Developing a Classroom Assessment Rubric
An Example From a Research-Based Undergraduate Course

By Chandrani Mishra, Loran Carleton Parker, and Kari L. Clase

The development and implementation of varied assessment practices is a major focus in higher education. Assessment benefits both students and teachers; it informs teachers about students’ learning and misconceptions, thereby helping teachers improve teaching practices and students assess their current state of understanding. In general, the field lacks a rubric or an assessment model for assessing students’ understanding of science content and their representations, which was the impetus for this study. The rubric developed in this study will help instructors assess students’ representational competence in a course-based research experience (CURE) and could also be adapted to assess students’ understanding of other scientific concepts and misconceptions. The rubric will enable teachers to collect evidence of students’ understanding so they can make their science teaching more authentic, support students’ learning of the core scientific concepts, and provide an opportunity for teachers to modify their instruction accordingly across science disciplines, benefitting science teaching and learning overall.

With a vision to transform and advance undergraduate biology education, the American Association for the Advancement of Science (2015) published Vision and Change in Undergraduate Biology Education: Chronicling Change, Inspiring the Future, which had a clear focus on how we can improve assessment of students’ understanding by developing new instruments for doing so. Specific recommendations included aligning assessments for the core concepts taught in the classroom, integrating multiple forms of assessments to evaluate students’ learning, using assessments to document student learning, and improving teaching and enhancing the learning environment using the assessment data. Additional recommendations from the National Research Council (2003) and other studies (Atkin & Black, 2003; Goubeaud, 2010; Richmond et al., 2008) on evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics (STEM) suggest using varied forms of assessments to both improve teaching and provide evidence of students’ learning. Scientists often use evidence for decision-making and developing new knowledge, and using a similar practice in their teaching would reap great benefits (DiCarlo, 2006; Tanner & Allen, 2004). Grades, a form of summative evidence, are still the most common form of evidence used in teaching to inform students about their progress at the end of the school year. Formative evidence, obtained through frequent assessments throughout a course, should be more widely used to assess students’ learning and misconceptions.

Role of assessments
Assessments help teachers ask questions about their own teaching, such as “How well are the students learning?” and “How should the teaching approaches be modified to better facilitate students’ learning and maximize student learning gains?” As illustrated in Figure 1, classroom assessment is an iterative approach in which teachers analyze students’ current understanding by evaluating the classroom assessment data, make instructional choices informed by their initial assessment, ask additional questions about students’ learning, and collect another round of assessment data to address the questions. Classroom assessments are thus critical in optimization of the process of teaching and learning and serve as a bridge between the instructional and learning outcomes. In addition to adopting assessment models to guide their instruction, teachers should provide students with opportunities to self-monitor their learning, such as giving them a simplified scoring
rubric. This monitoring is crucial for the development of students’ metacognitive awareness and facilitates their academic success (Kim & Ryu, 2013; National Research Council, 2001). The shift in the new assessment culture toward assessing more higher-order thinking processes and competencies of students rather than simple factual knowledge further demands the use of well-defined scoring rubrics (Dochy et al., 2006; Jonsson & Svingby, 2007). The use of scoring rubrics in formative assessments is gaining popularity among teachers across fields (Panadero & Jonsson, 2013). Additionally, teachers’ use of rubrics in grading students’ assignments is often perceived by students to be more fair and satisfactory (Powell, 2001; Reddy & Andrade, 2010).

Several researchers, however, have identified mismatches between the intentions of scoring rubric developers and the reality of what the scoring rubrics measure in practice (Baxter & Glaser, 1998). It is essential to develop assessment models and scoring rubrics based on established learning models or theories (National Research Council, 2001) to ensure they function as needed. The purpose of our study was to develop a rubric to assess students’ understanding of scientific concepts and their ability to represent the concepts through visual illustrations.

**Defining a scoring rubric**

A scoring rubric is defined as an evaluation scheme developed by teachers and evaluators to assess students’ efforts (Brookhart, 1999; Moskal, 2000). A rubric serves as a means of assessing students’ work because it provides guidelines of evaluation and scoring and informs an overall grading logic by defining the scoring levels (Zane, 2009). Scoring rubrics are most often used to assess students’ writing competency, but they have also been adapted to assess other competencies, including representational competence, argumentation ability, meta-representational competence, and mathematical competence. Scoring rubrics can also be successfully adapted to evaluate students’ projects, group activities, and poster and oral presentations across different fields of learning. The two main identified benefits of a scoring rubric are that it (i) helps instructors evaluate the extent to which students have met a particular criterion, and (ii) provides an opportunity for students to evaluate their own learning (Moskal, 2000). There are two primary types of scoring rubric: (i) the “holistic” rubric, used by instructors to gauge the overall performance of students without specifically judging any part of the performance, and (ii) the “analytic” rubric, also known as a diagnostic rubric, used to score individual parts of students’ performance using a specific criterion and then determine a total score by adding up the individual elements (Mertler, 2001; Moskal, 2000; Nitko, 2001). The purpose of the evaluation determines the type of rubric used by the instructors. Rubrics also inform instructors about their teaching and help them make new and modified instructional choices (Andrade, 2005; Schneider, 2006) based on evidence. In our study, we provide an example of an analytic scoring rubric that instructors can use to both assess multiple aspects of students’ performance and inform their own teaching.

