

RESEARCH

Open Access



Rubrics to assess critical thinking and information processing in undergraduate STEM courses

Gil Reynders^{1,2}, Juliette Lantz³, Suzanne M. Ruder², Courtney L. Stanford⁴ and Renée S. Cole^{1*} 

Abstract

Background: Process skills such as critical thinking and information processing are commonly stated outcomes for STEM undergraduate degree programs, but instructors often do not explicitly assess these skills in their courses. Students are more likely to develop these crucial skills if there is constructive alignment between an instructor's intended learning outcomes, the tasks that the instructor and students perform, and the assessment tools that the instructor uses. Rubrics for each process skill can enhance this alignment by creating a shared understanding of process skills between instructors and students. Rubrics can also enable instructors to reflect on their teaching practices with regard to developing their students' process skills and facilitating feedback to students to identify areas for improvement.

Results: Here, we provide rubrics that can be used to assess critical thinking and information processing in STEM undergraduate classrooms and to provide students with formative feedback. As part of the Enhancing Learning by Improving Process Skills in STEM (ELIPSS) Project, rubrics were developed to assess these two skills in STEM undergraduate students' written work. The rubrics were implemented in multiple STEM disciplines, class sizes, course levels, and institution types to ensure they were practical for everyday classroom use. Instructors reported via surveys that the rubrics supported assessment of students' written work in multiple STEM learning environments. Graduate teaching assistants also indicated that they could effectively use the rubrics to assess student work and that the rubrics clarified the instructor's expectations for how they should assess students. Students reported that they understood the content of the rubrics and could use the feedback provided by the rubric to change their future performance.

Conclusion: The ELIPSS rubrics allowed instructors to explicitly assess the critical thinking and information processing skills that they wanted their students to develop in their courses. The instructors were able to clarify their expectations for both their teaching assistants and students and provide consistent feedback to students about their performance. Supporting the adoption of active-learning pedagogies should also include changes to assessment strategies to measure the skills that are developed as students engage in more meaningful learning experiences. Tools such as the ELIPSS rubrics provide a resource for instructors to better align assessments with intended learning outcomes.

Keywords: Constructive alignment, Self-regulated learning, Feedback, Rubrics, Process skills, Professional skills, Critical thinking, Information processing

* Correspondence: renee-cole@uiowa.edu

¹Department of Chemistry, University of Iowa, W331 Chemistry Building,
Iowa City, Iowa 52242, USA

Full list of author information is available at the end of the article

Introduction

Why assess process skills?

Process skills, also known as professional skills (ABET Engineering Accreditation Commission, 2012), transferable skills (Danczak et al., 2017), or cognitive competencies (National Research Council, 2012), are commonly cited as critical for students to develop during their undergraduate education (ABET Engineering Accreditation Commission, 2012; American Chemical Society Committee on Professional Training, 2015; National Research Council, 2012; Singer et al., 2012; The Royal Society, 2014). Process skills such as problem-solving, critical thinking, information processing, and communication are widely applicable to many academic disciplines and careers, and they are receiving increased attention in undergraduate curricula (ABET Engineering Accreditation Commission, 2012; American Chemical Society Committee on Professional Training, 2015) and workplace hiring decisions (Gray & Koncz, 2018; Pearl et al., 2019). Recent reports from multiple countries (Brewer & Smith, 2011; National Research Council, 2012; Singer et al., 2012; The Royal Society, 2014) indicate that these skills are emphasized in multiple undergraduate academic disciplines, and annual polls of about 200 hiring managers indicate that employers may place more importance on these skills than in applicants' content knowledge when making hiring decisions (Deloitte Access Economics, 2014; Gray & Koncz, 2018). The assessment of process skills can provide a benchmark for achievement at the end of an undergraduate program and act as an indicator of student readiness to enter the workforce. Assessing these skills may also enable instructors and researchers to more fully understand the impact of active learning pedagogies on students.

A recent meta-analysis of 225 studies by Freeman et al. (2014) showed that students in active learning environments may achieve higher content learning gains than students in traditional lectures in multiple STEM fields when comparing scores on equivalent examinations. Active learning environments can have many different attributes, but they are commonly characterized by students "physically manipulating objects, producing new ideas, and discussing ideas with others" (Rau et al., 2017) in contrast to students sitting and listening to a lecture. Examples of active learning pedagogies include POGIL (Process Oriented Guided Inquiry Learning) (Moog & Spencer, 2008; Simonson, 2019) and PLTL (Peer-led Team Learning) (Gafney & Varma-Nelson, 2008; Gosser et al., 2001) in which students work in groups to complete activities with varying levels of guidance from an instructor. Despite the clear content learning gains that students can achieve from active learning environments (Freeman et al., 2014), the non-content-gains (including improvements in process skills) in these

learning environments have not been explored to a significant degree. Active learning pedagogies such as POGIL and PLTL place an emphasis on students developing non-content skills in addition to content learning gains, but typically only the content learning is assessed on quizzes and exams, and process skills are not often explicitly assessed (National Research Council, 2012). In order to fully understand the effects of active learning pedagogies on all aspects of an undergraduate course, evidence-based tools must be used to assess students' process skill development. The goal of this work was to develop resources that could enable instructors to explicitly assess process skills in STEM undergraduate classrooms in order to provide feedback to themselves and their students about the students' process skills development.

Theoretical frameworks

The incorporation of these rubrics and other currently available tools for use in STEM undergraduate classrooms can be viewed through the lenses of constructive alignment (Biggs, 1996) and self-regulated learning (Zimmerman, 2002). The theory of constructivism posits that students learn by constructing their own understanding of knowledge rather than acquiring the meaning from their instructor (Bodner, 1986), and constructive alignment extends the constructivist model to consider how the alignment between a course's intended learning outcomes, tasks, and assessments affects the knowledge and skills that students develop (Biggs, 2003). Students are more likely to develop the intended knowledge and skills if there is alignment between the instructor's intended learning outcomes that are stated at the beginning of a course, the tasks that the instructor and students perform, and the assessment strategies that the instructor uses (Biggs, 1996, 2003, 2014). The nature of the tasks and assessments indicates what the instructor values and where students should focus their effort when studying. According to Biggs (2003) and Ramsden (1997), students see assessments as defining what they should learn, and a misalignment between the outcomes, tasks, and assessments may hinder students from achieving the intended learning outcomes. In the case of this work, the intended outcomes are improved process skills. In addition to aligning the components of a course, it is also critical that students receive feedback on their performance in order to improve their skills. Zimmerman's theory of self-regulated learning (Zimmerman, 2002) provides a rationale for tailoring assessments to provide feedback to both students and instructors.

Zimmerman's theory of self-regulated learning defines three phases of learning: forethought/planning, performance, and self-reflection. According to Zimmerman, individuals ideally should progress through these three

phases in a cycle: they plan a task, perform the task, and reflect on their performance, then they restart the cycle on a new task. If a student is unable to adequately progress through the phases of self-regulated learning on their own, then feedback provided by an instructor may enable the students to do so (Butler & Winne, 1995). Thus, one of our criteria when creating rubrics to assess process skills was to make the rubrics suitable for faculty members to use to provide feedback to their students. Additionally, instructors can use the results from assessments to give themselves feedback regarding their students' learning in order to regulate their teaching. This theory is called self-regulated learning because the goal is for learners to ultimately reflect on their actions to find ways to improve. We assert that, ideally, both students and instructors should be "learners" and use assessment data to reflect on their actions, although with different aims. Students need consistent feedback from an instructor and/or self-assessment throughout a course to provide a benchmark for their current performance and identify what they can do to improve their process skills (Black & Wiliam, 1998; Butler & Winne, 1995; Hattie & Gan, 2011; Nicol & Macfarlane-Dick, 2006). Instructors need feedback on the extent to which their efforts are achieving their intended goals in order to improve their instruction and better facilitate the development of process skills through course experiences.

