JOURNAL OF CHEMICAL EDUCATION
ACS Editors' Choice
Article
pubs.acs.org/jchemeduc
Terms of Use

What Does a Student Know Who Earns a Top Score on the Advanced Placement Chemistry Exam?

Jennifer Claesgens,*,[†] Paul L. Daubenmire,[‡] Kathleen M. Scalise,[§] Scott Balicki,[∥] Perman Gochyyev,[⊥] and Angelica M. Stacy[#]

[†]Center for Science Teaching and Learning, Northern Arizona University, Flagstaff, Arizona 86011, United States

[‡]Graduate Group in Science and Math Education (SESAME), University of California at Berkeley, Berkeley, California 94720, United States

[§]College of Education, University of Oregon, Eugene, Oregon 97403, United States

[∥]Science Department, Boston Latin School, Boston, Massachusetts 02155, United States

[⊥]Policy, Organization, Measurement, and Evaluation (POME), University of California at Berkeley, Berkeley, California 94720, United States

[#]Department of Chemistry, University of California at Berkeley, Berkeley, California 94720, United States

ABSTRACT: This paper compares the performance of students at a high-performing U.S. public school (n = 64) on the advanced placement (AP) chemistry exam to their performance on the ChemQuery assessment system. The AP chemistry exam was chosen because, as the National Research Council acknowledges, it is the "perceived standard of excellence and school quality". In contrast to the nationally recognized AP chemistry exam, the ChemQuery assessment system is a research tool that uses item-response theory to map student progress on a scale of conceptual understanding in chemistry. Our findings indicate that the two types of assessments, traditional problem-solving skills and conceptual understanding, are highly correlated as measured here. However, student performance on the ChemQuery assessment is disappointingly low. On the basis of the data analysis, this paper discusses the implications of the findings with a focus on the current efforts to redesign the AP chemistry exam.

KEYWORDS: *High School/Introductory Chemistry, Chemical Education Research, Testing/Assessment, Learning Theories, Constructivism*

FEATURE: Chemical Education Research

INTRODUCTION

The goal of this research is to better understand how students learn chemistry by comparing student scores on the advanced placement (AP) chemistry exam to student scores on a measure of their conceptual understanding using the ChemQuery assessment system. To begin, a reasonable expectation for students who earn a top score of 5 on the AP chemistry exam is that they would be able to answer correctly the two limiting reactant questions in Box 1.

On the basis of the AP exam statistics, probably less than 50% of the students taking the AP exam would get full credit for their answers to the question posed in Box 1.[1] In comparison, students' responses to the ChemQuery question typically respond to the question about mole number correctly, but many are not able to answer the corresponding question on the number of molecules correctly. These results are consistent with the literature on problem-solving versus conceptual understanding in both chemistry and physics, which acknowledges that students can problem solve but maintain misconceptions that limit their conceptual understanding.[2−5]

The AP test in chemistry was chosen for the comparison to ChemQuery because it is an accepted standard to qualify students to earn credit for first-year, college-level chemistry coursework.[6] The value of AP exams is exemplified by the fact that students often earn extra points toward their grade point average by taking AP courses, and high schools are often measured by how many AP courses they offer to students.[6] At the college level, the AP exam scores are used for credit and placement, so an assumption of the exam is that students who do well meet the typical expectations for students who take an introductory college chemistry course.

In comparison, ChemQuery is an assessment system that uses a framework of the key ideas in the discipline and criterion-referenced analysis using item-response theory (IRT) to map student progress. The ChemQuery assessment system is designed to measure how student conceptual understanding develops primarily by asking for explanations to chemistry questions.[7] This is in contrast to the goals of the AP chemistry test, which emphasizes problem solving to demonstrate that students have mastered chemistry topics equivalent to college coursework.

This paper begins by describing the two types of assessments. The data analysis then compares students' scores on the ChemQuery assessment system to their self-reported AP chemistry exam scores to answer the question: Does student conceptual understanding of chemistry measured using the

Published: February 28, 2014

Box 1. Limiting Reactant Sample Questions

AP Chemistry Exam (College Board AP Program[1])

A 0.150 g sample of lead(II) nitrate is added to 125 mL of 0.100 M sodium iodide solution. Assume no change in volume of the solution.

$$Pb(NO_3)_2(s) + 2NaI(aq) \rightarrow PbI_2(s) + 2NaNO_3(aq)$$

(a) Calculate the number of moles of each reactant.
(b) Identify the limiting reactant. Show calculations to support your identification.

ChemQuery Item 88 (Adapted from Lythcott[2])

Hydrogen and oxygen react to form water vapor in the following reaction. Assume the reaction goes to completion.

$$2H_2(g) + O_2(g) \rightarrow 2H_2O(g)$$

(a) If you mix 2 moles of hydrogen and 2 moles of oxygen, how many moles of water vapor are produced?
(b) If you mix 2000 molecules of hydrogen and 2000 molecules of oxygen, how many molecules of water vapor are produced?

