Abstract Title Page

Title: Multilevel Models for Estimating the Effect of Implementing Argumentation-Based Elementary Science Instruction

Authors and Affiliations:

Mack Shelley
Christopher Gonwa-Reeves
Joan Baenziger
Ashley Seefeld
Iowa State University

Brian Hand
William Therrien
Mary Grace Villanueva
Jonte Taylor
University of Iowa

Abstract Body

Context:

The U.S. began a new national standards movement in the area of K-12 science education curriculum reform in the 1980s known as “Science for All” to develop a population that is literate in economic and democratic agendas for a global market focused on science, technology, engineering, and mathematics (STEM) (Duschl, 2008). The National Research Council (NRC) report, *Taking Science to School: Learning and Teaching Science in Grades K-8* (TSTS; NRC, 2007b) described shortages in attracting students to science learning and careers, and of science teachers (particularly women and minorities). More recently, researchers have focused on science reform that incorporates a cultural imperative in the teaching of science (Driver, Leach, Millar, & Scott, 1996; Millar, 2006; Osborne, Duschl, & Fairbrother, 2002). The sister NRC report, *Rising above the Gathering Storm* (RAGS, NRC, 2007a), describes four areas of needed proficiency for science students of how to: generate and evaluate scientific evidence and explanations; know, use, and interpret scientific explanations of the natural world; understand the nature and development of scientific knowledge; and participate productively in scientific practices and discourse. To this end, pedagogical skills in science education have moved from teaching students how to memorize what they need to know from science textbooks to developing an understanding of the knowledge-building process by learning how to develop explanations and predictions about our world.

The two NRC reports demonstrate changes in pedagogy and instruction that appear to be better suited to the evolving technological world. As members of society are expected to process information that is updated constantly and rapidly, it is critical to understand how ideas are developed and processed. Research in abstract reasoning teaches us that infants learn causal inference and differentiation of animate and inanimate objects, demonstrating that the learning ability of even the youngest children permits them to engage in complex decision making (Gelman & Brenneman, 2004; Mertz, 2004; Spelke, 2000). To do this, students require abstract deductive and inductive reasoning skills, including the ability to view with an open mind and a willingness to be aware of the world (Critical Thinking Co., 2011).

Measuring student success. Raudenbush (2008) argues that, in contrast to past models that describe conventional resources such as per pupil expenditures, teacher credentials, physical facilities, or class size (Cohen, Raudenbush, & Ball, 2003) as the direct cause for student outcomes, instruction is the proximal cause for student learning and thereby places the emphasis on the continuous classroom interplay of assessment and instruction. One such current pedagogical practice is the Science Writing Heuristic (SWH) approach, which combines current understandings of learning as a cognitive and negotiated process with the techniques of argument-based inquiry (Duschl & Grandy, 2007; Hand, 2007), critical thinking skills, and writing to strengthen student outcomes. This enables students to develop critical thinking, habits of mind, and communications, and these abilities result in cognitive and meta-cognitive attributes that foster understanding of the nature of science, scientific inquiry, and the big ideas of science.

Fostering scientific discourse. The importance of the use of language in science has become prominent in the science education literature (e.g., Dawes, 2004; Shelley, Yore, & Hand, 2009; Yore, Bisanz, & Hand, 2003). Argumentation, known as “the language of science” (Duschl, Ellenbogan, & Eurduran, 1999), is used currently by researchers to promote classroom learning. Students are able to construct meaning as they interpret and reinterpret events through the argumentation-driven lens of their prior knowledge (Berk & Winsler, 1995; Tippett, 2009).

The SWH approach. The SWH approach to learning is based on improving students' understanding of science by embedding science argument within typical inquiry lessons. Within this approach teachers are required to align their pedagogical practices with how children learn. Students are required to pose questions, generate claims and evidence, compare their answers to others, and reflect on changes in their understanding. Several studies have suggested that this type of inquiry approach promotes critical thinking and reasoning, as students are required to use oral and written language to negotiate their understanding of science. For example, Akkus, Gunel, and Hand (2007) demonstrated significant gains in student performance on science components of standardized tests such as the Iowa Tests of Basic Skills (ITBS) and the Iowa Tests of Educational Development (ITED) comparing student test performance between classrooms with high levels of traditional science teaching and those with high-quality SWH implementation. The effect size difference of 1.23 between high and low student achievers and only 0.13 for students in classrooms where teachers used high levels of SWH teaching indicates that the SWH approach is effective for all learners in the classroom.