In accordance with the aforementioned theoretical frameworks, tools used to assess undergraduate STEM student process skills should be tailored to fit the outcomes that are expected for undergraduate students and be able to provide formative assessment and feedback to both students and faculty about the students' skills. These tools should also be designed for everyday classroom use to enable students to regularly self-assess and faculty to provide consistent feedback throughout a semester. Additionally, it is desirable for assessment tools to be broadly generalizable to measure process skills in multiple STEM disciplines and institutions in order to increase the rubrics' impact on student learning. Current tools exist to assess these process skills, but they each lack at least one of the desired characteristics for providing regular feedback to STEM students.

Current tools to assess process skills

Current tests available to assess critical thinking include the Critical Thinking Assessment Test (CAT) (Stein & Haynes, 2011), California Critical Thinking Skills Test (Facione, 1990a, 1990b), and Watson Glaser Critical Thinking Appraisal (Watson & Glaser, 1964). These commercially available, multiple-choice tests are not designed to provide regular, formative feedback throughout a course and have not been implemented for this purpose. Instead, they are designed to provide summative

feedback with a focus on assessing this skill at a programmatic or university level rather than for use in the classroom to provide formative feedback to students. Rather than using tests to assess process skills, rubrics could be used instead. Rubrics are effective assessment tools because they can be quick and easy to use, they provide feedback to both students and instructors, and they can evaluate individual aspects of a skill to give more specific feedback (Brookhart & Chen, 2014; Smit & Birri, 2014). Rubrics for assessing critical thinking are available, but they have not been used to provide feedback to undergraduate STEM students nor were they designed to do so (Association of American Colleges and Universities, 2019; Saxton et al., 2012). The Critical Thinking Analytic Rubric is designed specifically to assess K-12 students to enhance college readiness and has not been broadly tested in collegiate STEM courses (Saxton et al., 2012). The critical thinking rubric developed by the Association of American Colleges and Universities (AAC&U) as part its Valid Assessment of Learning in Undergraduate Education (VALUE) Institute and Liberal Education and America's Promise (LEAP) initiative (Association of American Colleges and Universities, 2019) is intended for programmatic assessment rather than specifically giving feedback to students throughout a course. As with tests for assessing critical thinking, current rubrics to assess critical thinking are not designed to act as formative assessments and give feedback to STEM faculty and undergraduates at the course or task level. Another issue with the assessment of critical thinking is the degree to which the construct is measurable. A National Research Council report (National Research Council, 2011) has suggested that there is little evidence of a consistent, measurable definition for critical thinking and that it may not be different from one's general cognitive ability. Despite this issue, we have found that critical thinking is consistently listed as a programmatic outcome in STEM disciplines (American Chemical Society Committee on Professional Training, 2015; The Royal Society, 2014), so we argue that it is necessary to support instructors as they attempt to assess this skill.

Current methods for evaluating students' information processing include discipline-specific tools such as a rubric to assess physics students' use of graphs and equations to solve work-energy problems (Nguyen et al., 2010) and assessments of organic chemistry students' ability to "[manipulate] and [translate] between various representational forms" including 2D and 3D representations of chemical structures (Kumi et al., 2013). Although these assessment tools can be effectively used for their intended context, they were not designed for use in a wide range of STEM disciplines or for a variety of tasks.

Despite the many tools that exist to measure process skills, none has been designed and tested to facilitate frequent, formative feedback to STEM undergraduate students and faculty throughout a semester. The rubrics described here have been designed by the Enhancing Learning by Improving Process Skills in STEM (ELIPSS) Project (Cole et al., 2016) to assess undergraduate STEM students' process skills and to facilitate feedback at the classroom level with the potential to track growth throughout a semester or degree program. The rubrics described here are designed to assess critical thinking and information processing in student written work. Rubrics were chosen as the format for our process skill assessment tools because the highest level of each category in rubrics can serve as an explicit learning outcome that the student is expected to achieve (Panadero & Jonsson, 2013). Rubrics that are generalizable to multiple disciplines and institutions can enable the assessment of student learning outcomes and active learning pedagogies throughout a program of study and provide useful tools for a greater number of potential users.

Research questions

This work sought to answer the following research questions for each rubric:

1. Does the rubric adequately measure relevant aspects of the skill?
2. How well can the rubrics provide feedback to instructors and students?
3. Can multiple raters use the rubrics to give consistent scores?

Methods

This work received Institutional Review Board approval prior to any data collection involving human subjects. The sources of data used to construct the process skill rubrics and answer these research questions were (1) peer-reviewed literature on how each skill is defined, (2) feedback from content experts in multiple STEM disciplines via surveys and in-person, group discussions regarding the appropriateness of the rubrics for each discipline, (3) interviews with students whose work was scored with the rubrics and teaching assistants who scored the student work, and (4) results of applying the rubrics to samples of student work.

Defining the scope of the rubrics

The rubrics described here and the other rubrics in development by the ELIPSS Project are intended to measure process skills, which are desired learning outcomes identified by the STEM community in recent reports (National Research Council, 2012; Singer et al., 2012). In order to measure these skills in multiple STEM

disciplines, operationalized definitions of each skill were needed. These definitions specify which aspects of student work (operations) would be considered evidence for the student using that skill and establish a shared understanding of each skill by members of each STEM discipline. The starting point for this work was the process skill definitions developed as part of the POGIL project (Cole et al., 2019a). The POGIL community includes instructors from a variety of disciplines and institutions and represented the intended audience for the rubrics: faculty who value process skills and want to more explicitly assess them. The process skills discussed in this work were defined as follows:

- Critical thinking is analyzing, evaluating, or synthesizing relevant information to form an argument or reach a conclusion supported with evidence.
- Information processing is evaluating, interpreting, and manipulating or transforming information.

Examples of critical thinking include the tasks that students are asked to perform in a laboratory course. When students are asked to analyze the data they collected, combine data from different sources, and generate arguments or conclusions about their data, we see this as critical thinking. However, when students simply follow the so-called "cookbook" laboratory instructions that require them to confirm pre-determined conclusions, we do not think students are engaging in critical thinking. One example of information processing is when organic chemistry students are required to redraw molecules in different formats. The students must evaluate and interpret various pieces of one representation, and then they recreate the molecule in another representation. However, if students are asked to simply memorize facts or algorithms to solve problems, we do not see this as information processing.

Iterative rubric development

The development process was the same for the information processing rubric and the critical thinking rubric. After defining the scope of the rubric, an initial version was drafted based upon the definition of the target process skill and how each aspect of the skill is defined in the literature. A more detailed discussion of the literature that informed each rubric category is included in the "Results and Discussion" section. This initial version then underwent iterative testing in which the rubric was reviewed by researchers, practitioners, and students. The rubric was first evaluated by the authors and a group of eight faculty from multiple STEM disciplines who made up the ELIPSS Project's primary collaborative team (PCT). The PCT was a group of faculty members with

experience in discipline-based education research who employ active-learning pedagogies in their classrooms. This initial round of evaluation was intended to ensure that the rubric measured relevant aspects of the skill and was appropriate for each PCT member's discipline. This evaluation determined how well the rubrics were aligned with each instructor's understanding of the process skill including both in-person and email discussions that continued until the group came to consensus that each rubric category could be applied to student work in courses within their disciplines. There has been an ongoing debate regarding the role of disciplinary knowledge in critical thinking and the extent to which critical thinking is subject-specific (Davies, 2013; Ennis, 1990). This work focuses on the creation of rubrics to measure process skills in different domains, but we have not performed cross-discipline comparisons. This initial round of review was also intended to ensure that the rubrics were ready for classroom testing by instructors in each discipline. Next, each rubric was tested over three semesters in multiple classroom environments, illustrated in Table 1. The rubrics were applied to student work chosen by each PCT member. The PCT members chose the student work based on their views of how the assignments required students to engage in process skills and show evidence of those skills. The information processing and critical thinking rubrics shown in this work were each tested in at least three disciplines, course levels, and institutions.

