

Running head: HIGHLY RATED SCIENCE CURRICULUM AND STUDENT
DIVERSITY

Examining the Effects of a Highly Rated Science Curriculum Unit on Diverse Students:

Results from a Planning Grant

Sharon Lynch, Joel Kuipers, and Curtis Pyke

The George Washington University

Michael Szesze

Montgomery County Public School

This work is supported by the National Science Foundation, the U.S. Department of Education, and the National Institute of Health (REC-0228447). Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the position, policy of endorsement of the funding agencies.

Please address correspondence to slynch@gwu.edu.

Abstract

This paper reports on the results of a planning grant that studied the effects of a highly rated curriculum unit on a diverse student population. The treatment was introduced to 1500 8th grade students in five middle schools selected for their ethnic, linguistic, and socio-economic diversity. Students were given pre-, post- and delayed posttests on a Conservation of Matter Assessment and measures of motivation and engagement. This quasi-experiment found statistically significant posttest results for achievement, basic learning engagement, and goal orientation. Analysis of disaggregated data showed that subgroups of students in the treatment condition outscored their comparison group peers (n =1500) in achievement in all cases, except for students currently enrolled in ESOL. Analysis of video data of a diverse group of four students as the unit was enacted suggested that students entered a learning environment that permitted them to function in different, but consistent ways over time; i.e., the frequency of students' manipulation of objects showed a different pattern of engagement for each of the four students compared with patterns of verbal responses such as the use of scientific terms. The results of this planning grant paved the way for a large study of the scale-up of highly rated curriculum units.

Examining the Effects of a Highly Rated Science Curriculum Unit on Diverse Students:
Results from a Planning Grant

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. Because “science for all” does not necessarily mean that “one size fits all,” this research was designed to explore the effects of providing a high-quality curriculum unit to students in schools characterized by high levels of student diversity, as well as studying its effects on specific groups of students. Student “diversity” includes socioeconomic status (SES), ethnicity/race, gender, and students’ status as English language learners or having identified disabilities. Our goal was to measure students’ mastery of a specific concept (targeted by a content standard) as well as engagement and motivation outcomes. Moreover, through ethnographic study, we sought to understand how the curriculum unit functioned for diverse learners in one of the classrooms.

National and International Studies

Although there have been a variety of approaches to systemic science education reform, many efforts have lacked specific attention to the learning needs of diverse students, despite having made equity a goal (noteworthy exceptions exist and some are mentioned, herein). Unfortunately, the reform has not produced persuasive evidence of great improvements in student achievement overall, nor of narrowing achievement gaps. According to the Glenn Report (National Commission on Science and Mathematics

Teaching for the 21st Century, 2000), U.S. students' performance in science and mathematics is unacceptable and we are losing ground. Recent evidence from Third International Mathematics and Science Study-Repeat (TIMSS-R, 2000) showed U.S. students far from the goal of becoming first in the world in science and mathematics. The most recent science results for eighth graders on the National Assessment of Educational Progress (NAEP) showed that achievement gaps actually increased from 1996 to 2000. Students eligible for free/reduced price lunch programs (a proxy for socioeconomic status) decreased significantly in average score, while the scores of more affluent students increased significantly. For ethnicity/race, the picture for eighth graders in 2000 has not much improved from 1996, with about 40 point gaps between White and Black students, and 33 point gaps between White and Hispanic students (National Center for Educational Statistics, 2003).

Evaluation data from NSF-funded State Systemic Initiatives (SSIs) have not been able to demonstrate robust increases in science achievement nor closing of achievement gaps, despite the fact that these two goals were NSF "drivers" for the SSI reform efforts and federal funding. While one might argue that the NAEP data are not able to distinguish between states where science education reform was a major effort and those less active, neither have the SSIs been able to show progress (increased outcomes) for these drivers. An SSI evaluation (Laguarda, 1998) found that only seven of the 25 SSIs were able to generate credible evaluation evidence, and only four of those showed some modest evidence of shrinking the achievement gap—Ohio, Puerto Rico, Louisiana, and New Mexico. These states had made equity a central focus of their evaluations and collected data to examine gap questions by gender, ethnicity/race, and poverty level.

Why have the reforms not yielded more positive results? Admittedly, perhaps reform efforts have not had sufficient time to take effect, funding for such ambitious reforms was insufficient, and changing embedded educational practices in the U.S. seems a glacially slow process, with our decentralized system of education. However, given the number and scope of systemic change efforts (state, urban, local, and rural), one would hope to find improved outcomes. International comparisons reveal that the factors most often cited for poor performance in U.S. school science are curriculum structure and textbooks, instructional approaches (U.S. National Research Center for TIMSS, 1996) and the quality of our teaching staff (National Commission on Science and Mathematics Teaching for the 21st Century, 2000). TIMSS showed that not only did U.S. students' performance lag behind other countries in science and math, the quality of U.S. curricula and science textbooks is sorely lacking (U.S. National Research Center for TIMSS, 1996). At the same time, instruction in U.S. science classrooms is heavily dependent upon textbooks. For instance, 54% of eighth grade science teachers reported that they rely primarily on texts (National Educational Goals Panel, 1994). Assigned reading of textbooks is the most frequently given homework assignment, 77% of the time (Wenglinsky, 2000).

A factor that interacts with curriculum and instruction to stratify student achievement is the inability of the educational system to respond in an effective and timely manner to the changing diversity of the K-12 student population. There is a lack of credible, replicable research on effective instructional programs in science or mathematics for various diverse student populations (Lee & Luykx, 2004). However, it is possible to identify conditions that may affect this problem (Lynch, 2000): The less

skilled and experienced the teachers, the more they are likely to rely upon textbooks, and such teachers are concentrated in school districts where there is both more student diversity and poverty. Therefore, diverse learners are most apt to be at the mercy of poorly written texts and curriculum materials, and their teachers are least able to compensate for these weak materials due to lack of content knowledge and pedagogical skills (National Commission on Science and Mathematics Teaching for the 21st Century, 2000).

Results from the 1996 NAEP for 8th grade science students suggest that some classroom practices positively affect student outcomes (Wenglinsky, 2000). Practices most closely associated with high achievement were hands-on learning activities and high order thinking skills. Students whose science teachers received professional development in laboratory skills were 44% ahead of those whose teachers lacked such training. Perhaps most important, this study indicated that while SES was the single most influential predictor of achievement, .75 (consistent with the findings of the Coleman Report decades ago), *aspects of teacher quality taken together were about as strong an influence in science achievement, .74* (emphasis ours). This relationship was even higher for mathematics. However, the least common topics for professional development activities provided for teachers were methods for teaching limited-English-proficient, special needs, and culturally different students (Wenglinsky, 2000).

Although some may regard the citing of such statistics equivalent to a “deficit model” for viewing diverse learners—blaming the students—our position is that the problems reside in the educational system, rather than the students. Moreover, current state and federal high-stakes accountability systems make it virtually impossible to ignore

the immediate and consequential effects of achievement gaps on both students and educators. More important to us, however, is what the gaps in science outcomes signal for science literacy and their implications for students who deserve to be generally well-educated and fully as empowered citizens. Achievement gaps also reflect a skewed career opportunity structure for too many students after they complete K-12 education.

Current Efforts to Improve U.S. Science Curricula

The research cited above points to a compelling need to better understand how interactions between curricula, instruction, and student diversity affect student achievement within individual science classrooms. For the U.S., major barriers to improving science education include the quality of U.S. science curriculum overall—an unfocused and splintered vision; the low quality of textbooks and curriculum materials used by teachers and students (U.S. National Research Center for TIMSS, 1996); and, teachers' difficulties with implementing reform-based curricula (Lynch, 1997; Lynch, 2000) and in delivering instruction consistent with educational reform goals (Lynch, 2000; National Commission of Science and Mathematics Teaching for the 21st Century, 2000; Wenglinsky, 2000). National systemic reform initiatives point to the need for a more focused science curriculum and better curriculum materials for teachers to use (aligned with science standards, instructional methods, and assessment/accountability measures). U.S. science textbooks are larger than those of other countries, but they are also less comprehensible, more repetitious, and tend to ask students to do tasks that only require low-level kinds of thinking (U.S. National Research Center for TIMSS, 1996).

It is reasonable to assume that improved curriculum materials aligned with science education standards that encourage students to learn with understanding could

accelerate the reform process (given qualified teachers and reasonable resources).

Consequently, the National Science Foundation (NSF), the National Research Council/National Academy of Science (NRC/NAS), the United States Department of Education (USDOE), and the American Association for the Advancement of Science (AAAS) Project 2061 all are involved in efforts to stimulate the creation of improved curriculum materials or *to develop rating systems to identify promising, extant standards-based curricula*. These rating systems rely on experts who examine curriculum materials for critical, research-based attributes believed to contribute to successful teaching and learning. But the curriculum analyses may fall short of the goal of clarifying "What really works," especially for diverse learners, because they are conducted in settings divorced from classrooms.

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 & Roseman, 2002; 2003, but see also Holiday, 2003). The Project 2061 Curriculum Analysis consists of categories of questions for analyzing a unit of instruction (see AAAS, 1999). This analysis inquires if a curriculum unit: starts from ideas that are familiar or interesting to children; explicitly conveys a sense of purpose; takes into account student ideas, and conveys suggestions for teachers to find out what their students think about the phenomena related to the benchmark; provides for first-hand experiences with phenomena; and, has students represent their own ideas about phenomena and practice using the acquired knowledge and skills in varied contexts (Roseman, Kesidou, & Stern, 1996; Kesidou & Roseman, 2002).

These criteria seem to be at the heart of effective instruction for diverse learners. They are so comprehensive and sound that if curriculum materials met them, it seems likely that all students would be well on the way to learning targeted benchmarks. This seems consistent with equity concerns for diverse learners because it implies recognition of individual differences; and the need to provide students with explicit goals, hands-on experiences, and teacher modeling or scaffolding of scientific reasoning . One could hardly argue with the "just plain good teaching" embodied by the Curriculum Analysis, which may indeed promote equitable science achievement for all. However, there is a body of research, both quantitative and qualitative, that suggests that a "one-size-fits-all" approach to curriculum and instruction is not working for various diverse student populations.

Barriers to Understanding "What Really Works"

Is "best practice" in science education best under all circumstances (Bruckerhoff, 1997; Corcoran, 1997; Donmoyer, 1995; Lee & Luykx, 2004; Lynch et al., 1996; Lynch, 2000, Riechard, 1994)? Variations in classroom cultures and in the cultures of the students inhabiting U.S. classrooms may affect student outcomes. There is a small amount of research on the *effects of curriculum materials on diverse learners*. A recent comprehensive review of the research by Lee and Luykx (2004) found a dearth of work on the development, implementation and efficacy of curriculum materials. What exists seems to focus on four areas: how student diversity is represented in existing curriculum materials; the development of culturally and linguistically relevant materials; technology-based curriculum materials; and use of science curriculum materials with English

Language Learners. Few of these studies, however, directly examined the effects of curriculum materials on student achievement.

Lee and Luykx (2004) reviewed the few studies on curriculum and instruction that measured student outcomes: Matthews and Smith found higher outcome scores and more positive attitudes toward science when Native American students in Bureau of Indian Affairs schools used culturally relevant curriculum materials (compared with materials not so modified). Dixon found that English Language Learners who used a computer-based instructional environment outscored those who used a more traditional textbook approach, but the approaches did not vary with regards to modifications for student diversity. Other research reported gains in learning for students when exposed to curriculum materials modified to take into account the specific cultures of the students, or improved outcomes for “color-blind” curriculum materials delivered in an altered instructional environment to better accommodate diverse students’ needs (Lee & Luykx, 2004).

Two things are apparent given the current limited research base on the impact of science curriculum materials on learning outcomes for diverse students. First, few of the studies were designed to allow comparisons between students who received modified or innovative curriculum materials and a sufficiently large sample of students who did not receive it. Second, there are no clear trends in the types of successful modifications made to accommodate diverse learners. Some modifications have aimed at cultural continuity, or language adjustments (incorporating bilingualism), or novel curriculum approaches incorporating technology and emphasizing inquiry.

It seems difficult to generalize about modifications that *should* be made to *curriculum materials* for culturally and linguistically diverse students. Rather, the literature seems to suggest that the most promise lies in altering or tailoring the *instructional environment* in which curriculum materials are delivered to meet the needs of the students in a particular environment. This said, we find that studies that build theoretical perspectives for understanding how language and culture affect science learning, and that expose the advantages and impediments to learning wrought by both traditional and reform-based practices in science education, are illuminating (e.g., Aikenhead & Jegede, 1999; Allen, 1995; Gilbert & Derrick, 2001; Moje, 2001; Solano-Flores & Nelson-Barber, 2001; Warren et al., 2001).

Not only does the research base fall short in providing evidence for the wide-scale improvements in science curriculum and instruction sought by local and state school systems, and education policy makers at all levels (National Research Council, 2002), there is some evidence suggesting that implementing reform-based practices could be a double-edged sword. Von Secker & Lissitz' (1999) data analyses from the 1990 High School Effectiveness Study (NELS:90) suggested that while instructional policies aligned with national science standards (laboratory inquiry, emphasis on critical thinking, and less teacher-centered instruction) were associated with higher achievement overall, *"...they are equally likely to have the unintended consequence of contributing to higher achievement gaps among students with different demographic profiles"* (emphasis ours). The authors of this study caution that the disconcerting results must be interpreted against a backdrop that depicts the total school setting and require more sensitive, classroom-based measures of curricular and instructional impact on science achievement and equity.

A positive example, Ohio's SSI, Project Discovery, found that teachers who used standards-based, inquiry oriented teaching practices positively influenced urban, African American students' achievement and attitudes, especially that of boys (Kahle, Meece, & Scantlebury, 2000). Although achievement for Project Discovery students generally improved for all groups, for some cohorts, gaps widened, indicating that Project Discovery interventions may have been more successful for White students than for African Americans in some years (Laguarda, 1998). Overall, Project Discovery did succeed in both raising student achievement and closing gaps. It was a challenge to create the carefully structured evaluation design that allowed valid comparisons across schools (J. Kahle, personal communication, January 20, 2001).

In summary, large-scale studies measuring student outcomes have tended to show that reform efforts have not closed achievement gaps, but fail to explain why, or how to solve this problem. On the other hand, smaller scale studies and qualitative research provide explanations or specific examples of successful science experiences for some groups of diverse learners, but do little to suggest how to scale-up these interventions in order to reach more students for effective systemic reform. Thus, theoretical, methodological, and practical issues of generalizability and scalability to facilitate systemic reform remain unsolved (Suter & Frechtling, 2000).

