

Usable Assessments Aligned with Curriculum Materials:
Measuring Explanation as a Scientific Way of Knowing

David J. Lizotte, Christopher J. Harris, Katherine L. McNeill,
Ronald W. Marx, & Joseph Krajcik
University of Michigan

Paper presented at the annual meeting of the American Educational Research Association, April, 2003, Chicago, IL. Electronic version available at: www.hi-ce.org/aera2003

The research reported here was supported in part by the National Science Foundation (REC 0101780 and 0830310A605). Any opinions expressed in this work are those of the authors and do not necessarily represent either those of the funding agency or the University of Michigan.

Abstract

We present assessment rubrics developed through an assessment-driven design process for creating middle-school chemistry curriculum materials aimed at learning goals that intersect content and ways of knowing. Specifically, we focus on explanation as a scientific way of knowing. Student explanations in two successive enactments of our chemistry curriculum were analyzed using explanation rubrics aligned with key learning goals emphasized in the instructional materials. Data from the first enactment revealed specific student competencies and difficulties with components of explanation. These data were immediately usable as feedback for revising the instructional materials to better align with assessments and learning goals. In the second enactment, following curriculum revisions, students' explanations demonstrated greater pre-posttest gains relative to the first enactment. Reasoning remained the most challenging component of explanation for students in the second enactment, however, their reasoning statements were more advanced for certain types of science content. We suggest that students applied a developing explanation schema to assessment tasks that was consistent with our explanation rubric, but the results of which depended on content. More generally, we discuss advantages of using rubrics for multiple ways of knowing to align assessments with instructional materials and learning goals.

Usable Assessments Aligned with Curriculum Materials:
Measuring Explanation as a Scientific Way of Knowing

Classroom assessment can be used to enhance learning when assessments, curriculum materials, and instruction align to reinforce common learning goals (Pellegrino, Chudowsky, & Glaser, 2001). Whereas large-scale assessments are removed from classroom learning (Shepard, 2000), assessments that closely align with an enacted curriculum may be more immediately usable: students can use feedback from formative assessments for subsequent learning (Shepard, 2000), and teachers and researchers can obtain data that is sensitive to student performance attributable to a curriculum (Ruiz-Primo, Shavelson, Hamilton, & Klein, 2002) for evaluating materials and instruction.

We are currently developing usable assessments for an inquiry-based, middle-school chemistry curriculum aimed at specific national standards. Our assessment-driven design process is diagrammed in Figure 1. We began by identifying key middle-school content standards for chemistry (AAAS, 1993; NRC, 1996) and multiple ways of knowing the content standards that require diverse cognition from students. These ways of knowing encompass a range of cognitive indices of understanding advocated by the revised Bloom's Taxonomy (Anderson & Krathwohl, 2001) and others (Perkins, 1992; White & Gunstone, 1992), and scientific inquiry abilities emphasized in national standards (NRC, 1996, 2000). They include: identifying and describing phenomena, constructing explanations, designing and conducting experiments, analyzing data, and constructing and evaluating models.

Our assessments derive from these multiple ways of knowing science to support our goal of creating sensitive measures of student performance. Each way of knowing corresponds to a *base* assessment rubric (Harris, McNeill, Lizotte, Marx, & Krajcik, in press).

Base rubrics

A base rubric articulates the different components of a way of knowing and the scoring levels of those components. Base rubrics can be adapted to any science content at the middle-school level.

In this paper, we concentrate on the base rubric for explanation (Figure 2). We focus on explanation because national inquiry standards (NRC, 1996, 2000) and recent science education literature (Bell & Linn, 2000; Driver, Newton, & Osborne, 2000; Kuhn, 1993; Sandoval, 2003) treat explanation as an essential factor in student understanding of science content. Although

definitions of explanation vary considerably, congruent with many of these science educators we used an adapted version of Toulmin's (1958) model of argumentation for our explanation rubric. As such, the base rubric in Figure 2 entails three components: a claim (an assertion or conclusion for a problem), evidence (data that supports the claim – similar to Toulmin's grounds), and reasoning (an argument that justifies why the evidence supports the claim – a blending of Toulmin's warrants and backing).