After each semester, the feedback was collected from the faculty testing the rubric, and further changes to the rubric were made. Feedback was collected in the form of survey responses along with in-person group discussions at annual project meetings. After the first iteration of completing the survey, the PCT members met with the authors to discuss how they were interpreting each survey question. This meeting helped ensure that the surveys were gathering valid data regarding how well the rubrics were measuring the desired process skill. Questions in the survey such as "What aspects of the student work provided evidence for the indicated process skill?"

and "Are there edits to the rubric/descriptors that would improve your ability to assess the process skill?" allowed the authors to determine how well the rubric scores were matching the student work and identify necessary changes to the rubric. Further questions asked about the nature and timing of the feedback given to students in order to address the question of how well the rubrics provide feedback to instructors and students. The survey questions are included in the [Supporting Information](#). The survey responses were analyzed qualitatively to determine themes related to each research question.

In addition to the surveys given to faculty rubric testers, twelve students were interviewed in fall 2016 and fall 2017. In the United States of America, the fall semester typically runs from August to December and is the first semester of the academic year. Each student participated in one interview which lasted about 30 min. These interviews were intended to gather further data to answer questions about how well the rubrics were measuring the identified process skills that students were using when they completed their assignments and to ensure that the information provided by the rubrics made sense to students. The protocol for these interviews is included in the [Supporting Information](#). In fall 2016, the students interviewed were enrolled in an organic chemistry laboratory course for non-majors at a large, research-intensive university in the United States. Thirty students agreed to have their work analyzed by the research team, and nine students were interviewed. However, the rubrics were not a component of the laboratory course grading. Instead, the first author assessed the students' reports for critical thinking and information processing, and then the students were provided electronic copies of their laboratory reports and scored rubrics in advance of the interview. The first author had recently been a graduate teaching assistant for the course and was familiar with the instructor's expectations for the laboratory reports. During the interview, the students were given time to review their reports and the completed rubrics, and then they were asked about how well they understood the content of the rubrics and how accurately each category score represented their work.

Table 1 Learning environments in which the rubrics were implemented

Disciplines	Course Levels	Institution types ^a	Class sizes ^b	Pedagogies
Biology/HealthSciences	Introductory, intermediate, advanced	RU, CU	M, L, XL	Case study, lecture, laboratory, peer instruction, POGIL, Other
Chemistry	Introductory, intermediate, advanced	RU, PUI	S, M, L, XL	Case study, lecture, PBL, peer instruction, PLTL, POGIL
Computer Science	Introductory	CU	S, M	POGIL
Engineering	Advanced	PUI	M	Case study, flipped, other
Mathematics	Introductory, advanced	PUI	S, M	Lecture, PBL, POGIL
Statistics	Introductory	PUI	M	Flipped

^aRU Research University, CU Comprehensive University, and PUI Primarily Undergraduate Institution. ^bFor class size, S ≤ 25 students, M = 26–50 students, L = 51–150 students, and XL ≥ 150 students

In fall 2017, students enrolled in a physical chemistry thermodynamics course for majors were interviewed. The physical chemistry course took place at the same university as the organic laboratory course, but there was no overlap between participants. Three students and two graduate teaching assistants (GTAs) were interviewed. The course included daily group work, and process skill assessment was an explicit part of the instructor's curriculum. At the end of each class period, students assessed their groups using portions of ELIPSS rubrics, including the two process skill rubrics included in this paper. About every 2 weeks, the GTAs assessed the student groups with a complete ELIPSS rubric for a particular skill, then gave the groups their scored rubrics with written comments. The students' individual homework problem sets were assessed once with rubrics for three skills: critical thinking, information processing, and problem-solving. The students received the scored rubric with written comments when the graded problem set was returned to them. In the last third of the semester, the students and GTAs were interviewed about how rubrics were implemented in the course, how well the rubric scores reflected the students' written work, and how the use of rubrics affected the teaching assistants' ability to assess the student skills. The protocols for these interviews are included in the [Supporting Information](#).

Gathering evidence for utility, validity, and reliability

The utility, validity, and reliability of the rubrics were measured throughout the development process. The utility is the degree to which the rubrics are perceived as practical to experts and practitioners in the field. Through multiple meetings, the PCT faculty determined that early drafts of the rubric seemed appropriate for use in their classrooms, which represented multiple STEM disciplines. Rubric utility was reexamined multiple times throughout the development process to ensure that the rubrics would remain practical for classroom use. Validity can be defined in multiple ways. For example, the Standards for Educational and Psychological Testing (Joint Committee on Standards for Educational Psychological Testing, 2014) defines validity as "the degree to which all the accumulated evidence supports the intended interpretation of test scores for the proposed use." For the purposes of this work, we drew on the ways in which two distinct types of validity were examined in the rubric literature: content validity and construct validity. Content validity is the degree to which the rubrics cover relevant aspects of each process skill (Moskal & Leydens, 2000). In this case, the process skill definition and a review of the literature determined which categories were included in each rubric. The literature review was finished once the data was saturated: when no more new aspects were found. Construct validity is the degree to which the levels

of each rubric category accurately reflect the process that students performed (Moskal & Leydens, 2000). Evidence of construct validity was gathered via the faculty surveys, teaching assistant interviews, and student interviews. In the student interviews, students were given one of their completed assignments and asked to explain how they completed the task. Students were then asked to explain how well each category applied to their work and if any changes were needed to the rubric to more accurately reflect their process. Due to logistical challenges, we were not able to obtain evidence for convergent validity, and this is further discussed in the "Limitations" section.

Adjacent agreement, also known as "interrater agreement within one," was chosen as the measure of interrater reliability due to its common use in rubric development projects (Jonsson & Svingby, 2007). The adjacent agreement is the percentage of cases in which two raters agree on a rating or are different by one level (i.e., they give adjacent ratings to the same work). Jonsson and Svingby (2007) found that most of the rubrics they reviewed had adjacent agreement scores of 90% or greater. However, they noted that the agreement threshold varied based on the number of possible levels of performance for each category in the rubric, with three and four being the most common numbers of levels. Since the rubrics discussed in this report have six levels (scores of zero through five) and are intended for low-stakes assessment and feedback, the goal of 80% adjacent agreement was selected. To calculate agreement for the critical thinking and information processing rubrics, two researchers discussed the scoring criteria for each rubric and then independently assessed the organic chemistry laboratory reports.

Results and discussion

The process skill rubrics to assess critical thinking and information processing in student written work were completed after multiple rounds of revision based on feedback from various sources. These sources include feedback from instructors who tested the rubrics in their classrooms, TAs who scored student work with the rubrics, and students who were assessed with the rubrics. The categories for each rubric will be discussed in terms of the evidence that the rubrics measure the relevant aspects of the skill and how they can be used to assess STEM undergraduate student work. Each category discussion will begin with a general explanation of the category followed by more specific examples from the organic chemistry laboratory course and physical chemistry lecture course to demonstrate how the rubrics can be used to assess student work.

