Abstract Title Page

Not included in page count.

Title: A Multilevel Analysis of a Guided Inquiry Curriculum Unit on Motion and Force for a Diverse Student Population in "Ability-grouped" Classrooms

Author(s): Vasuki Rethinam & Sharon J. Lynch

Abstract Body

Limit 5 pages single spaced.

Background / Context:

Description of prior research and its intellectual context.

This study is an evaluation of *Exploring Motion and Forces: Speed, Acceleration, and Friction (M&F)* developed by ARIES of the Harvard-Smithsonian Center for Astrophysics (2001*)*. Although *M&F* has been used in many schools and field-tested, it has never been evaluated in rigorous experimental or quasi-experimental studies. This research also asks if *M&F* is equitable, inquiring about its effectiveness in different types of classrooms where 6^{th} grade students are grouped by ability, as well for demographic subgroups of students (ethnicity or by socioeconomic, ESOL, or disability status). .

This study is important for several reasons:

1. *M&F* has many (but not all) of the characteristics thought to be important in promoting student understanding of science concepts, as codified by the Project 2061 Curriculum Analysis (Kesidou & Roseman, 2002, 2003; Roseman, Stern, & Koppal, 2010, but see also Holliday, 2003).

2. *M&F* is a guided inquiry unit; there is considerable controversy over the effectiveness of inquiry as a method of teaching students science concepts (Coburn, Schuster & Adams, 2010; Hmelo-Silver, Duncan, & Chinn, 2007; Kirschner, Sweller, & Clark, 2006; Klahr, 2009). Moreover, few studies of inquiry in Minner, Levy and Century's (2009) meta-analysis of inquiry-based instruction were found to be methodologically adequate, and none used multi-level modeling to gauge impact (Minner, personal communication).

3. This study explores equity issues, describing *M&F*'s effects in different types of classrooms and for diverse subgroups (c.f., Secker & Lissitz, 1999). The classroom contextual variable of interest is "ability grouping", defined by "Classroom Achievement History" (CAH), or mean performance on prior standardized tests. Because students are nested in classrooms in which curriculum materials are enacted, one particularly salient contextual classroom variable is the placement of students in classrooms by administrative decision due to their perceived ability, prior standardized test scores, or other characteristics associated with ability such as special education or English Language Learner status (Spring, 2008).

The conceptual framework guiding this study is that curriculum materials shape what students learn and how teachers teach (Schoenfeld, 2004). In science education, curriculum materials are often the focus of both research and policy because they are seen to guide and improve teachers' practice (Banilower, Heck, & Weiss, 2007; Davis & Krajcik, 2005) and influence student outcomes. During the 1990's a new generation of curriculum materials was born, influenced by developments in cognitive science (Bransford, Brown, & Cocking, 1999). *M&F* has the instructional characteristics that modern cognitive research suggests would help students to learn about the challenging science ideas of motion and force, identified by using the Project 2061 Curriculum Analysis (Kesidou & Roseman, 2002; 2004). Project 2061 has used its Curriculum Analysis to identify curriculum materials *likely* to improve student learning in mathematics and in science, but hardly any science curriculum materials have passed muster (Kesidou & Roseman, 2002). This study asks if a curriculum unit such as *M&F* that meets many of the Project 2061 criteria is effective and equitable (Lynch, 2000), exploring the unit's effects in different types of classrooms and for diverse subgroups of students.

Purpose / Objective / Research Question / Focus of Study:
Description of the focus of the research.
The research questions are:

1. What is the impact of *M&F* (compared with classrooms using "business as usual" curriculum materials) on a curriculum-independent posttest of motion and force?
2. Does CAH (high vs. medium, and low vs. medium, at the class level) have an influence on mean student posttest scores (individual level)?
3. Does curriculum condition interact with CAH to influence student outcomes?
4. Does CAH interact with student demographic variables (ethnicity, ESOL, special education, or socio-economic status) to influence student-level outcomes?
5. Does curriculum condition interact with student demographic variables to influence student-level outcomes?

Description of the research location.
The setting for this study is a large metropolitan school system in the Central Atlantic region of the U.S. The system has no ethnic majority, and has substantial linguistic and cultural diversity with over 70 language groups represented. In addition, socioeconomic diversity is great and constantly increasing. This school district has consistently been one of the top performing districts in its state on measures of student achievement. The population for this effectiveness study of a middle school science curriculum unit was 6th grade students in the 2005-2006 school year, approximately 136,000 students overall, and 32,000 in grades 6-8.

