

# Reconsidering the Use of Scoring Rubrics in Biology Instruction

XIAOWEI TANG, JANET COFFEY,  
DANIEL M. LEVIN

## ABSTRACT

Scoring rubrics are widely employed across content areas and grade levels, including in high school biology classes. Besides regular external use within accountability systems, educators also have advanced their instructional use inside classrooms. In recent years, a consensus appears to be emerging in the educational literature that instructional use of rubrics is beneficial for student learning, and numerous examples in the research and practitioner literature establish their importance in teachers' planning, instruction, and assessment. We examine this assumption through close analysis of students' use of a scoring rubric in a high school biology classroom. We explore how instructional use of a scoring rubric influences biology teaching and learning activities, what messages about knowledge and learning such use conveys to students, and what influence such use may have on students' emergent understandings of what constitutes quality in biological thinking and practice. Our analysis suggests that instructional use of scoring rubrics can actually undermine the very learning it is intended to support. We discuss an alternative way to help students understand what constitutes high-quality work, and we draw implications for science teacher education.

**Key Words:** Biology teaching and learning; secondary biology; science teaching; science teacher education; science instruction; science assessment.

## ○ Introduction

The term “scoring rubric,” broadly defined, refers to any assessment tool that “lists criteria and provides some explanation for levels or ‘gradations’ of quality” (Popham, 1997). They are widely employed across content areas and grade levels, including in high school biology classes, which we discuss here. Besides regular external use within accountability systems, educators also have advanced their instructional use inside classrooms (e.g., Andrade, 2000). In recent years, a consensus appears to be

*The term “scoring rubric,” broadly defined, refers to any assessment tool that “lists criteria and provides some explanation for levels or ‘gradations’ of quality”.*

emerging in the educational literature that instructional use of rubrics, despite their limitations, is beneficial for student learning (e.g., Andrade & Valtcheva, 2009). There are numerous examples of such use in journals for biology instructors at all levels that establish their importance in teachers' planning, instruction, and assessment (Allen & Tanner, 2006; Siegel et al., 2011).

Our study examines this assumption through close analysis of students' use of a scoring rubric in a high school biology classroom. Specifically, we consider the following questions: (1) In what ways can instructional use of a scoring rubric influence biology teaching and learning activities? (2) What messages about knowledge and learning might such use convey to students? And (3) what influence might such use have on students' emergent understandings of what constitutes quality in biological thinking and practice? In brief, our analysis shows that instructional use of scoring rubrics can actually undermine the very learning it is intended to support. Before turning to our data, we first go over the arguments in support of instructional use of scoring rubrics.

## Use of Scoring Rubrics in Classroom Instruction: Support from the Literature

Formative assessment literature calls for making criteria and goals explicit to students (Black & Wiliam, 1998), as well as engaging students in self and peer assessments (White & Frederiksen, 1998). This literature suggests that clear criteria and shared goals can help focus students' attention on what “counts” (Coffey, 2003), building up learning motivation (Andrade & Du, 2005) and allowing for more focused and instructive

feedback (Black & Wiliam, 1998).

It makes sense that rubrics have emerged as a vehicle serving these purposes. Theoretically, with articulated levels and criteria,

rubrics would enable students to make “dependable judgments about the quality of their own work” (Stiggins, 2001) and gain insight into the tacit assumptions held by experts, including their teachers, about what counts as disciplinary quality (Andrade & Valtcheva, 2009).

Literature also suggests that students can be effective users of scoring rubrics and benefit from such use. For example, a quantitative research study showed that when a rubric was introduced and continuously used, college students’ assessments of their oral presentations on evolutionary biology were significantly correlated with that of the teacher (Hafner & Hafner, 2003). In another study, Andrade et al. (2008) conducted a quasi-experiment in which the treatment group first discussed the strengths and weaknesses of an exemplar and then applied a rubric to their own work. Students in the treatment group produced essays that received higher scores across dimensions than those of their peers. Studies also suggest that students report rubrics to be useful in terms of communicating teacher expectations and providing guidance for work (Andrade & Du, 2005).