Specific rubrics

Applying a way of knowing to a science content standard yields a “learning performance” (McNeill, Lizotte, Harris, Scott, Krajcik, & Marx, 2003; cf. Perkins, 1992) for the content standard. Each learning performance states succinctly the way a student will demonstrate “knowing” the standard.

A *specific* rubric (Harris et al., in press) articulates criteria for assessing students' completion of a learning performance. The specific rubric is an adapted version of the base rubric; it includes the same components as the base rubric but tailors them to the given learning performance. Because specific rubrics directly align with learning performances, they are connected to specific science content.

For example, we began the design of our chemistry unit with a national content standard about chemical reactions (*Science for All Americans*, AAAS, 1990, p.47). Applying explanation, one scientific way of knowing, to the content standard gave the following learning performance: ‘When various substances come in contact with each other, students identify whether a chemical reaction occurred [claim], provide evidence [evidence], and explain [reasoning].’ Using the base rubric for explanation as a guideline, we constructed a specific explanation rubric for the learning performance, shown in Figure 3.

Rubrics in practice

To summarize, an educator might start with key content goals (e.g. national standards) and multiple ways of knowing that content. Applying the different ways of knowing to each content standard would yield a set of learning performances. To assess students' completion of any given learning performance, the educator would use the base rubric for the appropriate way of knowing to construct a specific rubric for the learning performance.

Then the educator can use the learning performances and specific rubrics to structure learning and assessment tasks in the unit. The criteria of the specific rubric can be customized

for a specific learning or assessment task, facilitating alignment of assessments and instructional materials. For example, specific explanation rubrics for tasks concerning different chemical reactions (e.g. performing electrolysis of water and making soap) would involve different criteria for accurate and sufficient evidence that a chemical reaction occurred. Evaluation of student performances is consistent within the unit using the criteria of specific rubrics. Furthermore, a set of base rubrics encourages consistent evaluation of different ways of knowing across curriculum units, science content (e.g. chemistry, physics, biology), and middle-school grade levels.

Here, we report how we used explanation rubrics to assess students' progress relative to important explanation learning performances for our chemistry unit. We consider the development of our assessment strategy through two successive enactments of the unit.

Enactment 1

In the first enactment of the unit, students completed tasks in which they used evidence about substances to explain whether a chemical reaction occurred. We analyzed their explanations on pre- and posttest assessments using the specific explanation rubric for chemical reactions in Figure 3.

Methods

Participants. One teacher enacted the unit with three of her classes in a public middle school in Chicago. The student group was ethnically diverse and from lower-middle- to middle-income families. While the three classes consisted of a total of 88 students, only data from the 77 students who completed both pre- and posttest assessments were used in the present analyses.

Curriculum materials. The 4-week chemistry unit was inquiry rich and focused on a driving question (Krajcik, Berger, & Czerniak, 2002) that targeted the concepts of substance, property, and chemical reaction. Complete details of the unit are available in McNeill et al. (2003).

Assessment tasks. Students wrote explanations on identical pre- and posttest tasks, and during embedded unit tasks. A typical task provided data (or students collected data) from an experiment in which different substances were mixed together. The task required that students explain whether a new substance formed (alternatively, whether a chemical reaction occurred) when the substances were mixed. A series of questions tapped the components of explanation. For example:

1. Do you think a new substance formed after mixing the fat, rubbing alcohol, and sodium hydroxide? [Claim]
2. Provide 3 pieces of evidence to support your answer for question #1. [Evidence]
3. Explain why the evidence supports your answer. [Reasoning]

Scoring. Pairs of independent raters scored the explanations using the specific explanation rubric for chemical reactions (Figure 3); a third rater resolved disagreements. Average inter-rater reliability was greater than 90%. Each component of explanation (claim, evidence, and reasoning) received a score; additionally, component scores were weighted equally to compute a total score for the explanation.

Results and Discussion

We analyzed students' explanations for chemical reactions on the pre- and posttest tasks. Table 1 displays pre- and posttest mean scores, results of paired *t*-tests on the mean difference scores, and effect sizes.