Other studies of applications of SWH pedagogical teaching in chemistry (Anderson & Bodner, 2008; Bhattacharyya & Bodner, 2005; Rudd, Greenbowe, & Hand, 2007) compared to traditional formats, showed significant association between higher explanation scores and the SWH format. Hohenshell and Hand (2006) showed that 10th grade biology students scored significantly better than control students on conceptual questions after completing laboratory activities using traditional and SWH approaches ($F(1,43)=5.53$, $p=0.023$, partial $\eta^2=0.114$). Poock, Burke, Greenbowe, and Hand (2007) demonstrated benefits for female students from high-implementation use of the SWH approach. Gunel (2006) found larger effect size changes in students' scores (1.0, vs. 0.4 or less) from higher levels of implementation of the SWH approach on students' performance on ITBS/ITED science tests across a 3-year period.

Two-sample *t*-tests conducted on 5th grade Cornell Critical Thinking (CCT) data (Shelley et al., forthcoming; Villanueva et al., 2011) demonstrated that SWH students ($n=1{,}154$), compared to control students ($n=882$) had significantly higher gains overall ($p=.002$), as well as in levels of science induction ($p=.010$) and deduction ($p=.004$), but did not show significantly different gains on observation and credibility ($p=.322$) nor the assumptions underlying science deduction ($p=.191$). A structural equation model (SEM) (Shelley et al., forthcoming) also showed that ITBS achievement measures of Reading Comprehension (RC), Math Concepts/Estimation (M1), Math Problems and Data Interpretation (M2), Science Comprehension (NSI), and Science Inquiry (SI) form a single overall metric of student outcomes, and that each score is related significantly to overall student achievement. Hand et al. (2011) showed that the SWH treatment had a significantly positive effect in SEM results on student gains from pretest to posttest (IMP) on CCT test scores (see Figure 1; SWH=1, Control=0).

Purpose:

The purpose of this paper is to examine the impact of implementation of the SWH approach at 5th grade level in the public school system in Iowa as measured by CTT student test (Ennis & Millman, 2005) scores. This is part of a project that overall tests the efficacy of the SWH inquiry-based approach to build students' content knowledge, argumentation skills, and interest in science to construct the foundation of science literacy with elementary school children, so that all students "become familiar with modes of scientific inquiry, rules of evidence, ways of formulating questions and ways of proposing explanations" (National Research Council [NRC], 1996, p. 21).

Setting:

A description of the SWH study by letter, followed by an in-person meeting, was completed in the summer of 2009 with school district superintendents in Iowa to obtain permission for participation by elementary school buildings in the study. After obtaining consent from the district superintendents, a total of 48 schools were recruited into the study.

Participants:

The study was conducted on Iowa elementary school students, in grades 3-6, with 24 school buildings randomly assigned to treatment and 24 to control. Data for ITBS were collected on over 60,000 students; CCT scores, measured only at 5th grade, were obtained on over 2,000 students at pretest and posttest.

Intervention:

Teachers in school districts randomized to the intervention group were trained in the SWH technique during the summer of 2009 at workshops held at four geographic regions of Iowa. This training took part over three days and included specific training on the SWH approach including how to foster argumentation skills in students in the classroom. University of Iowa SWH staff obtained video recordings of individual teachers' performance in science classroom at different times over the academic year. Classroom implementation has continued since Fall 2009; all selected schools remain in the study.

Research Design:

An experimental design was employed, with random assignment of participating elementary school buildings to SWH treatment or Control condition. Once recruitment of buildings was completed, blocks were formed for the purposes of randomization. Blocks were either districts with multiple buildings or districts that were similar in enrollment based on percentage of students on free and reduced lunch or certified enrollment. Two exceptions to this randomization strategy were as follows: (1) two religious schools of comparable size were blocked together, and the other religious school, of very small size, was paired with another school of very small size; and (2) 10 schools not randomized initially because their data arrived later were randomized into districts as we received them.

Data Collection and Analysis:

ITBS student scores were obtained for all students from the years immediately prior to the start of the SWH study. These ITBS scores included all composite scores and a subgroup score in science. The CCT test was administered in a Fall 2009 pretest and spring 2010 posttest; these results were combined with ITBS scores for data analyses. A multilevel model was estimated to assess the relative contributions of individual student (Level 1) variables, classroom (Level 2), and randomization unit (Level 3) which is a single building except where clustering of buildings was necessary owing to small enrollment. We treat student as nested within classroom and classroom as nested within randomization unit.