Theoretical Perspectives for the Planning Grant Research

The contrast between the need to provide a single set of curriculum materials for all students within a classroom, school, or school district and the recommendations to modify curriculum and instruction to meet the needs of diverse learners, creates a quandary. On one hand, the constraints of schooling often require that a single set of

curriculum materials suffice for thousands of students and presumably help each student obtain scientifically appropriate understandings of science concepts. On the other hand, current theories of conceptual change and assumptions about the importance of understanding the social-historical contexts of schooling suggest that schools must be flexible and responsive to increasingly diverse student populations. In this planning grant research, we begin to study both horns of the dilemma.

The foundations build, in part, on theories of conceptual change, and link ideas about knowledge, curriculum materials, and student epistemology. Specifically:

- (a) conceptual knowledge is taught and learned in school;
- (b) curriculum materials can influence instruction and have unique effects on student concept development; and,
- (c) students' prior conceptions interact with curriculum materials, facilitating new learning and/or conceptual change.

These core beliefs focus the lens of our work sharply on the interactions between curriculum materials and students. We want to know if each learner within a diverse classroom is afforded the opportunity to interact with the curriculum materials to develop increasingly advanced conceptual knowledge in science.

Over the last 20 years, theories of conceptual change (Posner, Strike, Hewson, & Getzog, 1982; Strike & Posner, 1992) have influenced efforts both to develop and evaluate science curriculum materials. Conceptual change theory emphasizes the personal nature of concept development and concerns the conditions and mechanisms that support the rational formation, storage and use of concepts by students. More recently, revisions of the theory suggest the inclusion of constructs of engagement and

motivation as part of the conditions for conceptual change (Pintrich, Marx & Boyle, 1993; Strike & Posner, 1992). Thus, in this study we include measures of goal orientation (Ames & Ames, 1984; Midgely et. al., 2000; Pintrich, 2000) and engagement in learning activities (Bangert-Drowns & Pyke, 2001; Fredericks, Blumenfeld & Paris, 2004; Lee & Anderson, 1993; Marks, 2000) as important outcomes to be explored together with assessments of content knowledge.

However, in order to investigate *how* such curriculum materials function (Silverstein, 1993) in their social and cultural contexts of use by diverse learners, we take a different theoretical approach. As culturally defined activity systems (Engestrom et al., 1999; Halliday & Martin, 1993; Vygotsky & Kozulin, 1986), classrooms are dynamic environments in which children and teachers participate in a variety of intersecting communities of practice: social, school and scientific (Lave & Wenger, 1991). Adopting an ethnographic approach, we are able to provide a detailed description of a communicative system of culture and culturally defined situations in which relevant distinctions in that system occur (Conklin, 1962). Video technology allows ethnographers to trace the ways in which students move from peripheral forms of participation in science activities in the classroom to more central forms of participation.

Drawing from holistic models of communicative practice (Hymes, 1974; Hymes, 1980; Morris, 1946) we can find evidence of student participation/activity that corresponds to interactional, material, and linguistic features of classroom activity systems. Making sense of students' immediate responses to the learning tasks prompted by curriculum materials is important, especially in increasingly culturally and linguistically diverse classrooms. The instructional criteria set forth in the Project 2061

Curriculum Analysis provide a lens to capture the activities of students enacting a curriculum unit and suggest promising coding categories. For instance, we can record what students say about the scientific phenomena that they are studying, and how their ideas change over the course of the unit. We can see how the lab materials associated with the curriculum unit help students to learn or cause confusion (trouble). The structure of the student group as an emerging functional or dysfunctional community of science practice is revealed by the activities of each of its members, and the roles that they play.

In summary, theories of conceptual change permit researchers to speculate about what is going on in students' minds, and socio-cultural theory allows researchers to interpret the activities in which students are engaged as a "conceptual change" curriculum unit is enacted. While the theories are not commensurate, neither are they in conflict. Together, they hold promise for understanding the contribution of curriculum materials to improved outcomes for diverse learners.

EXAMINING A HIGHLY RATED SCIENCE CURRICULUM:

A PLANNING GRANT STUDY

This research reports the results from an Interagency Educational Research Initiative (IERI) planning grant, administered through the National Science Foundation. Planning grant research is similar to pilot study research. The aims are to establish the feasibility of a larger, long-term study; to demonstrate to the funding agency that a research collaborative is viable and that the partners can work well together to produce useful results; and, to show that an expanded research program of study has the potential

for making an impact on the field. After receiving a planning grant, the research team needs to analyze the data collected and report on it in a timely fashion, if the team is to apply for a more extensive research funding in the targeted program. We share the results of this planning grant here, which in turn, led to more extensive and ambitious proposal that has since been funded through 2007 by IERI. However, this study best represents our view *at the end of the planning grant year*, and does not take into account our entire research program, which will unfold over the next several years.

Study Goals

The study described here is the first step for a research program designed to understand how highly rated science curriculum materials, aligned with reform goals, improve educational outcomes for diverse student populations on a large scale. This work aims at the stated intent of the IERI—*to improve education achievement by providing scientifically-based knowledge and skills that lead to sustainable learning changes across diverse student populations* (emphasis ours) through interdisciplinary research that informs practice and can be implemented in real, complex, and varied educational environments (National Science Foundation, 2000). We chose to focus on “curriculum materials” as the intervention, because although the National Science Foundation has funded a great deal of innovative curriculum design, U.S science classrooms remain dominated by commercially published science textbooks. Studies indicate (and many educators suspect) that this is a reason for the mediocre science performance of U.S. students (Lynch, 2000). Additional evidence of the poor quality of U.S. curriculum materials exists in studies such as Project 2061’s Curriculum Analysis. The Project 2061

Analysis has identified very few acceptable science curriculum materials (Kesidou & Roseman, 2002).

This planning grant study examined the effects of one of the rare exceptions, a middle school science curriculum unit that received acceptable ratings from Project 2061 called Chemistry That Applies, (CTA) (State of Michigan, 1993). The intent was to study the implementation of CTA in highly diverse classrooms. An important ancillary goal was to develop research approaches for understanding student learning and culture as the new unit was enacted. This study can be viewed as a direct test of the unit's effectiveness, as well as an indirect exploration of the criteria contained in the Curriculum Analysis itself. We hoped to identify some attributes of the curriculum unit that might be most effective for diverse student populations.

An interdisciplinary team of university researchers (science and mathematics educators and a linguistic anthropologist) and science educators and evaluation specialists in a large public school system formulated three major goals:

--Goal 1: Conduct a pilot study that would test the efficacy of a highly rated curriculum unit for a diverse student population.

--Goal 2: Using ethnographic methods, develop a framework to explore how the unit functioned in a diverse classroom setting.

--Goal 3: Develop an understanding of what it might take to scale-up this type of research on curriculum implementation for a large number of schools. This goal is reported on in more detail in Lynch, Szesze, Pyke, & Kuipers (in press) and Lynch, (in press) and is beyond the scope of this paper.

The results for the first two goals will be discussed separately in the sections to follow.

Research Population, Samples, and Intervention

Population and Samples

Montgomery County Public Schools (MCPS) is a large Maryland school district (136,000 students) located in the Washington, DC metropolitan area, with a student population that is rapidly becoming more diverse, culturally, linguistically, and socioeconomically. By 1999, the student population had no ethnic majority, and African American, Hispanic American, and Asian American subpopulations are growing at faster rates than the White subpopulation. MCPS consistently occupies a position among the top-performing school districts in the State of Maryland. However, in order for MCPS to maintain its status, disaggregated achievement data suggest that MCPS must do more to identify and implement effective curricular and instructional practices in science and mathematics for diverse learners. In the last state assessment, there were slight decreases in achievement in 8th grade mathematics and science (e.g., in 1999 63.3% met the 8th grade state science standard, while the 2000 data showed that only 62.4% did) (Bulletin, 2000). Lower achievement was correlated with several diversity indicators.

The population studied in this planning grant included all 8th grade students from 10 MCPS middle schools, over-sampling from schools with the highest levels of student diversity. The demographics of the participants are presented in Table 1, showing that 8th graders in treatment and comparison groups are demographically similar. The demographic designations used in this study are those on file for each student in MCPS. Student eligibility for Free and Reduced-price Meals System (FARMS) is a proxy for socioeconomic status in this study. Within Table 1, the term “Never” refers to students who had never been eligible for a certain service; “Prior” refers to students who had once

been eligible, but were no longer; and, “Now” refers to students who were eligible for a service during the year of the study. Services include FARMS, special education, and English for Speakers of Other Languages (ESOL). Students in the treatment condition received the CTA curriculum unit. Students in the comparison condition were eligible for a variety of other approved curriculum materials that ranged from traditional and textbook-based, to new NSF-funded materials other than CTA, to teacher-designed materials.

[Insert Table 1 about here, please.]

The ethnographic portion of the study required that four students in one classroom be videotaped as the new CTA curriculum unit was enacted. Teachers in three of the five treatment schools volunteered their classes for videotaping. We selected the class of a middle-aged teacher who was a career changer in her third year of science teaching. She, in turn, suggested that the study focus on an ethnically and linguistically diverse group of four students who were seated at the same laboratory table. The remainder of this class of students had a similarly diverse profile. The students chosen for this portion of the study are given pseudonyms. “Kim” is an Asian American girl who was identified as “gifted and talented.” “Rafael” is a Latino who had recently transitioned out of ESOL (he was demographically coded as “Prior” ESOL status). He was an avid reader, as evidenced by his surreptitious reading of novels when he had the opportunity. “Angelique” is an African American girl who was very socially oriented (she was a cheerleader). “Mike” is a European American boy who had also been identified as “gifted and talented.” “Frankie,” their teacher, is an African American woman.

The Intervention

The curriculum unit chosen for this study was Chemistry That Applies (CTA), a 6-10 week long unit (State of Michigan, 1993). The unit is aimed at 8th to 10th grade students. It was written with a focus on teaching a standard/benchmark on the Structure of Matter (American Association for the Advancement of Science, 1993):

No matter how substances within a closed system interact with one another, or how they combine or break apart, the total weight of the system remains the same. The idea of atoms explains the conservation of matter: If the number of atoms stays the same no matter how they are rearranged, then their total mass stays the same. (p. 79)

CTA is a highly structured curriculum unit consisting of 24 lessons divided into four clusters. However, for this study, teachers were only asked to teach the first three clusters (18 lessons) because the fourth cluster (on catalysts) was beyond the scope of the 8th grade MCPS science curriculum framework. CTA's design was based upon conceptual change theory (A. Anderson, personal communication, January 8, 2001), and if implemented with fidelity seems congruent with the type of science instruction recommended in the National Science Education Standards (National Research Council, 1996). The unit is student-centered, hands-on, and phenomenon-based. It might be best described as "guided inquiry". Students explore four chemical reactions with increasing sophistication, with the aim of acquiring a deep understanding of the target standard/benchmark. Groups of students collect data, keep individual science notebooks

for analyzing results, and use evidence-based arguments in large and small groups to support their claims. At the close of each CTA lesson, there are question prompts (called “Think and Write”) that require students to use critical thinking skills. Complicated vocabulary is kept to a minimum, however.

Because the study was done in Maryland, we had to determine if CTA aligned well with Maryland Department of Education Learning Outcomes on conservation of matter (Grades 6-8 Chemistry), as well as with the Skills and Process Outcomes on Scientific Inquiry, Critical Thinking, and Applications of Science for grades 6-8 (Maryland School Performance Program, 2000). There was good alignment, but CTA did not cover all the chemistry topics listed in the Maryland curriculum framework. Teachers, therefore, would need to do additional chemistry instruction before or after the unit, or both. Nonetheless, it seemed likely that students who experienced CTA would not only learn conservation of matter, but also generally become more proficient with other chemistry concepts, processes, and thinking skills required by the state standards.

In this study, the CTA curriculum unit is compared a “comparison condition” which consisted of a menu of options approved by MCPS. Eighth grade science teachers could choose from a large variety of curriculum materials on a list of approved materials. These options could range from traditional 8th grade science textbooks from companies such as Prentis Hall to NSF-funded, reform-based curriculum units on chemistry. In some cases, it appeared that teachers drew from curriculum materials in combination, and created their own units. However, MCPS 8th grade science teachers all were required to follow the State of Maryland’s curriculum framework and indicators, as well as MCPS’ local standards. Because MCPS was just completing a Local Systemic Change grant

focused on middle school science, teachers had routinely and systematically been exposed to reform-based instructional practices through professional development, and had pilot tested a number of reform-based curriculum materials. However, none of these curriculum materials had been evaluated with experimental or quasi-experimental designs, so their efficacy was inconclusive. Teachers had a variety of opinions about which was best. In summary, during the course of this research, the 8th grade teachers in all ten of the study schools (both CTA and comparison) were asked to teach the same curriculum standards within the same framework during the same quarter and had routinely attended professional development meetings to encourage the use of the reform-based strategies. However, for comparison teachers at the time of this study, there was no way of ascertaining which curriculum materials they were using, beyond saying that they were not using CTA (see Lynch, Szesze, Pyke, and Kuipers, in press, and Lynch, in press, for a more extensive discussion).

Goal 1: Test the Efficacy of CTA in a Diverse Student Population

Questions, Design, and Procedures

This part of the planning grant study is a quasi-experiment designed to compare outcomes for a large number of students using CTA with outcomes for students who used other approved curriculum materials. It was specifically designed to answer two research questions:

1. Does the CTA curriculum unit result in higher mean scores (on achievement, engagement, and motivation measures) for students in the treatment group compared with those in the comparison group?

2. Does disaggregating outcome data and testing for interactions between groups (based upon demographic characteristics) and the curriculum conditions reveal important patterns not captured in the reporting of mean scores of treatment and comparison groups?

All 8th grade students from 10 of the 35 MCPS middle schools participated (see Table 1), over-sampling schools with the highest levels of student diversity. In order to create equivalent treatment and comparison groups of student participants, the 10 schools were matched on demographics, resulting in five matched pairs. One school from each pair was randomly selected to implement the highly rated CTA curriculum unit and the other school used curriculum materials normally available. This sampling method produced two equivalent samples, rich in diversity, with enough students to provide power for significance tests on disaggregated subsamples.

Students were given a motivation and engagement questionnaire and a science content assessment as pretests, posttests, and delayed posttests (administered about four months after the chemistry unit was completed). The questionnaire contained Likert-type items targeting five scales (at least three items per scale) producing continuous scores for the five motivation and engagement constructs. The Conservation of Matter content assessment also produced a continuous variable generated from 10 items (the characteristics of the outcome variables are discussed in more detail in the following sections). The database maintained by MCPS allowed the disaggregating of data according to FARMS, ethnicity, gender, and status as English language learner. The available demographic characteristics guided how the data were analyzed, as we ascertained whether CTA had differential effects on subgroups of students. Given a set of

continuous outcome variables, Analysis of Variance (ANOVA) and Analysis of Covariance (ANCOVA) techniques were used to analyze the data with pretest scores as the covariate when analyzing the disaggregated data. Assumptions for these analyses were evaluated using the guideline established in the ANOVA/ANCOVA literature (c.f. Tabachnick & Fidell, 1989). First, the data were screened for descriptive values (e.g., mean, SD, minimum and maximum) in the expected range with missing data and sample size differences noted. Next, assumptions of normality, linearity, and multicollinearity/singularity were investigated. Finally, homogeneity of variance and regression were considered. In most cases, assumptions were met, with some exceptions for homogeneity due to unexpected small sub-samples with levels of ESOL. We believe the analyses are robust, but results must be interpreted with caution because no corrective action was taken prior to the analyses presented here. In addition to ANOVA/ANCOVA analyses, when significant main effects and interactions were found, exploratory follow-up contrasts were conducted to explain the effects. All analyses were performed using SPSS for Windows, Version 12.