Population / Participants / Subjects:
Description of the participants in the study: who, how many, key features or characteristics.
The Study Sample
The sample for the study consisted of 78 6th grade classrooms in 8 schools in a Central Atlantic metropolitan school district. Students in the sample included a total of 1,841 6th grade students. Treatment students were similar to comparison students in mean student scale scores in math and reading and distribution of student demographic characteristics (in percentages) in treatment and comparison.

Intervention / Program / Practice:
Description of the intervention, program or practice, including details of administration and duration.
Exploring Motion and Forces: Speed, Acceleration, and Friction (M&F), was developed for the ARIES curriculum program by the Science Education Department at the Harvard-Smithsonian Center for Astrophysics (2001*)*. *M&F* is a six-week physical science curriculum unit designed for grades 5-8 consisting of 18 "Explorations" or lessons. *M&F* consists of a *Teacher Manual* and a student *Science Journal*. There is no traditional student textbook. Each Exploration begins with an introspective query about the student's ideas about motion or force, asking for a written response. Next, typically, the teacher might introduce the day's activity. Then students working in groups, conduct the laboratory Explorations, guided by the *Science Journal*, and respond to question prompts in the *Journal*. The largest proportion of time spent on an Exploration can be described as student-group-centered lab work; groups of students set up equipment and assemble tools such as sliding disks; run trials and collect data; create tables and graphs guided by the *Journal;* and, summarize group data (Lynch, 2008). During the Exploration, students construct individual responses to question prompts in their *Journals* that asks them to think about their conceptual understanding about motion and force, but they may

discuss their ideas and results in groups. At the end of an Exploration, the teacher may engage the entire class in a discussion about their results and how the findings fit into their developing ideas. A substantial amount of lab materials are required to conduct the Explorations (sliding disks, ramps, marbles, bells, rolling carts), and teachers must be well-organized to keep the equipment in ready supply.

Curriculum materials in use in comparison 6th grade classrooms varied. Teachers in the comparison condition chose from a menu of available instructional materials for teaching the school district's curriculum framework about motion and force. Teachers reported that they used textbooks, handouts, Internet resources, and video clips. Some reported using a published curriculum unit developed by a former district teacher, and most used a Bill Nye video on motion and forces.

Research Design:

Description of research design (e.g., qualitative case study, quasi-experimental design, secondary analysis, analytic essay, randomized field trial).

In this quasi-experiment, despite the fact that the aim was to focus on classroom level effects, schools had to be used as the sampling unit because of school district policy that required all science teachers within a middle school to have the same professional development experiences and use the same curriculum materials. Schools in treatment and comparison conditions matched on socio-economic and other demographic variables and on prior standardized test scores. Different types of schools were matched and then randomly assigned to treatment or business-as-usual comparison curriculum condition, *so all the 6th grade science classrooms within a school* were also so designated. This process resulted in a sample of treatment and comparison classrooms that were demographically very similar to one another for each type of school, as well as across schools, overall. The goal of this assignment method was to create two similar samples of classrooms because the curriculum materials were delivered at the classroom level, and are one unit of analysis in this study. Within each school, the two science classrooms with the highest mean scores for reading and math and prior science grades were selected from all 6th grade science classrooms for that school, and these classrooms were designated *high achievement history classrooms* by the authors for the purpose of this study. A similar process was used to designate two *low achievement history classrooms* for each middle school. The science classrooms sandwiched in between the high and low achievement history classrooms were designated *medium achievement history classrooms*, no matter their number, which varied according to the size of the middle school. In all, there were 16 science classrooms in the low achievement history category; 16 in the high achievement history category, and 46 in the medium achievement history category. Treatment and comparison classrooms had similar demographic characteristics for low, medium and high achievement history classrooms.

Data Collection and Analysis:

Description of the methods for collecting and analyzing data.

The treatment and comparison units on motion and force ideas were implemented at about the same time across all schools in the study. Teachers in both treatment and comparison conditions agreed to be observed in their classrooms as the units were implemented and to be interviewed by researchers after the units were completed. Treatment and comparison teachers reported no significant difference in the mean number of hours spent teaching about motion and forces ideas. During this study, there was emphasis placed on fidelity of implementation (FOI) to ensure that the *M&F* curriculum unit was actually enacted with students in treatment classrooms (O'Donnell et al., 2007). Findings indicated most of *M&F's* lesson components observed were implemented with a high degree of fidelity to the lesson's intent across virtually all of the treatment classrooms. In addition, trained observers also visited about half of the comparison classrooms and all of the comparison teachers. Researchers interviewed all the comparison teachers.