Results:

The model predicting change in CCT scores between pretest and posttest was estimated using R software by a linear mixed model fit by restricted maximum likelihood (Table 1). The estimated model, using R notation, is:

IMP ~ 1 + Trt + ASN + WHT + SED + GAT + FRL + T1M + T1L + PreTest + ASN:T1M + SED:T1L + Trt:T1L + Trt:SED + (1 | TID) + (1 | Unit1)

where IMP is posttest CCT scores minus pretest CCT score, Trt is SWH treatment vs. control, ASN is Asian American, WHT is white non-Hispanic, SED is special education, GAT is gifted and talented, FRL is free and reduced lunch, T1M is Type 1 remedial mathematics, T1L is Type 1 remedial language, PreTest is the pretest CCT test score, ASN:T1M is the interaction between Asian-American and Type 1 mathematics, SED:T1L is the interaction between special education and Type 1 language, Trt:T1L is the interaction between treatment/control and Type 1 language, Trt:SED is the interaction between treatment/control and special education, (1 | TID) is the Level 2 effect of classroom, and (1 | Unit1) is the Level 3 effect of randomization unit. The model was estimated on 2,009 5th grade students who completed both the CCT pretest and posttest at Level 1, 81 classrooms at Level 2, and 47 randomization units at Level 3. The random effects results show that classroom (TID) accounts for only about 2% of explained variance, and randomization unit (Unit1) accounts for about 7%. Effects not in the model are Asian American (AI), Hawaiian or Native Alaskan (HAW), English language learner (ELL), Section G504 special education status (G504), and African American (BLK), as well as related interactions, which were not significant in the presence of the other effects in preliminary models.

Fixed effect results for the model indicate significantly positive effects at Level 1 for SWH treatment compared to control (Trt), Asian-American student ethnicity (ASN), non-Hispanic white student ethnicity (WHT), gifted and talented status (GAT), and the interaction of Asian-American ethnicity with Type 1 Math (ASN:T1M); significantly negative Level 1 effects were estimated for special education status (SED), free and reduced lunch status (FRL), Type 1 Math (T1M), pretest score (PreTest), and the interaction between treatment and Type 1 Language (Trt:T1L). The main effect of Type 1 Language (T1L), the interaction between special education and Type 1 Language (SED:T1L) and the interaction between treatment and special education (Trt:SED) were not significant.

Conclusions:

These multilevel results clearly support the conclusion of significant CCT improvements due to SWH implementation compared to control, after accounting for appropriate covariates. The relatively minimal variance components attributable to randomization unit and classroom support the validity of the randomization strategy and the lack of confounding attributable to innate differences in teacher characteristics.

The results reported here are for a multilevel model that uses individual student differences between posttest and pretest CCT scores to measure “improvement.” Subsequent analysis will incorporate level of teacher implementation based on classroom observations, use maximum likelihood estimates for missing data replacement to enhance model power, and compare different missing data substitution methods (mean overall, student mean, question mean, a combination of student and question means). We will try using the above as probabilities in Bernoulli trails and bootstrap estimate to see how variance estimates change. We also will investigate subcategories of CCT and ITBS data, by addressing the nested data structure instead of treating the sum of scores across subcategories as approximately normal.

Appendices

Appendix A. References

Akkus, R., Gunel, M., & Hand, B. (2007). Comparing an inquiry-based approach known as the science writing heuristic to traditional science teaching practices: Are there differences? *International Journal of Science Education, 29*(14), 1745-1765.

Anderson, T.L., & Bodner, G.N. (2008). What can we do about 'Parker'? A case study of a good student who didn't 'get' organic chemistry, *Chemistry Education Research and Practice, 9*, 93-101. DOI: 10.1039/B806223B.

Berk, L.E., & Winsler, A. (1995). *Scaffolding children's learning: Vygotsky and early childhood education.* Washington, DC: National Association for the Education of Young Children.

Bhattacharyya, G., & Bodner, G.M. (2005). It gets me to the product: How students propose organic mechanisms. *Journal of Chemistry Education, 82*(9), 1402. DOI: 10.1021/ed082p1402.

Cohen, D.K., Raudenbush, S.W., & Ball, D.B. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25,* 119-142.

Critical Thinking Co. (2011). URL www.criticalthinking.com/company/articles.jsp?code=c. Accessed July 20, 2011.

Dawes, L. (2004). Talking and learning in classroom science. *International Journal of Science Education, 26,* 677- 695.

Driver, R., Leach, J., Millar, R., & Scott. P. (1996). *Young people's images of science*. Philadelphia, PA: Open University Press.

Duschl, R.A., Ellenbogan, K., & Eurduran, S. (1999, March 28-31). *Promoting argumentation in middle school science classrooms: A Project SEPIA evaluation.* Paper presented at the Annual Meetings of the National Association for Research in Science Teaching, Boston, MA.