The Conservation of Matter Assessment

A new assessment analysis and development procedure was developed in collaboration with the AAAS Project 2061 to establish the validity of The Conservation of Matter Assessment (AAAS, 2003a; AAAS, 2003b; Stern & Ahlgren, 2002). The procedure focuses on the alignment of items with middle grades standards and the use of language and illustrations that allow the assessment to be read and understood by a maximum number of students. In this study, the conservation of matter benchmark

(stated earlier in the paper) is the target. The assessment development team included science content experts, science educators, science teachers, and assessment specialists.

The assessment asks students to consider four different physical phenomena and respond to multiple probes that require both selected and written responses. These four phenomena include closed vs. open systems, appearance or disappearance of substances, and chemical or physical changes. The final set of 10 items was obtained after piloting and evaluating the measurement qualities of a larger number of phenomena and items. Coding categories for student responses that indicate scientifically appropriate understanding, and those indicating alternative ideas were determined by the development team's analysis of pilot data. A small group of science educators, science teachers, and chemistry graduate students were trained to use the codes and rated all the assessments. Approximately 2% of the assessments were rated by multiple raters for reliability purposes and across all pairs of raters, inter-rater reliability Kappa scores ranged from .70 to .81. A Kappa greater than .7 is considered satisfactory. The 10 items in the Conservation of Matter Assessment reliably target the construct with Cronbach's $\alpha = .71$.

A panel including science educators and an assessment expert determined the weighting scheme for combining the 10 items into a single score. The weighting scheme distributes the contribution of students' ideas about (a) conservation of mass (approximately 60% of the total score) and (b) the idea that atoms explains conservation of mass (approximately 40% of the total score). In addition, the scheme balances the contribution of selected response and constructed response items in the total score. In the end, the weighting scheme transforms the items into a score that ranges from 0-100. The

expert panel then analyzed information from exploratory test samples of students, setting cut scores to distinguish four levels of student understanding of the “conservation of matter” benchmark:

- Scores in the 0-23 range (raw scores of 0 to 2) show no consistent evidence of understanding the benchmark ideas.
- Scores in the 24-50 range (raw scores of 3 to 5) show some evidence of understanding in specific contexts.
- Scores of 51-70 (raw scores of 6-8) reveal some fluency with the ideas, but also misconceptions in certain contexts.
- Scores of 71-100 (raw scores of 9 or 10) were found to demonstrate a flexible understanding of, and commitment to, the benchmark ideas, with few errors or misconceptions.

Measuring Engagement and Motivation (Goal Orientation)

This study measures engagement and motivation. *Engagement* is defined as the deployment of strategies (cognitive, affective, and behavioral) for sustaining involvement in activities aimed at conceptual learning (Bangert-Drowns & Pyke, 2001). *Motivation* is defined as *goal orientation*, which is an individual’s personal orientation toward learning that influences directly affect, behavior, strategy use, attention, self-regulation, and performance.

Engagement is a multi-level construct (basic and advanced) that involves students’ use of strategies for sustaining learning activity, assessed by the Basic and Advanced Learning Engagement scales. *Basic learning engagement* is often characterized by student behavioral strategies for sustaining active participation in

classroom work, such as: following instructional directions, completing assignments and homework, and attending to the explicit behavioral demands of a task or lab activity. Researchers studying patterns of disengagement among adolescents, particularly students not succeeding in school, view this level of engagement as a positive step. *Advanced learning engagement* is characterized by deeply involved interaction, self-regulated learning, and learning for understanding. Such strategies involve creating personal challenges, testing hypotheses about how things work, persisting on difficult tasks, and trying different solution paths to problems. Intrinsic motivation to learn is at the heart of this type of engagement and builds from extensive understanding of the task at hand, when students purposefully participate in activities related to the content or classroom context. As such, advanced modes of engagement require environments that allow for choice, planning, testing, and exploring knowledge (see Pyke, Lynch, & Kuipers, 2003 for a more detailed discussion).

To assess motivation, three scales for goal orientation were used. The *Mastery Goal Orientation* scale indicates how students value the development of their personal understandings and skills when given a learning task. The *Performance Orientation/Approach* scale assesses students' values for external rewards, such as the personal recognition that they are better than others or knowing more than others. The *Performance Orientation/Avoid* scale assesses if students' purposes or goals in an achievement settings are aimed primarily at avoiding looking incompetent.

In this study, engagement in science learning was measured by responses to the Basic Learning Engagement and Advanced Learning Engagement scales, adapted from similar scales developed and validated on the general school population by Helen Marks

(2000) of Ohio State University. Goal orientation was assessed by scales from Patterns of Adaptive Learning Scales (PALS) developed and validated with middle school students by researchers at the University of Michigan (Midgley et al., 2000). Each scale has a reliability calculated by using Cronbach's alpha and produced an acceptable rating (Basic Learning Engagement = .66, Advanced Learning Engagement = .81, Mastery Goal Orientation = .89, Performance Orientation/Approach = .88, and Performance Orientation/Avoid = .74).

Question 1 Results: Does CTA Result in Higher Scores for Achievement, Engagement, and Motivation?

The Conservation of Matter Assessment

As expected, due to the quasi-experimental design, no significant differences on the Conservation of Matter Assessment were found on the pretest. Prior to instruction, both the CTA and comparison groups demonstrated little accurate or flexible understanding of conservation of matter, with low mean pretest scores of 21.66 (CTA) and 21.72 (comparison).

The effectiveness of the CTA curriculum was tested by comparing posttest and delayed posttest scores across curriculum conditions. Data analyses for the posttest and delayed posttest revealed statistically significant differences in favor of the CTA curriculum: $F(1, 1857) = 59.80, p < .05$ for posttest scores, and $F(1, 1888) = 72.16, p < .05$ for delayed posttest scores. On the posttest, CTA students improved their mean score by 20 points, but comparison group students gained only about 11 points. Considering that both groups began with very little understanding of conservation of matter ideas, it seems each group made some progress in understanding the target concept.

[Insert Table 2 about here, please.]

Follow-up comparisons on the frequencies of posttest score levels show that students in the CTA condition seemed to make more progress in moving out of the lowest level (scores below 23) than the comparison group. Only 22% of CTA students remained at a low level of understanding, while 38% of comparison group scored in that range on the posttest. However, the proportions of the most successful students (who had scores above 70) were about equal for both groups (21% for the CTA group versus 20% for the comparison group). These data suggest that about one fifth of the students “got” the target concept, no matter the curriculum condition. But for those who are still struggling to understand conservation of matter, CTA moved them farther along in understanding than the comparison condition.

Engagement and Goal Orientation Scales

On two of the five scales assessing engagement and goal orientation, significant effects in favor of the CTA curriculum were found. ANCOVA was used to control for pretest differences in all the engagement and goal orientation analyses. In the analysis of Basic Learning Engagement, CTA students had significantly higher mean posttest scores after adjusting for pretest differences, $F(1, 1724) = 9.8, p < .05$. Basic Learning Engagement scores indicate the extent to which students use basic learning strategies (i.e., taking notes, paying attention, etc.). The CTA curriculum also produced significantly higher Mastery Learning Orientation scores at posttest when controlling for pretest differences, $F(1, 1689) = 10.26, p < .05$. Higher mastery scores are associated with the most adaptive goals for learning. Although it seemed reasonable to expect that CTA would also foster higher scores on the Advanced Learning Engagement scale

because CTA prompts students to make sense of science ideas in labs through experiences with physical phenomena, the data did not reflect this.

In summary, moderate, significant effects were found in favor of CTA on the Conservation of Matter Assessment (Cohen's $d = .35$). Controlling for pretest scores, small effects were found for Basic Learning Engagement (Cohen's $d = .13$) and Mastery Goal Orientation (Cohen's $d = .12$). While it was beyond the scope of the planning grant study to test relationships among these variables, there is some evidence that CTA students were more engaged, at least to the extent that they attended to the curriculum unit's demands. Also, CTA to some extent encourages students to learn for understanding. Perhaps as a result of these changes in motivation and engagement, they learned more than if they had received the standard curriculum. As we replicate the study, we intend to make minor but crucial adjustments to the wording of the items on these scales that prompt students to respond to them for the chemistry curriculum unit just completed, rather than to capture students' generic views of science classes.

Question 2 Results: Do Overall Mean Scores Hide Different Effects Across Groups?

Comparing differences in overall mean scores for the curriculum condition does not reveal whether scores improved in a similar fashion for different groups of students who learned with CTA. The intent of the study was to understand if gaps existed and to use this information to help close them. Therefore, the second research question asked if CTA had a similar impact on groupings of students defined by various demographic variables included in the MCPS database (i.e., ethnicity; eligibility for FARMS--a proxy for socioeconomic status; gender; and, ESOL status). This suggested the need to test for

interactions between the curriculum condition and different groupings of students. For each demographic variable, a Curriculum Condition X Subgroup between-groups Analysis of Covariance (ANCOVA) was performed on posttest scores. Independent variables consisted of curriculum condition (CTA and comparison) and subgroup. The covariates were pretest scores. For all of the analyses, results of the evaluation of assumptions for ANCOVA were satisfactory (see Tabachnick & Fidell, 1989).

The Conservation of Matter Assessment and Group Differences

Mean scores for the Conservation of Matter Assessment across each of the subgroups are presented in Table 3. Examining Table 3, we can see that pretest score patterns were similar across curriculum conditions (CTA and comparison). Within groups, pretest scores indicate initial differences in understanding of conservation of matter for subgroups. These differences follow patterns that are often found in the research literature and described earlier in this manuscript: There are within group differences by ethnicity, SES (FARMS), and ESOL status. The exception is gender—boys and girls have similar pretest scores.

[Insert Table 3 here about here please]

Ethnicity. The columns of Table 3 that display the results for ethnicity show that various ethnic subgroups had different pretest scores. White and Asian American students initially demonstrated more knowledge of conservation of matter than Hispanic and African American students. Figure 1 (comparing pre- and posttest scores by ethnicity) shows that after instruction, the scores of Asian Americans caught up with White students' scores in the CTA condition, but not in the comparison condition where gaps

were maintained. African American and Hispanic students taking CTA outperformed their peers in the comparison group.

[Insert Figure 1 here, please]

However, statistical analysis (2 X 4 between-subgroups ANCOVA) showed no statistically significant interaction between ethnicity and curriculum condition, suggesting the *effect* of CTA was similar across all four ethnic groups. As expected, statistically significant main effects for curriculum condition and ethnicity were found. The main effect for CTA is discussed earlier in this paper, and the main effect for ethnicity, $F(3, 1689) = 63.0, p < .05$, is consistent with the pattern in the pretest scores (see Table 3). The combined main effects are indicated in Figure 1. Across ethnic subgroups, students with CTA learned more about the concept than their comparison group counterparts.

FARMS. This demographic variable allows the comparison of students never eligible for FARMS (Never) with those who were now eligible for FARMS (Now) and those who had once been eligible (Prior). As such, it is the best available estimate of students' socioeconomic status. We infer that the Prior group is in some way intermediate in SES to the Never and Now groups. Table 3 shows that the pretest scores across subgroups for CTA and comparison conditions are similar. The Never and Prior groups have higher pretest scores than Now students, those currently eligible for FARMS. Figure 2 show gains for all three subgroups of students, favoring CTA.

[Insert Figure 2 about here please.]

However, as with ethnicity, the interaction of FARMS with curriculum condition was not statistically significant, suggesting the effect of CTA was similar across all three

groups. Again, as expected, there were significant main effects for curriculum condition and for level of FARMS, $F(2, 1219) = 23.7, p < .05$. However, Figure 2 highlights the practical significance of the CTA curriculum. CTA students in the Prior and Now FARMS categories improved their understanding of conservation of matter to the level of “some understanding, with misconceptions.” In contrast, their comparison group peers remained close to the “no understanding” level, after instruction. This implies that CTA moved lower income students further along in understanding the conservation of matter ideas than the comparison curriculum condition.

ESOL. This grouping compares students who were never eligible for ESOL services (Never) with those who were currently eligible (Now) and those students who have transitioned from ESOL to the regular classroom (Prior). The data displayed in Table 3 and Figure 3 show similar patterns of mean pretest scores within subgroups across both curriculum conditions. As expected, students Never eligible for ESOL services had higher pretest scores than the Prior group, and both have higher scores than the Now group (currently eligible for ESOL). Figure 3 shows an increase for the Prior group taking CTA (as well as for the Never group). However, the scores of the Now groups in both the CTA and comparison conditions are nearly flat, from pre- to posttest.

[Place Figure 3 about here please.]

In this case, a statistically significant interaction from the 2 X 3 ANCOVA was found between curriculum condition and ESOL status, $F(2, 1377) = 3.3, p < .05$. Figure 3 data suggest that CTA Never and Prior students outperformed their peers in the comparison condition. But students currently receiving ESOL services (Now) did not receive the same benefit. Despite the concern for current ESOL students who either don't

learn well with CTA or who do not demonstrate their learning on the assessment, students who have transitioned out of ESOL made noteworthy gains with CTA.

Gender. Table 3 shows that boys and girls scores did not vary with gender and no statistically significant interaction was found between curriculum condition and gender.

In summary, the results of this series of ANCOVAs for the Conservation of Matter Assessment for various demographic variables across curriculum conditions yielded only one significant interaction, for ESOL status. In all other analyses, there were significant main effects for curriculum condition and for groupings, as predicted by the literature (except for gender). The implication is that with the exception of ESOL (in particular those currently receiving ESOL services) CTA had a consistent positive effect across the various subgroups. CTA improved their understanding of conservation of matter, but could not close gaps that may be the result of within group differences in initial understanding. When mean scores are analyzed according to where they fall on the various levels of the Conservation of Matter Assessment, there is more relative improvement of the CTA students at the lower levels of understanding. To put it another way, in the comparison curriculum condition, some subgroups of students started out with little understanding and remained at that low level, but their counterparts in CTA moved to a higher level, implying that they were beginning to learn this difficult concept.

Engagement, Goal Orientation, and Group Differences

Tables 4 and 5 display mean scores by demographic variables for the learning engagement and goal orientation scales. In general, mean scores across pre- and posttests were in the moderate range (i.e., indicating “somewhat true of me” responses were typical), suggesting students in both curriculum conditions commonly reported being

engaged both at a basic level and at more advanced levels; they were interested in learning concepts and skills; and, they focused on their own personal performance in comparison with their peers.