**Conceptual framework**

An analytic rubric has the ability to diagnose individual components of students’ learning and is often based on principles guided by the constructivist theory of learning (National Research Council, 2001; Taylor, 1997). These three principles guided the development of our rubric (National Research Council, 2001):

1. **Performance tasks should use a measurement model that sup-**
ports learning and measurement. Analytic scoring rubrics have the potential to support students’ learning and measure their performances, as students can use them to analyze their current state of understanding and self-reflect on their learning. Moreover, they are designed in a way to help the instructors provide informative feedback to students on their learning as well as identify ways to improve their teaching to obtain the desired outcomes.

2. Rubrics should measure all applicable facets of competence. This principle provides the foundation for the development of an analytic rubric that accounts for all aspects of multifaceted learning. For example, the rubric should not only measure the overall performance or content knowledge but also be able to identify the underlying skills and competencies that are crucial to overall success.

3. Rubrics should contain absolute measures of success. An analytic scoring rubric should be composed of both quantitative and qualitative measures of assessment for absolute measure of students’ learning. This principle is also important for highlighting the clear expectations for a student to be successful and efficiently differentiating among different scoring levels.

The rubric developed in this study is informed by these three principles and was designed for the purpose of supporting students’ learning.

Example from a research-based undergraduate course on phages

Visual representation, specifically within a STEM domain, is a mode of communication of ideas and thoughts that is widely used in teaching and learning (Quillin & Thomas, 2015). Communication in science demands extensive use of visual representations; one critical aspect of being a scientist is the ability to communicate with representations or the development of representational competence (Kozma & Russell, 2005). For students to acquire representational competence, they need guidance from their instructors. To provide students with proper scaffolding, teachers need to be able to assess students’ current representational competence (Tippett, 2016), and students need to know what is expected of them regarding scientific practices in the classroom and beyond. Thus, instructors need specific tools to assess how students reason with representations, integrate information, and develop argumentation. To address this need, our study aims to develop a rubric for assessing students’ understanding of scientific concepts using representations in a biology course involving research experience. As noted, classroom assessments can be used to improve both teaching and learning. We describe in this article how one can develop an assessment rubric for their own class and also how to identify students’ misconceptions as part of that process as a way to understand where students are lacking in knowledge or any misconceptions they may have. Additionally, the rubric can be adapted to assess students’ understanding of any scientific concept across disciplines.

Rubric development

Course context

Course content was designed and implemented according to the Science Education Alliance Phage Hunters Advancing Genomics and Evolutionary Science (SEA-PHAGES) project supported by the Howard Hughes Medical Institute (HHMI). The project provides undergraduates with a platform to experience the process of scientific discovery as part of this course by discovering new bacteriophages. The entire project is distributed across two semesters in a two-course series, the first being the wet lab, followed by the second, the bioinformatics lab. In the wet lab, students isolate and characterize bacteriophages from the environment, purify the phages using aseptic techniques, and visualize and name their phage. At the end of the wet lab semester, purified phages are archived in a public HHMI database, and the genomes are submitted for sequencing at a facility. In the bioinformatics lab, students work to annotate the phage genomes using different bioinformatics software, such as Phamerator and DNA Master. The entire course is designed to provide students with an authentic science research-based experience.

Data collection

For the development of the assessment rubric, data were collected from a research-based undergraduate course at a midwestern university during the semesters spanning fall 2016 through spring 2018, which involved 145 students (n = 83 in 2016–17; n = 62 in 2017–18). At the end of each school year, in spring 2017 and spring 2018, students were asked to respond to the following two questions to evaluate their understanding of genomes at the end of the course: (i) Consider what you have learned about genomes. Use the space below to draw a visual representation that demonstrates how you understand and visualize a genome. (ii) Write a
Rubric design

The rubric is designed to evaluate students’ understanding of a genome after participation in this course-based research experience. For the development of the rubric, we carefully examined the objectives of the task with the course instructor and identified the expected student attributes (Figure 2). This step is crucial for the development of a rubric that measures what it is designed to measure (Airasian, 2001; Nitko, 2001).