Students' explanation total scores improved significantly ($p < .001$) from pre- to posttest, with a moderate effect size. Furthermore, each component of their explanations improved significantly ($ps < .05$). The overall improvement in explanation (i.e. total score) is largely attributable to a gain in the evidence component, which had an effect size comparable to that of the total score. Yet the mean posttest evidence score ($M = 1.21$) was somewhat low on an absolute scale (max = 2.50 for each component). More troubling was the low mean posttest reasoning score ($M = 0.29$). Most students provided accurate claims at pretest; consequently the effect size of the gain in that component was low.

These data suggest that a detailed analysis of student *performances* can uncover rich information about student "knowing." Our assessment tasks required students to use science content to explain new phenomena. Consequently, explanation scores for the open-ended tasks should measure both recall and application of science content, as opposed to multiple choice items concerning the same content that would measure recall exclusively.

Moreover, carving up explanation into its components revealed intricacies of students' explanations. The component scores showed that students' claims were strong at the start of the unit, their evidence improved the most but can stand to improve more, and their reasoning did not develop well. These competencies and difficulties with explanation are concealed by the

explanation total score, yet they are essential to the immediate usability of the assessments, in other words, as feedback for revising curriculum materials and instruction.

Curriculum Revision

Hindsight tells us that it is not surprising that students had difficulty with the evidence and especially reasoning components of explanation. Our assessment criteria (i.e. the base explanation rubric) were not transparent (Frederiksen & Collins, 1989) either to the students or the teacher. Although we used explanation terminology in the curriculum materials (e.g. conclusion, data, evidence, reasoning, explain), the materials did not define criteria for explanation and make them available as guidance for constructing explanations. Furthermore, an external review group (Project 2061 of AAAS) noted that some of the learning performances addressing explanations failed to articulate what was required of students.

It is possible that the teacher tried to compensate for the lack of transparency. In Harris et al. (in press), we reported qualitative analyses of a different teacher's enactment of a portion of the unit.¹ That teacher discussed with her students what "counts" as evidence in an explanation; student artifacts collected before and after the discussions, and analyzed with appropriate explanation rubrics, demonstrated improvement in evidence. The teacher in Enactment 1 could have addressed evidence similarly, accounting for gains in her students' evidence scores. Still, reasoning was a problem across classrooms. Clearly, we needed to make our explanation rubric more transparent to students.

Specific rubrics revisited

To better align the curriculum materials and the explanation rubric, we revised relevant learning performances to more accurately reflect the components of the base explanation rubric. For example, the learning performance for explaining chemical reactions became: 'When various substances come in contact with each other, students make a claim about whether a chemical reaction occurred, provide evidence for their claim, and articulate reasoning that links the evidence to their claim.'

In conjunction with our revision of learning performances, we revised the specific explanation rubrics to obtain a finer-grained analysis of students' explanations. The specific

¹ The enactment reported in Harris et al. (in press) differed from the present Enactment 1 in terms of student demographics as well as percentage of the unit covered. Pretest-posttest data on explanations ($N = 12$) from that site statistically were not suitable for analysis separately or combined with data from Enactment 1. However, we collected other informative data (e.g. classroom field notes and student artifacts) from that site which we did not collect for the present site.

rubrics include more detailed criteria for differentiating student responses for each of the components: claim, evidence, and reasoning.

Teaching explanation

Our revised curriculum materials include a lesson on explanation in science. The lesson provides support for teaching “scientific explanation” as a way of knowing that comprises a claim, evidence, and reasoning; student and teacher materials reinforce this idea consistently through the unit. Thus, criteria of the base explanation rubric become transparent to students. Our goal is to help students develop an explanation “schema” (cf. “argument schema,” Reznitskaya & Anderson, 2002) that they can apply to different science content.

Below, we consider students’ progress relative to two explanation learning performances for the revised chemistry unit.

Enactment 2

In the second enactment of the unit, students completed explanation tasks aimed at broader range of content standards. We examined their explanations on pre- and posttest tasks in which they either: 1) used evidence about properties of materials to explain whether any of the materials were the same substance, or 2) used evidence about substances to explain whether a chemical reaction occurred. We analyzed the explanations using appropriate specific rubrics for the tasks.