In sum, there seems to be substantial agreement that using scoring rubrics for pedagogical purposes is productive in conveying ideas about the qualities of good work and improving the quality of student work. Inter-rater reliability between teacher and student has commonly been adopted as a measure of effective rubric use (e.g., Hafner & Hafner, 2003), despite the caution that such agreement depends heavily on the number of criteria (Falchikov & Goldfinch, 2000). This caution suggests that the finding that students can effectively use a rubric to evaluate in ways similar to the teacher does not necessarily demonstrate that such use is supportive of student learning. It is also important to note that how rubrics are used in classrooms across the research studies differs in significant ways, yet this variation is not fully explored.

This is where we enter the conversation. We agree that it is at the very core of teachers’ responsibilities to help students understand what constitutes disciplinary quality and engage them in meaningful assessment activities. What we call into question is the particular role rubrics can and should play in this.

### Instructional Use of Scoring Rubric: A Case from Ms. H’s High School Biology Classroom

At the heart of our analysis lies a case of high school biology students assessing sample responses to a short-essay prompt on natural selection, using a scoring rubric designed for use on a high-stakes state biology test. The case, drawn from a research project examining what biology teachers attended to in classrooms, typifies how rubrics are used in biology classrooms, according to our observations among the teachers in this and other projects in which we have worked with biology teachers. (All names used here are pseudonyms.)

The case took place in a ninth-grade biology class on evolution. Ms. H divided the students into groups, providing each group a scoring rubric and a sample response to an essay prompt on natural selection. (They had studied evolution and natural selection and were familiar with conceptual content as well as the focus of the question.) Ms. H then instructed the groups to evaluate the response and justify a score using the rubric (Figure 1). Five minutes later, groups were called in turn to explain their evaluations.

According to Ms. H, this activity was for students to see “how difficult it is to grade [responses] using the rubric,” and for her to check on students’ understandings of natural selection through their comments and assessments. Many teachers from her school district made similar use of that rubric, intending to convey to students

Science Rubric  
Brief Constructed Response

	Level of Understanding	Synthesis of Information	Use of Supporting Details	Use of Accurate Scientific Terminology	Application of Information*
4	The student has a <b>full and complete understanding</b> of the question or problem.	The response reflects a complete synthesis of information.	Pertinent and complete supporting details demonstrate an integration of ideas.	The use of accurate scientific terminology enhances the response.	An effective application of the concept to a practical problem or real-world situation reveals an insight into scientific principles.
3	The student has a <b>good understanding</b> of the question or problem.	The response reflects some synthesis of information.	The supporting details are generally complete.	The use of accurate scientific terminology strengthens the response.	The concept has been applied to a practical problem or real-world situation.
2	The student has <b>basic understanding</b> of the question or problem.	The response provides little or no synthesis of information.	The supporting details are adequate.	The use of accurate scientific terminology may be present in the response.	The application of the concept to a practical problem or real-world situation is inadequate.
1	The student has <b>some understanding</b> of the question or problem.	The response addresses the question.	The supporting details are only minimally effective.	The use of accurate scientific terminology is not present in the response.	The application, if attempted, is irrelevant.
0	The student has <b>no understanding</b> of the question or problem. The response is completely incorrect or irrelevant.				

\*On the High School Assessment, the bullet that defines application criteria will only be used when application is requested in the item stem.

Figure 1. Scoring rubric used in Ms. H’s class.

the criteria their responses would be evaluated against on the biology test, which high school students need to pass in order to graduate.

The essay prompt reads as follows:

One of the birds found on the Galápagos Islands is the medium ground finch. These birds prefer to eat small seeds, which are easier to eat than large seeds. However, when food is scarce, such as during a drought, some of the birds can eat larger seeds. The ability to crush and eat larger seeds is related to beak thickness, an inherited characteristic. Birds with thick beaks can crush large seeds more easily.

- Describe the changes that would occur in the medium ground finch population during a long period of drought when food is scarce.
- Explain how this set of changes is an example of the process of natural selection.

The rubric employed consists of four dimensions, four performance levels under each dimension, and specific criteria corresponding to each level (for details, see Figure 1). The language used to articulate the criteria is notably vague. For example, under the dimension “use of supporting details,” criteria vary from “only minimally effective” to “adequate” to “pertinent and complete.” What constitutes “adequate” and how this differs from “pertinent and complete” is left to users’ interpretations. Similarly, level 4 in the “application of information” dimension requires that the response shows “effective application of the concept” that “reveals an insight into scientific principles,” without more clarity on what counts as “effective application” and how such application suggests “insight into scientific principles.”

