
The "pHunger Games": Manuscript Review to Assess Graduating Chemistry Majors

David J. Gorin*, Elizabeth R. Jamieson, K. T. Queeney, Kevin M. Shea, Carrie G. Read
Spray^a

5 Department of Chemistry, Smith College, Northampton, MA 01063, USA

ABSTRACT

Numerous options exist to assess student performance using standardized, multiple-choice exams at the course and department levels. This paper describes the development and implementation of an alternative department-level assessment for
10 graduating chemistry majors. The assessment detailed here evaluates students' ability to transfer chemical knowledge from their classes to a real life application, namely the review of a scientific paper. Working in groups of three with full access to reference materials, students review a paper intentionally doctored by the faculty to contain a variety of errors. Student groups identify and correct mistakes in a paper with content
15 spanning numerous chemistry subdisciplines. To motivate student effort, a prize is awarded to the group submitting the most thorough review. The data collected from the "pHunger Games" will inform curricular reform and innovation throughout the department.

KEYWORDS

20 General Public, Interdisciplinary/Multidisciplinary, Curriculum, Testing/Assessment, Student-Centered Learning, Applications of Chemistry

BACKGROUND AND INTRODUCTION

Inspired by college-wide discussions of assessment and prodded by requirements for accreditation,¹ faculty in the Smith College Department of Chemistry became interested
25 in department-level assessment of student learning upon completion of a degree in chemistry.² We had previously defined departmental learning outcomes³ (Box 1), and first considered well-established ACS assessment instruments. The ACS Diagnostic of

Undergraduate Chemistry Knowledge (DUCK) multiple-choice exam assesses broad student learning at the end of the undergraduate experience and enables nationwide comparisons among students.⁴

Box 1. Learning Outcomes for Graduating Chemistry Majors

1. Read and understand a published scientific paper in the chemistry literature.
2. Write about chemistry clearly both to experts and non-experts, in a way that “tells a story” with chemistry.
3. Locate and use valid, peer-reviewed sources and distinguish those sources from less reliable ones.
4. Access chemical knowledge and skills from the full range of our curriculum to analyze problems outside the context of a particular course or subfield.
5. Use knowledge and skills gained from our curriculum to understand new techniques, phenomena, and problems within chemistry.
6. Interpret chemical data. This has a least two components: interpretation of raw data, which can include quantitative manipulation of those data, and then using that interpretation to draw larger conclusions about phenomena.
7. “Think like a chemist”; break relevant problems down to the fundamental molecular/atomic level, and recognize when (and why) understanding is incomplete.

After careful consideration, we decided that a multiple-choice test like the DUCK would not address most of our desired learning outcomes (Box 1). The DUCK is most closely linked to learning outcome 4 (Box 1), but the types of chemistry problems offered on this and other standardized exams tend to be limited to one subfield/topic at a time and therefore do not in our view even completely address that outcome, which aspires to have students draw on content and skills from more than one course to solve a given problem. More practically, we did not want to offer an assessment in which student performance would be highly dependent on how much they studied immediately prior to the test. We also recognize that a full picture of student learning is

sometimes obscured on standardized tests where issues like anxiety and stereotype threat come into play.⁵

We evaluated available models of department-level assessment, including portfolios of course artifacts, tests of laboratory skills, in-house generated comprehensive exams, and capstone seminar presentations.⁶ Based on this survey, we decided to focus on an open-ended assessment that required students to solve problems, that allowed them access to outside resources (e.g. print and online) to do so, that stood apart from the content and assignments of any individual course, and that mimicked as much as possible the work of real chemists. Recent assessments designed to measure students' abilities to transfer and link concepts as they progress through general and organic chemistry courses⁷ were particularly helpful as we thought about how to assess students' ability to apply content from throughout the chemistry curriculum in a department-level capstone examination.

Our assessment is built around the scholarly review of a scientific paper. Students work in groups to identify and correct mistakes in a paper doctored by faculty to include spurious data analysis, unsupported conclusions, and other substantive errors. This exercise intrinsically provides assessment of learning outcomes 1, 4 and 6 (Box 1); by purposely including techniques and/or topics that are not explicitly covered in our curriculum, we have also used it to address outcome 5. The other learning outcomes in our list (which is essentially a "greatest hits" list culled by consensus from a much longer list of desired outcomes) are better addressed through other exercises that we will not detail here.

IMPLEMENTATION PLAN

The assessment is conducted at the end of the academic year outside of any specific course to allow seniors to make use of material and skills learned in their final

semester. Fearing that adding another requirement might discourage students considering a chemistry major, we decided to encourage voluntary participation^{8,9} by making the assessment into a competition with prizes¹⁰ for the best review. Thus, the
70 “pHunger Games” was born.