Duschl, R.A., & Grandy, R. (Eds.). (2007). *Establishing a consensus agenda for K-12 science inquiry*. Rotterdam, NL: Sense.

Duschl, R. (2008). Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. *Review of Research in Education, 32*(1), 268-291.

Ennis, R.H., & Millman, J. (2005). *Cornell Critical Thinking Test: Level X (5th ed.)*. Seaside, CA.

Gelman, R., & Brenneman, K. (2004). Science learning pathways for young children. *Early Childhood Research Quarterly*, *19*, 150–158.

Gunel, M. (2006) *Investigating the impact of teachers' practices of inquiry and non-traditional writing on students' academic achievement of science during longitudinal professional development program.* Unpublished doctoral dissertation, Iowa State University, Ames, IA.

Hand, B. (ed.). (2007). *Science inquiry, argument and language: A case for the science writing heuristic.* Rotterdam, NL: Sense.

Hand, B., Therrien, W., Villanueva, M.G., Taylor, T., Shelley, M., Gonwa-Reeves, C., & Baenziger, J. (2011, September 8-10). *Using an argument-based inquiry approach to learn science: Year 1 results of the science writing heuristic (SWH).* Presented at the Fall 2011 Conference of the Society for Research on Educational Effectiveness, Washington, DC.

Hohenshell, L.M., & Hand, B. (2006). Writing-to-learn strategies in secondary school cell biology: A mixed method study. *International Journal of Science Education*, *28*(2/3), 261-289.

Mertz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction*, *22*, 219–290.

Millar, R. (2006). Twenty first century science: Insights from the design and implementation of a scientific literacy approach in school science. *International Journal of Science Education, 28*(13), 1499-1521. DOI:10.1080/09500690600718344.

National Research Council. (2007a). *Rising above the gathering storm: Employing America for a brighter economic future.* Washington, DC: National Academies Press.

National Research Council. (2007b). *Taking science to school: Learning and teaching science in grades K–8.* Washington, DC: National Academies Press.

National Research Council. (1996). *The national science education standards*. Washington, DC: National Academies Press.

Osborne, J.F., Duschl, R., & Fairbrother, R. (2002). *Breaking the mould: Teaching science for public understanding*. London: Nuffield Foundation.

Poock, J.A., Burke, K.A., Greenbowe, T.J., & Hand, B.M. (2007). Using the science writing heuristic to improve students' academic performance. *Journal of Chemical Education, 84,* 1371–1379.

Raudenbush, S. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206-230.

Shelley, M., Gonwa-Reeves, C., Baenziger, J., Seefeld, A., Hand, B., & Therrien, W. (forthcoming). Modeling outcomes for elementary science education: The science writing heuristic project. *Joint Statistical Meeting 2011 Proceedings, Social Statistics Section* [CD-ROM]. Alexandria, VA: American Statistical Association.

Shelley, M.C., II, Yore, L.D., & Hand, B. (Eds.), (2009). *Quality research in literacy and science education: International perspectives and gold standards.* Dordrecht, NL: Springer.

Spelke, E. (2000). Core knowledge. *American Psychologist*, *55*, 1233–1243.

Rudd, J., Greenbowe, T.J., & Hand, B.M. (2007). Using the science writing heuristic to improve students' understanding of general equilibrium. *Journal of Chemical Education, 84*(12), 2007-2011.

Tippett, C. (2009). Argumentation: The language of science. *Journal of Elementary Science Education*, 21(1), 17-25. DOI: 10.1007/BF03174713.

Villanueva, M.G., Hand, B., Therrien, W., Shelley, M., & Taylor, J. (2011, September 8-10). *Questions, claims, & evidence: Lessons learned from students and teachers engaging in the science writing heuristic (SWH) approach*. Presented at the Fall 2011 Conference of the Society for Research on Educational Effectiveness, Washington, DC.

Yore, L.D., Bisanz, G.L., & Hand, B.M. (2003). Examining the literacy component of science literacy: 25 years of language arts and science research. *International Journal of Science Education, 25,* 689-725.