The criteria include “the use of accurate scientific terminology” (see highlighted region of Figure 1), a straightforward judgment requiring little analysis into the meaning of the criteria or of the work being evaluated. This criterion became the entry point for students’ use of the rubric. The major reason, we suspect, lies in its time-saving features, and the fact that it is relatively uncomplicated to pick out the terminology.

We see the tendency to gravitate toward these criteria even in the teacher. As Ms. H rotated around the classroom while students worked in groups, she frequently brought up the question “Did he use any scientific terminology?” when sensing difficulties in analyzing and scoring the responses. Much of the group discussions focused on terms and vocabulary.

In a subsequent meeting, Ms. H confirmed that she was trying to steer the students toward identifying terms, hoping that they would go from there to think about whether the sample responses “even make sense using these words.” This could have been a productive goal, as it would require reasoning through the response to understand and evaluate the meanings of the terms used. However, in the rest of that class period, there was little evidence that this goal was realized. The focus on terminology continued to dominate, as the following episodes demonstrate.

### Episode 1: “But They Didn’t Use Scientific Terms!”

One sample response reads: “After a long drought, birds with thicker beaks would survive; birds with thinner beaks would die

off. And as a result, the population would decrease.” The assigned group gave it a score of 2 because “it basically answered the first bullet.” The following exchanges ensued:

- Mark: And, it didn’t answer bullet number two, ‘cause it asked him the –, how did they –
- Ken: [in a low voice] He didn’t explain it as an example of –
- Ms. H: What is that you’re mumbling?
- Ken: Natural selection.
- Mark: The selection. He didn’t mention natural selection yet.
- Ms. H: They never said natural selection, but they alluded to the fact that there is going to be a decrease in the population, they just don’t really explain it elaborately.
- Mark: [responding quickly] But, they didn’t use scientific terms.
- Ken: Yeah.
- Ms. H: Ok. Good.

In justifying their scoring, the group offered that the term “natural selection” was not used. By pointing out that the idea of natural selection was partly implied in the “decrease in the population” prediction, Ms. H challenged this conclusion, pushing for engagement with the meaning of the text. However, her challenge was quickly set aside when the students referred to “use scientific terms” as the foundation of their assessment.

Such exchanges highlight how a focus on terminology criteria suppressed an opportunity to critically analyze an idea. A close analysis of the text could lead students to agree with Ms. H’s point, or to argue that “decrease in population” is simply an inference of population change after the drought, offering no explanation why such change served as an account of “natural selection.”

Reasoning at that depth was not evident in the conversations. Starting with the quick response “But,” Mark suggested that regardless of whether the meaning of “natural selection” was implied or not, their assessment would stay the same, since that term was literally absent. The fact that “natural selection” also appeared in the prompt – and thus its presence or absence could not constitute solid evidence for assessment – was not even mentioned. Yet, in light of the rubric criteria on terminology, Ms. H’s positive evaluation (“Good”) validated this rationale.

### Episode 2: “Did They Explain Natural Selection Correctly?”

Another sample response reads as follows:

Finches with thick beaks would live on and breed. And breed to produce more thick beaked finches. This shows natural selection. The most adapted finches were the only ones who successfully survive the draughts and were able to produce more of their kind. Natural selection occurs when a natural trait of an organism makes it too weak for its environment and so it dies off.

The assigned group gave it a score of 3. Instead of a general rationale, Ms. H asked for “evidence of good understanding,” which merited a 3 in the “level of understanding” column. In support of their score, the students suggested that “they explained the term ‘natural selection,’” referring to the last sentence in that response: “Natural selection occurs when a natural trait of an organism makes it too weak for its environment and so it dies off.” Ms. H continued

to pursue their reasoning by asking “Did they explain natural selection correctly?” In response, Chara, the class “natural selection expert,” compared the response to the explanation recorded in her notebook and concluded that “they kind of did it.”