A two-level hierarchical linear model was used. HLM, version 6.0.3 (Raudenbush, Bryk, Cheong, & Congdon, 2004) was employed for data analyses. SPSS, version 12.0.1 was used to enter student- and classroom-level composite variables, and create the SSM (sufficient statistics matrix) file for HLM analysis. Categorical variables were dummy coded as 0 and 1. Student demographic variables, e.g., ESOL or special education status, were modeled at the individual level and fixed at the classroom level. The relationship between the individual level variables and outcomes was fixed across classrooms. When fixed, the outcome scores were predicted only by an intercept (see Raudenbush & Bryk, 2002). Classroom-level covariates (LAH, HAH, and curriculum condition) were entered to help reduce the unexplained variance attributed to the classroom in outcomes.

Student understanding of the motion and force was determined by the Motion and Forces Assessment (MFA), a curriculum-independent test with credible psychometric properties. Student scores on MFA were transformed into weighted scale scores ranging from 0 to 100. Student level (or level 1) predictors were FARMS, SPED, ESOL, African American, and Hispanic. Classroom level (or level 2) predictors were Curriculum Condition (CC) and Classroom Achievement History including Low, Medium and High Achievement History Classrooms (LAH, MAH, and HAH, respectively). Using the HLM program (Raudenbush, Bryk, Cheong, & Congdon, 2004), hierarchical linear modeling technique was employed for the data analyses.

Findings / Results:

Description of the main findings with specific details.

First, a fully unconditional model was estimated, without any predictors at individual- or classroom-level. Results showed that the student-level variance (σ^2) was 422.29, and the classroom-level variance (τ) was 85.72. SD of classroom means ($\sqrt{\tau}$) was 9.26. This value is used for determining effect sizes for the main effects. The infraclass correlation (ICC, calculated as $\tau/(\tau+\sigma^2)$ was 0.17. The ICC is the proportion of variance in the posttest scores that exists between classrooms. Most of the variance exists between students (83%) within classrooms. Because the variance accounted for at the classroom (higher) level is greater than 10%, a multilevel analysis was justified (Lee, 2000).

Next, the fully-conditional model was estimated. The non-significant variables were not removed from the model. The final model had student-level variance (σ^2) of 377.91 and classroom-level variance (τ) of 77.99 so that 10.7 % of the variance in the data was explained by all predictors. Analysis of the results, shown in Table 4, revealed no significant influence of LAH on the mean posttest scores of students (γ_{02}); no significant interaction between CC and CAH, and no significant interaction between CC and demographic variables such as special education status (γ_{11}), ESOL(γ_{21}), FARMS(γ_{31}), or ethnicity/race (γ_{41} and γ_{51}). The variables that were significant were: main effect of curriculum condition (γ_{01}), main effects of HAH (γ_{03}), interaction effects of LAH and ESOL status (γ_{22}), interaction effects of LAH and Hispanic (γ_{52}) and interaction effects of HAH and Hispanic (γ_{53}).

For the main effect of curriculum condition (CC) on student performance (γ_{01}), the estimated coefficient was 6.03 (SE=2.77, $p<.05$), indicating that treatment classrooms had higher post-test scores—6.03 points higher, on average—than comparison classrooms, ES = 0.65. The main effects of other student-level demographic variables were negative and statistically significant, indicating that such subgroups scored lower than the reference students. For example, special education status (SPED) had a coefficient of -4.32 for its main effect (γ_{10}; SE=1.27, p<.001). Student's eligible for special education had posttest scores 4.32 points *lower*, on average, than students who were not. This difference had ES of 0.68 (the value is negative in Table 4 because the coefficient is negative).

For the interaction effect of LAH and ESOL status (γ_{22}), the estimate was -9.93 (SE=2.93, p<.001). This indicates that ESOL students in LAH classes had, on average, scores 9.93 points *lower* than ESOL students in classes in MAH, with ES = 0.77.