Methods

Participants. Two teachers enacted the unit in different public middle schools in Detroit; they had a total of seven classes. Student groups at the two schools were demographically similar; most students were African American and from lower- to lower-middle-income families. While a total of 230 students participated in the enactment, only data from 155 students who completed both pre- and posttest assessments were used in the present analyses.

Curriculum materials. The revised chemistry unit was 8 weeks in duration and targeted a larger number of concepts than the unit in Enactment 1: substance, property, chemical reaction, particulate nature of matter, and conservation of mass. The driving question and emphasis on inquiry were consistent with the unit in Enactment 1. Revisions relevant to the present argument were discussed under Curriculum Revision; further detail on revisions is available in McNeill et al. (2003).

Assessment tasks. Students wrote explanations on identical pre- and posttest tasks, and during embedded unit tasks. Tasks resembled those in Enactment 1 in that students considered data from an experiment. However, Enactment 2 included two types of tasks: 1) Substance task. This type of task involved data on properties of several materials and required that students explain whether any of the materials were the same substance. 2) Chemical Reaction task. This type of task was equivalent to that in Enactment 1.

Tasks in Enactment 2 did not involve a series of questions like those in Enactment 1. Instead, students were asked to ‘write a scientific explanation.’ For example, ‘Write a scientific explanation that answers the question: Are any of the solids the same substance?’ (Substance task), or ‘Write a scientific explanation that answers the question: Did a chemical reaction occur when Maya mixed hexane and ethanol?’ (Chemical Reaction task).

Additionally, some tasks provided prompts for students to write a claim, evidence, and reasoning, while others did not provide prompts. For the pretest and posttest, each student received an equal number of tasks, of each task type, with and without prompts. Across students, tasks with and without prompts were counterbalanced. Present analyses do not consider effects of the prompts.

Scoring. Explanations were scored as in Enactment 1, using an appropriate specific rubric for each type of task (Substance or Chemical Reaction). Average inter-rater reliability was greater than 90%.

Results and Discussion

We analyzed students’ explanations on pre- and posttest tasks. Tables 2 and 3 display pre- and posttest mean scores, results of paired *t*-tests on the mean difference scores, and effect sizes for Substance and Chemical Reaction tasks, respectively.

Each component of students’ explanations showed a significant gain from pre- to posttest, for both Substance and Chemical Reaction tasks ($p < .001$). Effect sizes of the claim score gains were similar for the two types of tasks and of respectable magnitude. Mean posttest claim scores did not differ by task type, $t(154) = .96$, *ns*. Thus, given sets of experimental data, students’ abilities to accurately differentiate substances and identify chemical reactions were comparable at posttest, both superior relative to pretest.

Evidence and reasoning scores differed by task type. For each component, effect sizes of the pretest-posttest gains were similar across type of task; however, mean posttest scores for the

components were higher for Substance than Chemical Reaction tasks, $t(154) = 7.82, p < .001$ for evidence and $t(154) = 9.53, p < .001$ for reasoning. The mean posttest scores for Chemical Reaction tasks were somewhat low on an absolute scale ($M_s = 0.72$ for evidence and 0.41 for reasoning; $\text{max} = 2.50$ for each component); scores for Substance tasks were more convincing ($M_s = 1.30$ for evidence and 0.96 for reasoning; $\text{max} = 2.50$). Overall, students performed better on Substance than Chemical reaction tasks, and better on evidence than reasoning.

Despite curriculum revisions aimed at making criteria for appropriate claims, evidence, and reasoning transparent to students, reasoning remained especially challenging for them in Enactment 2. However, reasoning scores for Substance tasks were promising. To explore why students' reasoning scores were higher for Substance than Chemical Reaction tasks, we examined the distribution of posttest reasoning responses by level of the specific explanation rubric, separately for each type of task. Figures 4 and 5 show the distributions for Substance and Chemical Reaction tasks, respectively.

Higher levels of reasoning (i.e. levels 1b and 2) require a justification for evidence, in other words, a reason why the evidence provided “counts” as appropriate evidence to support a claim. Students' posttest reasoning responses were far more likely to include such a justification for Substance (40% of responses) than Chemical Reaction (11%) tasks. But students did provide reasoning for Chemical Reaction tasks, even if that reasoning did not meet higher criteria. For Chemical Reaction tasks, 35% of responses included some form of accurate reasoning (i.e. levels 1a, 1b, or 2); for comparison, 53% of responses for Substance tasks did. Overall, students were less likely to include reasoning in explanations when identifying chemical reactions than when differentiating substances, and the reasoning they did provide was more likely to be at a lower level.