Information processing rubric

The definition of information processing and the focus of the rubric presented here (Fig. 1) are distinct from

Information Processing		Evaluating, interpreting, and manipulating or transforming information.				
Category	0	1	2	3	4	5
Evaluating		Minimally determined the significance or relevance of information/data needed for the task		Partially determined the significance or relevance of information/data needed for the task		Completely determined the significance or relevance of information/data needed for the task
Interpreting		Inaccurately provided meaning to data, made inferences and predictions from data, or extracted patterns from data		Provided meaning to data, made inferences and predictions from data, or extracted patterns from data with some errors		Accurately provided meaning to data, made inferences and predictions from data, or extracted patterns from data
Manipulating or Transforming (Extent)		Minimally converted information/data from one form to another		Partially converted information/data from one form to another		Completely converted information/data from one form to another
Manipulating or Transforming (Accuracy)		Inaccurately converted information/data from one form to another		Converted information/data from one form to another with some errors		Accurately converted information/data from one form to another
Comments:						

Fig. 1 Rubric for assessing information processing

cognitive information processing as defined by the educational psychology literature (Driscoll, 2005). The rubric shown here is more aligned with the STEM education construct of representational competency (Daniel et al., 2018).

Evaluating

When solving a problem or completing a task, students must evaluate the provided information for relevance or importance to the task (Hanson, 2008; Swanson et al., 1990). All the information provided in a prompt (e.g., homework or exam questions) may not be relevant for addressing all parts of the prompt. Students should ideally show evidence of their evaluation process by identifying what information is present in the prompt/model, indicating what information is relevant or not relevant, and indicating why information is relevant. Responses with these characteristics would earn high rubric scores for this category. Although students may not explicitly state what information is necessary to address a task, the information they do use can act as indirect evidence of the degree to which they have evaluated all of the available information in the prompt. Evidence for students inaccurately evaluating information for relevance includes the inclusion of irrelevant information or the omission of relevant information in an analysis or in completing a task. When evaluating the organic chemistry laboratory reports, the focus for the evaluating category was the information students presented when identifying the chemical structure of their products. For students who received a high score, this information included their measured value for the

product's melting point, the literature (expected) value for the melting point, and the peaks in a nuclear magnetic resonance (NMR) spectrum. NMR spectroscopy is a commonly used technique in chemistry to obtain structural information about a compound. Lower scores were given if students omitted any of the necessary information or if they included unnecessary information. For example, if a student discussed their reaction yield when discussing the identity of their product, they would receive a low Evaluating score because the yield does not help them determine the identity of their product; the yield, in this case, would be unnecessary information. In the physical chemistry course, students often did not show evidence that they determined which information was relevant to answer the homework questions and thus earned low evaluating scores. These omissions will be further addressed in the "Interpreting" section.

Interpreting

In addition to evaluating, students must often interpret information using their prior knowledge to explain the meaning of something, make inferences, match data to predictions, and extract patterns from data (Hanson, 2008; Nakhleh, 1992; Schmidt et al., 1989; Swanson et al., 1990). Students earn high scores for this category if they assign correct meaning to labeled information (e.g., text, tables, graphs, diagrams), extract specific details from information, explain information in their own words, and determine patterns in information. For the organic chemistry laboratory reports, students received high scores if they accurately interpreted their measured values and NMR peaks. Almost every student obtained

melting point values that were different than what was expected due to measurement error or impurities in their products, so they needed to describe what types of impurities could cause such discrepancies. Also, each NMR spectrum contained one peak that corresponded to the solvent used to dissolve the students' product, so the students needed to use their prior knowledge of NMR spectroscopy to recognize that peak did not correspond to part of their product.

In physical chemistry, the graduate teaching assistant often gave students low scores for inaccurately explaining changes to chemical systems such as changes in pressure or entropy. The graduate teaching assistant who assessed the student work used the rubric to identify both the evaluating and interpreting categories as weaknesses in many of the students' homework submissions. However, the students often earned high scores for the manipulating and transforming categories, so the GTA was able to give students specific feedback on their areas for improvement while also highlighting their strengths.

Manipulating and transforming (extent and accuracy)

In addition to evaluating and interpreting information, students may be asked to manipulate and transform information from one form to another. These transformations should be complete and accurate (Kumi et al., 2013; Nguyen et al., 2010). Students may be required to construct a figure based on written information, or conversely, they may transform information in a figure into words or mathematical expressions. Two categories for manipulating and transforming (i.e., extent and accuracy) were included to allow instructors to give more specific feedback. It was often found that students would either transform little information but do so accurately, or transform much information and do so inaccurately; the two categories allowed for differentiated feedback to be provided. As stated above, the organic chemistry students were expected to transform their NMR spectral data into a table and provide a labeled structure of their final product. Students were given high scores if they converted all of the relevant peaks from their spectrum into the table format and were able to correctly match the peaks to the hydrogen atoms in their products. Students received lower scores if they were only able to convert the information for a few peaks or if they incorrectly matched the peaks to the hydrogen atoms.

Critical thinking rubric

Critical thinking can be broadly defined in different contexts, but we found that the categories included in the rubric (Fig. 2) represented commonly accepted aspects of critical thinking (Danczak et al., 2017) and suited the

needs of the faculty collaborators who tested the rubric in their classrooms.

Evaluating

When completing a task, students must evaluate the relevance of information that they will ultimately use to support a claim or conclusions (Miri et al., 2007; Zohar et al., 1994). An evaluating category is included in both critical thinking and information processing rubrics because evaluation is a key aspect of both skills. From our previous work developing a problem-solving rubric (manuscript in preparation) and our review of the literature for this work (Danczak et al., 2017; Lewis & Smith, 1993), the overlap was seen between information processing, critical thinking, and problem-solving. Additionally, while the Evaluating category in the information processing rubric assesses a student's ability to determine the importance of information to complete a task, the evaluating category in the critical thinking rubric places a heavier emphasis on using the information to support a conclusion or argument.

When scoring student work with the evaluating category, students receive high scores if they indicate what information is likely to be most relevant to the argument they need to make, determine the reliability of the source of their information, and determine the quality and accuracy of the information itself. The information used to assess this category can be indirect as with the Evaluating category in the information processing rubric. In the organic chemistry laboratory reports, students needed to make an argument about whether they successfully produced the desired product, so they needed to discuss which information was relevant to their claims about the product's identity and purity. Students received high scores for the evaluating category when they accurately determined that the melting point and nearly all peaks except the solvent peak in the NMR spectrum indicated the identity of their product. Students received lower scores for evaluating when they left out relevant information because this was seen as evidence that the student inaccurately evaluated the information's relevance in supporting their conclusion. They also received lower scores when they incorrectly stated that a high yield indicated a pure product. Students were given the opportunity to demonstrate their ability to evaluate the quality of information when discussing their melting point. Students sometimes struggled to obtain reliable melting point data due to their inexperience in the laboratory, so the rubric provided a way to assess the student's ability to critique their own data.

