Abstract Title Page

Not included in page count.

Title: Balancing self-directed learning with expert mentoring: The Science Writing Heuristic approach

Authors and Affiliations:

Mack Shelley
Luke Fostvedt
Christopher Gonwa-Reeves
Joan Baenziger
Michael McGill
Ashley Seefeld
Iowa State University

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

Abstract Body

Limit 4 pages single-spaced.

Background / Context:

Description of prior research and its intellectual context

A key element of productive and effective partnerships in science education is establishing and maintaining linkages between teachers and researchers that can eventuate in enhanced student outcomes. Such partnerships between practitioners and the research community are a natural outgrowth of developments during the past two or three decades in science education policy in the United States. 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).

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.

The focus of this study is implementation of the Science Writing Heuristic (SWH) curriculum (Hand, 2007), which combines current understandings of learning as a cognitive and negotiated process with the techniques of argument-based inquiry, critical thinking skills, and writing to strengthen student outcomes. Success of SWH is dependent on the teachers who are implementing the curriculum in the classroom. Central to the SWH philosophy is the emphasis on self-direction. An often mistaken assumption regarding self-direction is that students are doing all of the work. A key to mastery is what is put in front of the student. As important as the emphasis on self-direction is the selection of appropriate materials leading to achievable end-points. The degree of success of the SWH curriculum relies on an appropriate balance of self-direction and expert mentoring, which necessarily is the blend of the students and the work they are doing. SWH requires that the teacher adapts, stops, redirects, responds, and so on. Great problems often are solved by beginners or novices to the field. The best teachers appreciate the beginner's mind and do not get in the way, even when something appears like a false start.

Purpose / Objective / Research Question / Focus of Study:

Description of the focus of the research.

The purpose of this paper is to examine the impact of implementation of the SWH approach at 5th grade in the public school system in Iowa as measured by the Cornell Critical Thinking (CCT) student test (Ennis & Millman, 2005) scores, Reformed Teaching Observation Protocol (RTOP; Adamson, Banks, Burtch, Cox, Judson, Turley, Benford, & Lawson, 2003; Lawson, Benford, Bloom, Carlson, Falconer, Hestenes, Judson, Piburn, Sawada, Turley, & Wyckoff, 2002; Piburn, Sawada, Falconer, Turley, Benford, & Bloom, 2000; Sawada, Piburn, Judson, Turley, Falconer, Benford, & Bloom, 2002) teacher ratings.

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 with the purpose of constructing 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; [a new set of science education standards, the Next Generation Science Standards, currently is under review; see http://www.nsta.org/about/standardsupdate/default.aspx]).

Setting:

Description of the research location.

The study was conducted with Iowa elementary school students, in grades 3-6, with 24 school buildings randomly assigned to treatment and 24 to control. 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.

Population / Participants / Subjects:

Description of the participants in the study: who, how many, key features, or characteristics.

CCT test scores, measured only at 5th grade, were obtained on over 2,000 students in elementary schools throughout the state of Iowa at pre-test and post-test. Videos from 150 teachers were received and rated using the RTOP instrument, of which 37 were for the fifth grade teachers on whom this analysis is based. CCT and video data were obtained from students and teachers in both treatment and control schools following randomization of buildings to treatment/control condition.

Intervention / Program / Practice:

Description of the intervention, program, or practice, including details of administration and duration.

Teachers in school districts randomized to the intervention group were trained in the SWH technique during the summer of 2009 at three-day workshops held at four geographic regions of Iowa. The workshops provided specific training on the SWH approach, including how to foster argumentation skills for students in the classroom. Video recordings were obtained of individual teachers' performance in science classrooms at different times over the academic year. Classroom implementation has continued since Fall 2009, and all 48 selected schools remain in the study.

Research Design:

Description of the research design.

A cluster randomized 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:

Description of the methods for collecting and analyzing data.

The CCT test was administered in a Fall 2010 pretest and a Spring 2011 posttest. Assessment of teacher efficacy was determined using the RTOP rating instrument with scores based on watching videos submitted by teachers. A multilevel model was estimated to assess the relative contributions of individual student (Level 1), Teacher (Level 2), and School (Level 3) variables. We treat student as nested within Teacher and Teacher as nested within School.