[Insert Tables 4 and 5 about here, please]

We performed ANCOVAs similar to those described above for all five engagement and goal orientation scales. There were no significant interactions found for the Advanced Learning Engagement or Performance Orientation/Approach or Performance Orientation/Avoid across the curriculum conditions for each of the demographic variables (ethnicity, gender, FARMS, or ESOL status). Having found statistically significant main effects favoring CTA for the Basic Learning Engagement and Mastery Goal Orientation for the entire group, we were interested in patterns that might emerge when those data were disaggregated. The ANCOVAs revealed statistically significant interactions for two demographic groupings, FARMS and ESOL status, described in more detail below.

Basic Engagement and FARMS. After adjustment by the covariate (pretest mean = 3.8), a statistically significant interaction was found between curriculum condition and FARMS, $F(2, 1500) = 3.7, p < .05$. Mean scores in Table 4 suggest that the CTA students currently eligible for FARMS (Now) and those that were once eligible (Prior) were more engaged in following directions, paying attention and completing tasks than their peers in all other conditions. Figure 4 displays the interaction by showing the adjusted posttest mean scores for each FARMS level across the comparison and CTA conditions. The Prior group appears to cause the disordinal interaction by showing a greater increase Basic Engagement than found in the other conditions.

[Insert Figure 4 about here, please.]

Basic Engagement and ESOL. No statistically significant interaction was found between curriculum condition and ESOL level, after adjustment for the covariate. As expected, there was a main effect for curriculum condition favoring CTA as reported previously, and there was a main effect for ESOL grouping, $F(1, 1500) = 7.3, p < .05$. Interestingly, adjusted marginal mean scores show that Now ESOL students currently eligible for ESOL had the highest Basic Learning Engagement scores (mean = 3.8) with no differences for the other two categories (means = 3.6).

Mastery Goal Orientation and FARMS. A statistically significant interaction was found between curriculum condition and FARMS, $F(2, 1470) = 2.9, p = .056$. Figure 5 displays the interaction by showing the adjusted mean scores for each FARMS level across the comparison and CTA conditions. As can be seen in Figure 5, the Prior group (i.e., those once eligible for FARMS) appears to cause the disordinal interaction by showing a greater increase on this scale.

[Insert Figure 5 about here, please.]

Mastery Goal Orientation and ESOL. There was no statistically significant interaction between ESOL status and scores on the Mastery Goal Orientation scale. There was a main effect for ESOL status, $F(2, 1470) = 4.1, p < .05$. The adjusted marginal means show that students currently eligible for ESOL (Now) produced the highest mastery scores (mean = 4.3) with no differences between the other two categories (mean = 4.1)

In summary, the disaggregation of engagement and goal orientation data revealed significant interactions for students eligible for FARMS, especially for those students at

the lower ends of the SES continuum. In the CTA condition, students reported more basic engagement and mastery goal orientation than students who had never been eligible for FARMS. This is a particularly encouraging, if statistically modest result, because it suggests that CTA reached students in a way that the comparison curriculum did not.

Summary of Results for Conservation of Matter Assessment, Engagement, and Goal Orientation

Does CTA, a curriculum unit given a high rating by AAAS Project 2061, produce better outcomes than the curricular options typically offered to MCPS students who attend middle schools with highly diverse student populations? The data from this planning grant study suggest that the CTA curriculum unit (implemented for the first time by participating teachers) resulted in higher mean scores for the Conservation of Matter Assessment, as well as for the Mastery Goal Orientation and Basic Learning Engagement scales.

The results show that all students, no matter the curriculum condition, on average have high scores on Performance Goal Orientations and Advanced Learning Engagement strategies. However, this fairly positive view of students' orientation toward learning is moderated by the finding that students on average do not master "conservation of matter" ideas, as evidenced by their scores on that assessment. Conservation of matter is difficult for 8th grade students, and many do not have a very accurate or committed understanding of this concept, even after instruction (38% of the comparison group remained in the "low range" after instruction, compared with 22% of the CTA group). The number of students who do not seem to learn the target concept seems puzzling, because they generally report positive motivation and engagement. This raises the question: Why do

students that believe they are motivated and engaged, particularly while learning with a highly rated curriculum, still fail to demonstrate sophisticated and flexible knowledge of the target concept?

Disaggregated data show that the CTA curriculum does seem to help students learn about conservation of matter. Most important, subgroups of students who frequently do not score well on tests of science achievement and who had lower pretest scores (low SES students, African American, Hispanic, and English Language Learners) all surpassed their peers in the corresponding comparison subgroup. Moreover, in many cases, their scores approached or surpassed subgroups of students who generally score well in science and had higher pretest scores (higher SES, Whites, Asian American, and students who did not require ESOL classes). The data show that more African American and Hispanic students, and students eligible for FARMS, moved from a level of “no understanding” on the Conservation of Matter pretest to middle levels of understanding on the posttest, while comparison group peers on average remained in the “no understanding” range. The lack of significant statistical interactions (for ethnicity and FARMS) implies that, even with disparities in prior knowledge, CTA increased understanding across the board, without increasing gaps. (The exception to this trend is for ESOL students, discussed below.)

The variables Mastery Goal Orientation, Basic Learning Engagement, and the Conservation of Matter Assessment form a consistent pattern of mean scores for the FARMS and ESOL subgroups. For instance, although the Never FARMS group clearly outperforms the other two FARMS levels on the Conservation of Matter Assessment, the Prior FARMS group makes large, practical gains. Concomitantly, Prior FARMS students

also reported greater increases in Mastery Goal Orientation and Basic Learning Engagement strategies after CTA than their peers. It is reasonable to conjecture that CTA provided these students with clearer goals and encouraged them to be more task-focused and involved in their learning than the comparison condition, contributing to their improved learning.

The Conservation of Matter Assessment data show an interaction between curriculum condition and ESOL level, apparently caused by the relative lack of improved achievement for students Now in ESOL. However, interestingly, this same subgroup had very high Mastery Orientation and Basic Engagement scores. These students reported that they were focused on mastery of ideas and were working actively toward that goal. Nonetheless, they had difficulty learning the target concept (or responding adequately to the assessment). One might speculate that these data reflect the attitudes of students who wanted to learn both how to speak English and science. However, when science is taught in English, learning a difficult target concept becomes elusive. Another interpretation is that the Conservation of Matter Assessment did not capture ESOL students' increased understanding, due to its literacy demands.

Goal 2: Using Ethnographic Methods to Explore how CTA Functioned in a Diverse

Classroom Setting

Research Focus, Design, and Procedures

In order to understand more about how CTA actually functioned for a group of learners in a classroom, we selected a single classroom for more intensive, ethnographic analysis. The approach to the study of classroom communication used here draws from ethnographic and linguistic analysis of classroom interaction (Cazden,2001; Erickson,

1992; Heath 1983). By linking the specific, patterned nature of communicative practices with their sociocultural functions in context, this approach seeks to identify and document the processes by which educational outcomes are produced. By joining ethnographic perspectives with advanced techniques of video analysis, this approach provides a number of unique advantages:

- It offers new techniques for looking at behavior patterns that are virtually invisible without advanced playback technology.
- Like all video analysis, it is able to reveal aspects of worldview and "native" categories by recording what actors are responding to in their environment.
- It allows for systematically and reliably coding complex and simultaneous behaviors directly onto the video images for easy storage and retrieval.
- It provides a permanent and easily shareable corpus of data.
- It can expose *mechanisms* and *processes* in observed behavior rather than analyzing "snapshots" of selected moments.

To capture the classroom activity on video, two cameras were placed in one selected CTA classroom each day as the curriculum was enacted, from the beginning until the completion of the unit. One camera focused on a *table* of four children, while another was placed at the back of the classroom and recorded the activity of the class overall. The students' speech and selected activities were carefully recorded, transcribed, and digitally linked to the video. The focus of the camera and the subsequent analysis was thus not on the teacher and her implementation of the unit, but on the diverse ways in which the children interpreted the curriculum unit through their questions, their actions, and their orientations toward one another.

After digitizing and storing 45 hours of video data on a central disk in the university's discourse lab, each video file was loaded into a visual qualitative data analysis software program, Atlas.ti, for transcription and coding. Approximately 1830 pages of transcribed data were coded by lesson segment, student utterance, utterance type, and manipulation activity. Coders were evaluated for inter-rater reliability. After extensive discussions on the theoretical, methodological and operational definitions of the coding categories (discussed below), trained coders examined and coded identical segments of transcribed videotapes. They compared their codes, and any differences were discussed until consensus was reached. An explicit written code manual was developed in this way. The process was repeated for each coding category until substantial (~80%) agreement in both theory and actual practice was achieved.

We report here on some of the methods that we have explored to locate systematic variations among students in their verbal behavior and some nonverbal (hands-on) behavior over the course of the enactment of this highly rated curriculum unit. The results that follow focus on generating inferences about the Project 2061 Curriculum Analysis criteria in action. We use evidence from discourse and activities where scientific thinking was likely to be present, coding for use of scientific terms, clarification of ideas and procedures, and manipulation of objects in the CTA lessons.

Results

A Sample Vignette

To get a sense of how the Project 2061 Curriculum Analysis criteria (AAAS, 1999) function in an actual classroom setting, it is helpful to examine closely a short stretch of activity in which some relevant behaviors occur. In this scene, transcribed from

a video, Kim, Mike, Rafael, and Angelique are getting ready to write up the results of a lesson in which Alka Seltzer and water were weighed before and after they were mixed in open and closed systems. The purpose of this lesson is to determine if gases have weight. For the closed system, a bottle cap was placed over the mouth of a bottle to capture the evolving gas. After collecting the data for the closed system, Angelique asks whether she can weigh the bottle and cap again after the gas had been released.

Transcription:

Kim: It'd be two...

Mike: It didn't lose or gain

Kim: It didn't gain or it lose...

Rafael: So what do we do with this? [gestures towards a bottle]

Kim: Not when it stayed the same. The weight stayed the same....because...

Angelique: Can I weigh the bottle again?

Kim: Yeah but now the gas ran out.

Angelique: I know. I want to weigh it now.

[For the next 45 seconds Angelique weighs the bottle and records the results]

Angelique: It *lost* weight.

Mike: No it didn't. It stayed the same.

Kim: It didn't lose or gain.

Angelique: No, right *now*.

Mike: No. It's the same.

Angelique: No it's not.

Mike: I mean the end result's the same. Gas doesn't have weight.

Angelique: I'm talking about this.

After the group discusses briefly the results of the previous experiment, Rafael wonders about what the group should do next. Angelique, earlier praised by Mike as an expert at using the scale, decides to weigh the bottle and cap again after the cap had been removed. Kim implies that Angelique will get different results due to the loss of gas. Angelique weighs the bottle and reports that the flask and its contents *lost* weight. Mike says, no it didn't. At first Angelique thinks he must be referring to the earlier experiment, because he seems to be contradicting her own direct experience. She re-frames her observations and he contradicts her again, apparently uninterested in the actual results. He seems to frame an answer aligned with his perception of the lesson purpose, rather than explaining Angelique's actual experience. In the process, he reveals his (mis) conception that this stasis occurs because "gas doesn't have weight."

To understand this complex event better, we are building interpretations of each child's actions and constantly comparing our inferences across the coded data, using the Project 2061 curriculum analysis criteria as frames—with each specific criterion referred to by category and number according to Project 2061's classification system (AAAS, 1999). The transcript above provides a good example for presenting our preliminary ideas about the students. In the case of Mike, a child classified as "gifted and talented", an important focus of his interpretive energy is determining and responding to a lesson's *purpose* (see Project 2061 Category I, Criterion #IB). In this instance, he interprets the

purpose as “conservation of matter,” which to him means that “everything stays the same.” He is less interested in the particulars of his table’s activity. When inputs from the teacher, the curriculum unit, peers, and even the experiment itself seem contradictory with respect to conveying his idea of the conservation of matter, Mike adopts a strategy of producing answers that take the rhetorical form “it stays the same.” He is often correct. Mike frames his (false) conclusion authoritatively *using abstract, stylistically “marked” scientific terms* (see Project 2061 Category IV) by stating that “gas doesn’t have weight.” The incident exposes his *misconception*, but does not (yet) correct it (see Project 2061 Category II).

For Angelique, a self-identified African-American girl placed in “regular” classes (she is not identified as gifted and talented or having a disability or eligible for ESOL), the table itself presents many opportunities for *engagement with relevant phenomena* (see Project 2061 Category III), an important criterion for evaluating instructional support. She participates verbally and manually with high frequency. She seldom uses *scientific terms* (see Project 2061 Category IV, criterion #IVA) in her descriptions, but she initiates her own problem solving and a line of thought not initially shared by the other students at that time. She apparently engages in *scientific thinking* (see Project 2061 Category V). For Angelique, this curriculum unit has created a learning environment that stimulates her to ask creative questions that are appropriate to the lesson purpose.

Rafael is an English language learner who has transitioned out of ESOL and into mainstream science. He seems oriented to lesson purpose also, but his way of accessing this lesson purpose is not by internalizing the rhetorical phrases of CTA and teacher, but by concentrating on procedures and sequences (see Project 2061 Category I Criterion

#IC). One of the criteria for evaluating the quality of a curriculum unit's instructional support in the Project 2061 Curriculum Analysis is whether the lessons proceed in a meaningful and logical sequence. Rafael devotes much of his energy to knowing the portion of the lesson segment on which the group is focused. His ultimate goal for following the lesson sequence, however, appears to be not so much whether he gets the answers right from the teacher's or CTA's point of view, but whether he can get enough information about the task at hand to remove barriers for his further participation. In short, it seems that Rafael does not want to misunderstand a situation occurring in what is, for him, a second language, and lose face in front of other students. Accordingly, he is oriented to doing the manual procedures with objects, and experiences the relevant science phenomena in the laboratory investigations in a procedural way (see Project 2061 Category III). However, it is less clear whether other aspects of the lesson's logic are revealing themselves to him.

For Kim, a self-identified Asian-American student classified as “gifted and talented,” the lesson presents multiple challenges. She is oriented flexibly to the table's practices, the unit purpose, the CTA written text, and the teacher. She consults all four for information, and is the member of the group most likely to seek out the teacher. She has internalized the logic of the lesson, in this case applying it to the new situation proposed by Angelique, and engages in scientific thinking by making a correct prediction (see Project 2061 Category V).

Coding for Scientific Thinking

To explore whether these apparent differences in the interpretive strategies of these four students corresponded to systematic differences in behavior over the course of

the entire unit, we looked for confirming and disconfirming evidence in the transcribed corpus of 18771 utterances (i.e., turns at talk), examining all instances of: use of scientific terminology, clarification requests, and object manipulation.