Analyzing

In tandem with evaluating information, students also need to analyze that same information to extract

Critical Thinking		Evaluating, analyzing, and/or synthesizing relevant information to form an argument or reach a conclusion supported with evidence.				
Category	0	1	2	3	4	5
Evaluating		Minimally determined the relevance and reliability of information that might be used to support a conclusion or argument		Partially determined the relevance and reliability of information that might be used to support a conclusion or argument		Extensively determined the relevance and reliability of information that might be used to support a conclusion or argument
Analyzing		Inaccurately interpreted information to determine meaning and to extract relevant evidence		Interpreted information to determine meaning and to extract relevant evidence with some errors		Accurately interpreted information to determine meaning and to extract relevant evidence
Synthesizing		Inaccurately connected or integrated information to support an argument or reach a conclusion		Connected or integrated information to support an argument or reach a conclusion with some errors		Accurately connected or integrated information to support an argument or reach a conclusion
Forming Arguments (Structure)		Made a claim and provided incomplete evidence to support it.		Made a claim and provided partial evidence to support it.		Made a claim and provided complete evidence to support it.
Forming Arguments (Validity)		The claim, evidence, and reasoning were minimally consistent with accepted disciplinary ideas and practices		The claim, evidence, and reasoning were partially consistent with accepted disciplinary ideas and practices		The claim, evidence, and reasoning were fully consistent with accepted disciplinary ideas and practices
Comments:						

Fig. 2 Rubric for assessing critical thinking

meaningful evidence to support their conclusions (Bailin, 2002; Lai, 2011; Miri et al., 2007). The analyzing category provides an assessment of a student's ability to discuss information and explore the possible meaning of that information, extract patterns from data/information that could be used as evidence for their claims, and summarize information that could be used as evidence. For example, in the organic chemistry laboratory reports, students needed to compare the information they obtained to the expected values for a product. Students received high scores for the analyzing category if they could extract meaningful structural information from the NMR spectrum and their two melting points (observed and expected) for each reaction step.

Synthesizing

Often, students are asked to synthesize or connect multiple pieces of information in order to draw a conclusion or make a claim (Huitt, 1998; Lai, 2011). Synthesizing involves identifying the relationships between different pieces of information or concepts, identifying ways that different pieces of information or concepts can be combined, and explaining how the newly synthesized information can be used to reach a conclusion and/or support an argument. While performing the organic chemistry laboratory experiments, students obtained multiple types of information such as the melting point and NMR spectrum in addition to other spectroscopic data such as an infrared (IR) spectrum. Students

received high scores for this category when they accurately synthesized these multiple data types by showing how the NMR and IR spectra could each reveal different parts of a molecule in order to determine the molecule's entire structure.

Forming arguments (structure and validity)

The final key aspect of critical thinking is forming a well-structured and valid argument (Facione, 1984; Glassner & Schwarz, 2007; Lai, 2011; Lewis & Smith, 1993). It was observed that students can earn high scores for evaluating, analyzing, and synthesizing, but still struggle to form arguments. This was particularly common in assessing problem sets in the physical chemistry course.

As with the manipulating and transforming categories in the information processing rubric, two forming arguments categories were included to allow instructors to give more specific feedback. Some students may be able to include all of the expected structural elements of their arguments but use faulty information or reasoning. Conversely, some students may be able to make scientifically valid claims but not necessarily support them with evidence. The two forming arguments categories are intended to accurately assess both of these scenarios. For the forming arguments (structure) category, students earn high scores if they explicitly state their claim or conclusion, list the evidence used to support the argument, and provide reasoning to link the evidence to their

claim/conclusion. Students who do not make a claim or who provide little evidence or reasoning receive lower scores.

For the forming arguments (validity) category, students earn high scores if their claim is accurate and their reasoning is logical and clearly supports the claim with provided evidence. Organic chemistry students earned high scores for the forms and supports arguments categories if they made explicit claims about the identity and purity of their product and provided complete and accurate evidence for their claim(s) such as the melting point values and positions of NMR peaks that correspond to their product. Additionally, the students provided evidence for the purity of their products by pointing to the presence or absence of peaks in their NMR spectrum that would match other potential side products. They also needed to provide logical reasoning for why the peaks indicated the presence or absence of a compound. As previously mentioned, the physical chemistry students received lower scores for the forming arguments categories than for the other aspects of critical thinking. These students were asked to make claims about the relationships between entropy and heat and then provide relevant evidence to justify these claims. Often, the students would make clearly articulated claims but would provide little evidence to support them. As with the information processing rubric, the critical thinking rubric allowed the GTAs to assess aspects of these skills independently and identify specific areas for student improvement.

Validity and reliability

The goal of this work was to create rubrics that can accurately assess student work (validity) and be consistently implemented by instructors or researchers within multiple STEM fields (reliability). The evidence for validity includes the alignment of the rubrics with literature-based descriptions of each skill, review of the rubrics by content experts from multiple STEM disciplines, interviews with undergraduate students whose work was scored using the rubrics, and interviews of the GTAs who scored the student work.

The definitions for each skill, along with multiple iterations of the rubrics, underwent review by STEM content experts. As noted earlier, the instructors who were testing the rubrics were given a survey at the end of each semester and were invited to offer suggested changes to the rubric to better help them assess their students. After multiple rubric revisions, survey responses from the instructors indicated that the rubrics accurately represented the breadth of each process skill as seen in each expert's content area and that each category could be used to measure multiple levels of student work. By the end of the rubrics' development, instructors were writing responses such as "N/A" or "no suggestions" to indicate that the rubrics did not need further changes.

Feedback from the faculty also indicated that the rubrics were measuring the intended constructs by the ways they responded to the survey item "What aspects of the student work provided evidence for the indicated process skill?" For example, one instructor noted that for information processing, she saw evidence of the manipulating and transforming categories when "students had to transform their written/mathematical relationships into an energy diagram." Another instructor elicited evidence of information processing during an in-class group quiz: "A question on the group quiz was written to illicit [sic] IP [information processing]. Students had to transform a structure into three new structures and then interpret/manipulate the structures to compare the pKa values [acidity] of the new structures." For this instructor, the structures written by the students revealed evidence of their information processing by showing what information they omitted in the new structures or inaccurately transformed. For critical thinking, an instructor assessed short research reports with the critical thinking rubric and "looked for [the students'] ability to use evidence to support their conclusions, to evaluate the literature studies, and to develop their own judgements by synthesizing the information." Another instructor used the critical thinking rubric to assess their students' abilities to choose an instrument to perform a chemical analysis. According to the instructor, the students provided evidence of their critical thinking because "in their papers, they needed to justify their choice of instrument. This justification required them to evaluate information and synthesize a new understanding for this specific chemical analysis."

Analysis of student work indicates multiple levels of achievement for each rubric category (illustrated in Fig. 3), although there may have been a ceiling effect for the evaluating and the manipulating and transforming (extent) categories in information processing for organic chemistry laboratory reports because many students earned the highest possible score (five) for those categories. However, other implementations of the ELIPSS rubrics (Reynders et al., 2019) have shown more variation in student scores for the two process skills.

To provide further evidence that the rubrics were measuring the intended skills, students in the physical chemistry course were interviewed about their thought processes and how well the rubric scores reflected the work they performed. During these interviews, students described how they used various aspects of information processing and critical thinking skills. The students first described how they used information processing during a problem set where they had to answer questions about a diagram of systolic and diastolic blood pressure. Students described how they evaluated and interpreted the graph to make statements such as "diastolic [pressure] is our y-intercept" and "volume is the independent

