COORDINATING THEORY AND EVIDENCE:  
THE BENEFITS OF A MODEL-BASED SCIENCE CURRICULUM

Corinne Zimmerman  
Kalyani V. Raghavan  
Learning Research and Development Center  
University of Pittsburgh  
Pittsburgh PA 15260

Paper presented at the American Educational Research Association Annual Meeting  
New Orleans, LA  
April 2002

Abstract
The Model-Assisted Reasoning in Science (MARS) project seeks to promote model-centered instruction as a means of improving middle-school science education. As part of the evaluation of the sixth-grade curriculum, performance of MARS and non-MARS students was compared on a curriculum-neutral task. Fourteen students participated in structured interviews in which they experimented with a balance to identify and test potentially relevant factors. The apparatus provided three manipulable variables (two affected balance, one was a non-causal distractor variable). Although both groups were equally able to identify and test variables, all MARS students discovered a quantitative rule to describe the operation of the balance, whereas only one non-MARS student did so. MARS students discovered this numerical relationship through experimentation, regardless of their scientific reasoning profile (i.e. experimenters or theorists). Critical components of instruction that may foster scientific reasoning, in particular the ability to flexibly coordinate theory and evidence, include multiple opportunities to draw conclusions from data and an emphasis on the successive refinement of models.

Acknowledgments
This research was supported by a grant from the Office of Educational Research & Improvement (OERI), U.S. Department of Education. Address correspondence to Kalyani Raghavan (kalyani@pitt.edu).
COORDINATING THEORY AND EVIDENCE: 
THE BENEFITS OF A MODEL-BASED SCIENCE CURRICULUM

Objectives and Theoretical Perspective

Since 1991, the Model-Assisted Reasoning in Science (MARS) project team has been investigating and promoting the use of curricula that emphasize model-centered instruction as a means of improving middle-school science education (e.g., Raghavan & Glaser, 1995). In MARS instruction, students learn to use models the way scientists use them, as communication and reasoning tools that can help them describe physical-world phenomena, depict and test ideas about underlying causes, and identify and explore relationships between ideas.

Topics covered in the sixth-grade curriculum include states and properties of matter, area, volume, mass, density, interaction properties, force, and net force. These topics correspond to concepts recommended for middle school students in Benchmarks and National Science Education Standards (AAAS, 1993; NRC, 1996, 2000). They were selected for the MARS curriculum because they are well suited to the design of a coherent instructional system that supports the development of model-assisted reasoning. Researchers have identified the intuitive ideas individuals have about many of these topics, enabling curriculum developers to challenge misconceptions and build on preconceptions, and to create instruction that is intelligible, relevant, and leads to an appreciation of scientific theory building as an incremental and corrective process (Camp & Clement, 1994; de Vos & Verdonk, 1996; Driver et al., 1985; Duit et al., 1992; Jones et al., 2000; NRC, 1999; Solomon, 1992).

In addition to being based in science standards and the misconceptions literature, curricular development has been empirically grounded in ongoing research documenting students’ prior ideas and the impact of MARS instruction on their emerging conceptual understanding of scientific phenomena (Raghavan, Sartoris, & Glaser, 1998a, 1998b). Whereas previous research focused on students’ developing concepts, in the present study we explore students’ performance on a task that allows us to examine scientific reasoning as it occurs. The ability to coordinate theory and evidence is critical to the development of scientific reasoning (e.g., Kuhn, 1989). The balance scale is an ideal task in which to explore students’ different approaches to coordinating theory and evidence. Like many scientific
phenomena, balance concepts often are based on prior conceptions and can be resistant to change. Therefore, students come to the task with correct and incorrect ideas about the concept of balance.

Two objectives guided the design of the present study. First, we were interested in documenting the co-development of scientific reasoning skills and the acquisition of conceptual knowledge on a performance task involving the evaluation of experimental evidence. Second, we were interested in comparing the performance and problem-solving approaches of students receiving MARS instruction to that of students in a nationally-recognized, standards-based curriculum. Students were interviewed individually and verbal protocols were used to assess specific skills (e.g., using controlled tests) and to explore whether there were differences among students in their problem-solving approaches. Individual differences in scientific inquiry have been documented for children and adults (e.g., Dunbar & Klahr, 1989; Schauble et al., 1991; Stanovich, 1999). These studies were used as a guide for analyzing student protocols. The interviews provided an opportunity to observe students’ thinking as it occurred, and to explore students’ different approaches to coordinating experimentation, theory, and evidence. Differences between MARS and non-MARS students might shed light on whether, and how, model-centered instruction has an impact on student thinking.