Findings / Results:

Description of the main findings with specific details.

The modified version of the RTOP instrument provided ratings on teacher/classroom characteristics for use as predictors in a linear mixed effects model. Using these predictors it is possible to determine the extent to which teacher/classroom characteristics affect critical thinking in either a positive or negative manner. The three levels in the design of the study (Student, Teacher, and School) represent three sources of variation in the data. The model predicting improvement in CCT scores was estimated using R software by a linear mixed model fit using restricted maximum likelihood. The estimated model, using R notation, is:

Improvement ~ Pre-score + Curriculum + Average RTOP Score + White Student + Black Student + Hispanic Student + Asian Student +Special Education Student + Free and Reduced Lunch Status + Gifted and Talented Student + English Language Learner Status + (1| Teacher) + (1|School)

Results show interesting differences when the SWH curriculum (TRTSWH) and RTOP scores are alternated as predictors. Table 1 incorporates TRTSWH, Table 2 incorporates RTOP, and Table 3 incorporates both TRTSWH and RTOP into the model containing all other predictors.

While neither of these predictors is statistically significant in Table 3, when both variables are in the model together, each is statistically significant when included separately. These results indicate that the SWH intervention is effective at increasing student CCT outcomes, a higher level of teacher implementation measured by RTOP enhances CCT outcomes, and SWH professional development efforts are effective at enhancing teacher preparation to provide argument-based classroom inquiry. Figure 1 displays histograms of the RTOP ratings, and Table 4 provides average teacher RTOP ratings by curriculum and semester.

Based on the fixed effects shown in Table 3, the model results indicate an increase of about 1.3 points from pre-test to post-test for students receiving the SWH curriculum compared to students receiving the control curriculum (TRTSWH in Table 3). A one-point increase in the average RTOP rating corresponds to an average increase of about 0.6 points from pre-test to post-test (RTOP in Table 3). The demographic variables Asian (ASN) and White (WHT) had significant positive coefficients, as did Gifted and Talented status (GAT). Special Education status (SED) had a significant negative coefficient. The coefficient for the Pre-score covariate (Prescore) was negative, which reflects the “ceiling” effect whereby the maximum score on the test limits the scope for improvement for students scoring high on the pre-test. The remaining coefficients in the model, Black (BLK) and Hispanic (HSP) students, English Language Learner (ELL) status, and Free and Reduced Lunch (FRL) status, were not statistically significant but were included for completeness as their coefficient estimates may be useful for comparison in future studies.

Conclusions:

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

For a curriculum to be successful, it must be implemented effectively. The SWH curriculum relies on a partnership with teachers to make their own lesson plans. The ideal SWH lesson plan allows students to experience their education with the teacher acting as a resource. These multi-level results indicate that the efficacy of the SWH curriculum is affected by the quality of implementation. The RTOP instrument measures the quality of implementation of the SWH curriculum. Higher RTOP ratings correspond with greater CCT improvements. Teachers who were able to act as a resource and effectively direct student investigations had the greatest increases in critical thinking scores.

The ratings from the RTOP instrument did suffer from low inter-rater reliability, with some raters having rather large discrepancies in the rating behaviors (Table 5 and Figures 2 and 3). This results in adding noise to measurement of the teacher ratings, which in turn increases the standard errors for many of the coefficients and raises the level of Type II error. We are working on improving the consistency of the teacher ratings for future analyses. Table 6 displays the items employed in the RTOP instrument used for this study.

Subsequent analyses will create indices based on aggregate classroom teacher behavior rather than analyzing an average RTOP score. We will also consider the use of statistical imputation methods to improve estimation, due to the fact that videos were not available for all teachers. We will also work to improve the inter-rater reliability of the teacher ratings and consider a way to account for the high rater variability, and thereby reduce Type II error. In addition, after receiving the second year of Iowa Test of Basic Skills data in July 2012 it will be possible to evaluate the effect of the SWH curriculum and teacher characteristics on student performance in multiple subject content areas.

Appendices

Not included in page count.

Appendix A. References

References are to be in APA version 6 format.