Use of scientific terminology. Use of scientific terminology, after experiencing the relevant phenomena, is a key component of developing and using scientific ideas (see Project 2061 Category IV). As Latour and Woolgar (1986) have shown in science labs and Lemke (1990) has documented in high school lab science classrooms, use of science terminology reflects as much about the social situation in which it is used as about the physical phenomena being referred to. To investigate the use of scientific terminology, we searched for science terms relevant to the unit. Terms such as “mass,” “gas,” “observe,” “condensation,” “substance,” and “reaction” were relatively low-frequency but are stylistically distinct from everyday language. Although the use of such terms tended to be referentially appropriate, the social aspects of the term usage were of interest to us. They tended to be used appropriately and in context, but in ways that seemed to have social consequences.

Figure 6 shows that Mike used scientific terms the most often. Although Mike did not necessarily always exhibit a better understanding of the target concept than the other students, (as we can see from the sample transcript), one way of interpreting this variation in students’ use of science terms may be their interpretive orientations. In the intensely peer-oriented adolescent culture of the table, use of scientific terms can be likened to the use of a specialized “code,” and their use amounts to a kind of “code-switching.” Since acts of code-switching are often acts of social identification (Gumperz, 1982), Mike’s frequent use of scientific terms may signal an identification with the unit’s goals and

lesson purpose, and a corresponding diminution of his identification with the colloquial orientation of the table activity system.

[Insert Figure 6 about here please.]

Clarification episodes: "Trouble" and remedies. The verbal practices by which children seek clarification of concepts, procedures, and roles are powerful indicators of the social organization of knowledge in both classrooms and everyday life (Darling, 1989; Heath, 1983; Ochs, 1984; Pica, 1987). From an ethnographic perspective, such "clarification episodes" can be re-framed to focus on the patterns by which the curriculum results in conceptual, communicative, or interactional "trouble" (Jefferson, 1980). "Trouble" is an interactional event in which a participant signals that there is a disruption to the communicative flow, due to difficulties with procedure, concepts, roles, or linguistic comprehension. Students indicate "trouble" in various ways, including asking explicit questions, long pauses, facial expressions, and other means of indicating disruptions in the communicative process. Following these indications of communicative "trouble," students may request clarification from a peer, the teacher, or the text. There are usually "responses," or remedies, to the clarification requests, including a verbal event in which a student either self clarifies, receives peer remedy, solicits a teacher remedy, or resorts to a remedy written in the CTA lesson text. Together "troubles" and "remedies" result in clarification episodes (Ochs, 1984). These commonly occurring behavioral practices can provide an ethnographic window into the communicative processes by which the "purpose" of a lesson or unit is "conveyed," and its problems are interpreted and resolved.

Discourse analysis showed how responses to clarification requests differ among students, and suggests that as a percentage of his overall speech in the class, Mike is the most likely to correct his peers. Figure 7 represents the percentage of the child's verbal participation. While Mike often offers remedies for others, particularly when they exhibit conceptual difficulties, he is, as we have seen, not always correct. However, he is nonetheless oriented to the lesson purpose.

[Insert Figure 7 about here please.]

Object Manipulation. Because the CTA curriculum, according to its authors, is designed to be “hands-on, minds-on” the task-related manipulation of relevant objects is an integral component of the unit. The frequency of task-related acts of object manipulation are also an act of interpretation on the part of the children at the table, insofar as such behaviors reflect an allocation of attention and cognitive resources, and a choice of social participatory style. Such interpretations can thus be viewed as ethnographic indicators of the Project 2061 Curriculum Analysis Category III, engaging students with relevant phenomena. Videos from particularly lab-rich periods of the unit implementation (November 16-20, 2000 and January 8-12, 2001) were coded for object manipulation by the students.

Overall, Figure 8 shows that Mike and Angelique participate *manually* more often than Kim or Rafael. Not only do they engage more frequently with objects, they appear to spend more time with the objects when they do engage, although the evidence for this is only preliminary. In addition, Mike performs object manipulation by proxy, providing other students with detailed instructions as to how, when, and where to manipulate.

[Insert Figure 8 about here please.]

Preliminary Conclusions on Ethnographic Data

Preliminary findings suggest important differences in how the four students at the table approach CTA, as it is enacted. Emphatically, each student in the video is not intended to “represent” a demographic group. However, each child does show a distinctive pattern of interacting with his/her peers and the teacher in the context of CTA. Students’ real-time interpretations of curriculum lessons provide an insider’s perspective to 2061 criteria. Quantitative analysis of video data reveals that students’ interpretive strategies differ systematically; for example, for Mike, achieving the “lesson purpose” is a matter of adopting a rhetorical stance in line with the curriculum; for Angelique and Rafael, achieving lesson purpose is more embedded in lesson practices and procedures.

CONCLUSIONS

The first research goal of this planning grant study was to test the efficacy of a highly rated curriculum unit for a diverse student population. We anticipated that CTA would be somewhat more effective than the “standard fare” due to its instructional qualities. These were well defined through the Project 2061 Curriculum Analysis. However, we were unsure whether some groups of students might be adversely affected by CTA. This was because the research literature on culturally and linguistically diverse populations suggested that the backgrounds that students bring to science classrooms might not match the kinds of “doing, thinking, talking science” required by CTA. We were also interested in exploring the relative importance of the curriculum rating criteria.

We developed the Conservation of Matter Assessment, guided by conceptual change theory (i.e., students’ gradual cognitive movement from naive conceptions to more scientifically accurate ones through experiences with physical phenomena).

Encouragingly, in nearly all analyses of the Conservation of Matter Assessment, students who experienced CTA had significantly higher scores than their comparison peers.

Although achievement gaps did not shrink among CTA groups from pre- to posttests, in the comparison condition, gaps appeared to widen. With the CTA curriculum, most students made progress toward a flexible use and understanding of “conservation of matter.” We also measured student motivation and engagement, and in two of the five scales found main effects favoring CTA. This suggests that the instructional characteristics of CTA supported the conceptual learning of the majority of students, and were more effective than the standard fare.

The only exception was that the CTA and Comparison group students currently eligible for ESOL services showed similar results for the Conservation of Matter Assessment. Perhaps the literacy demands of CTA are too great and need modification, or perhaps our assessment failed to capture the learning gains of these students. The goal orientation and engagement results indicate that most of these students viewed CTA about the same as the regular curriculum, with surprisingly high scores reported on these scales. These results require more exploration.

The second goal of this study, a highly ambitious one, was to develop a framework to analyze how CTA functioned in one diverse classroom setting. The ethnographic portion of our study sought to identify how one diverse group of four students worked with one another as the CTA curriculum was enacted. Video data were analyzed to tease out how the students experienced the instructional characteristics attributed to CTA. We found that CTA offered diverse students varied, numerous entries to understanding the target concept. Students seemed to form a functional group with the

aim of marching through the CTA lessons. Clearly, the students assumed different roles as the curriculum was enacted. Overall, our data show students with measurably distinctive but consistent patterns of verbal and non-verbal interaction as they encountered CTA curriculum materials and responded to its requirements. As the unit progressed, all four students exhibited increasing levels of involvement and “science talk” (Lee, 2000), discussion consistent with science habits of mind (AAAS, 1993). We documented variations in the amount of talking, manipulating of materials, use of scientific terms, seeking clarification, giving or taking direction, and many other factors.

Our goal is to continue to make connections between the ethnographic and experimental portions of the study to explain how CTA functions for diverse learners. For instance, the experimental data show that students who formerly were eligible for ESOL services made great gains with CTA. The video data show that Rafael, a former ESOL student, did not speak much, but manipulated the science materials frequently. This provided him with an active role in the group, although we infer that his developing language skills made CTA a challenge. Certainly Rafael’s experience with CTA cannot be generalized to all English Language Learners, but the video record of the daily activity provides a possible explanation for the efficacy of CTA.

The third goal of the planning grant study was to develop an understanding of what it might take to scale-up this research on curriculum implementation for a large number of schools across a school district (cf., Lynch, Szesze, Pyke & Kuipers, in press; Lynch, in press). As a result of the present study, we have received IERI funding for an additional five years to continue the inquiry discussed in this paper. The planning grant study reported here positioned us well for the next phases of our research: the further

exploration and development of theories about student diversity and science learning; exploring classroom ecologies as new curriculum units with specified instructional characteristics are provided; and, examining factors that affect student outcomes as highly rated curriculum materials are introduced on increasingly large scales.

References

- Aikenhead, G. S., & Jegede, O. J. (1999). Cross-cultural science education: A cognitive explanation of a cultural phenomenon. *Journal of Research in Science Teaching*, 36, 269-288.
- Allen, N. J. (1995, April). "Voices from the bridge" Kickapoo Indian students and science education: A worldview comparison. Paper presented at the annual meeting of the National Association for Research in Science Teaching, San Francisco, CA.
- American Association for the Advancement of Science. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.
- American Association for the Advancement of Science, (1999). *Project 2061 evaluation of science and mathematics Textbooks*. Retrieved October 27, 2003 from the web site: <http://www.project2061.org/research/textbook/default.htm>.
- American Association for the Advancement of Science (2003a, July). *Project 2061 analysis of science and mathematics assessment tasks*. Manuscript in preparation.
- American Association for the Advancement of Science (2003b, July). *Project 2061 development of science and mathematics assessment tasks*. Manuscript in preparation.
- Ames & Ames. (1984). Goals structures and motivation. *Elementary School Journal*, 85, 39-52.

- Bangert-Drowns, R. L., & Pyke, C. (2001). Student engagement with educational software: An exploration of literate thinking with electronic literature. *Journal of Educational Computing Research*, 24(3), 213-234
- Bruckerhoff, C. (1997, March). *Lessons learned in the evaluation of statewide systemic initiatives*. Paper presented to the Evaluation and Policy Studies Team at the National Institute for Science Education, University of Wisconsin, Madison, WI.
- Bulletin. (2000, Dec. 11). *Middle school gains fuel county rise on MSPAP*. Rockville, MD: Montgomery County Public Schools.
- Cazden, C. B. (2001). *Classroom discourse : the language of teaching and learning*. Portsmouth, NH: Heinemann.
- Corcoran, T. B. (1997, February). *Evaluating systemic reform*. Paper presented at the NISE Forum, Washington DC.
- Darling, A. L. (1989). Signaling Non-Comprehensions in the Classroom: Toward a Descriptive Typology. *Communication Education*, 38(1),34-40.
- Donmoyer, R. (1995). Rhetoric and reality of systemic reform: A critique of the proposed national science education standards. *Cognosis*, 4, 1-3, 7-8.
- Engestrom, Y., & Miettinen, R., et al. (1999). *Perspectives on activity theory*. Cambridge; NY: Cambridge University Press.
- Erickson, Frederick (1992). Ethnographic Microanalysis of Interaction. In M. LeCompte, et al (Ed.), *The Handbook of Qualitative Research in Education..* (pp. 201-225). New York: Academic Press.
- Fredericks, J. A., Blumenfeld, P.C. & Paris, A.H. (2004).School engagement: Potential of the concept, state of the evidence. *Review of Educational Research*, 74, 59-109.

Gilbert, A. & Yerrick, R. (2001). Same school, separate worlds: A sociocultural study of identity, resistance, and negotiation in a rural, lower track science classroom.

Journal of Research in Science Teaching, 38(5),574-598.

Gumperz, J. J. (1982). *Discourse strategies*. Cambridge Cambridgeshire ; New York: Cambridge University Press.

Halliday, M. A. K., & Martin, J.R. (1993). *Writing science : literacy and discursive power*. Pittsburgh, PA: University of Pittsburgh Press.

Heath, S. B. (1983). *Ways with words: Language, life, and work in communities and classrooms*. Cambridge Cambridgeshire, New York: Cambridge University Press.

Holliday, W.G. (2003). Methodological concerns about AAAS's Project 2061 study of science textbooks. *Journal of Research in Science Teaching*, 40(5), p. 529-534.

Hymes, D. H. (1974). *Foundations in sociolinguistics; An ethnographic approach*. Philadelphia, PA: University of Pennsylvania Press.

Hymes, D. H. (1980). *Language in education: Ethnolinguistic essays*. Washington, D.C.: Center for Applied Linguistics

Jefferson, G. (1980). On "Trouble-Premonitory" Response to Inquiry. *Sociological Inquiry*, 50(3-4),153-185.

Kahle, J.B., Meece, J. & Scantlebury, K. (2000). Urban African-American middle school science students: Does standards-based teaching make a difference? *Journal of Research in Science Teaching*, 37, 1019-1041.

Kesidou, S. & Roseman, J.E. (2002). How well do middle school science programs measure up? Findings from Project 2061's curriculum review. *Journal of Research in Science Teaching*, 39, 522-549.

Kesidou, S., & Roseman, J.E. (2003). Project 2061 analyses of middle-school science

- textbooks: A response to Holliday. *Journal of Research in Science Teaching*, 40(5), p. 535-543.
- Laguarda, K. G. (1998). *Assessing SSI's impact on student achievement: An imperfect science*. Menlo Park, CA: SRI Associates.
- Latour, B. & Woolgar, S. (1986). *Laboratory life: The construction of scientific facts*. Princeton, N.J.: Princeton University Press.
- Lave, J., & Wenger, E. (1991). *Situated learning : legitimate peripheral participation*. Cambridge, NY: Cambridge University Press
- Lee, O. (2000). *Equity for Culturally and Linguistically Diverse Students in Science Education: Recommendations for a Research Agenda*. Paper presented at the National Institute for Science Education Forum, Detroit, May 2000.
- Lee, O. & Anderson, C. W. (1993). Task engagement and conceptual change in middle school science classrooms. *American Educational Research Journal*, 30, 585-610.
- Lee, O. & Luykx, A. (2004). Science education and student diversity: Synthesis and research agenda. A monograph supported by the Center for Research on Education Diversity, and Excellence (CREDE) at the University of California at Santa Cruz and the National Center for Improving Student Learning and Achievement (NCISLA) in Mathematics and Science at the University of Wisconsin in Madison.
- Lemke, J. L. (1990). *Talking science : language, learning, and values*. Norwood, N.J.: Ablex Pub. Corp.

- Lynch, S. (1997). Novice teachers' encounters with national science education reform: Entanglements or intelligent interconnections? *Journal for Research in Science Teaching*, 34 (1), 3-17.
- Lynch, S. (2000). *Equity and science education reform*. Mahwah, NJ: Lawrence Erlbaum and Associates.
- Lynch, S. (in press). ISO metaphor and theory for scale-up research: Eagles in the Anacostia and activity systems. Article submitted for publication to *Educational Researcher*, March, 2004.
- Lynch, S., Atwater, M., Cawley, J., Eccles, J., Lee, O., Marrett, C., Rojas-Medlin, D., Secada, W., Stephanich, G., & Willetto, A. (1996). *An equity blueprint for Project 2061 science education reform (second draft)*. Washington, DC: American Association for the Advancement of Science, Project 2061.
- Lynch, S., Szesze, M., Pyke, C. & Kuipers, J. (in press). Scaling-up highly rated middle science curriculum units for diverse student populations: Features that affect collaborative research, and vice versa. Forthcoming in B. Schneider (Ed.) *Conceptualizing Scale-Up: Multidisciplinary Perspectives*.
- Marks, H. (2000). Student engagement in instructional activity: patterns in the elementary, middle and high school years. *American Educational Research Journal*, 37(1), 153-184.
- Maryland School Performance Program, (2000, Oct. 5). Maryland State Performance Standards-MSPP School Improvement in Maryland: *What do our students need to know and be able to do?* <http://www.mdk12.org/mspp/standards/science/>.