Participants, Materials, and Methods

Fourteen sixth-grade students from the same school were interviewed by a researcher uninvolved with the project. Students were selected by their teachers to provide diversity in terms of race, gender, and scholastic ability. Seven students were instructed in the MARS curriculum and seven were instructed in a nationally-recognized science curriculum.

Students participated in a curriculum-neutral performance task (the balance-scale task), an adaptation of a teacher-training task used by McDermott (1996). Three variables could be manipulated on the balance apparatus: the number of weights (washers) and their horizontal and vertical placements (i.e., three rows of six hooks on each side of the fulcrum). Of the three variables, two had an effect on the balance and one did not. Vertical position or row (i.e., the location of the washers up or down on three rows of hooks) is a distractor and does not have an impact on the balance beam. Both children and adults have naive conceptions that lower weights exert more pull or force (Aoki, 1991).
Students were asked to (a) identify and test the three relevant variables (prompts ensured that students noticed all three), (b) explore multiple ways to balance the scale when the weights on one side were fixed, (c) generate a “general rule” to explain how the scale works (free-form experimentation was encouraged, if needed), and (d) apply their rule to four test situations depicted on paper (some with multiple weights distributed across multiple pegs).

**Results and Discussion**

There were no differences between the MARS and non-MARS students with respect to their ability to *identify* the three critical variables, either before or after prompting (see Table 1). After ensuring that all students were aware of the three focal variables, the interviewer asked the students to test each variable to determine if it made a difference in whether the beam tips or balances. The MARS students, on average, used controlled tests for 2.29 (sd = 1.11) of three tests and non-MARS students used controlled tests for 1.71 (sd = .95). Although the MARS students were more likely to used controlled tests, the difference was not statistically reliable ($p = 0.32$).

**Table 1  Comparing MARS and Non-MARS Students on the Balance Task**

<table>
<thead>
<tr>
<th>Interview Component</th>
<th>Mean Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MARS</td>
</tr>
<tr>
<td>Variable identification</td>
<td>81.0</td>
</tr>
<tr>
<td>Controlled tests of focal variables</td>
<td>76.3</td>
</tr>
<tr>
<td>Successful solutions (FQ1)</td>
<td>71.2</td>
</tr>
<tr>
<td>Successful solutions (FQ2)</td>
<td>89.5</td>
</tr>
<tr>
<td>Application questions</td>
<td>78.6</td>
</tr>
</tbody>
</table>

*Note. FQ = focus questions. See text for further description of the Interview Components.

*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$
Two “focus questions” involved students generating as many ways as possible to make the scale balance when the weights on one side were fixed. The number of successful solutions students generated was expressed as a percentage of the total number of attempts. The MARS students outperformed the non-MARS students on both focus questions (see Table 1). The mean percentages for MARS and non-MARS students were 71.2 and 39.5, respectively, for the first focus question ($t(12) = 3.19, p < .01$). For the second focus question, mean percentages were 89.5 and 45.2 for MARS and non-MARS, respectively ($t(12) = 4.89, p < .001$).

After actively exploring the apparatus during the focus questions, students were asked to come up with a “general rule” to describe the behavior of the balance beam. They were given the option of conducting additional experiments before answering. All seven MARS students were able to express a quantitative rule. Quantitative rules made reference to information about the number of washers and the number of the hook. For example, one student’s rule was: “You have to have an equal amount of weight on each side, like if you have 2 weights on the 3, that makes 6, so you have to have 6 on the other side.” Additionally, during the application questions, students (either verbally or on paper) multiplied the number of washers (mass) by the position number (distance) on each side and compared the products (i.e., the mathematically correct solution).

The rules generated by six of the non-MARS students were qualitative in nature. For example: “Always try to have the same amount of washers on one side, that can make it even . . . And sometimes you can use one more washer. I keep my extra washer in the inside just to keep it balanced, so try to keep the extra washer near the inside [i.e., close to the fulcrum].” Only one non-MARS students generated the correct quantitative rule.