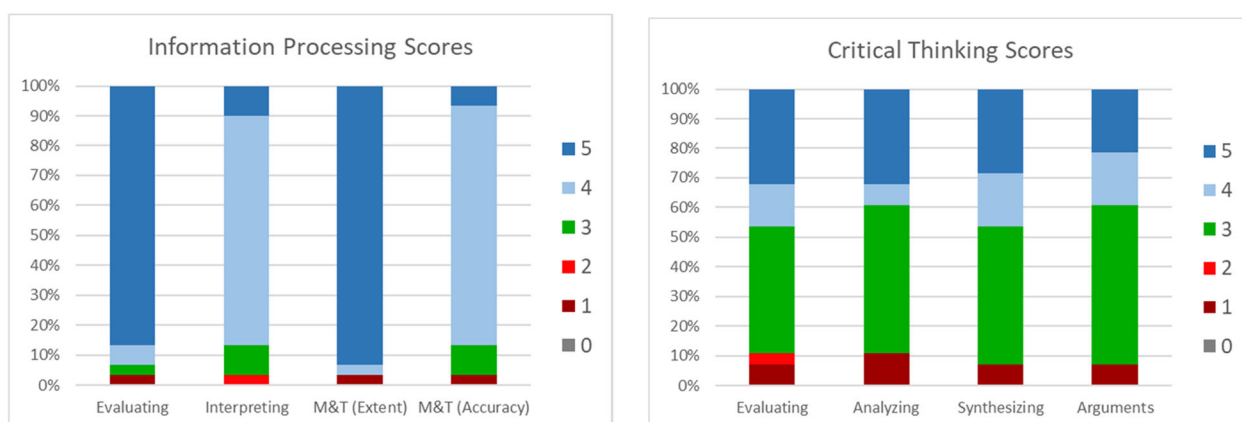


Fig. 3 Student rubric scores from an organic chemistry laboratory course. The two rubrics were used to evaluate different laboratory reports. Thirty students were assessed for information processing and 28 were assessed for critical thinking

variable.” The students then demonstrated their ability to transform information from one form to another, from a graph to a mathematic equation, by recognizing “it’s a linear relationship so I used Y equals $M X$ plus B ” and “integrated it cause it’s the change, the change in V [volume]. For critical thinking, students described their process on a different problem set. In this problem set, the students had to explain why the change of Helmholtz energy and the change in Gibbs free energy were equivalent under a certain given condition. Students first demonstrated how they evaluated the relevant information and analyzed what would and would not change in their system. One student said, “So to calculate the final pressure, I think I just immediately went to the ideal gas law because we know the final volume and the number of moles won’t change and neither will the temperature in this case. Well, I assume that it wouldn’t.” Another student showed evidence of their evaluation by writing out all the necessary information in one place and stating, “Whenever I do these types of problems, I always write what I start with which is why I always have this line of information I’m given.” After evaluating and analyzing, students had to form an argument by claiming that the two energy values were equal and then defending that claim. Students explained that they were not always as clear as they could be when justifying their claim. For instance, one student said, “Usually I just write out equations and then hope people understand what I’m doing mathematically” but they “probably could have explained it a little more.”

Student feedback throughout the organic chemistry course and near the end of the physical chemistry course indicated that the rubric scores were accurate representations of the students’ work with a few exceptions. For example, some students felt like they should have received either a lower or higher score for certain categories, but they did say that the categories themselves

applied well to their work. Most notably, one student reported that the forms and supports arguments categories in the critical thinking rubric did not apply to her work because she “wasn’t making an argument” when she was demonstrating that the Helmholtz and Gibbs energy values were equal in her thermodynamics assignment. We see this as an instance where some students and instructors may define argument in different ways. The process skill definitions and the rubric categories are meant to articulate intended learning outcomes from faculty members to their students, so if a student defines the skills or categories differently than the faculty member, then the rubrics can serve to promote a shared understanding of the skill.

As previously mentioned, reliability was measured by two researchers assessing ten laboratory reports independently to ensure that multiple raters could use the rubrics consistently. The average adjacent agreement scores were 92% for critical thinking and 93% for information processing. The exact agreement scores were 86% for critical thinking and 88% for information processing. Additionally, two different raters assessed a statistics assignment that was given to sixteen first-year undergraduates. The average pairwise adjacent agreement scores were 89% for critical thinking and 92% for information processing for this assignment. However, the exact agreement scores were much lower: 34% for critical thinking and 36% for information processing. In this case, neither rater was an expert in the content area. While the exact agreement scores for the statistics assignment are much lower than desirable, the adjacent agreement scores do meet the threshold for reliability as seen in other rubrics (Jonsson & Svingby, 2007) despite the disparity in expertise. Based on these results, it may be difficult for multiple raters to give exactly the same scores to the same work if they have varying levels of content knowledge, but it is important to note that the

rubrics are primarily intended for formative assessment that can facilitate discussions between instructors and students about the ways for students to improve. The high level of adjacent agreement scores indicates that multiple raters can identify the same areas to improve in examples of student work.

Instructor and teaching assistant reflections

The survey responses from faculty members determined the utility of the rubrics. Faculty members reported that when they used the rubrics to define their expectations and be more specific about their assessment criteria, the students seemed to be better able to articulate the areas in which they needed improvement. As one instructor put it, “having the rubrics helped open conversations and discussions” that were not happening before the rubrics were implemented. We see this as evidence of the clear intended learning outcomes that are an integral aspect of achieving constructive alignment within a course. The instructors’ specific feedback to the students, and the students’ increased awareness of their areas for improvement, may enable the students to better regulate their learning throughout a course. Additionally, the survey responses indicated that the faculty members were changing their teaching practices and becoming more cognizant of how assignments did or did not elicit the process skill evidence that they desired. After using the rubrics, one instructor said, “I realize I need to revise many of my activities to more thoughtfully induce process skill development.” We see this as evidence that the faculty members were using the rubrics to regulate their teaching by reflecting on the outcomes of their practices and then planning for future teaching. These activities represent the reflection and forethought/planning aspects of self-regulated learning on the part of the instructors. Graduate teaching assistants in the physical chemistry course indicated that the rubrics gave them a way to clarify the instructor’s expectations when they were interacting with the students. As one GTA said, “It’s giving [the students] feedback on direct work that they have instead of just right or wrong. It helps them to understand like ‘Okay how can I improve? What areas am I lacking in?’” A more detailed account of how the instructors and teaching assistants implemented the rubrics has been reported elsewhere (Cole et al., 2019a).

Student reflections

Students in both the organic and physical chemistry courses reported that they could use the rubrics to engage in the three phases of self-regulated learning: forethought/planning, performing, and reflecting. In an organic chemistry interview, one student was discussing how they could improve their low score for the

synthesizing category of critical thinking by saying “I could use the data together instead of trying to use them separately,” thus demonstrating forethought/planning for their later work. Another student described how they could use the rubric while performing a task: “I could go through [the rubric] as I’m writing a report...and self-grade.” Finally, one student demonstrated how they could use the rubrics to reflect on their areas for improvement by saying that “When you have the five column [earn a score of five], I can understand that I’m doing something right” but “I really need to work on revising my reports.” We see this as evidence that students can use the rubrics to regulate their own learning, although classroom facilitation can have an effect on the ways in which students use the rubric feedback (Cole et al., 2019b).

Limitations

The process skill definitions presented here represent a consensus understanding among members of the POGIL community and the instructors who participated in this study, but these skills are often defined in multiple ways by various STEM instructors, employers, and students (Danczak et al., 2017). One issue with critical thinking, in particular, is the broadness of how the skill is defined in the literature. Through this work, we have evidence via expert review to indicate that our definitions represent common understandings among a set of STEM faculty. Nonetheless, we cannot claim that all STEM instructors or researchers will share the skill definitions presented here.

There is currently a debate in the STEM literature (National Research Council, 2011) about whether the critical thinking construct is domain-general or domain-specific, that is, whether or not one’s critical thinking ability in one discipline can be applied to another discipline. We cannot make claims about the generalness of the construct based on the data presented here because the same students were not tested across multiple disciplines or courses. Additionally, we did not gather evidence for convergent validity, which is “the degree to which an operationalized construct is similar to other operationalized constructs that it theoretically should be similar to” (National Research Council, 2011). In other words, evidence for convergent validity would be the comparison of multiple measures of information processing or critical thinking. However, none of the instructors who used the ELIPSS rubrics also used a secondary measure of the constructs. Although the rubrics were examined by a multidisciplinary group of collaborators, this group was primarily chemists and included eight faculties from other disciplines, so the content validity of the rubrics may be somewhat limited.