Had students focused their assessment on evidence of understanding, there was much to be discussed. For example, can natural selection be applied to the relationship between “a natural trait of an organism” and “its environment,” or should it always refer to a population? Is natural selection a one-time event in which one phenotype dies off and the other survives, or should it be understood as happening over generations? Discussions like these, if they occur, would provide Ms. H with useful information about her students’ understandings regarding natural selection, and convey to students specific notions about what kind of details they should look for in order to evaluate the quality of such biological explanations.

Similar to how the use of terminology was reduced to the *presence* of terminology in the first excerpt, here “evidence of good understanding” was reduced to whether the description of the terminology was aligned with its notebook definition. Again, an “acceptable” assessment was made without critical analysis of response content or disciplinary logic.

### Episode 3: “But I Don’t See Reproduce”

The assigned group gave the following sample response a score of 4:

Over time a change would occur in the finch population. Over a long period of draught when food is scarce, the birds that only eat the smaller seeds will begin to die off because there will not be enough small seeds for all of the birds. The birds with thicker beaks have an inherited advantage and will thrive even during the period of draught. Therefore eventually, most of the whole population will soon consist of birds with thicker beaks because thicker beaks are necessary to survive. This inherited characteristic will be passed on from each generation to keep the species alive. This is considered natural selection because those organisms without an inherited advantage simply die off whereas the other finches survive because of their genetic variation.

Following Ms. H’s request for evidence that the response addressed both bullets, Max, the speaker of the group, first read supporting texts from the original response, and then brought up an additional point:

Max: And they use the terms.

Ms. H: But what terms? Let’s see ’em.

Max: They use “variation,” they use “natural selection,” they use “survive,” but I don’t see “reproduce.”

When justifying their score by reading from the original response, Max pointed out that the sentence “Therefore eventually, most of the whole population will soon consist of birds with thicker beaks” indicated “a trait passing on through reproduction.” Only six conversational turns later, in the above exchanges, she took it as a problem that the term “reproduce” was missing. Obviously, checking of terminology was done in separation from interpretation and analysis of the response content; and while level 4 of the terminology dimension emphasizes that the use of terminology

should “enhance the response,” here the assessment does not address how the terms are used at all.

The above episodes illustrate the following possible problems with instructional use of scoring rubrics:

- Use of scoring rubrics as instructional tools does little to raise holistic thinking about the quality of work and can encourage literal acceptance of an authoritative sense of quality.
- Criteria are commonly articulated in a language that, while appearing accessible, is difficult for students to translate into practice.
- The most superficial and operable criteria in a rubric receive the most attention in classroom assessment practice. This tendency risks conveying a distorted message about what constitutes quality in biology explanations. Such a limited view of quality is reinforced by the reductionism inherent in many rubrics.

One might argue that Ms. H was using a bad rubric. Popham (1997) summarized several common flaws that plague many teacher-made and commercially made rubrics: criteria being excessively task-specific or general; containing dysfunctional details; and equating a test of the skill with the skill itself. The rubric in our case avoids many of these flaws and actually adheres to the attributes of a good rubric – containing three to five evaluative criteria, representing key attributes of the skills being assessed, and being “teachable” in the sense that teachers can help students develop these skills (Popham, 1997). The vagueness of its language, while causing difficulties for students to understand, is appropriate and necessary for external scoring. Removing such vagueness requires inclusion of more task-specific details, which, in turn can increase the rigidity of the assessment.

It might also be argued that the scoring rubric was used in a problematic manner. On some accounts, this is a legitimate criticism. Literature does suggest that students need to learn how to use rubrics, which could take much of a school year (Fairbrother et al., 1995). That said, the literature does not make clear the nature of beneficial usage. It is also worth noting that Ms. H’s use of the rubric was aligned with recommendations from her district curriculum guide, and similar implementations have been repeatedly observed across our work with high school teachers.

Other challenges include claims that target the inaccessibility of the language, or ones that emphasize that the description may be unique to Ms. H’s classroom. To confront either or both of these notions, we draw on an example from another high school biology class. A similar piece of the state scoring rubric was summarized by the teacher into three general criteria and translated into a language that was more easily accessible for the students (“Answer all the bullets,” “Use and explain vocabulary,” and “Use evidence and details to support statements”).