The interview ended with four paper-and-pencil questions. Students were shown drawings of an equal-arm balance with upright, equidistant pegs and blocks to represent weights. Students were asked to predict whether the scale would tip to the left, tip to the right, or balance. The MARS students scored 78.6% on the application questions, compared to 39.3% for the non-MARS students ($t(12) = 2.14, p = 0.05$). The success of the MARS students on the application questions was due, at least in part, to their ability to apply the quantitative rule they had generated to describe the behavior of the balance scale.
Verbal protocols were further analyzed to ascertain whether students from the different curricula were using different approaches to experimentation (see Table 2). Individual differences in approaching scientific reasoning tasks have been noted by other researchers (e.g., Dunbar & Klahr, 1989). Most notably, there is a distinction between “theorists” and “experimenters.” Klahr and Dunbar (1988) described theorists as individuals who tend to generate hypotheses and then test the predictions of the hypotheses. That is, they have a theory in mind and then collect data or observations to test the theory. Experimenters, in contrast, tend to make data-driven discoveries by generating data and finding the hypothesis that best summarizes or explains that data.

Table 2

*Individual Difference Profiles and Number of Students Fitting Each Profile for Each Curriculum*

<table>
<thead>
<tr>
<th>Profile</th>
<th>Approach to Experimentation</th>
<th>Evidence Evaluation</th>
<th>Theory/Evidence Relationship</th>
<th>MARS (n = 7)</th>
<th>Non-MARS (n=7)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Theorists</td>
<td>Mixed (controlled &amp; try-and-see)</td>
<td>Mixed (correct when theory matches)</td>
<td>Theory revised based on evidence</td>
<td>3*</td>
<td>6</td>
</tr>
<tr>
<td>Experimenters</td>
<td>Controlled tests</td>
<td>Correct</td>
<td>Rule derived from evidence</td>
<td>4*</td>
<td>1*</td>
</tr>
</tbody>
</table>

* Students generating the correct quantitative rule.

Students classified as theorists spontaneously and explicitly stated their expectations about the role of the different variables. Most students had correct intuitions about the role of weight, but held correct and incorrect conceptions about the horizontal and vertical placement of the washers. The approach to experimentation was mixed – students used both controlled and “try-and-see” tests. As
such, the manner in which they evaluated evidence depended on whether or not the test was confounded, and whether they realized the confound. When the results of a controlled test were correctly assessed, the student’s theory about the role of the variable was updated to match the pattern of evidence. For example, a student who believed that the horizontal dimension did not make a difference, but found evidence that it did, would refine his or her theory. Six of the seven non-MARS students, and three of seven MARS students matched this profile.

Experimenters did not spontaneously verbalize expectations about the effect of variables during the variable testing phase of the interview. They usually experimented by using controlled tests to determine the effect of each variable. Moreover, they were systematic in exploring solutions during the two focus questions requiring asymmetric balance configurations. All of the students fitting this profile (four MARS and one non-MARS) discovered or induced the correct quantitative rule from the generation and evaluation of evidence. These students approached the task without stated preconceived notions but rather they generated the theory (i.e., the rule for balance) based on the pattern of data they obtained during their experimentation.

**Educational/Scientific Importance**

As science curricula increasingly are becoming empirically grounded in educational and psychological research and anchored to the educational reform promoted by national standards, it is critical to evaluate their effectiveness. It is also important to document the ways in which such curricula have an impact on student thinking.

In the current study, we explored whether or not students in an alternate, model-based curriculum performed as well as those in a nationally-recognized curriculum on a scientific reasoning task. Both groups did well on a curriculum-neutral performance task, displaying basic inquiry skills such as identifying and testing variables. There were, however, ways in which the MARS students excelled: (a) All MARS students, regardless of approach to inquiry (i.e., classification as experimenters or theorists), were able to discover the correct quantitative rule for determining balance; (b) MARS
students were more likely to take an experimental approach, using the patterns in the data to induce a mathematical rule rather than approaching the task with a naive theory; (c) MARS students were more successful in applying what they learned with the physical apparatus to solve related paper-and-pencil problem situations.

To make sense of these data, we reflected on the critical differences in the types of instruction in the two science curricula. Both groups were involved in extended investigations of scientific phenomena. Two differences were apparent: (a) MARS students had several opportunities to collect data and draw conclusions about the same topic, whereas the non-MARS students would collect data over longer periods of time with limited opportunities to draw conclusions; (b) the multiple opportunities that MARS students had to draw conclusions required successive refinement of a model. That is, model-revision is an integral part of the learning sequence and instruction necessitates building upon previous knowledge. Moreover, the emphasis on model revision, rather than just model generation, may be the crucial component of instruction that allows students to develop skills to flexibly coordinate theory and evidence.
References


Jones, M. G., Carter, G., & Rua, M. J. (2000). Exploring the development of conceptual ecologies:
Communities of concepts related to convection and heat. *Journal of Research in Science Teaching, 37*(2), 139-159.