Adamson, A.E., Banks, D., Burtch, M., Cox III, F., Judson, E., Turley, J.B., Benford, R. & Lawson, A.E. (2003). Reformed undergraduate instruction and its subsequent impact on secondary school teaching practice and student achievement. *Journal of Research in Science Teaching, 40*(10), 939-958.

Cohen, D.K., Raudenbush, S.W., & Ball, D.B. (2003). Resources, instruction, and research. *Educational Evaluation and Policy Analysis*, *25*(2), 119-142.

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

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

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.

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

Lawson, A.E., Benford, R., Bloom, I., Carlson, M.P., Falconer, K.F., Hestenes, D.O., Judson, E., Piburn, M.D., Sawada, D., Turley, J., & Wyckoff, S. (2002). Evaluating college science and mathematics instruction: A reform effort that improves teaching skills. *Journal of College Science Teaching, 31*(6), 388-393.

Mertz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction*, *22*(2), 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.

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.

Piburn, M., Sawada, D., Falconer, K., Turley, J., Benford, R., & Bloom, I. (2000). Reformed Teaching Observation Protocol (RTOP). ACEPT IN-003. The RTOP rubric form, training manual and reference manual containing statistical analyses, are all available from <http://PhysicsEd.BuffaloState.Edu/AZTEC/rtop/RTOP_full/PDF/>. IN-001

contains the RTOP rubric alone, IN-002 contains rubric and training manual, IN-003 adds the statistical analyses.
Raudenbush, S. (2008). Advancing educational policy by advancing research on instruction. *American Educational Research Journal, 45*(1), 206-230.
Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. *School Science and Mathematics, 102*(6), 245-253.
Spelke, E. (2000). Core knowledge. *American Psychologist*, *55*(11), 1233–1243.

Appendix B. Tables and Figures

Not included in page count.

Table 1: Model with Curriculum Effect and no Teacher Effect

AIC BIC logLik deviance REMLdev
6908 6978 -3440 6891 6880

Random Effects:

Group	Variance	Standard Deviation
School	1.8023	1.3425
Teacher	1.3434	1.1590
Residual	34.9701	5.9135

Number of obs: 1073, groups: Teacher 37; School, 30

Fixed Effects				90% HPD Interval	
Parameter	**Estimate**	**Standard Error**	**t-value**	**Lower**	**Upper**
(Intercept)	16.94949	2.21497	7.652	13.1537375	20.5130786
Prescore	-0.43827	0.02676	-16.380	-0.4806945	-0.3919577
TRTSWH	1.66173	0.76407	2.175	0.3386640	2.9097182
WHT	3.22114	1.86227	1.730	0.1554076	6.3063951
SED	-4.35975	0.63312	-6.886	-5.4034332	-3.3290900
ASN	6.45804	2.61801	2.467	2.1243187	10.7347785
BLK	0.07080	1.51721	0.047	-2.4494839	2.5429463
HSP	-0.31966	1.20631	-0.265	-2.2981637	1.7078957
GAT	3.03438	0.58525	5.185	2.0612661	4.0004284
FRL	-0.42632	0.40520	-1.052	-1.1104794	0.2258389
ELL	-1.37421	1.59929	-0.859	-3.9771432	1.2671998

Table 2: Model with Teacher Effect and no Curriculum Effect

AIC BIC logLik deviance REMLdev
6910 6980 -3441 6892 6882

Random effects:

Group	Variance	Standard Deviation
School	1.8481	1.3595
Teacher	1.5236	1.2343
Residual	34.9526	5.9121

Number of obs: 1073, groups: Teacher, 37; School, 30

Fixed Effects				90% HPD Interval	
Parameter	**Estimate**	**Standard Error**	**t-value**	**Lower**	**Upper**
(Intercept)	16.14300	2.37548	6.796	12.23795409	20.0679278
Prescore	-0.44234	0.02678	-16.520	-0.48582614	-0.3976667
RTOP	1.03314	0.56898	1.816	0.06103774	1.9241450
WHT	3.28918	1.86277	1.766	0.24514193	6.4073286
SED	-4.40387	0.63313	-6.956	-5.42527539	-3.3397906
ASN	6.42259	2.61829	2.453	2.07992485	10.6956751
BLK	0.11176	1.51707	0.074	-2.42003196	2.5916580
HSP	-0.38356	1.20667	-0.318	-2.38422951	1.5702153
GAT	3.05681	0.58580	5.218	2.09545720	4.0310137
FRL	-0.43491	0.40527	-1.073	-1.10704651	0.2233088
ELL	-1.39641	1.60058	-0.872	-3.99996283	1.2199184