Rather than articulating criteria for each level along numerous dimensions, students were given anchor papers with graders’ notes as a reference and were asked to assess each other’s responses by making a judgment and providing “examples of what was done correctly or what could be improved.” This modified version of the state rubric proved to be easier for students to understand, and – by asking for examples for each dimension – more oriented toward the substance of ideas reflected in the response. However, many of the students did not provide examples on their written assessments, and among those who did, few provided solid

reasoning about how certain evidence or detail supported their overall judgment. Students' assessments were still focused on "correctness" and use of terminology, rather than on the quality of the scientific reasoning and explanation. So, while the assessment literature asserts that accessibility of language is important to consider (Stiggins, 2001), accessibility of language alone is not enough to ensure meaningful and productive use.

Observations in Ms. H's classroom also offered evidence that she could be responsive to student reasoning (Levin, 2008), and that these very students could engage in reasoning through the plausibility and reasonableness of biological ideas and reasoning. Therefore, the lack of cognitive engagement in this activity was not an issue of natural ability, nor of the teacher's perception of her students' abilities, but was largely shaped by the nature of the activity and the actual rubric tool.

## ○ Discussion

Our analysis highlights the distracting influences that instructional use of scoring rubrics can bring to a biology classroom. Stripped from an underlying rationale for what constitutes quality and why, and how the different dimensions interrelate, rubrics may not only fail to help students better understand the quality of biological ideas and reasoning, but also undermine their intuitive sense of good scientific explanations.

### Roots of the Problem

Expert scorers understand how different aspects of quality interrelate, and why they are worth considering in the first place. With such knowledge, they can interpret the rubric in a more nuanced, flexible way that captures the dynamics of quality, making prompt decisions on whether a criterion is applicable to a particular piece of work. For example, if a rubric requires the use of terminology, and yet the meaning of an absent term was present in a response; or if a rubric requires graphic representation of data, and yet the particular data set does not lend itself to that, expert scorers could make accommodations rather than adhere to literal interpretations and judgments.

Instructional use of rubrics typically assumes the converse to be true – that is, making criteria explicit helps students develop an understanding of quality and of how dimensions of quality are interrelated. This assumption overlooks the fact that experts themselves do not achieve an understanding and application of such dynamics by simply drawing on the rubric, but through interpretation of substantive meanings and sense-making of that experience.

Starting with an explicit, preexisting rubric can draw students' attention away from such fundamental experience but toward discrete criteria that may not make sense to them yet. This is where things get problematic with instructional use of scoring rubrics. We saw this in Ms. H's class, and it is commonly played out across classrooms.

Furthermore, external assessment is a quite different activity from the production of the work being assessed. What scorers should be looking for (and at) to make a judgment about quality is not necessarily what students should be striving for in producing that quality. In this sense, rubrics useful for scoring may not be helpful for student learning.

To help advance this point, we turn to an example from outside the classroom. Almost every second-grader understands that the success of a basketball player's shot is whether or not it goes through the net. However, handing a second-grader a ball and telling them to get the ball through the net is not necessarily going to result in an improved shot. Certain instructions can help improve their shot, such as keeping shoulders square to the basket, using their non-shooting hand to guide the ball, using legs rather than pushing with their arms, etc. These are features of a shot that will normally increase the probability for a ball to go through the net. Some may, therefore, claim that these are exactly the type of features a rubric assessing shots should include.

However, we would not expect such a rubric to drive our evaluation of an expert's shot. Reggie Miller, a leading scorer in the history of the National Basketball Association, had an unconventional shot, which basketball announcers and sports bloggers frequently mentioned. He shot with his hand behind the ball, his elbow away from his body and his legs apart, which would not fare well if examined in light of any rubric. However, it would be absurd to deny that he is a good shooter. The rubric used to identify a successful shot is straightforward: does it go through the net or not? The rubric to help a young basketball player improve his shot is very different in nature. One should also note that if Reggie Miller tried to adjust his shot to the articulated features of the "rubric," his shooting very likely would suffer. He would no longer be the individual on the court who you would want to have the ball in his hands as time on the clock is winding down.