This assessment has been conducted for four consecutive years (see Supporting Information for a detailed description of logistics). Each year a team of faculty write a new paper for review, using a published paper as both inspiration and template.¹¹ They doctor and shorten the paper so that it contains approximately ten significant errors of
75 varying complexity. An effort is made to find a paper that will force students to draw on content from more than one chemistry subdiscipline. On the day of the competition, student teams are given five hours to review the paper, consult any print or online resources (though they are prohibited from searching for the original paper), and write their review. The faculty authors grade the reviews using a prepared rubric. Declaring
80 victors enables us to motivate student participation and effort, but the competition is a means to an end – the value of the pHunger Games arises from analysis of all student answers, which provides insight into how well student capabilities at graduation map onto our learning outcomes.

85 THE ASSESSMENT – EXAMPLE PROBLEMS AND RESULTS

In four iterations of the pHunger Games assessment, students do not seem paralyzed by the task of manuscript review, and every team has successfully identified and explained at least some of the errors. Although all student groups have found some success, there has been great variability in students’ ability to identify and/or correct
90 specific errors in each assessment.

To illustrate how the assessment works and how it is graded, three excerpts from a single year’s exam are reviewed below. Each problematic area in the paper is defined as

a “rubric item” and represents a specific mistake that students can note and correct (For a discussion of how we generate rubric items by altering an original research manuscript, please see the Supporting Information). To grade student answers, faculty look for each rubric item within the students’ review and assign the explanation a score (excellent, good, fair, poor) depending upon the quality (or complete absence) of answer. The excerpts below are accompanied by brief summaries of the rubric items and the scoring rubrics used to evaluate student responses. The complete doctored paper and grading rubrics are available to any interested reader through direct correspondence with the authors.

The 2013-2014 pHunger Games paper¹² described the use of small organic molecules as anion sensors. Detection was based upon a change in UV/Vis absorbance of the sensor in the presence of various anions (cyanide, hydroxide, fluoride, etc.). The introduction and conclusion made strong claims about the potential applicability of these sensors for the detection of cyanide in aqueous environmental samples (Box 2). However, the data within the paper demonstrated that the sensor was not selective for cyanide and functioned only in organic solvents; the prospects for cyanide detection in water were therefore dim.

For this rubric item 4 of the 7 teams offered no criticism, and were therefore graded as “poor” in accord with the grading rubric. Only 1 of the 7 offered a substantive, well-reasoned objection, which was deemed “excellent”. Notably, an excellent critique in this particular case does not require deep chemical content knowledge - instead, it demands that students pause to consider the chemical implications of a proposed application (e.g. detecting pollutants in water requires a sensor that works in water). In the year 2011-2012 exam, students were faced with a similar problem – the paper claimed to develop a Cu sensor for use in living systems, yet no cellular or other in vivo studies were done – and students’ answers were largely unsatisfactory. Taken together, the

responses to these rubric items suggest that students may focus on the experimental
120 details in the paper without considering the broader scientific context and applicability.
Since several departmental electives are literature-based seminar courses, further effort
is being devoted in electives to encouraging students to pause and consider the “big
picture” as they read papers, rather than exclusively focusing on experimental
techniques and data. This can be as simple as requiring students, in either written or
125 oral presentations about the literature they read, to begin by summarizing the main
goals and findings of the paper.

Box 2. Example #1

Rubric item: Although introduction and conclusion emphasize the detection of cyanide in aqueous environmental samples, the reported sensor is not shown to function in water and is not selective for cyanide. All experiments were run in chloroform and similar signal responses are reported for cyanide, acetate, and fluoride anions.

Excerpts: Introduction: *The design of anion sensors is a fast emerging research field due to the importance of detecting anions in medicine, chemical industry, and the environment. Chloride and nitrate ions, for example, are tracers of water pollution while cyanide, a highly toxic anion, is a dangerous environmental contaminant associated with leakage from electroplating and herbicide industries. Therefore, the development of fast, easy, and sensitive detection techniques for the continuous monitoring of anions is important.*

Conclusion: *In summary, two dibenzophenazine-based sulfonamide molecules have been synthesized and shown to be effective fluorescent turn-on sensors for several anions. Compounds 1 and 2 are ideal candidates for the detection of cyanide in waste water or other environmentally derived samples.*

Grading rubric:

Poor: Issue not noticed.