Differences in students' explanation responses and scores across content may underscore the importance of content knowledge in constructing explanations. Students could reason at higher levels; they did so for Substance tasks. Yet higher levels of reasoning were rare for Chemical Reaction tasks. It is possible that students generally had more difficulty with curriculum content about chemical reactions than with content about substances. Supporting this possibility, students' scores on a posttest multiple-choice measure were significantly lower for the subset of items targeting chemical reaction concepts ($M = 3.06$ points, $SD = 0.98$; $\text{max} = 5.00$) than for the equally-weighted subset targeting substance concepts ($M = 3.86$ points, $SD =$

0.78; max = 5.00), $t(154) = 9.55, p < .001$. Less knowledge about chemical reaction concepts could have left students ill-equipped to provide reasoning in their explanations of chemical reaction phenomena. Perhaps differences in explanation across content were most pronounced for reasoning because reasoning requires greater mastery of the target content than do claim and evidence components.

Another possibility is that differences in explanation across content were a manifestation of differences in the specific rubrics for the tasks. Our criteria for higher level reasoning for Chemical Reaction tasks may be more stringent than those for Substance tasks because the content is more difficult, making high reasoning scores for Chemical Reaction tasks more difficult to obtain.

Nevertheless, that many students could provide some form of reasoning for Chemical Reaction tasks, despite the difficult content, suggests that they realized that a “scientific explanation” necessitates reasoning. The students may have begun developing an explanation “schema” (cf. Reznitskaya & Anderson, 2002) that they applied to the tasks.

General Discussion

Enactment 1 to Enactment 2

The effect sizes of pretest-posttest gains in evidence and reasoning scores were more impressive in Enactment 2 than in Enactment 1. Revision of the curriculum appeared to move it in the right direction in terms of teaching explanation. Yet posttest scores for each component were lower than expected for Enactment 2 suggesting that further revisions are necessary. Reasoning demands special attention because compared to claim and evidence, reasoning scores were relatively low across enactments, albeit higher on an absolute scale for Enactment 2 than Enactment 1.

Future curriculum revisions

The revised curriculum included explicit instruction in the criteria for explanation. However, the criteria may have remained elusive to the student, especially those for reasoning. For example, the curriculum materials defined reasoning as ‘a statement that tells how your evidence supports your claim.’ This definition may not be accessible to students, presenting them with not only the task of constructing an appropriate reasoning statement but the additional challenge of figuring out the purpose of such a statement. Currently, we are improving the curriculum materials to make criteria for explanation more explicit and accessible to students.

For example, defining reasoning as a justification of ‘why evidence “counts” as evidence for a claim’ could be more accessible terminology, and guidelines for how to apply this idea may make the component more explicit. Likewise, we are addressing the components claim and evidence. Our goal is to make the criteria of the explanation rubric more transparent.

Students’ higher-level reasoning for Substance relative to Chemical Reaction tasks is an intriguing and informative finding. If appropriate, high-level reasoning on a task requires high mastery of content, we might expect such a finding based on students’ higher mean score for multiple-choice items targeting substance relative to chemical reaction concepts. When faced with content that was less familiar (e.g. identifying chemical reactions), a number of students simply repeated their evidence as reasoning, a low-level response that accounted for most of the accurate reasoning responses for Chemical Reaction tasks in Enactment 2. The challenging content seemed to intensify students’ difficulty with reasoning, indicating that revisions of science content in the instructional materials must accompany revisions addressing criteria for explanation. But their attempts to provide reasoning are encouraging. These attempts suggest that they realize reasoning is important and are trying to incorporate the component in their explanations; in other words, possibly trying to apply a developing explanation “schema” to the task at hand.