### Ongoing & Evolving Assessment Conversations: An Alternative

Supporting learning of high-quality biological explanations and preparing students for tests are the dual goals Ms. H and many other biology teachers pursued when employing scoring rubrics for instructional use. The two overlap to some extent, since it is generally assumed that external scorers base their assessment on a deep understanding of disciplinary quality. For either goal, we argue for an alternative: starting with activities that allow students to critically analyze and reflect on the quality of work and letting criteria grow from there. Though time-consuming and requiring much more effort, this alternative can provide students with the kind of experience through which external scorers develop their expertise in assessment. Andrade et al. (2008), for instance, suggested that in comparison with abstractly constructed criteria, criteria generated through students' analysis of a model writing sample, when applied in their self-assessment of drafts, led to greater gains in essay-writing assignments.

Coffey (2003) provided a detailed case illustrating what such an alternative looks like in the science classroom. Through "Connections," a middle school science program, a class of sixth-graders continuously engaged in assessment-related conversations. Starting from stating what they liked or did not like about particular work, asking for clarifications, generating their own evaluation sheets, reflecting on and incorporating feedback from peers and the teacher, and so on, the students gradually constructed shared meanings of "high quality" for scientific presentations. Although initial evaluations typically drew on surface traits such as speaking loudly and making eye contact, over time the focus shifted toward content-related issues, such as "clarity of information,

organization of information, source of information, and even accuracy of information” (p. 9).

This example flips current rubric practice on its head, because even very good rubrics that articulate sophisticated versions of expertise and scientific practice risk artificially demarking notions of quality and undermining students’ scientific reasoning when set out in advance as a roadmap for reasoning. The notion of starting with the work students do and their own ideas of disciplinary quality is aligned with Duschl & Gitomer’s (1997) emphasis on engaging students over time in an ongoing assessment conversation about what constitutes high-quality work.

The example also demonstrated multiple ways of scaffolding such long-term engagement. The teacher in this case made reflective discussions about work a classroom routine, identified areas for feedback that students did not attend to, directed students to comment on specific issues, and organized regular discussions about what makes something high-quality work. Providing such scaffolding requires adaptive expertise of localizing deep understandings of disciplinary quality to various contexts, of interpreting from students’ remarks their nascent or developing criteria of quality, and of choosing the appropriate discursive strategies to facilitate quality-related meaning-making through ongoing conversations. Implementing the alternative thus requires much effort from teacher educators’ end. Novice and practicing teachers should have chances to experience this type of activities from the learner’s perspective. They need supports to develop the responsiveness and thoughtfulness that will allow them to guide their students in talking and setting criteria for quality. On top of that, it is also important to help them see that this alternative provides an even more powerful avenue for test preparation, because it provides access to the contextualized and holistic way the students are most likely to be assessed in exams.

## ○ Limitations & Conclusions

It might be argued that it is important for students to be familiar with the rubrics that are going to be used on the high-stakes tests they are required to take, and that our alternative does not allow for that. It could, however. If students were allowed to spend some time developing their own understandings of what constitutes high-quality work, they could then analyze the official test rubric – looking at how it is aligned, or not, with their ideas – and consider how their deeper understanding of quality would translate into the expectations of the rubric.

Additionally, readers might be concerned that we have not shown how the students’ learning was affected by the instructional use of the rubric. After all, perhaps they *did* learn how to use the rubric in a productive way that helped them perform better on an assessment. We did not collect any data to examine students’ performance on the writing tasks before and after the rubric discussion. However, our descriptions and analysis of the classroom interactions – which, according to Vygotsky (1978), is where learning first takes place – showed little evidence of productive learning as we have described above, in terms of either developing a sense of quality or demonstrating a deep understanding of the biological concepts involved.

Finally, it might also be suggested that things could be different if more time were allotted for this activity, which would give the

students greater opportunity to have deeper conversations and to better unpack the rubric. Sadly, however, in the many classrooms we have observed, we have rarely seen such longer discussions. These rubric discussions are usually limited to a portion of one class period. Most likely, with the pronounced emphasis on high-stakes tests, spending more time on the scoring rubric would distract teachers from issues of quality and deep conceptual understanding, and focus them on quickly giving the students a sense of how to respond to the rubric. Additionally, constructivist philosophy implies that learning is more productive when it begins with what people already know. Thus, devoting the instructional time available to engaging students in discussions of disciplinary quality, tapping into their intuitive and emerging sense of quality work, is more likely to help them understand normative disciplinary standards of quality than spending that time on responding directly to the rubric.