- Midgley, C. et al., (2000). *Manual for the Patterns of Adaptive Learning Scales*. Ann Arbor, MI: University of Michigan.
- Moje, E.B. (2001). “Maestro, what is quality?”: Examining competing discourses in project based science. *Journal of Research in Science Teaching*, 34 (4), 469-498.
- Morris, C. W. (1946). *Signs, language and behavior*. New York: Prentice-Hall, Inc.
- National Center for Educational Statistics, (2003, September 12 update). *Science 2000 major results: Subgroup results for the nation*. Washington, DC: U.S. Government Printing Office.
- (<http://nces.ed.gov/nationsreportcard/science/results/index.asp>).
- National Commission of Science and Mathematics Teaching for the 21st Century, (2000). *The Glenn Commission Report: Before it's too late*. Washington, DC: Author.
- National Education Goals Panel. (1994). *Data volume for the National Education Goals Report (Vol.1)*. Washington, DC: Author.
- National Science Foundation. (1994). *Women, minorities, and persons with disabilities in science and engineering (1994 ed.)*. Arlington, VA: Author.
- National Science Foundation, (2000). *Interagency education research initiative (IERI), Program solicitation*. Arlington, VA: Author
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2002). *Scientific research in education. Committee on Scientific Principles for Education*. Shavelson, R.J., & Towne, L., (Eds.). Center for Education. Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

- Ochs, Elinor (1984). Clarification and Culture. In D. Schiffrin (Ed). *Georgetown University Round Table on Languages & Linguistics, 1984*. (pp. 325-341). Washington DC: Georgetown University Press.
- Pica, Teresa (1987). Second-language Acquisition, Social Interaction, and the Classroom. *Applied Linguistics* 8(1),3-21.
- Pintrich, P. R., (2000). Multiple goals, multiple pathways: The role of goal orientation in learning and achievement. *Journal of Educational Psychology*, 92, 544-555
- Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: The role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research*, 63(2), 167-199.
- Pyke, C., Lynch, S., & Kuipers, J. (2003). *Learning Science with Curriculum that Supports Instruction: A Study of Motivation Engagement and Achievement Among Learners Within a Diverse Population*. Paper presented at the 2003 Annual Meeting of the American Educational Research Association, Chicago, Il..
- Posner, G.J., Strike, K.A., Hewson, P.G., & Gertzog, W.A. (1982). Accommodation of scientific conception: Toward a theory of conceptual change. *Science Education*, 66, 211-227.
- Riechard, D.E. (1994). National science education standards: Around the reform bush...again? *Clearinghouse*, 67, 135-36.
- Roseman, J.E., Kesidou, S., & Stern, L. (1996, November). *Identifying curriculum materials for science literacy: A Project 2061 evaluation tool*. Paper presented for the National Research Council's Colloquium "Using the National Science

- Education Standards to Guide the Evaluation, Selection, and Adaptation of Instructional Materials," Washington, DC.
- Silverstein, M. (1993). Metapragmatic discourse and metapragmatic function. In J. Lucy (Ed.), *Reflexive language: Reported speech and metapragmatics*, (pp. 33-58). Cambridge: Cambridge University Press.
- Solano-Flores, G. & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching*, 38(5), 553-573.
- State of Michigan, (1993). *Chemistry That Applies*, Lansing, MI: Author.
- Stern, L. & Ahlgren, A. (2002). Analysis of students' assessments in middle school curriculum materials: Aiming precisely at benchmarks and standards. *Journal of Research in Science Teaching*, 39, 889-910.
- Strike, A. A., & Posner, G. J. (1992). A revisionist theory of conceptual change. In R. J. Hamilton (Ed.), *Philosophy of science, cognitive psychology, and educational theory and practice* (pp. 147-176). New York: State University of New York Press.

Table 1

Demographics: Number of Participants by Group and Curriculum Condition

Sample by Category	n	Curriculum Condition	
		Comparison	CTA
Gender	2259		
Male		533	595
Female		519	612
Ethnicity	2259		
American Indian		3	3
African American		343	361
Asian American		141	163
Hispanic American		199	281
White/European American		366	399
Free and Reduced-price Meals System	2243		
Never		552	608
Prior		220	234
Now		272	357
Special Education Status	2243		
Never		927	1047
Prior		57	71
Now		60	81
English for Speakers of Other Languages Status	2243	843	930
Never		140	191
Prior		61	78
Now			

Note. Total n varies by group due to missing values in the data as reported by MCPS.

Table 2

Means of Achievement, Engagement, and Goal Orientation by Curriculum Condition

Variable	Test	M	SD	Curriculum Condition				
				Comparison		CTA		
				n	M	SD	n	
Conservation of Matter								
	Pre	21.72	17.85	813	21.66	18.58	1119	
	Post	32.71	25.84	809	41.68	29.55	1087	
	Delayed	29.60	23.92	877	39.68	28.17	1023	
Learning Engagement								
Basic		Pre	3.67	0.77	1053	3.75	0.75	1140
		Post	3.62	0.76	923	3.71	0.74	1037
Advanced		Pre	2.88	0.80	1037	2.93	0.78	1119
		Post	2.91	0.84	900	2.96	0.81	1024
Goal Orientation								
Mastery Goal		Pre	4.04	0.86	1036	4.22	0.79	1129
Orientation		Post	4.02	0.88	919	4.22	0.80	1028
Performance		Pre	2.72	0.99	1047	2.76	0.99	1128
Orientation/Approach		Post	2.67	1.05	921	2.68	1.03	1042
Performance		Pre	2.94	0.95	1046	3.00	0.95	1130
Orientation/Avoid		Post	2.84	0.98	923	2.83	1.00	1038

Note. The reported n is the number of students that completed all tests: pretest, posttest and delayed posttest in each subgroup. Individual scores are derived from an unweighted average Likert items from the engagement and motivation questionnaire. 1 = Not true at all, 3 = Somewhat true, 5 = Very true.

* Differences in means for curriculum conditions controlling for pretest differences $p < .05$.

Table 3

Means and Standard deviations of Content Assessment Scores for Subgroups Partitioned Within Demographic Categories

	Demographic Category											
	Gender		Ethnicity				FARMS			ESOL		
	Male	Female	AfAm	AsAm	Hisp	White	Never	Prior	Now	Never	Prior	Now
n	Comparison											
Pretest	200	208	149	53	69	137	205	91	109	329	57	19
(SD)	20.95	21.24	16.91	23.12	16.79	27.73	24.16	20.01	16.72	22.14	19.47	12.79
Posttest	(17.54)	(16.59)	(14.48)	(16.59)	(13.65)	(19.47)	(18.13)	(16.55)	(14.53)	(17.12)	(17.94)	(12.35)
(SD)	32.85	33.89	25.17	39.58	22.62	45.44	41.31	28.50	21.98	34.71	29.13	21.53
Delayed	(28.26)	(23.75)	(22.56)	(24.70)	(19.39)	(27.57)	(27.79)	(22.99)	(18.84)	(26.48)	(22.91)	(23.19)
Posttest	31.53	31.76	26.06	35.38	24.74	39.92	37.39	27.50	23.63	32.71	28.57	20.97
(SD)	(24.87)	(24.00)	(22.31)	(26.89)	(19.72)	(25.38)	(26.61)	(21.22)	(18.81)	(24.68)	(23.92)	(15.82)
n	Chemistry That Applies											
Pretest	425	440	246	132	180	307	463	156	240	688	136	35
(SD)	22.50	21.95	16.89	25.54	17.69	28.32	25.77	19.74	17.39	23.47	18.17	15.17
Posttest	(19.30)	(18.02)	(15.58)	(17.96)	(17.08)	(20.20)	(19.73)	(17.68)	(15.77)	(19.28)	(15.69)	(13.53)
(SD)	43.37	43.57	36.05	51.78	34.41	51.63	50.39	39.36	33.84	46.26	37.34	21.87
Delayed	(29.32)	(28.89)	(26.68)	(30.45)	(26.88)	(28.55)	(29.32)	(27.72)	(26.10)	(29.05)	(26.64)	(24.04)
Posttest	40.65	40.46	31.49	46.43	32.91	49.75	45.89	38.32	31.78	42.74	33.76	25.86
(SD)	(27.98)	(27.76)	(25.14)	(29.08)	(24.88)	(27.58)	(28.61)	(26.91)	(24.39)	(28.53)	(23.73)	(19.40)

Note. The reported n is the number of students that completed all tests: pretest, posttest and delayed posttest in each subgroup.

Table 4

Means of Learning Engagement for Subgroups Partitioned Within Demographic Categories

Comparison	Demographic Category											
	Gender		Ethnicity				FARMS			ESOL		
	Male	Female	AfAm	AsAm	Hisp	White	Never	Prior	Now	Never	Prior	Now
Basic Learning Engagement												
Pretest	3.60	3.68	3.62	3.82	3.47	3.72	3.75	3.56	3.54	3.64	3.60	3.78
Posttest	3.58	3.61	3.59	3.66	3.49	3.63	3.68	3.48	3.50	3.58	3.53	3.89
CTA												
Pretest	3.70	3.81	3.72	3.91	3.69	3.77	3.78	3.75	3.71	3.74	3.77	3.92
Posttest	3.67	3.77	3.68	3.84	3.66	3.75	3.73	3.77	3.68	3.69	3.76	4.07
Advanced Learning Engagement												
Comparison												
Pretest	2.89	2.77	2.88	2.76	2.79	2.83	2.88	2.72	2.84	2.83	2.79	2.95
Posttest	2.96	2.83	2.89	3.01	2.80	2.91	2.97	2.81	2.80	2.89	2.84	2.98
CTA												
Pretest	2.98	2.90	3.00	2.94	2.85	2.94	2.95	2.88	2.94	2.94	2.90	2.91
Posttest	3.03	2.91	3.00	3.03	2.87	2.97	2.98	2.97	2.95	2.97	2.98	2.86

Note. Individual scores are derived from an unweighted average Likert items from the engagement and motivation questionnaire. 1 = Not true at all, 3 = Somewhat true, 5 = Very true.

* Differences in means for curriculum conditions controlling for pretest differences p<.05.

Table 5

Means of Goal Orientation for Subgroups Partitioned Within Demographic Categories

Test	Demographic Category											
	Gender			Ethnicity			FARMS			ESL		
	Male	Female	AfAm	AsAm	Hisp	White	Never	Prior	Now	Never	Prior	Now
Mastery Goal Orientation												
Comparison												
Pre												
Post	4.04	3.96	4.08	4.02	3.97	3.92	4.04	3.92	3.99	3.98	4.10	3.98
CTA	4.05	3.92	4.06	3.96	3.99	3.92	4.02	3.90	3.97	3.96	4.00	4.19
Pre												
Post	4.23	4.23	4.36	4.31	4.22	4.10	4.21	4.29	4.23	4.22	4.28	4.19
	4.21	4.23	4.27	4.30	4.22	4.14	4.19	4.35	4.20	4.22	4.18	4.32
Performance Orientation/Approach												
Comparison												
Pre												
Post	2.86	2.60	2.75	2.84	2.77	2.64	2.66	2.72	2.85	2.70	2.79	2.96
CTA	2.85	2.54	2.66	3.03	2.74	2.59	2.69	2.68	2.69	2.66	2.73	2.94
Pre												
Post	2.86	2.68	2.77	2.77	2.74	2.78	2.81	2.60	2.80	2.76	2.75	2.83
	2.77	2.57	2.65	2.76	2.68	2.64	2.68	2.56	2.72	2.66	2.67	2.77
Performance Orientation/Avoid												
Comparison												
Pre												
Post	3.04	2.87	3.02	3.06	2.94	2.85	2.92	2.87	3.06	2.93	3.03	3.04
CTA	2.99	2.73	2.87	3.12	2.87	2.75	2.83	2.89	2.84	2.83	2.85	2.99
Pre												
Post	3.05	2.96	3.05	2.98	2.98	2.99	3.02	3.00	2.99	3.00	2.95	3.24
	2.87	2.78	2.81	2.79	2.87	2.82	2.85	2.73	2.84	2.82	2.75	3.07

Figure 1. CTA and comparison group unadjusted pre- and posttest means on Conservation of Matter Assessment by Ethnicity

Figure 2. CTA and comparison group unadjusted pre- and posttest means on Conservation of Matter Assessment by Free and Reduced-price Meals System (FARMS) status

Figure 3. CTA and comparison group unadjusted pre- and posttest means on Conservation of Matter Assessment by English for Speakers of Other Languages (ESOL) status

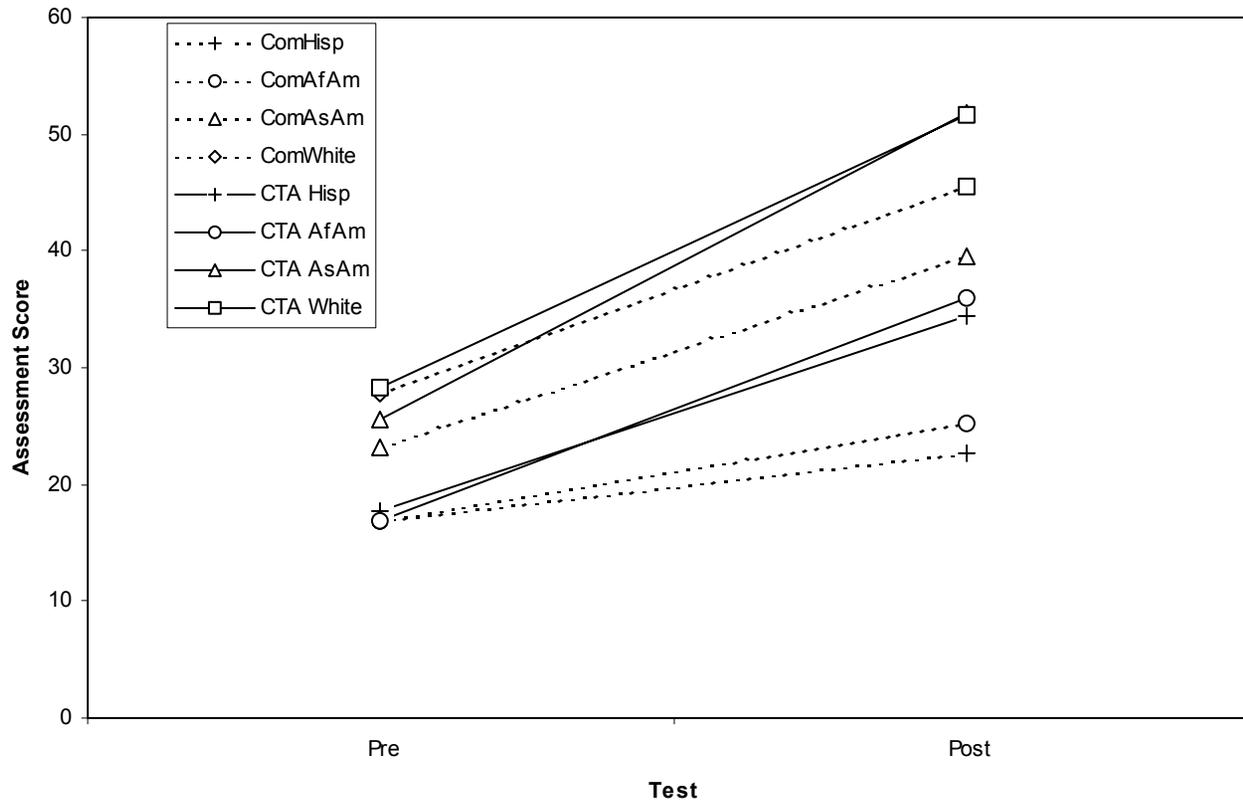
Figure 4. Adjusted posttest mean scores for Mastery Goal Orientation among Free and Reduced-price Meals System (FARMS) status by curriculum condition subgroups. Means adjusted by the pretest covariate, pretest mean = 4.2