Table 3: Model with Both a Teacher Effect and a Curriculum Effect

AIC	BIC	logLik	deviance	REMLdev
6908	6983	-3439	6890	6878

Random Effects

Group	Variance	Standard Deviation
School	1.9691	1.4032
Teacher	1.2414	1.1187
Residual	34.9596	5.9127

Number of obs: 1073, groups: Teacher, 37; School, 30

Fixed Effects				90% HPD Interval	
Parameter	Estimate	Standard Error	t-value	Lower	Upper
(Intercept)	16.09443	2.36517	6.805	12.2941198	20.1248351
Prescore	-0.44002	0.02680	-16.416	-0.4817935	-0.3938022
TRTSWH	1.30284	0.86025	1.514	-0.1826180	2.6329769
RTOP	0.63669	0.61988	1.027	-0.3860891	1.6332863
WHT	3.26405	1.86264	1.752	0.1181678	6.2710767
SED	-4.37506	0.63320	-6.909	-5.4222702	-3.3455828
ASN	6.41933	2.61788	2.452	2.0033165	10.6140525
BLK	0.06059	1.51709	0.040	-2.4616032	2.5664872
HSP	-0.33411	1.20626	-0.277	-2.3255556	1.6381444
GAT	3.05378	0.58555	5.215	2.0966869	4.0225680
FRL	-0.42529	0.40519	-1.050	-1.0956714	0.2316348
ELL	-1.40144	1.59977	-0.876	-4.0331012	1.2356220

Figure 1: Histograms of Reformed Teaching Observation Protocol (RTOP) ratings for teachers who submitted a video of themselves teaching during the fall and/or spring semester. Of the 150 teachers who submitted videos, 50 submitted one for each semester.