In conclusion, we argue that rather than continuing to engage in unquestioned construction and use of rubrics, the biology education community should explore more critically how rubrics operate in classrooms. The field needs greater evidence-based understanding of how the use of rubrics influences the ways teachers teach and students learn, and biology teachers need to think critically and systematically about how they are using rubrics in instruction.

## ○ Acknowledgments

Work on this project was supported by a grant from the twelfth Five-year National Plan on Educational Research (grant BFA110052, “How sociocultural learning environments influence the effectiveness of classroom scientific inquiry”) and by a grant from the National Science Foundation (ESI 0455711, “What influences teachers’ modification of curriculum?”).

## References

- Allen, D. & Tanner, K. (2006). Rubrics: tools for making learning goals and evaluation criteria explicit for both teachers and learners. *CBE Life Sciences Education*, 5, 197–203.
- Andrade, H. (2000). Using rubrics to promote thinking and learning. *Educational Leadership*, 57(5), 13–18.
- Andrade, H. & Du, Y. (2005). Knowing what counts and thinking about quality: students report on how they use rubrics. *Practical Assessment, Research and Evaluation*, 10(4).
- Andrade, H., Du, Y. & Wang, X. (2008). Putting rubrics to the test: the effect of a model, criteria generation, and rubric-referenced self-assessment on elementary school students’ writing. *Educational Measurement: Issues and Practice*, 27(2), 3–13.
- Andrade, H. & Valtcheva, A. (2009). Promoting learning and achievement through self-assessment. *Theory Into Practice*, 48, 12–19.
- Black, P. & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education*, 5, 7–74.
- Coffey, J.E. (2003). Making connections: students engaging in and with assessment. Ph.D. dissertation, Stanford University, Palo Alto, CA.
- Duschl, R.A. & Gitomer, D.H. (1997). Strategies and challenges to changing the focus of assessment and instruction in science classrooms. *Educational Assessment*, 4, 37–73.

Fairbrother, R., Black, P.J. & Gill, P. (Eds.) (1995). *Teachers Assessing Pupils: Lessons from Science Classrooms*. Hatfield, UK: Association for Science Education.

Falchikov, N. & Goldfinch, J. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research*, 70, 287–322.

Hafner, J.C. & Hafner, P.M. (2003). Quantitative analysis of the rubric as an assessment tool: an empirical study of student peer-group rating. *International Journal of Science Education*, 25, 1509–1528.

Levin, D.M. (2008). What secondary science teachers pay attention to in the classroom: situating teaching in social and institutional systems. Ph.D. dissertation, University of Maryland, College Park, MD.

Popham, W.J. (1997). What's wrong – and what's right – with rubrics. *Educational Leadership*, 55(2), 72–75.

Siegel, M.A., Halverson, K., Freyermuth, S. & Clark, C.G. (2011). Beyond grading: a series of rubrics for science learning in high school biology courses. *Science Teacher*, 78, 28–33.

Stiggins, R.J. (2001). *Student-Centered Classroom Assessment*. Columbus, OH: Merrill Prentice Hall.

Vygotsky, L.S. (1978). *Mind in Society: The Development of Higher Mental Process*. Cambridge, MA: Harvard University Press.

White, B.Y. & Frederiksen, J.R. (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cognition and Instruction*, 16, 3–118.

XIAOWEI TANG (xiaowei.tang@gmail.com) is an Associate Professor at Southwest University, Basic Education Research Center. JANET COFFEY (janet.coffey@moore.org) is a Program Officer in Science Learning, Gordon and Betty Moore Foundation, 1661 Page Mill Rd., Palo Alto, CA 94304. DANIEL M. LEVIN (dlevin2@umd.edu) is a Clinical Assistant Professor in the Department of Teaching and Learning, Policy and Leadership, College of Education, University of Maryland, College Park.

**DISCOVER  
NEW WONDERS  
OF DNA**

**AT THE  
DNA STORE!**

**toys, balloons, neckties,  
art, earrings, mugs,  
models, coins, stamps,  
cards, roadsigns, jewelry,  
puzzles, most anything  
you can imagine!**

**thednastore.com      dna.science**