Implications for using rubrics in practice

We presented data garnered through use of a base rubric for one way of knowing, explanation. The base rubric offers a number of advantages: 1) An instrument for breaking down student “knowing” to reveal key student competencies and difficulties with ways of knowing science content. 2) Provides data about student performances that are immediately usable to teachers and researchers for revising curriculum and instruction. 3) Adaptable to any science content via specific rubrics, promoting consistent assessment of ways of knowing demonstrated by students across content and over time (in this case, over two enactments). 4) Alignment of assessments with instructional materials and learning goals derived from content and ways of knowing, providing sensitive measures of what students understand. We hope to extend these conclusions to other ways of knowing shortly, and offer a set of base rubrics to the science education community.

References

- American Association for the Advancement of Science. (1993). *Benchmarks for science literacy*. New York: Oxford University Press.
- American Association for the Advancement of Science. (1990). *Science for all Americans: A Project 2061 report on the literacy goals in science, mathematics, and technology*. New York: Oxford University Press.
- Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Bell, P., & Linn, M. C. (2000). Scientific arguments as learning artifacts: Designing for learning from the Web with KIE. *International Journal of Science Education*, 22 (8), 797-817.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classrooms. *Science Education*, 84 (3), 287-312.
- Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher*, 18 (9), 27-32.
- Harris, C. J., McNeill, K. L., Lizotte, D. J, Marx, R. W., & Krajcik, J. (in press). Aligning standards-based assessment and curriculum materials for teaching inquiry science. *PEERs Matter*.
- Krajcik, J., Berger, C. F., & Czerniak, C. M. (2002). *Teaching science in elementary and middle school classrooms: A project-based approach* (2nd ed.). New York: McGraw Hill.
- Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education*, 77 (3), 319-337.
- McNeill, K. L., Lizotte, D. J, Harris, C. J., Scott, L. A., Krajcik, J., & Marx, R. W. (2003, March). *Using backward design to create standards-based middle-school inquiry-oriented chemistry curriculum and assessment materials*. Paper presented at the annual meeting of the National Association for Research in Science Teaching, Philadelphia, PA.
- National Research Council. (1996). *National science education standards*. Washington, DC: National Academy Press.
- National Research Council. (2000). *Inquiry and the National Science Education Standards: A guide for teaching and learning*. Washington, DC: National Academy Press.

- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (Eds.). (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Academy Press.
- Perkins, D. (1992). *Smart schools: From training memories to educating minds*. New York: The Free Press.
- Reznitskaya, A., & Anderson, R. C. (2002). The argument schema and learning to reason. In C. C. Block, & M. Pressley (Eds.), *Comprehension instruction: Research-based best practices* (pp. 319-334). New York: The Guilford Press.
- Ruiz-Primo, M. A., Shavelson, R. J., Hamilton, L., & Klein, S. (2002). On the evaluation of systemic science education reform: Searching for instructional sensitivity. *Journal of Research in Science Teaching*, 39 (5), 369–393.
- Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *The Journal of the Learning Sciences*, 12 (1), 5-51.
- Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29 (7), 4-14.
- Toulmin, S. (1958). *The uses of argument*. Cambridge, UK: Cambridge University Press.
- White, R., & Gunstone, R. (1992). *Probing understanding*. London: The Falmer Press.

Table 1

Enactment 1 Data and Inferential Statistics for Students' (N=77) Chemical Reaction Explanations

Score Type	Maximum	Pretest <i>M</i> (<i>SD</i>)	Posttest <i>M</i> (<i>SD</i>)	<i>t</i> (76) ^a	Effect Size ^b
Total	7.50	2.85 (1.49)	3.71 (1.25)	4.60 ^{***}	0.57
Component					
Claim	2.50	1.95 (1.04)	2.21 (0.81)	1.92 [*]	0.25
Evidence	2.50	0.74 (0.72)	1.21 (0.83)	4.15 ^{***}	0.65
Reasoning	2.50	0.16 (0.42)	0.29 (0.53)	2.04 [*]	0.31

^a One-tailed paired *t*-test^b Effect size was calculated as the difference between pretest *M* and posttest *M* divided by pretest *SD*.^{***} $p < .001$ ^{*} $p < .05$

Table 2

Enactment 2 Data and Inferential Statistics for Students' (N=155) Substance Explanations