There were significant cross-level interaction terms for HAH and LAH for Hispanic students. Hispanic students scored lower, on average (γ_{50}= -3.71; ES of 0.40); there were positive coefficients for Hispanic students in both LAH (γ_{52}) and HAH classes (γ_{53}). This indicates that Hispanic students in LAH classes had on average, scores 9.60 points higher than Hispanic students in classes in MAH, with ES = 0.77. Likewise Hispanic students in HAH classes had on average, scores 7.27 points higher than Hispanic students in classes in MAH, with ES = 0.59 (see Table 1).

[Insert Table 1.]

A second set of HLM analyses was conducted after removing the non-significant variables (see Table 2) and variables that were significant in the full model remained significant in the reduced model except for HAH variable. A deviance statistic was used to compare the difference in fit between the two models and the analysis showed as non-significant, indicating that the reduced model is as good a fit of the data as the full model. However, for reporting purposes we use the full conditional model.

[Insert Table 2.]

Conclusions:

Description of conclusions, recommendations, and limitations based on findings.

This study showed that the *M&F* guided inquiry curriculum unit (with many of the important attributes identified by the Project 2016 Curriculum Analysis as important for effective instruction and learning) yielded higher mean classroom level outcome scores on an assessment of motion and force concepts, ES = .65. This study also explored classroom level effects of *M&F* for classrooms with low, medium or high achievement histories, but showed no significant interactions between curriculum condition and classroom achievement history. This indicates that the positive effects of *M&F* were spread more or less evenly across classroom types, an important finding because some types of curriculum materials are often assumed to be better suited for some types of classrooms (and students) than others.

The finding that there was a significant interaction in a negative direction for student's eligible for ESOL services when placed in low achievement history classrooms, irrespective of curriculum condition, is particularly worrisome. Students eligible for ESOL services scored lower on MFA compared to their peers not eligible for ESOL services. Moreover, ESOL students in the low achievement history classrooms were at a substantial disadvantage compared with their ESOL peers in medium achievement history classrooms. This suggests that *M&F* alone is not making up for placement in low achievement history classrooms, and that some other services are needed.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

American Association for the Advancement of Science (AAAS). (2003). *Project 2061 middle grades science textbooks: A Benchmarks-based evaluation*. Retrieved June 1, 2004, from http://www.project2061.org/tools/textbook/mgsci/INDEX.HTM.

Banilower, E. R.*, Heck, D. J. & Weiss, I.R. (2007). Can professional development make the vision of the standards a reality? The impact of the national science foundation's local systemic change through teacher enhancement initiative. *Journal of Research in science teaching, 44*, 375-395.

Embryo, K. L., & Schumacher, R. E. (2003, April). *Investigating classroom effectiveness measure over time*. Paper presented at the Annual meeting of the American Educational Research Association. San Diego, California.

Bransford, J., Brown, A., & Cocking, R. (1999). How people learn: Brain, mind, experience, and school. Washington, DC: National Academy Press.

Coburn, W. Schuster, D. & Adams, B. (2010). *Experimental Comparison of Inquiry and Direct Instruction in Science.* Paper presented at the 2010 annual conference of the Society for Research Effectiveness, Washington DC March 4-6.

Davis, E. A. & Krajcik, J. S. (2005). Designing Educative Curriculum Materials to Promote Teacher Learning. *Educational Researcher*, *34*, 3-14.

Harvard Smithsonian Center for Astrophysics. (2001). *ARIES-Exploring Motion and Forces: Speed, Acceleration, and Friction*. Watertown, MA: Charlesbridge Publishing

Hmelo-Silver, C. E., Duncan, R. G. & Chinn, C. A. (2007). Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, and Clark (2006). *Educational Psychologist, 42*(2), 99–107.

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

Kirschner, P.A., Sweller, J. & Clark, R.E. (2006). Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching. *Educational Psychologist*, *41*(2), 75 – 86.

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, 535-43.

Kesidou, S., & Roseman, J.E. (2004). Can middle-school science textbooks help students learn important ideas? Findings from project 2061's curriculum evaluation study: Life science**. J***ournal of Research in Science Teaching,* 41, 538-568.

Klahr, D. (2009) "To every thing there is a season, and a time to every purpose under the heavens": What about Direct Instruction? In S. Tobias and T. M. Duffy (Eds.) *Constructivist Theory Applied to Instruction: Success or Failure?* Taylor and Francis.

Lee, V. E. (2000). Using Hierarchical Linear Modeling to study social contexts: The case of school effects. *Educational Psychologist, 35*, 2, 125-141.

Lee, V. E., & Loeb, S. (2000). School size in Chicago elementary schools: Effects on teachers' attitudes and students' achievement. *American Educational Research Journal, 37*, 1, 3-31.