Fair: Note at least one problem with conclusion, which might be: 1) compounds not tested in water or 2) no evidence of cyanide selectivity.

Good: Note both issues listed in the “fair” answer.

Excellent: Notice both issues listed in the “fair” answer and suggest experiments to address them.

One important learning outcome is that students interpret graphed and/or
130 processed data (Box 1, Item 6). Incorrect conclusions drawn from graphed data are
therefore incorporated into every iteration of the pHunger Games, with special emphasis
on including experiments or data representations not explicitly covered in the
curriculum. In the 2013-2014 exam, a Job's plot was used to draw conclusions about
the binding stoichiometry of the small molecule sensor to the anion analyte (Box 3, see
135 Supporting Information for a discussion of how we created this rubric item). Job's plots
are not currently part of the Smith chemistry curriculum but can reasonably be
interpreted based upon foundational concepts that students could understand, perhaps
in consultation with a thermodynamics or analytical chemistry textbook. Students did
quite well on this rubric item, with five teams earning a "good" score and one team each
140 receiving "excellent" and "fair" scores. In this case, identical conclusions are drawn from
two graphs that present significantly different data. This may have helped students to
identify that one of the Job's plots was interpreted incorrectly.

Box 3. Example #2

Issue: Incorrect conclusion for binding stoichiometry drawn from Job's plots.

Excerpts: *The binding isotherms of acetate, benzoate, cyanide, and fluoride anions were generated from the change in absorption of the sensors. To calculate the binding constants of the sensors to these anions, the binding isotherms were fitted to a 1:1 binding model as suggested by the Job plots, examples of which are shown in Figures 4 and 5.*

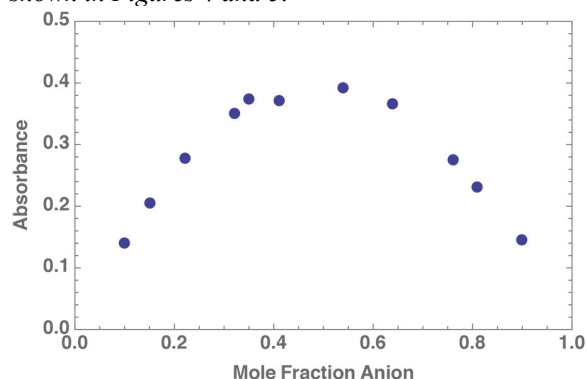


Figure 4: Job's plot for compound 1 reacting with TBAOBn

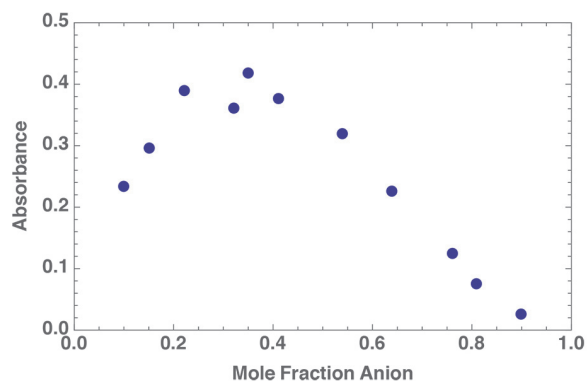


Figure 5: Job's plot for compound 2 reacting with TBACN

Grading rubric:

Poor: Issue not noticed.

Fair: Notice discrepancy between the two plots.

Good: State that the Fig 5 data doesn't allow for the conclusion of a 1:1 binding stoichiometry.

Excellent: Explain that the Fig 4 data suggests 1:1 stoichiometry, while the Fig 5 data is noisy but suggests 2:1 stoichiometry.

The learning outcome on interpreting chemical data (Box 1, item 6) also requires
145 students to analyze and draw conclusions from numerical (non-graphical) data.

Therefore, at least one numerical data set is included in each year's test, and one of the rubric items always requires students to manipulate and/or plot quantitative data to corroborate conclusions in the manuscript. In the 2013-2014 paper, UV-Vis absorbance at a specific wavelength as a function of analyte concentration was tabulated (Box 4,
150 see Supporting Information for a discussion of how we created this rubric item). These data were then ostensibly used to determine the equilibrium constant for the sensor-analyte interaction.

In the table of binding constants, the value for binding of sensor **2** with cyanide is a clear outlier, differing from all other values by 2 log units (~100-fold). This is
155 inconsistent with the tabulated absorbance data, which is very similar for the response of sensors 1 and 2 and cyanide. Despite this hint, student groups generally performed poorly on this rubric item. Four teams gave "poor" responses, one team gave a minimal ("fair") response, and only two teams offered a substantive ("good") correction. To address this, qualitative evaluation of quantitative data sets for outliers or unexpected
160 results has been incorporated into class assignments for our Advanced General Chemistry course, and this may be extended to other courses.