Score Type	Maximum	Pretest <i>M</i> (<i>SD</i>)	Posttest <i>M</i> (<i>SD</i>)	<i>t</i> (154) ^a	Effect Size ^b
Total	7.50	1.32 (1.68)	3.57 (2.33)	11.91 ^{***}	1.34
Component					
Claim	2.50	0.61 (0.82)	1.31 (0.99)	7.96 ^{***}	0.85
Evidence	2.50	0.58 (0.78)	1.30 (0.96)	8.60 ^{***}	0.92
Reasoning	2.50	0.13 (0.30)	0.96 (0.84)	12.40 ^{***}	2.76

^a One-tailed paired *t*-test^b Effect size was calculated as the difference between pretest *M* and posttest *M* divided by pretest *SD*.^{***} $p < .001$

Table 3

Enactment 2 Data and Inferential Statistics for Students' (N=155) Chemical Reaction Explanations

Score Type	Maximum	Pretest <i>M</i> (<i>SD</i>)	Posttest <i>M</i> (<i>SD</i>)	<i>t</i> (154) ^a	Effect Size ^b
Total	7.50	0.93 (1.14)	2.53 (1.58)	13.16 ^{***}	1.40
Component					
Claim	2.50	0.70 (0.86)	1.40 (0.85)	8.42 ^{***}	0.82
Evidence	2.50	0.20 (0.41)	0.72 (0.61)	10.57 ^{***}	1.29
Reasoning	2.50	0.03 (0.13)	0.41 (0.56)	8.69 ^{***}	2.98

^a One-tailed paired *t*-test^b Effect size was calculated as the difference between pretest *M* and posttest *M* divided by pretest *SD*.^{***} $p < .001$

Figure 1. Model of Assessment-Driven Design Process

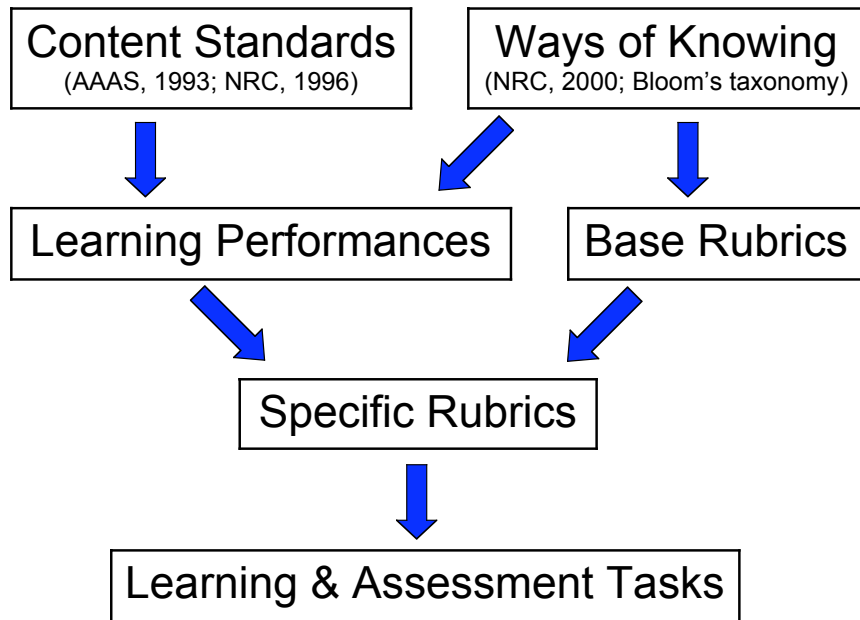


Figure 2. Base explanation rubric.

Component	Level		
	0	1	2
Claim – Assertion or conclusion for a problem	Does not make a claim, or makes an inaccurate claim.	Makes an accurate but incomplete claim.	Makes an accurate and complete claim.
Evidence – Data that supports claim	Does not provide evidence, or only provides evidence that does not support claim.	Provides accurate but insufficient evidence to support claim. May include some evidence that does not support claim.	Provides accurate and sufficient evidence to support claim
Reasoning – Argument that links evidence to claim	Does not provide reasoning, or only provides reasoning that does not link evidence to claim.	Provides accurate but incomplete reasoning that links evidence to the claim. May include some reasoning that does not link evidence to claim.	Provides accurate and complete reasoning that links evidence to claim.