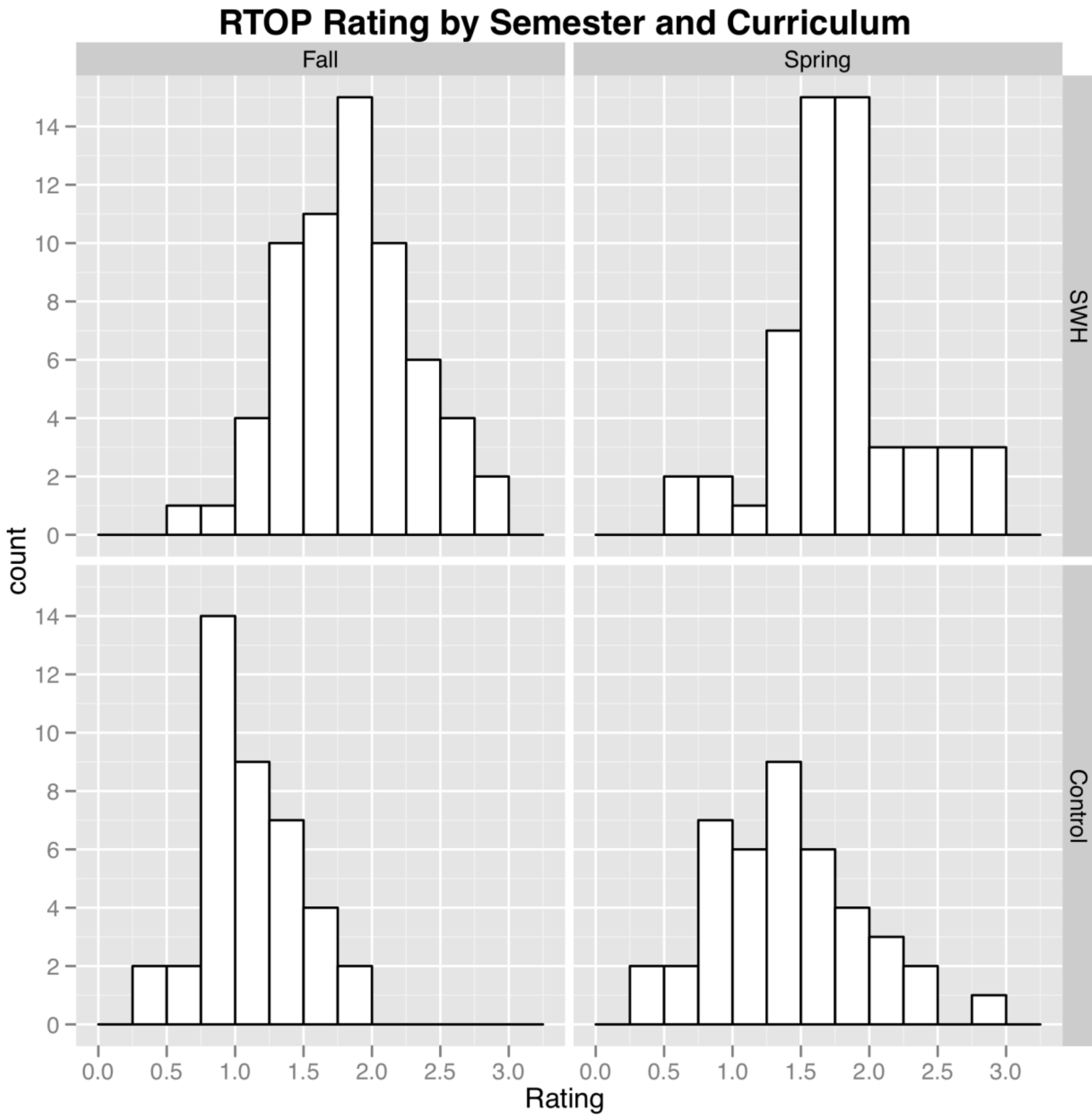

Table 4: Average teacher RTOP ratings by curriculum and semester.

Curriculum	Semester	Average Rating	Sample Size
Control	Fall	1.114435	40
Control	Spring	1.400794	42
Science Writing Heuristic	Fall	1.855655	64
Science Writing Heuristic	Spring	1.787006	54

Table 5: Inter-rater Reliability (IRR; Krippendorff's Alpha) and Intraclass Correlation (ICC) of the RTOP scores.

Question	IRR	ICC
Q1	0.035	0.242
Q2	0.447	0.082
Q3	0.244	0.154
Q4	0.395	0.076
Q5	0.419	0.074
Q6	0.239	0.050
Q7	0.271	0.212
Q8	0.110	0.185
Q9	0.078	0.320
Q10	-0.022	0.494
Q11	0.182	0.159
Q12	0.144	0.323
Q13	0.196	0.161
Q14	0.169	0.330
Average	0.347	0.183

Figure 2: Using a linear mixed effects model with rater as a random component, a term is added to account for the bias specific to each rater. The estimates of the coefficients for the rater effects are plotted, illustrating differences among raters.