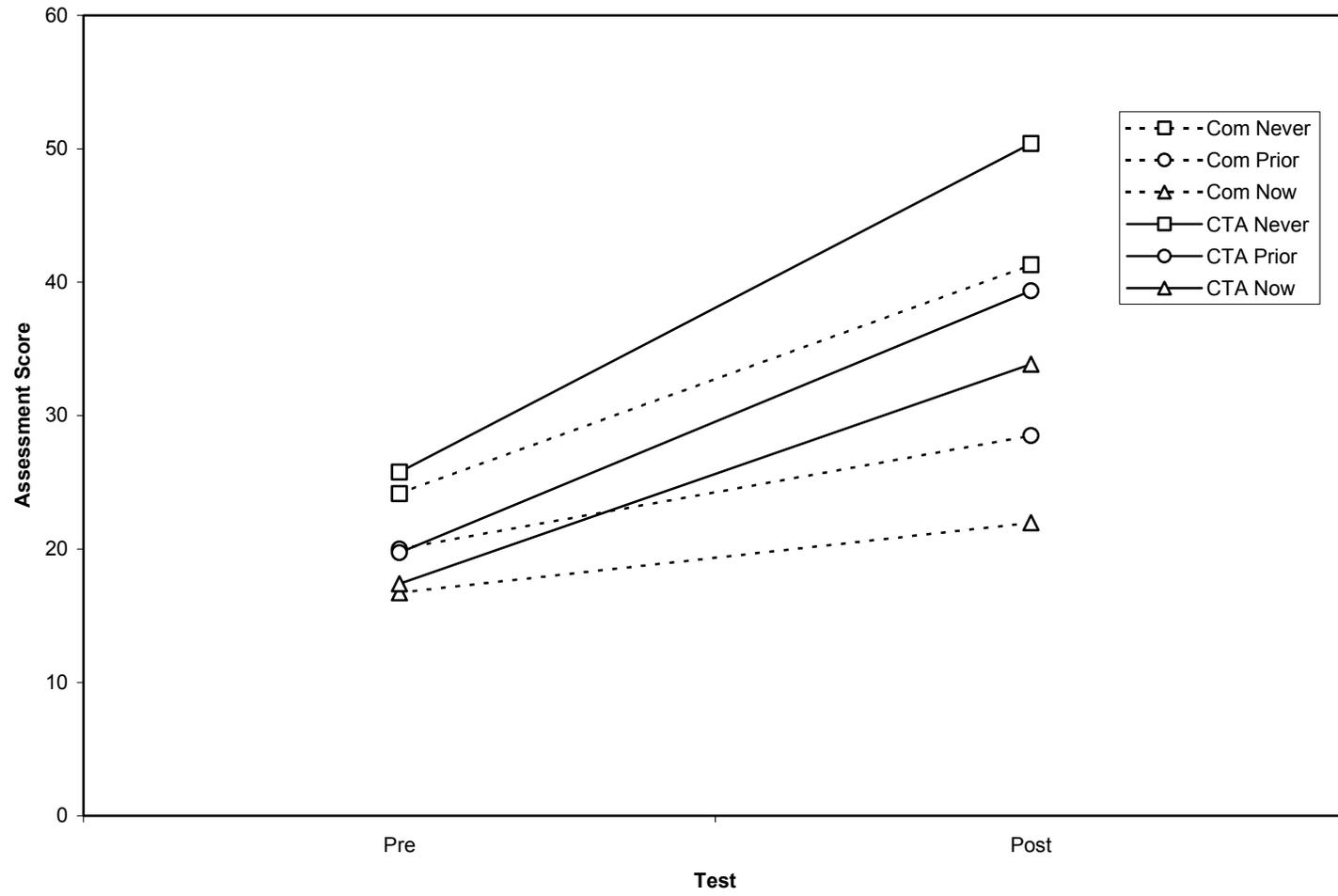
Figure 5. Adjusted posttest mean scores for Basic Learning Engagement among Free and Reduced-price Meals System (FARMS) status by curriculum condition subgroups. Means adjusted by the pretest covariate, pretest mean = 3.7

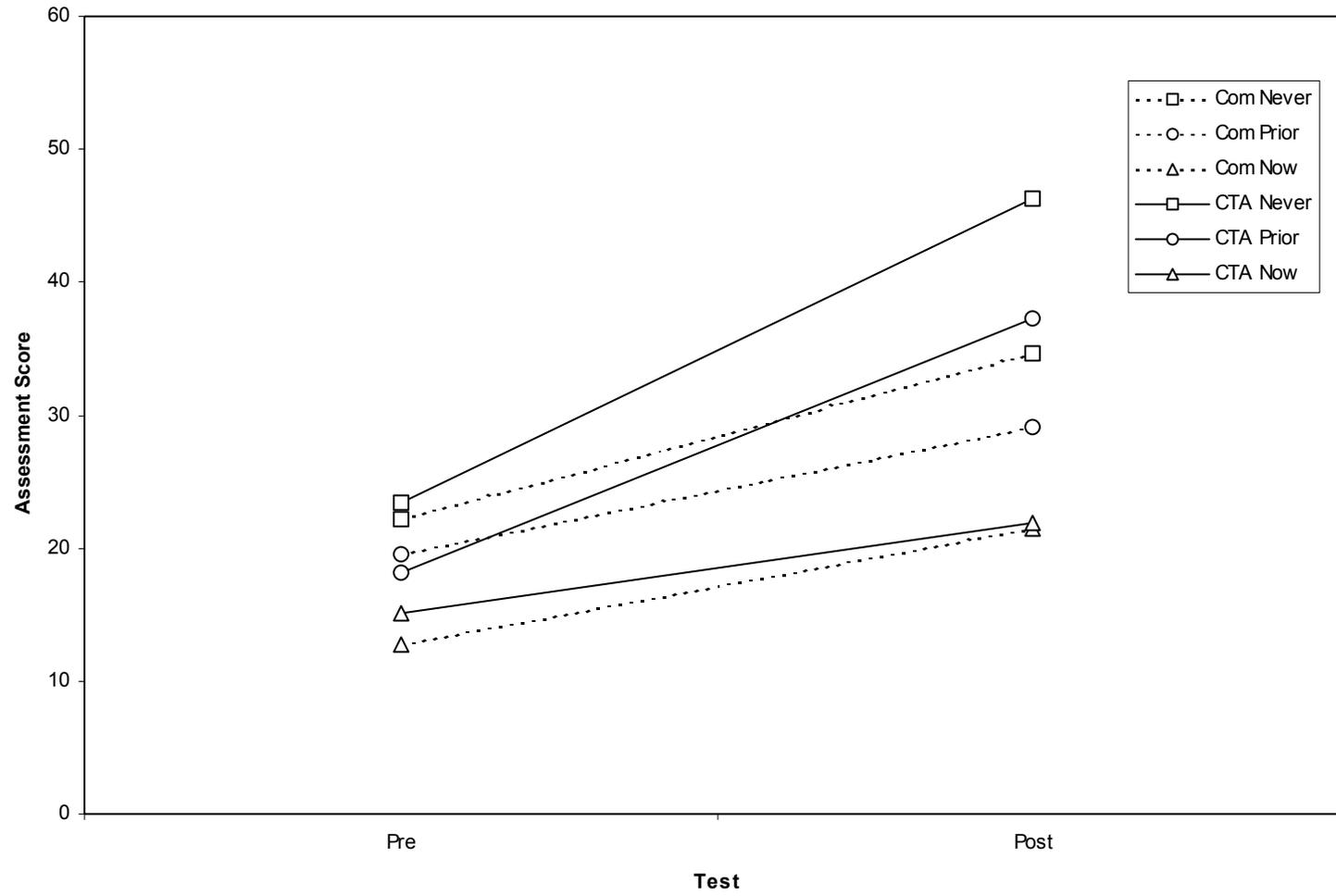
Figure 6. Use of scientific terms by students as a percentage of each student's individual total utterances

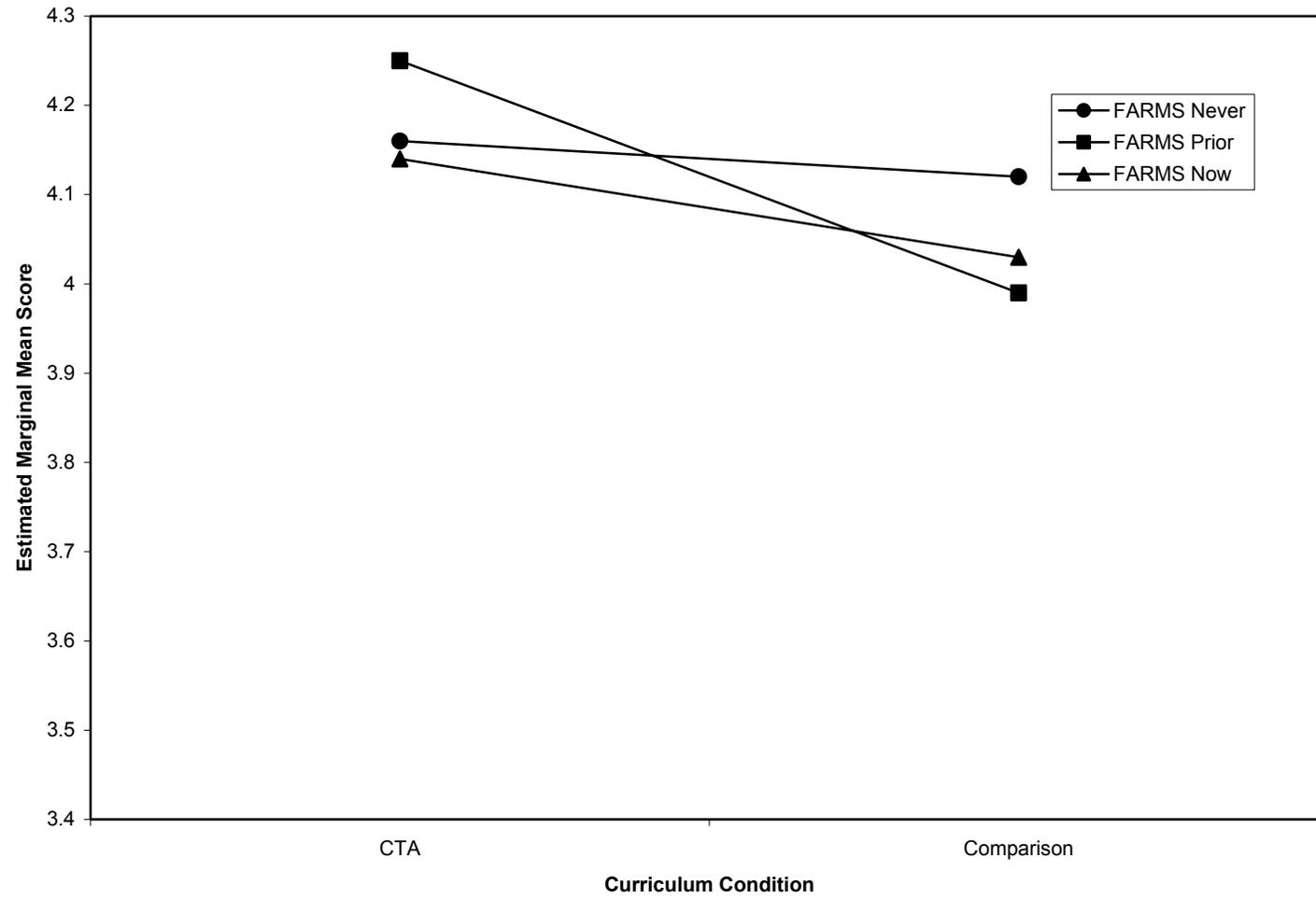
Figure 7. Percentage of each student's utterances that clarify peers' trouble statements