Following data collection, we used a deductive approach to coding (Patton, 2002) to code students’ representations and associated explanations. To develop the rubric, scoring categories that included five different dimensions were adopted from Niemi (1996). The original scoring dimensions for the rubric were used to analyze mathematical representations, so we adapted the rubric to analyze students’ representations and/or responses in a science context. The dimensions include the following:

1. **General content quality** measures students’ knowledge about a topic and is rated on a scale from 1 to 5 (global rating: 1 = no knowledge, 5 = highest level of understanding).

2. **Concepts and principles** records the number of general and abstract ideas incorporated in students’ responses.

3. **Facts and procedures** records the number of facts or procedures students mention in their responses.

4. **Misconceptions/Errors** is rated on a scale of 1 to 3 (1 = one or more serious misconceptions, 2 = one or more factual or procedural errors, 3 = no errors or misconceptions).

5. **Integration and argumentation** was used to understand how well a student developed a conceptual argument by integrating concepts, principles, facts, and procedures. It is rated on a scale of 1 to 5 (global rating: 1 = no integration, 5 = highest level of integration).

We used NVivo 12 software to analyze students’ responses for the development of the rubric. To ensure trustworthiness, we used interrater reliability throughout the analysis (Saldaña, 2013). For example, although part of the coding was done by a single researcher, another half of the students’ responses were coded by three researchers, and all raters shared their reasoning process when there were discrepancies until they could reach 100% consensus.

An example of scoring rubric is displayed in Figure 3. The next section provides a brief description of each scoring category, with examples of students’ responses.

**Description of rubric**

**General content quality**

In this category, we identified students’ overall understanding about the genome as represented in both their drawing and their accompanying explanation. We gave students a rating of 1 (having “no knowledge”) if they clearly mentioned that they did not have any substantive idea about genomes. Students whose ideas about genomes were mostly incorrect were given a rating of 2, or labelled as having “low level of understanding.” Similarly, students who had some idea about genomes but whose ideas were incomplete or incorrect were given a rating of 3, or labelled as having “moderate level of understanding.” If students showed a good but not very descriptive or detailed understanding of genomes, they were given a rating of 4, or labelled as having “high level of understanding.” Finally, students whose representations and associated explanations showed an in-depth understanding of genomes were given a rating of 5, or labelled as having the “highest level of understanding.”

**Concepts and principles**

In this scoring category, we identified the different concepts and principles in students’ responses; some sample responses included “All genes or
### FIGURE 3

**Example of a rubric.**

<table>
<thead>
<tr>
<th>Student responses</th>
<th>General content quality</th>
<th>Concepts and principles</th>
<th>Facts and procedures</th>
<th>Mis-conceptios/errors</th>
<th>Integration and argumentation</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1" alt="genome" /> “Genome is a pair of chromosomes, one of which inherit from mother and one from father”</td>
<td>2 = Low level of understanding</td>
<td>Wrong conceptual understanding</td>
<td>Serious misconception</td>
<td>1 = No</td>
<td></td>
</tr>
<tr>
<td><img src="image2" alt="DNA" /> “DNA makes up a genome”</td>
<td>3 = Moderate level of understanding</td>
<td>A. Genomes contain DNA</td>
<td>No</td>
<td>1 = No</td>
<td></td>
</tr>
<tr>
<td><img src="image3" alt="genome" /> “Genome is the entire DNA sequence of an organism including all genes and their location”</td>
<td>4 = High level of understanding</td>
<td>A. Genomes are all genes B. Genomes contain DNA</td>
<td>No</td>
<td>2 = Low</td>
<td></td>
</tr>
<tr>
<td><img src="image4" alt="genome" /> “A genome is composed of specific DNA sequences (genes) that code for proteins via transcription and translation. There are introns and exons in a genome. Introns do not code for proteins and are cut out of mature mRNA while exons are kept and translated into proteins. The promoter region contains genes that either promote or slow the production of the genes past the 1+ arrow”</td>
<td>5 = Highest level of understanding</td>
<td>A. Genomes contain DNA B. Genes code for proteins C. Contains introns and exons each with different functions</td>
<td>A. Introns do not code for proteins and are cut out. B. Exons are kept and translated into proteins C. The promoter region controls the expression of genes.</td>
<td>No</td>
<td>4 = High</td>
</tr>
</tbody>
</table>
complete inventory of our DNA,” “Connects all DNA,” “Genomes are genetic sequences,” and “Genomes contain DNA, which has genes made of protein.”

Facts and procedures
In this category, we identified the different facts and procedural information provided by students, such as “Boxes along the number line represents the genes”; “Introns do not code for proteins and are cut out of mature mRNA, while exons are kept and translated into proteins”; and “Protein contains a start and stop site.”