Box 4. Example #3

Issue: The reported binding affinities of sensors **1** and **2** for cyanide differ by a factor of 100 (Table 2), which is inconsistent with the raw absorbance data (Table 1).

Excerpts: *Typical data for the cyanide titration of 1 and 2 are given in Table 1. The obtained binding constant results, given in Table 2, were in accordance with previously observed and reported results on similar systems.*

Equiv CN ⁻	Abs for 1	Abs for 2
0.15	0.020	0.015
0.30	0.035	0.040
0.55	0.081	0.071
0.80	0.126	0.145
1.05	0.221	0.174
1.45	0.251	0.190
1.80	0.238	0.191

Table 1: Absorbance data for the titration of **1** and **2** (both 5 μ M) with cyanide.

	1	2
Anion	log K	log K
OAc ⁻	5.93	5.94
OBn ⁻	6.73	6.43
CN ⁻	6.01	3.89
F ⁻	5.72	5.95

Table 2: Binding constants of sensors **1** and **2** with several anions.

Grading rubric:

Poor: Issue not noticed.

Fair: Note that log K for **2** with cyanide seems anomalous.

Good: Note that the data in Table 1 suggests that **1** and **2** bind similarly to cyanide, which is inconsistent with the reported difference in binding affinities.

Excellent: Build upon a “good” answer by plotting the data to determine the correct log K.

The evaluation of rubric items relevant to specific learning outcomes is underway.

165 Preliminary assessment of items testing data analysis (Box 1, item 6) suggests that student performance diverges on questions that require the interpretation of processed or graphed data (as in Box 3) versus quantitative or unprocessed data (as in Box 4). Student performance on these rubric items is presented in Table 1, broken down by

competition year; the percentage of excellent and good answers submitted by student groups for all rubric items that test a specific skill (manipulate quantitative data, interpret processed data) are shown. The number of rubric items on each test that assess students' ability to either manipulate quantitative data (row 1) or interpret processed data (row 2) are shown in parenthesis.

Although all teams did well in year 1, this most likely reflects that the first iteration of the exam was too easy; the doctored data was obvious and straightforward to assess. In subsequent years, the data sets were incorporated to resemble native (undoctored) data. This required students to identify the existence of a problem and then solve it, rather than only to solve a clearly presented problem. On these more difficult exams,¹³ students consistently did poorly in evaluating quantitative data, suggesting that they often accept quantitative conclusions without manipulating the original data.

Table 1. Comparison by Year of Excellent/Good Answers on the pHunger Games Manuscript Assessment

Learning Outcome	Percentage of Responses Rated as "Excellent" or "Good" Answers (Number of Rubric Items Within Each Test Addressing the Specified Learning Outcome)		
	Year 1	Year 2	Year 3
Manipulate quantitative data	100 (1)	13 (2)	29 (1)
Interpret graphed/processed data	88 (2)	17 (6)	71 (2)

As further iterations of the pHunger Games enable the collection of more data, additional trends in student performance may emerge. Interestingly, students with the highest GPA have not consistently won the competition. This correlation will be evaluated further in the future.

CONCLUSION

In summary, an open-ended, competitive assessment of student learning based upon review of a scientific manuscript has been developed. Tracking of trends and patterns in student performance across multiple test years is enabled by evaluation of rubric items that assess the same learning outcome or specific content area. To date,

the curricular changes implemented in response to the pHunger Games have largely arisen from qualitative assessment of student responses to individual rubric items. Specific strategies that students should use to meet our desired learning outcomes have emerged, such as holding the “big picture” point of a paper in mind while evaluating results and qualitatively evaluating quantitative data for inconsistencies. We have sought to incorporate these strategies into various courses. The “pHunger Games” manuscript review offers departments¹⁴ an open-ended, competition-based option for the capstone assessment of students.

ASSOCIATED CONTENT

Supporting Information

Detailed descriptions of logistics for running the pHunger Games, faculty’s role during the competition day, and the process for generating rubric items within the manuscript for review.

AUTHOR INFORMATION

Corresponding Author

*E-mail: dgorin@smith.edu

Current Affiliation

α: Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170, USA.

ACKNOWLEDGMENTS

We thank Minh Ly and Cate Rowen (Smith College Office of Institutional Research) for help with analysis and data collection, the Smith College Provost’s Office for funding, chemistry department colleagues for their contributions, and student participants. D.J.G. gratefully acknowledges support from the NSF CAREER program (1554814).