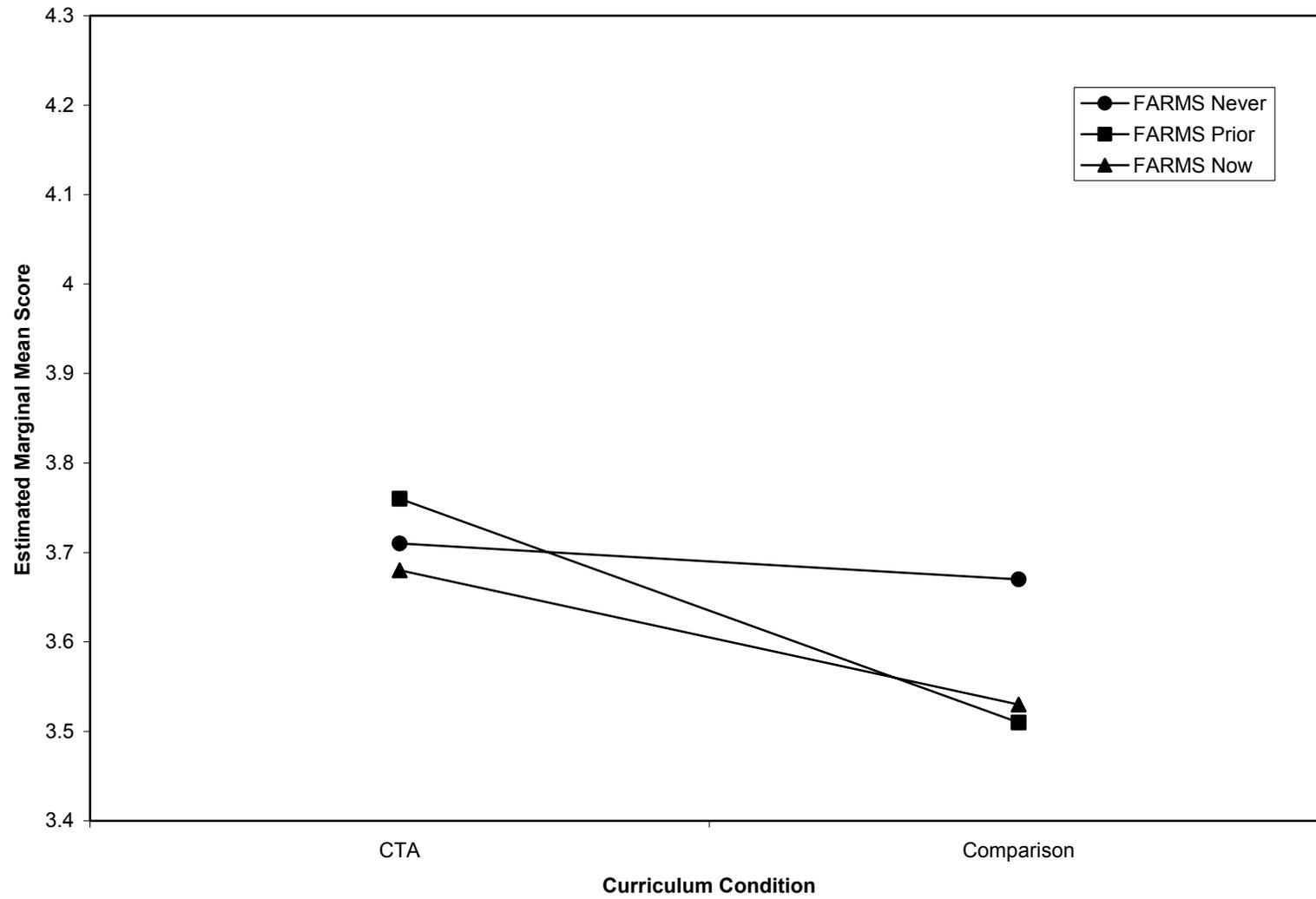
Figure 8. Number of manipulative events for each student

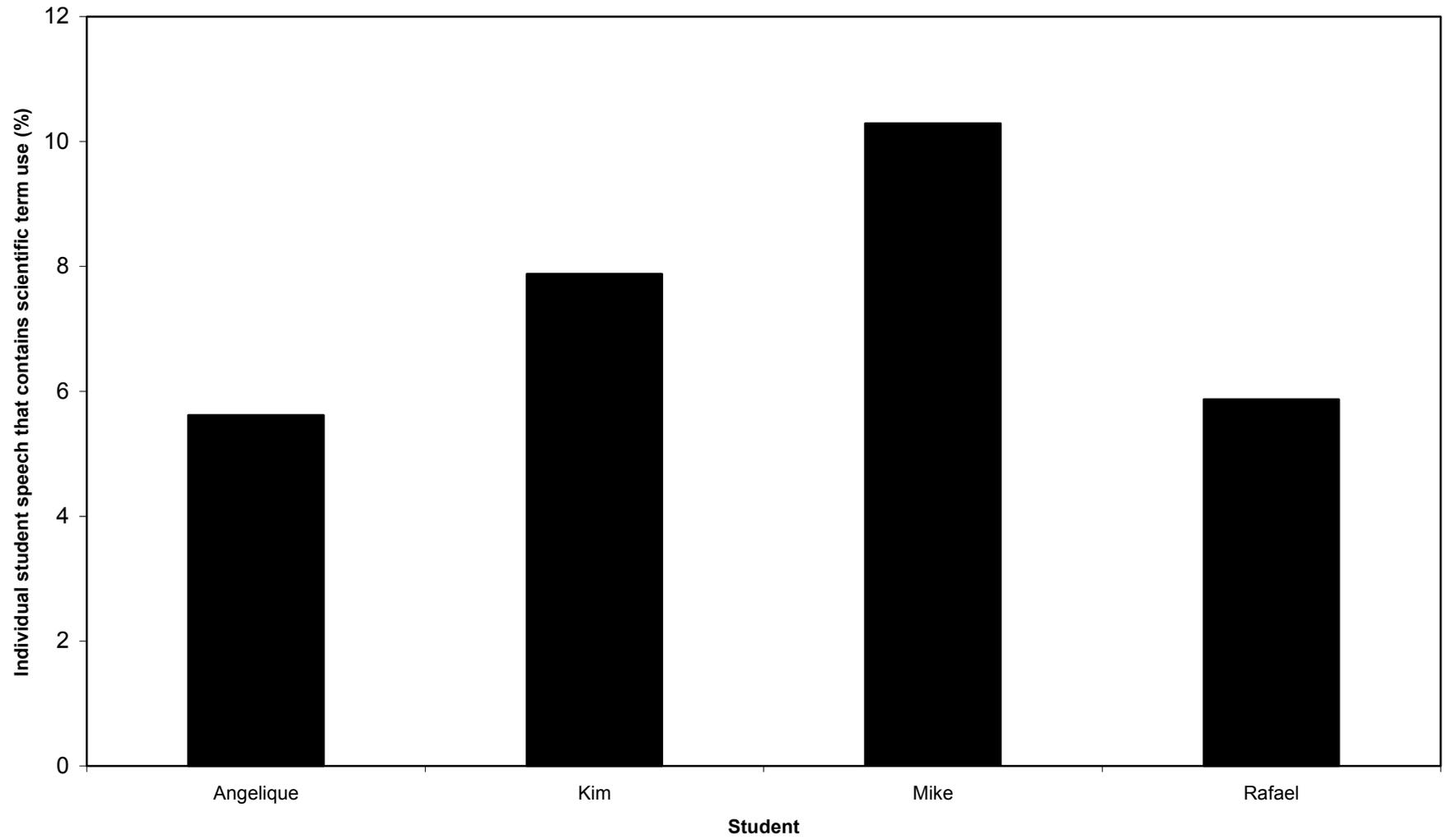


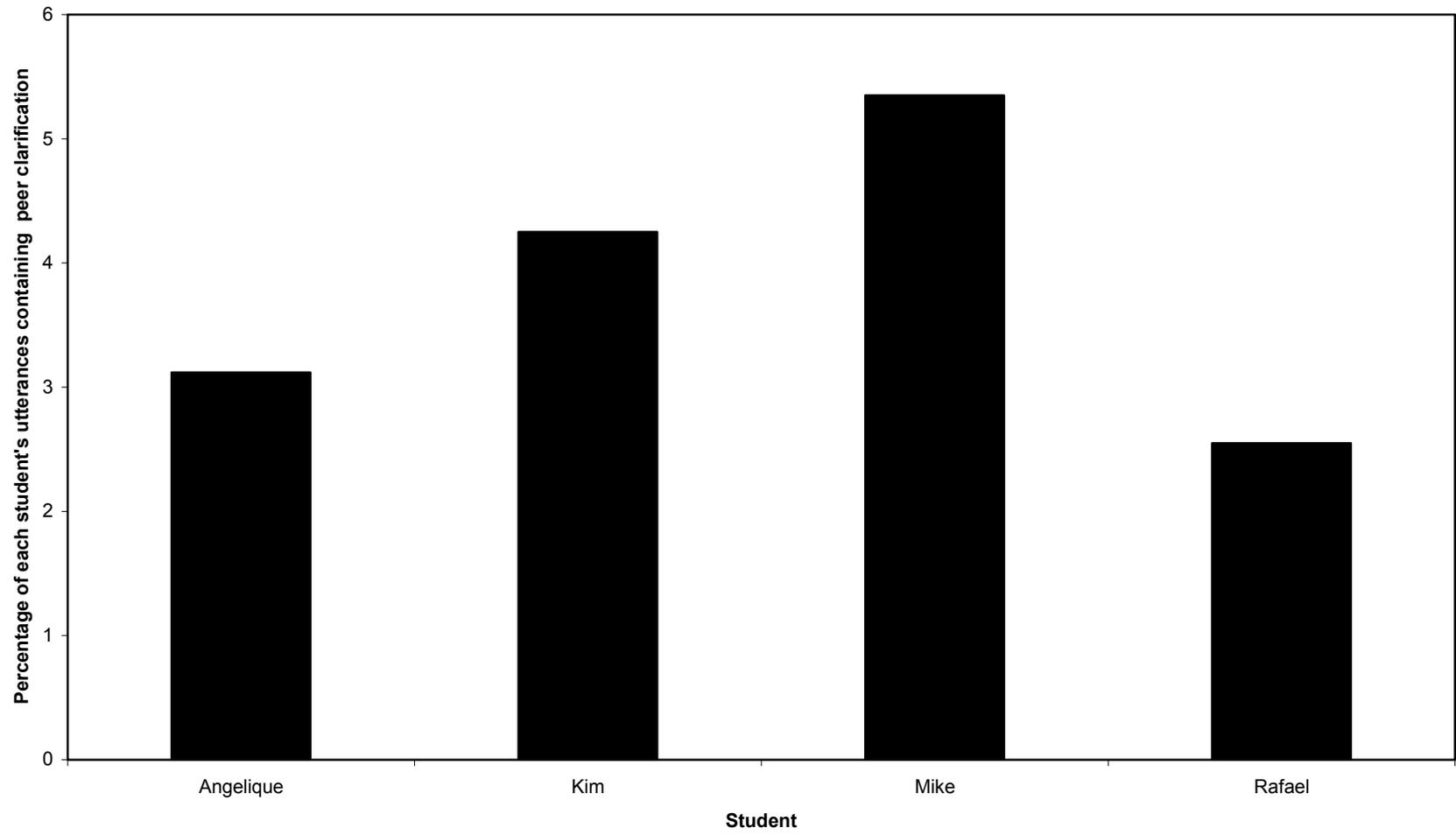


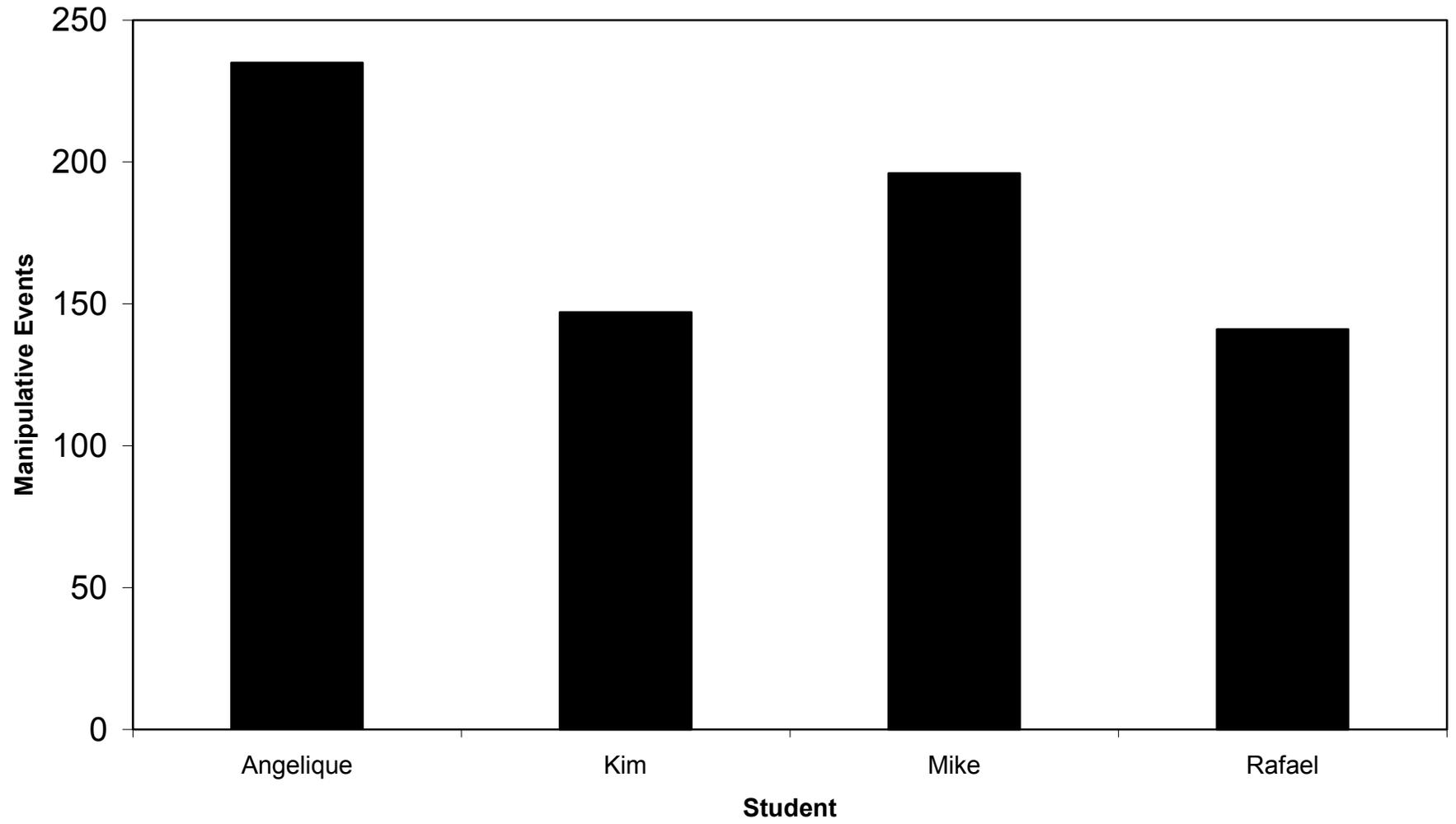












Author Notes

The authors of this study are: Sharon Lynch, Professor, Department of Teacher Preparation and Special Education, The George Washington University; Joel Kuipers, Professor, Department of Anthropology, The George Washington University; Curtis Pyke, Associate Professor, Department of Teacher Preparation and Special Education, The George Washington University; and, Michael Szesze, Director of preK-12 Science Education, Montgomery County Public Schools, Montgomery County Maryland.

This work is supported by the National Science Foundation, the U.S. Department of Education, and the National Institute of Health (REC-0228447). Any opinions, findings, conclusions, or recommendations are those of the authors and do not necessarily reflect the position, policy of endorsement of the funding agencies. We would like to acknowledge research assistants Betsy Moyer and Drew Johnson who contributed to the collection of data during the planning grant, and of research assistant Annie Knight Hansen, who assisted in the preparation of this manuscript.

Please address correspondence concerning this article to Sharon Lynch, Department of Teacher Preparation and Special Education of The George Washington University, 2134 G. St. NW, Washington DC 20052. E-mail: slynch@gwu.edu.