Finally, the generalizability of the rubrics is limited by the relatively small number of students who were interviewed about their work. During their interviews, the students in the organic and physical chemistry courses each said that they could use the rubric scores as feedback to improve their skills. Additionally, as discussed in the “Validity and Reliability” section, the processes described by the students aligned with the content of the rubric and provided evidence of the rubric scores’ validity. However, the data gathered from the student interviews only represents the views of a subset of students in the courses, and further study is needed to determine the most appropriate contexts in which the rubrics can be implemented.

Conclusions and implications

Two rubrics were developed to assess and provide feedback on undergraduate STEM students’ critical thinking and information processing. Faculty survey responses indicated that the rubrics measured the relevant aspects of each process skill in the disciplines that were examined. Faculty survey responses, TA interviews, and student interviews over multiple semesters indicated that the rubric scores accurately reflected the evidence of process skills that the instructors wanted to see and the processes that the students performed when they were completing their assignments. The rubrics showed high inter-rater agreement scores, indicating that multiple raters could identify the same areas for improvement in student work.

In terms of constructive alignment, courses should ideally have alignment between their intended learning outcomes, student and instructor activities, and assessments. By using the ELIPSS rubrics, instructors were able to explicitly articulate the intended learning outcomes of their courses to their students. The instructors were then able to assess and provide feedback to students on different aspects of their process skills. Future efforts will be focused on modifying student assignments to enable instructors to better elicit evidence of these skills. In terms of self-regulated learning, students indicated in the interviews that the rubric scores were accurate representations of their work (performances), could help them reflect on their previous work (self-reflection), and the feedback they received could be used to inform their future work (forethought). Not only did the students indicate that the rubrics could help them regulate their learning, but the faculty members indicated that the rubrics had helped them regulate their teaching. With the individual categories on each rubric, the faculty members were better able to observe their students’ strengths and areas for improvement and then tailor their instruction to meet those needs. Our results indicated that the rubrics helped instructors in multiple STEM disciplines and at multiple institutions reflect on

their teaching and then make changes to better align their teaching with their desired outcomes.

Overall, the rubrics can be used in a number of different ways to modify courses or for programmatic assessment. As previously stated, instructors can use the rubrics to define expectations for their students and provide them with feedback on desired skills throughout a course. The rubric categories can be used to give feedback on individual aspects of student process skills to provide specific feedback to each student. If an instructor or department wants to change from didactic lecture-based courses to active learning ones, the rubrics can be used to measure non-content learning gains that stem from the adoption of such pedagogies. Although the examples provided here for each rubric were situated in chemistry contexts, the rubrics were tested in multiple disciplines and institution types. The rubrics have the potential for wide applicability to assess not only laboratory reports but also homework assignments, quizzes, and exams. Assessing these tasks provides a way for instructors to achieve constructive alignment between their intended outcomes and their assessments, and the rubrics are intended to enhance this alignment to improve student process skills that are valued in the classroom and beyond.

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s40594-020-00208-5>.

Additional file 1. Supporting Information

Abbreviations

AAC&U: American Association of Colleges and Universities; CAT: Critical Thinking Assessment Test; CU: Comprehensive University; ELIPSS: Enhancing Learning by Improving Process Skills in STEM; IR: Infrared; LEAP: Liberal Education and America’s Promise; NMR: Nuclear Magnetic Resonance; PCT: Primary Collaborative Team; PLTL: Peer-led Team Learning; POGIL: Process Oriented Guided Inquiry Learning; PUI: Primarily Undergraduate Institution; RU: Research University; STEM: Science, Technology, Engineering, and Mathematics; VALUE: Valid Assessment of Learning in Undergraduate Education

Acknowledgements

We thank members of our Primary Collaboration Team and Implementation Cohorts for collecting and sharing data. We also thank all the students who have allowed us to examine their work and provided feedback.

Supporting information

- Product rubric survey
- Initial implementation survey
- Continuing implementation survey

Authors’ contributions

RC, JL, and SR performed an initial literature review that was expanded by GR. All authors designed the survey instruments. GR collected and analyzed the survey and interview data with guidance from RC. GR revised the rubrics with extensive input from all other authors. All authors contributed to reliability measurements. GR drafted all manuscript sections. RC provided extensive comments during manuscript revisions; JL, SR, and CS also offered comments. All authors read and approved the final manuscript.

Funding

This work was supported in part by the National Science Foundation under collaborative grants #1524399, #1524936, and #1524965. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Department of Chemistry, University of Iowa, W331 Chemistry Building, Iowa City, Iowa 52242, USA. ²Department of Chemistry, Virginia Commonwealth University, Richmond, Virginia 23284, USA. ³Department of Chemistry, Drew University, Madison, New Jersey 07940, USA. ⁴Department of Chemistry, Ball State University, Muncie, Indiana 47306, USA.

