of public welfare beyond science and engineering, such as commercial technology, the economy, cost-efficient environmental protection, solution of social problems;
- the principal discipline(s) of the project;

Principle Discipline of the Project

In the last year, SCALE-uP has made three principle contributions to the field of science education research:

1. Results from this project have been in high demand for professional development for university professors of science education or state level science educators in
the area of “evidence-based research”. This is one of the few studies using both quasi-experimental evidence and video data to demonstrate the effectiveness of science curriculum materials. Because the first unit studied has consistent evidence of effectiveness and is a guided inquiry unit heavily dependent on laboratory experiences, this has also raised some interest in the field and we anticipate more attention from the field of special education, given our recent publication efforts.

2. In Year 4, we completed a major effort to explore the “comparison conditions” in our quasi-experimental study. This includes both teachers and students in the comparison condition. While this was prompted by unexpected results of our study, it should focus attention on the importance of understanding the comparison condition in quasi-experimental research.

3. Sharon Lynch organized a paper set to summarize the research of the SCALE-uP grant and presented the papers at NARST, which was well-received.

**Other Discipline of Science or Engineering**

None

**Development of Human Resources**

See the list of GWU researchers for a list of doctoral students and masters students involved in our grant. In addition, we have supported the academic work of one Post-Doctoral Researcher, Gail Viechnicki, Ph.D., Linguistics, University of Chicago; and, one graduate student:

1. Raluca Teodorescu, Graduate Student, Physics Department, George Washington University

**Resources that Form the Infrastructure for Research and Education**

**Blackboard**

During Year 4, SCALE-uP continued to monitor a project communication site through Blackboard as a way to share research across the university. The site includes video clips categorized by topic, articles and other publications that serve as the foundation of our research base, and discussion boards. This site is password protected.

**Writers' Workshops**

During Year 4, Dr. Sharon Lynch, Dr. Joel Kuipers, and Dr. Curtis Pyke instituted the SCALE-uP Writers' Workshops for SCALE-uP staff. These workshops entailed small working groups around specific conference papers intended for publication. Papers ready for consideration were presented to the full SCALE-uP staff for feedback and critique. The Writers' Workshops
have included faculty and students from GSEHD DTPSE (~8), two faculty members, and a post doc; a faculty member from CSAS Department of Anthropology, a post doc, and two graduate students. Workshops were held throughout the Year 4 academic year from 10:30 – noon (unless otherwise noted) in the Discourse Lab of GWU. All students and faculty with a general interest in the study of curriculum through either quantitative or qualitative methods were encouraged to attend. The workshop topics and presentations included:

February 2, 2006: The Impact of a Middle School Motion and Forces Curriculum Unit on Student Outcomes: Results From Consecutive Quasi-Experimental Studies. Rob Ochsendorf, Curtis Pyke, Sharon Lynch, Bill Watson.
March 30, 2006: Are "Highly-Rated" Middle School Science Curriculum Materials Effective and For Whom? Results From a Set of Implementation Studies. Sharon Lynch.

These workshops have enhanced the productivity and interdisciplinary nature of the grant. Several articles have been submitted for publication, and there have been a raft of conference presentations as a result of the writers' workshops. Please see peer reviewed conference presentations for results of this endeavor.

**Aspects of Public Welfare Beyond Science and Engineering**

None

**Special Requirements**

None
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

Change in Objectives of Scope

During Year 4 of the grant, on March 30, 2006, the Co-Principal Investigator, Michael Szesze, of MCPS, accepted a new position with the Department of Defense Schools and left the MCPS school system. The interim Co-PI is Project Specialist Bonnie Hansen-Grafton of MCPS.

Un-obligated funds:

None

Animal, Human Subjects, Biohazards

All necessary informed consent forms and IRB applications have been submitted to the IRB office and approved by IRB as needed. Copies of all approved IRB forms are available upon request.
Scaling up Curriculum for Achievement, Learning, and Equity Project

Year 4 Annual Report: 2005 - 2006
July 1, 2006

References