Misconceptions/Errors
In this category, we screened students’ responses for any misconceptions and errors. If a misconception was identified, it was categorized as a factual or procedural error—for example, “Genetic sequence on a larger scale” and “Genome is found in DNA”—or as a serious misconception, such as “Genomes are chromosomes containing DNA in eukaryotes and just DNA in prokaryotes.”

Integration and argumentation
In this category, students’ responses were categorized as having no, low, moderate, high, or the highest level of integration and argumentation of thoughts and ideas for a detailed description of a genome. For example, student responses such as “Genome has sections of DNA that forms genes. The boxes along the number line represents the genes” were labelled as low integration/argumentation. In contrast, other student responses—such as “A genome is all of genomes that make our chromosome. It is a complete set of data basically explaining us. A genome is a complete inventory of our DNA”—clearly show how multiple ideas were combined to develop a coherent idea about a genome and its function.

Appropriateness of instrument for a rubric
An important step toward the development of a scoring rubric is to accumulate evidence to support the appropriateness of the instrument (Moskal & Leydens, 2000; Mertler, 2001). Limited content, construct, and face validities were obtained by reviewing the rubric with the instructor of the course to ensure the rubric met the intended goals. We also worked with the instructor to confirm that the rubric measures what it is designed to measure and incorporates the necessary knowledge and skills before we revised it accordingly. An important step toward obtaining these validities was to identify the learning objectives and the intended attributes we expected to observe in students. We mention it as limited validity, however, because the rubric needs to be further reviewed for content, construct, and also criterion validity—that is, to understand if the rubric’s assessment of students’ performance on a given task can be generalized to other relevant activities. Furthermore, we made every attempt to improve the interrater and intrarater reliability of the rubric, which is key to the development of a valid assessment tool (Moskal & Leydens, 2000). We made sure we had well-defined scoring categories and clear differences between the categories to avoid any confusion. As mentioned in an earlier section, interrater reliability obtained by the three researchers during the coding process contributed to the overall reliability of the rubric. Further enhancement of the reliability rubric, however, can be achieved if other teachers can assess a sample set of responses using the rubric and provide their feedback (Moskal & Leydens, 2000).

It is important to mention that this rubric could be adopted and modified to assess students’ understanding of any science content. Additionally, the scoring categories could be used independently (e.g., if an instructor needs to assess misconceptions and errors or integration and argumentation, they could use only the relevant scoring categories). The rubric can also be used as a pre- and post-assessment tool to assess changes in students’ understanding following an intervention.

Discussion
In an effort to make classroom curricula more student centered and engage students in authentic inquiry, biology educators and researchers are emphasizing the introduction of a scientific perspective in the curricula that relates to the scientific research world (American Association for the Advancement of Science, 2015; Handelsman et al., 2004; Labov et al., 2010). Such scientific teaching requires the collection of evidence to revise current teaching practices so instructors can help students learn core scientific concepts and develop competence of the same (American Association for the Advancement of Science, 2015). The rubric described in this article will help teachers identify gaps between their teaching and students’ learning and modify their teaching approaches accordingly (Cotner et al., 2008). The rubric created in our study will help students develop representational competence and learn core scientific concepts, as the rubric helps instructors learn how students differ from experts and identify the potential misconceptions that hinder students’ understanding of core concepts. The rubric could be adapted for assessing across var-
ried content areas, so we hope it will be useful to instructors across disciplines for providing valuable formative feedback to students. An analytic scoring rubric is also crucial for students to use to assess their own learning and become more responsible for their own learning (National Research Council, 2001). Students perceive rubrics as a useful resource for planning ways to approach an assignment, check their work for quality, and minimize overall anxiousness (Andrade, 2005; Bolton, 2006; Reddy & Andrade, 2010).

Further investigation, however, is necessary to unravel the application of the rubric in other content areas and assess students’ understanding of concepts in other science disciplines. Additionally, just handing the rubric to students cannot reap benefits unless they are taught how to properly use the rubric for self-assessment. Additional research is thus needed on how students can use the rubric for maximum benefit, and future studies on the role of a rubric in altering teachers’ instruction will be helpful. To conclude, rubrics have been identified as having great potential for informing instructional guidance and assessing students’ performance, and we believe our rubric can help achieve these goals.

References


**Chandrani Mishra** (chandranimishra@gmail.com) is a postdoctoral researcher in the School of Agricultural and Biological Engineering, **Loran Carleton Parker** (carleton@purdue.edu) is the associate director and principal scholar at the Evaluation and Learning Research Center, and **Kari L. Clase** (kclase@purdue.edu) is a professor in the School of Agricultural and Biological Engineering and director of the Biotechnology Innovation and Regulatory Science Center, all at Purdue University in West Lafayette, Indiana.