Lynch, S.J. (2000). *Equity and science education reform.* Mahwah, NJ: Lawrence Erlbaum and Associates.

Lynch, S.J. (2008, March). How do curriculum materials improve student outcomes at the class level? Study construct validity, fidelity of implementation, and the comparison group. Paper presented at the 2008 annual meeting of the American Educational Research Association, New York, NY.

Minner, D.D. Levy, A. J., & Century, J. (2010). Inquiry-based science instruction - what is it and does it matter? Results from a research synthesis years 1984 to 2002**.** *Journal of Research in Science Teaching, 47, 474-496.*

Raudenbush, S. W., Bryk, A. S., Cheong, Y. F., & Congdon, R. T. (2004). *HLM: Hierarchical Linear and Nonlinear Modeling*, Chicago: Scientific Software International, Inc.

Raudenbush, S. W., Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and data analysis methods* (2nd ed.). Thousand Oakes, CA: Sage Publications.

Roseman, J.E., Stern, L., & Koppal, M. (2010). A method for analyzing the coherence of high school biology textbooks. *Journal of Research in Science Teaching, 47*, 47-70.

Rowe, K. J., & Hill, P. W. (1998). Modeling educational effectiveness in classrooms: The use of multi-level structural equations to model student's progress. *Educational Research and Evaluation, 4*, 4, 307-347.

Schoenfeld, A. H. (2004). The math wars. *Educational Policy,* 18(1), 253.

Secker, C.E., & Lissitz, R.W. (1999). Estimating the impact of instructional practices on student achievement in science. *Journal of Research on Science Teaching, 36*, 1110-1126.

Spring, J. H. (2008). *American Education*. New York: McGraw-Hill

Stern, L. & Roseman, J.E. (2004). *Journal of Research on Science Teaching, **41***, **538 – 568.**

Appendix B. Tables and Figures
Not included in page count.

Table 1.

Results from the fully conditional HLM analysis

Source		Coeff.	SE	p≤ .05	ES[a]
Intercept (γ_{00})		53.15	1.07	0.0001	
	CC (γ_{01})	6.03	2.77	0.03	.65
	LAH (γ_{02})	-4.64	4.41	0.30	--
	HAH (γ_{03})	6.30	3.15	0.05	.68
	LAH X CC (γ_{04})	0.51	5.30	0.92	--
	HAH X CC (γ_{05})	-4.59	5.19	0.38	--
SPED (γ_{10})		-4.32	1.27	0.001	-.47
	CC (γ_{11})	2.28	2.45	0.35	--
	LAH (γ_{12})	0.02	2.83	1.00	--
	HAH (γ_{13})	5.71	3.30	0.08	--
ESOL (γ_{20})		-7.74	1.46	0.0001	-.84
	CC (γ_{21})	1.46	2.77	0.60	--
	LAH (γ_{22})	-9.93	2.93	0.001	-.77
	HAH (γ_{23})	-2.97	4.50	0.51	--
FARMS (γ_{30})		-6.96	1.23	0.0001	-.75
African American (γ_{40})		-6.90	1.54	0.0001	-.75
	CC (γ_{41})	3.57	2.87	0.21	--
	LAH (γ_{42})	-1.13	3.31	0.73	--
	HAH (γ_{43})	-1.72	3.35	0.61	--
Hispanic (γ_{50})		-3.71	1.40	0.01	-.40
	CC (γ_{51})	1.31	2.70	0.63	--
	LAH (γ_{52})	9.60	3.72	0.01	.78
	HAH (γ_{53})	7.27	3.21	0.02	.59

[a] Effect sizes for cross-level interactions are computed by dividing each interaction coefficient by the standard deviation of the slope, calculated by multiplying the respective standard error by the square root of the sample size (for further discussion, see Lee, Loeb, & Lubeck, 1998). For example, the ESOL standard deviation is calculated as $1.46 \times \sqrt{78} = 12.89$.

Table 2.

Results from the Reduced Model

Source		Coeff.	SE	p≤ .05	ES[a]
Intercept (γ_{00})		53.15	1.07	0.0001	
	CC (γ_{01})	5.17	2.09	0.02	.56
	LAH (γ_{02})	-4.33	2.67	0.11	--
	HAH (γ_{03})	4.00	2.64	0.13	--
SPED (γ_{10})		-4.33	1.27	0.001	-.47