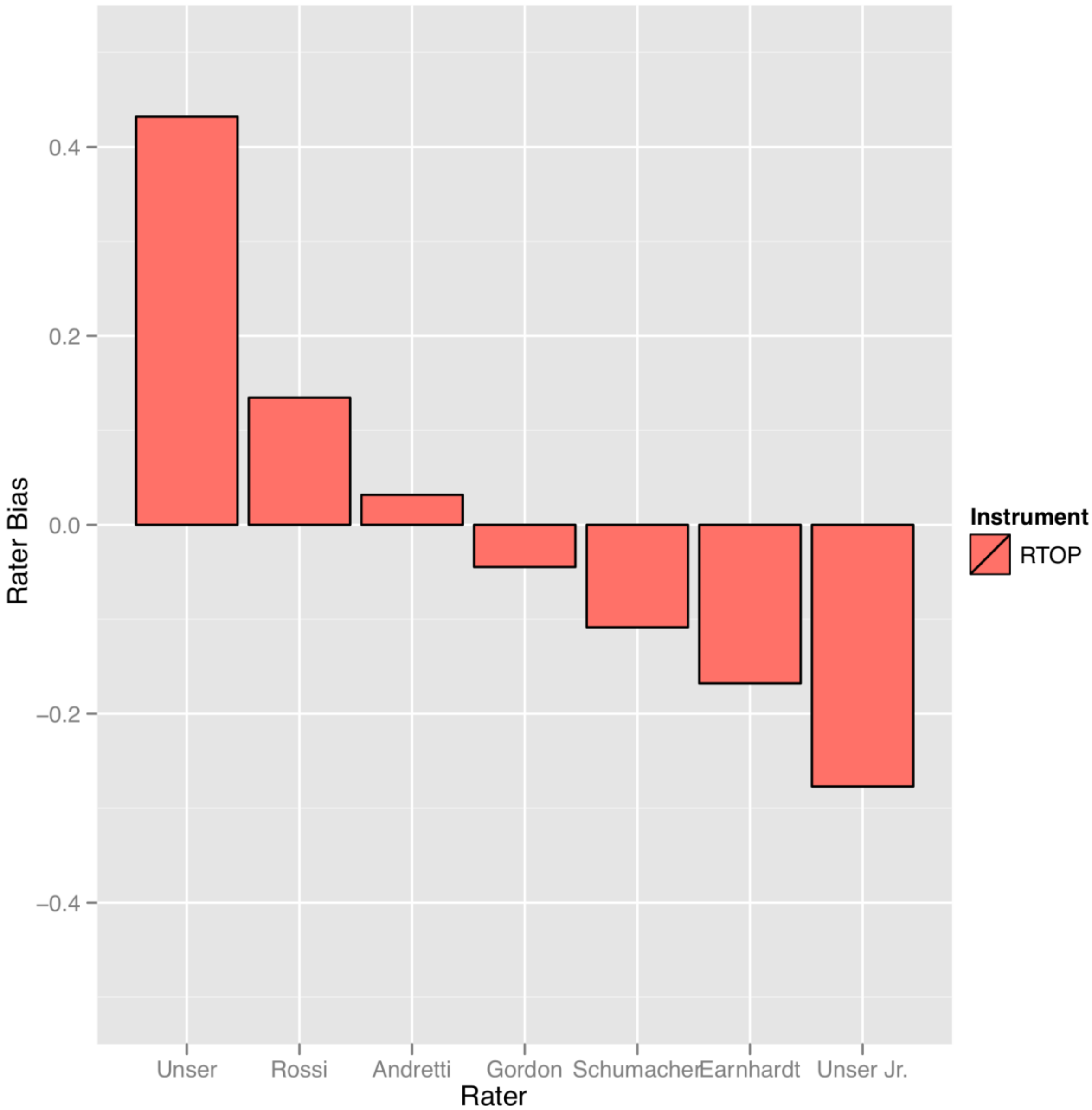

Figure 3: Average ratings assigned from each rater for each of the 10 questions in the RTOP instrument. Of the 150 teachers rated, 58 were rated multiple times