REFERENCES

¹ a) *Standards for Accreditation*, Commission on Institutions of Higher Education, New England Association of Schools and Colleges, 2016. b) *Undergraduate Professional Education in Chemistry*:

ACS Guidelines and Evaluation Procedures for Bachelor's Degree Programs, American Chemical Society, Washington, DC, 2015.

² Bretz, S. L. Navigating the Landscape of Assessment. *J. Chem. Educ.* **2012**, 89, 689-691.

³ Elmgren, M.; Ho, F.; Akesson, E.; Schmid, S.; Towns, M. Comparison and Evaluation of Learning Outcomes from an International Perspective: Development of a Best-Practice Process. *J. Chem. Educ.* **2015**, 92, 427-432.

⁴ a) American Chemical Society Division of Chemical Education Examinations Institute. <http://chemexams.chem.iastate.edu/> (accessed Feb 2016). b) Home, T. Assessment Data and Decision Making in Teaching. *J. Chem. Educ.* **2011**, 88, 1017.

⁵ Steele, C. M. *Whistling Vivaldi: How Stereotypes Affect Us and What We Can Do*, Norton: New York, 2010.

⁶ a) Towns, M. H. Developing Learning Objectives and Assessment Plans at a Variety of Institutions: Examples and Case Studies. *J. Chem. Educ.* **2010**, 87, 91-96. b) Kirton, S. B.; Al-Ahmad, A.; Fergus, S. Using Structured Chemistry Examinations (SCHEMs) As an Assessment Method To Improve Undergraduate Students' Generic, Practical, and Laboratory-Based Skills. *J. Chem. Educ.* **2014**, 91, 648-654. c) Galloway, K. R.; Bretz, S. L. Development of an Assessment Tool To Measure Students' Meaningful Learning in the Undergraduate Chemistry Laboratory. *J. Chem. Educ.* **2015**, 92, 1149-1158.

⁷ a) Ye, L.; Oueini, R.; Lewis, S. E. Developing and Implementing as Assessment Technique To Measure Linked Concepts. *J. Chem. Educ.* **2015**, 92, 1807-1812. b) DeFever, R. S.; Bruce, H.; Bhattacharyya, G. Mental Rolodexing: Senior Chemistry Majors' Understanding of Chemical and Physical Properties. *J. Chem. Educ.* **2015**, 92, 415-426.

⁸ Student participation has always been high: 2012 – 71% (12 out of 17 seniors), 2013 – 92% (11 out of 12 seniors), 2014 – 89% (17 out of 19 seniors), and 2015 – 83% (19 out of 23 seniors).

⁹ Although we have not done so, this assessment could be made mandatory by incorporating it into a capstone or seminar course required for seniors within the major.

¹⁰ Each member of the winning team is awarded \$100.

¹¹ The following papers served as inspiration for our pHunger Games papers: a) Zeng, L.; Miller, E. W.; Pralle, A.; Isacoff, E. Y.; Chang, C. J. A Selective Turn-On Fluorescent Sensor for Imaging Copper in Living Cells. *J. Am. Chem. Soc.* **2006**, 128, 10-11. b) Gonsalvi, L.; Adams, H.; Sunley, G. J.; Ditzel, E.; Haynes, A. A Dramatic Steric Effect on the Rate of Migratory CO Insertion on Rhodium. *J. Am. Chem. Soc.* **1999**, 121, 11233-11234. c) El-Ballouli, A. O.; Zhang, Y.; Barlow, S.; Marder, S. R.; Al-Sayah, M. H.; Kaafarani, B. R. Fluorescent detection of anions by dibenzophenazine-based sensors. *Tetrahedron Lett.* **2012**, 53, 661-665. d) Chisholm, M. H.; Durr, C. B.; Gustafson, T. L.; Kender, W. T.; Spilker, T. F.; Young, P. J. Electronic and Spectroscopic Properties of Avobenzene Derivatives Attached to Mo₂ Quadruple Bonds: Suppression of the Photochemical Enol-to-Keto Transformation. *J. Am. Chem. Soc.* **2015**, 137, 5155-5162.

¹² This paper was inspired by reference 11c.

¹³ In year 2 (2012-2013), a concerted effort was made to significantly increase the test's difficulty. Student performance was unsatisfactory on many rubric items, perhaps due to the large number of errors present and the complexity of the arguments and data sets.

¹⁴ In 2015, Wellesley College conducted a version of the pHunger Games for their graduating seniors.