Received: 1 October 2019 Accepted: 20 February 2020

Published online: 09 March 2020

References

- ABET Engineering Accreditation Commission. (2012). *Criteria for Accrediting Engineering Programs*. Retrieved from <http://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2016-2017/>.
- American Chemical Society Committee on Professional Training. (2015). *Undergraduate Professional Education in Chemistry: ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs*. Retrieved from <https://www.acs.org/content/dam/acsorg/about/governance/committees/training/2015-acg-guidelines-for-bachelors-degree-programs.pdf>
- Association of American Colleges and Universities. (2019). VALUE Rubric Development Project. Retrieved from <https://www.aacu.org/value/rubrics>.
- Bailin, S. (2002). Critical Thinking and Science Education. *Science and Education*, 11, 361–375.
- Biggs, J. (1996). Enhancing teaching through constructive alignment. *Higher Education*, 32(3), 347–364.
- Biggs, J. (2003). Aligning teaching and assessing to course objectives. *Teaching and learning in higher education: New trends and innovations*, 2, 13–17.
- Biggs, J. (2014). Constructive alignment in university teaching. *HERDSA Review of higher education*, 1(1), 5–22.
- Black, P., & William, D. (1998). Assessment and Classroom Learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7–74.
- Bodner, G. M. (1986). Constructivism: A theory of knowledge. *Journal of Chemical Education*, 63(10), 873–878.
- Brewer, C. A., & Smith, D. (2011). *Vision and change in undergraduate biology education: a call to action*. American Association for the Advancement of Science. DC: Washington.
- Brookhart, S. M., & Chen, F. (2014). The quality and effectiveness of descriptive rubrics. *Educational Review*, 1–26.
- Butler, D. L., & Winne, P. H. (1995). Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 65(3), 245–281.
- Cole, R., Lantz, J., & Ruder, S. (2016). Enhancing Learning by Improving Process Skills in STEM. Retrieved from <http://www.elipss.com>.
- Cole, R., Lantz, J., & Ruder, S. (2019a). PO: The Process. In S. R. Simonson (Ed.), *POGIL: An Introduction to Process Oriented Guided Inquiry Learning for Those Who Wish to Empower Learners* (pp. 42–68). Sterling, VA: Stylus Publishing.
- Cole, R., Reynders, G., Ruder, S., Stanford, C., & Lantz, J. (2019b). Constructive Alignment Beyond Content: Assessing Professional Skills in Student Group Interactions and Written Work. In M. Schultz, S. Schmid, & G. A. Lawrie (Eds.), *Research and Practice in Chemistry Education: Advances from the 25th IUPAC International Conference on Chemistry Education 2018* (pp. 203–222). Singapore: Springer.
- Danczak, S., Thompson, C., & Overton, T. (2017). 'What does the term Critical Thinking mean to you?' A qualitative analysis of chemistry undergraduate, teaching staff and employers' views of critical thinking. *Chemistry Education Research and Practice*, 18, 420–434.
- Daniel, K. L., Bucklin, C. J., Leone, E. A., & Idema, J. (2018). Towards a Definition of Representational Competence. In *Towards a Framework for Representational Competence in Science Education* (pp. 3–11). Switzerland: Springer.
- Davies, M. (2013). Critical thinking and the disciplines reconsidered. *Higher Education Research & Development*, 32(4), 529–544.
- Deloitte Access Economics. (2014). Australia's STEM Workforce: a survey of employers. Retrieved from <https://www2.deloitte.com/au/en/pages/economics/articles/australias-stem-workforce-survey.html>.
- Driscoll, M. P. (2005). *Psychology of learning for instruction*. Boston, MA: Pearson Education.
- Ennis, R. H. (1990). The extent to which critical thinking is subject-specific: Further clarification. *Educational researcher*, 19(4), 13–16.
- Facione, P. A. (1984). Toward a theory of critical thinking. *Liberal Education*, 70(3), 253–261.
- Facione, P. A. (1990a). *The California Critical Thinking Skills Test—College Level*. In *Technical Report #1. Experimental Validation and Content: Validity*.
- Facione, P. A. (1990b). *The California critical thinking skills test—college level*. In *Technical Report #2. Factors Predictive of CT: Skills*.
- Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23), 8410–8415.
- Gafney, L., & Varma-Nelson, P. (2008). *Peer-led team learning: evaluation, dissemination, and institutionalization of a college level initiative* (Vol. 16). Springer Science & Business Media, Netherlands.
- Glassner, A., & Schwarz, B. B. (2007). What stands and develops between creative and critical thinking? Argumentation? *Thinking Skills and Creativity*, 2(1), 10–18.
- Gosser, D. K., Cracolice, M. S., Kampmeier, J. A., Roth, V., Strozak, V. S., & Varma-Nelson, P. (2001). *Peer-led team learning: A guidebook*. Prentice Hall Upper Saddle River, NJ.
- Gray, K., & Koncz, A. (2018). The key attributes employers seek on students' resumes. Retrieved from <http://www.nacweb.org/about-us/press/2017/the-key-attributes-employers-seek-on-students-resumes/>.
- Hanson, D. M. (2008). A cognitive model for learning chemistry and solving problems: implications for curriculum design and classroom instruction. In R. S. Moog & J. N. Spencer (Eds.), *Process-Oriented Guided Inquiry Learning* (pp. 15–19). Washington, DC: American Chemical Society.
- Hattie, J., & Gan, M. (2011). Instruction based on feedback. *Handbook of research on learning and instruction*, 249–271.
- Huitt, W. (1998). Critical thinking: an overview. In *Educational psychology interactive*. Retrieved from <http://www.edpsycinteractive.org/topics/cogsys/critthnk.html>.
- Joint Committee on Standards for Educational Psychological Testing. (2014). *Standards for Educational and Psychological Testing*. American Educational Research Association.
- Jonsson, A., & Svingby, G. (2007). The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, 2(2), 130–144.
- Kumi, B. C., Olimpo, J. T., Bartlett, F., & Dixon, B. L. (2013). Evaluating the effectiveness of organic chemistry textbooks in promoting representational fluency and understanding of 2D-3D diagrammatic relationships. *Chemistry Education Research and Practice*, 14, 177–187.
- Lai, E. R. (2011). Critical thinking: a literature review. *Pearson's Research Reports*, 6, 40–41.
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into Practice*, 32, 131–137.
- Miri, B., David, B., & Uri, Z. (2007). Purposely teaching for the promotion of higher-order thinking skills: a case of critical thinking. *Research in Science Education*, 37, 353–369.
- Moog, R. S., & Spencer, J. N. (Eds.). (2008). *Process oriented guided inquiry learning (POGIL)*. Washington, DC: American Chemical Society.
- Moskal, B. M., & Leydens, J. A. (2000). Scoring rubric development: validity and reliability. *Practical Assessment, Research and Evaluation*, 7, 1–11.
- Nakhleh, M. B. (1992). Why some students don't learn chemistry: Chemical misconceptions. *Journal of Chemical Education*, 69(3), 191.
- National Research Council. (2011). *Assessing 21st Century Skills: Summary of a Workshop*. Washington, DC: The National Academies Press.
- National Research Council. (2012). *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*. Washington, DC: The National Academies Press.
- Nguyen, D. H., Gire, E., & Rebello, N. S. (2010). Facilitating Strategies for Solving Work-Energy Problems in Graphical and Equational Representations. *2010 Physics Education Research Conference*, 1289, 241–244.

- Nicol, D. J., & Macfarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31(2), 199–218.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: a review. *Educational Research Review*, 9, 129–144.
- Pearl, A. O., Rayner, G., Larson, I., & Orlando, L. (2019). Thinking about critical thinking: An industry perspective. *Industry & Higher Education*, 33(2), 116–126.
- Ramsden, P. (1997). The context of learning in academic departments. *The experience of learning*, 2, 198–216.
- Rau, M. A., Kennedy, K., Oxtoby, L., Bollom, M., & Moore, J. W. (2017). Unpacking “Active Learning”: A Combination of Flipped Classroom and Collaboration Support Is More Effective but Collaboration Support Alone Is Not. *Journal of Chemical Education*, 94(10), 1406–1414.
- Reynders, G., Suh, E., Cole, R. S., & Sansom, R. L. (2019). Developing student process skills in a general chemistry laboratory. *Journal of Chemical Education*, 96(10), 2109–2119.
- Saxton, E., Belanger, S., & Becker, W. (2012). The Critical Thinking Analytic Rubric (CTAR): Investigating intra-rater and inter-rater reliability of a scoring mechanism for critical thinking performance assessments. *Assessing Writing*, 17, 251–270.
- Schmidt, H. G., De Volder, M. L., De Grave, W. S., Moust, J. H. C., & Patel, V. L. (1989). Explanatory Models in the Processing of Science Text: The Role of Prior Knowledge Activation Through Small-Group Discussion. *J. Educ. Psychol.*, 81, 610–619.
- Simonson, S. R. (Ed.). (2019). *POGIL: An Introduction to Process Oriented Guided Inquiry Learning for Those Who Wish to Empower Learners*. Sterling, VA: Stylus Publishing, LLC.
- Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (Eds.). (2012). *Discipline-Based education research: understanding and improving learning in undergraduate science and engineering*. Washington D.C.: The National Academies Press.
- Smit, R., & Birri, T. (2014). Assuring the quality of standards-oriented classroom assessment with rubrics for complex competencies. *Studies in Educational Evaluation*, 43, 5–13.
- Stein, B., & Haynes, A. (2011). Engaging Faculty in the Assessment and Improvement of Students' Critical Thinking Using the Critical Thinking Assessment Test. *Change: The Magazine of Higher Learning*, 43, 44–49.
- Swanson, H. L., Oconnor, J. E., & Cooney, J. B. (1990). An Information-Processing Analysis of Expert and Novice Teachers Problem-Solving. *American Educational Research Journal*, 27(3), 533–556.
- The Royal Society. (2014). *Vision for science and mathematics education: The Royal Society Science Policy Centre*. London: England.
- Watson, G., & Glaser, E. M. (1964). *Watson-Glaser Critical Thinking Appraisal Manual*. New York, NY: Harcourt, Brace, and World.
- Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice*, 41(2), 64–70.
- Zohar, A., Weinberger, Y., & Tamir, P. (1994). The Effect of the Biology Critical Thinking Project on the Development of Critical Thinking. *Journal of Research in Science Teaching*, 31, 183–196.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)