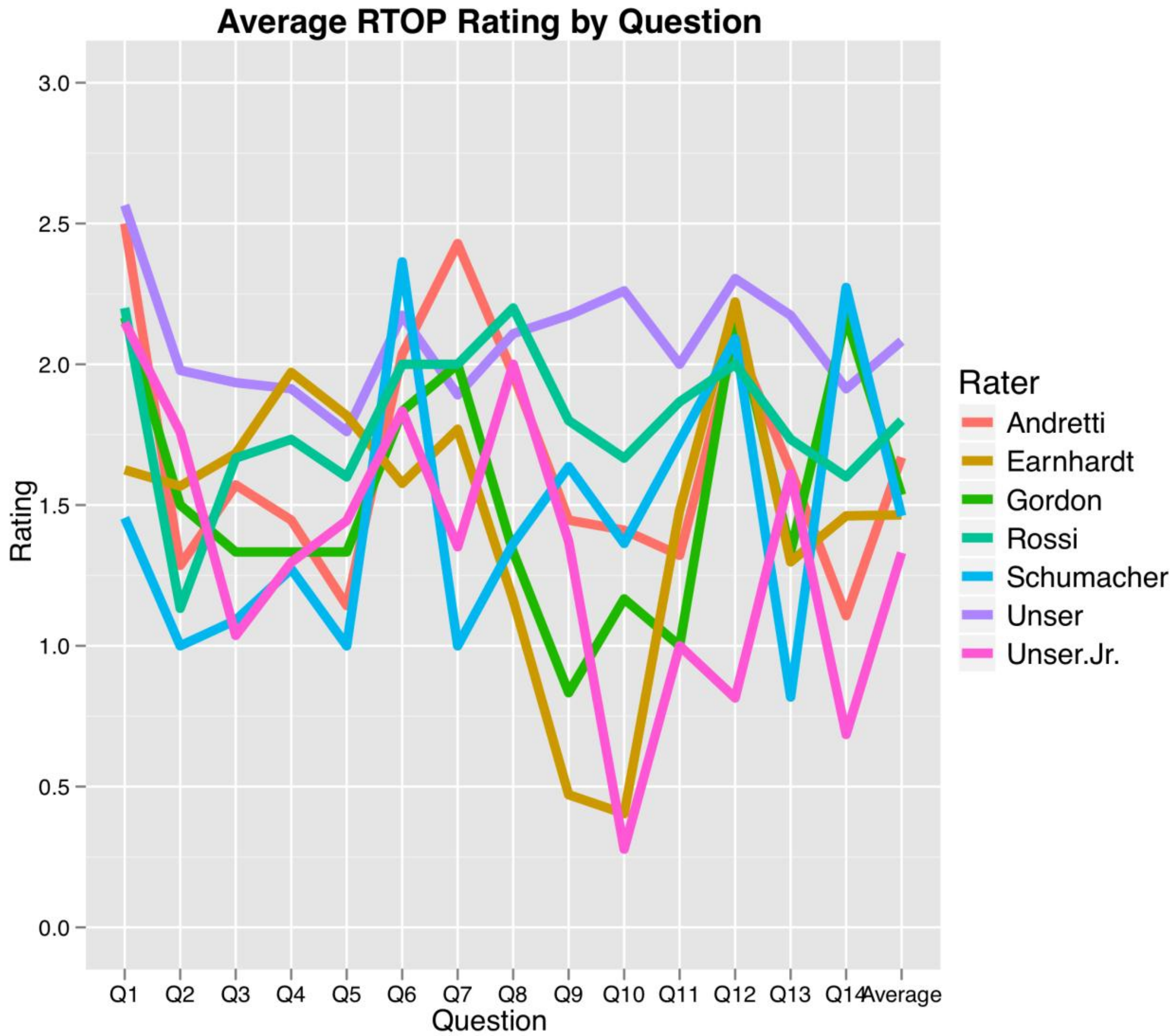

Table 6: Selected Questions from the RTOP instrument used to evaluate teachers in the study.

	Modified RTOP Instrument	
RTOP # (original)	**Descriptor**	**Score 0-4**
	Lesson Design and Implementation	
1 (1)	The instructional strategies and activities respected students' prior knowledge and the preconceived notions inherent therein.	
2 (4)	This lesson encouraged students to seek and value alternative modes of investigation or of problem solving.	
3 (5)	The focus and direction of the lesson was often determined by ideas originating with students.	
	Procedural Knowledge	
4 (13)	Students were actively engaged in thought provoking activity that often involved the critical assessment of procedures.	
5 (14)	Students were reflective about their learning.	
6 (15)	Intellectual rigor, constructive criticism and the challenging of ideas were valued.	
	Communicative Interactions	
7 (16)	Students were involved in the communication of their ideas to others using a variety of means and media.	
8 (17)	The teacher's questions triggered divergent modes of thinking.	
9 (18)	There was a high proportion of student talk and a significant amount occurred between and among students.	
10 (19)	Student questions and comments often determined the focus and direction of classroom discourse.	
	Student/Teacher Relationships	
11 (21)	Active participation by students was encouraged and valued.	
12 (22)	Students were encouraged to generate conjectures, alternative solution strategies, and ways of interpreting evidence.	
13 (24)	The teacher acted as a resource person, working to support and enhance student investigations.	
14 (25)	The metaphor "teacher as listener" was very characteristic of this classroom.	