Appendix B. Tables and Figures

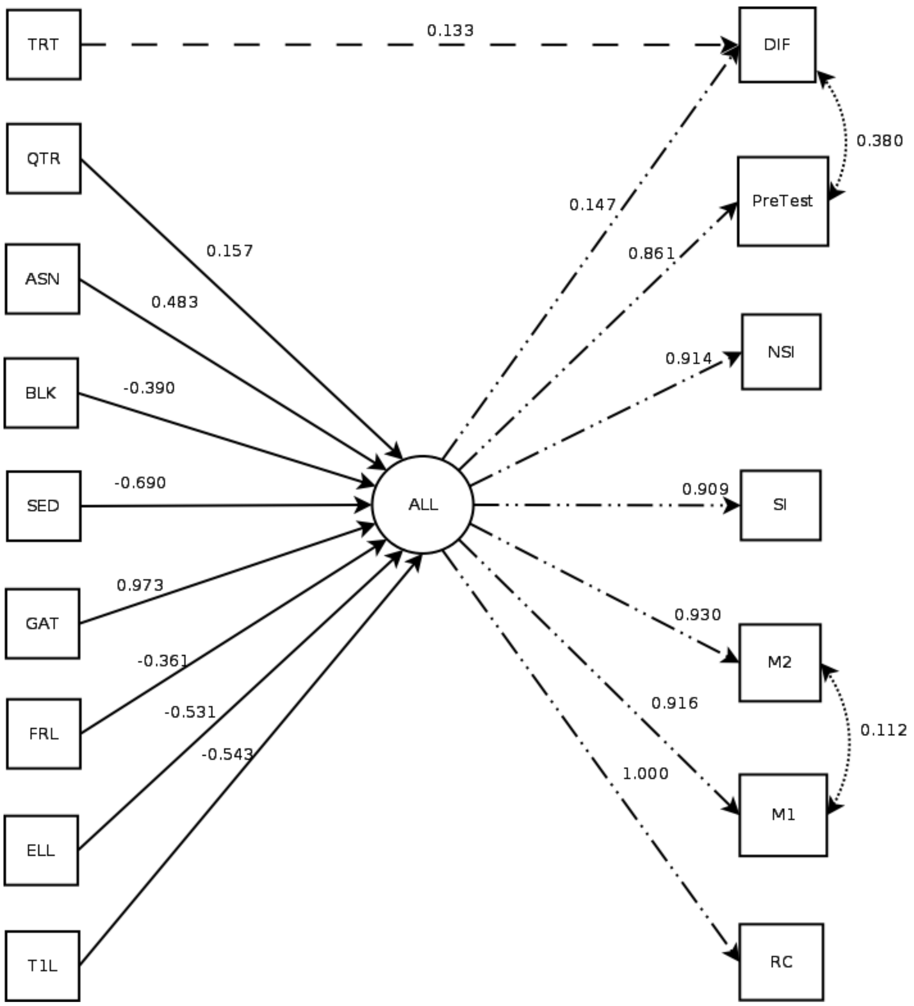

Figure 1. Preliminary Structural Equation Model

Table 1. Summary of Results of Multilevel Model for Pretest to Posttest Changes in Cornell Critical Thinking Test

AIC=12773; BIC=12868; log likelihood=-6370; deviance=12749; REMLdev=12739

Random effects:

Groups	Name	Variance	Std.Dev.
TID	(Intercept)	0.71943	0.8482
Unit1	(Intercept)	2.41478	1.5540
Residual		32.29859	5.6832

Fixed effects:

	Estimate	Std. Error	*t* value
(Intercept)	16.0662	1.0055	15.978
Trt	1.5960	0.5950	2.683
ASN	1.5677	0.9947	1.576
WHT	1.7977	0.6448	2.788
SED	-3.2987	0.7016	-4.701
GAT	3.0565	0.4165	7.339
FRL	-0.8768	0.2924	-2.998
T1M	-2.9743	0.9493	-3.133
T1L	-0.3999	0.9917	-0.403
PreTest	-0.4094	0.0193	-21.211
ASN:T1M	16.9620	5.8986	2.876
SED:T1L	3.0970	1.6634	1.862
Trt:T1L	-2.3192	1.1409	-2.033
Trt:SED	-1.6615	0.8755	-1.898

Analysis of Variance Table

	Df	SumSq	Mean Sq	F value
Trt	1	167.0	167.0	5.1714
ASN	1	32.0	32.0	0.9909
WHT	1	142.7	142.7	4.4184
SED	1	968.5	968.5	29.9864
GAT	1	20.2	20.2	0.6260
FRL	1	27.5	27.5	0.8525
T1M	1	175.5	175.5	5.4341
T1L	1	15.2	15.2	0.4696
PreTest	1	14783.3	14783.3	457.7076
ASN:T1M	1	271.7	271.7	8.4120
SED:T1L	1	149.5	149.5	4.6272
Trt:T1L	1	138.2	138.2	4.2800
Trt:SED	1	116.3	116.3	3.6015