ChemQuery assessment system correlate to students' scores on the AP chemistry test, and if so, what does this mean in terms of interpreting of student understanding of chemistry based on an AP score?

STRUCTURE OF THE ASSESSMENTS

The AP chemistry exam and the ChemQuery assessment system are two quite distinct tests to measure student ability in chemistry. The AP chemistry exam is a large-scale examination to measure students' knowledge of college-level general chemistry topics, while the ChemQuery assessment system is designed to measure how students' conceptual understanding of chemistry develops over the course of instruction. Both serve different purposes but are designed to measure what students "know" about chemistry. There are two main differences in the assessment design between the exams: the first is the choice of question type and the second is how students' responses are scored and interpreted.

The goal of the AP chemistry exam is to "prepare students to attain a depth of understanding of fundamentals and reasonable competence in dealing with chemical problems".[1] A secondary goal is to "contribute to the development of the students' abilities to think clearly and to express their ideas orally and in writing".[1] The AP chemistry exam is designed to measure and compare students' knowledge of a broad range of topics that include, among many others: stoichiometry; gas laws; oxidation−reduction reactions; equilibrium; and thermodynamics. Topics are distributed randomly in the exam and student understanding is measured as a total score; these scores are then norm-referenced.

The AP chemistry exam consists of three sections administered over 3 h. The first half of the exam is made up of 75 multiple-choice items covering a variety of topics to be answered in 90 min. The second section begins with two open-ended, free-response questions to be answered in 40 min. In the final component, students choose five from among eight prompts to "translate" from written text into chemical equations within 50 min. As a large-scale test, over half of the exam consists of answering multiple-choice items. Students, or schools districts on their behalf, pay a fee to take the exam. Postsecondary institutions often reward high performance on the exam by awarding course credit or class waivers for first-year-level chemistry courses.[8]

The ChemQuery assessment system consists of a set of free-response conceptual questions that are selected and administered as a high school test in a 50-min period by the students' instructor. While the AP chemistry exam covers a broad range of topics and compares student performance based off of raw scores that are summed up across the exam, the ChemQuery assessment system uses a framework of the key ideas in the discipline and criterion-referenced analysis with item response theory to map student progress, which will be described in the next section. This exam is not high-stakes and has been used solely for research.[7,9−11] In this study, a subset of the ChemQuery assessment items were administered during class time as tests twice during the school year, once as a pretest and once as a posttest to measure student learning. Only ChemQuery posttest scores were used in the analysis to compare with students' AP chemistry exam scores.

The ChemQuery Construct: The Perspectives of Chemists

Development of the ChemQuery assessment system has used both qualitative and quantitative methods, including a grounded theory approach,[12] to see what types of student thinking emerged that could then be described and measured statistically using Rasch modeling.[13] An initial step entailed developing a model to organize the overarching ideas of the discipline of chemistry into a framework that is described as the Perspectives of Chemists.[7] The Perspectives is a multidimensional construct map, which describes a hierarchy of conceptual understanding of chemistry ranging from novice to graduate levels along three sets of progress variables: Matter, Change and Energy. We consider these "big ideas" as the variables to measure students' emerging understanding in chemistry within the assessment system. The Matter variable can be described as the properties associated with "stuff", while the Change variable focuses on accounting for making "new stuff". Energy is still in development, but the concept seems to develop from early ideas of temperature on the pathway to enthalpy and entropy.

To then measure how student understanding develops along each variable necessitated thinking about how domain knowledge is organized in students' heads, so within each variable is a scale to describe a proposed progression of how students learn chemistry over the course of instruction. The levels within the proposed variables are constructed such that students give more complex and sophisticated responses as they develop from describing their initial ideas in Level 1 (Notions), to relating the language of chemists to their view of the world in Level 2 (Recognition), to formulating connections between several ideas in Level 3 (Formulation), and eventually reaching Level 4 (Construction).

The resulting Perspectives framework is intended to describe and measure how students learn and reason using models of chemistry to predict and explain phenomena, as summarized in Table 1. The horizontal axis of the table relates the three dimensions of domain knowledge in chemistry referred to as the progress variables: Matter, Change, and Energy. The perceived progression of explanatory reasoning that develops as students gain understanding in chemistry is along the vertical axis as Notions, Recognition, Formulation, and Construction with more sophisticated levels of understanding at the top.

The purpose of the Perspectives framework is to make explicit the relationship between domain knowledge and how students make meaning of the ideas as they learn chemistry

Table 1. Perspectives of Chemists Framework Relating Conceptual Understanding for Three Progress Variables

Student Levels of Understanding		Matter	Change	Energy[a]
Expert	Level IV: Construction Examining assumptions, comparing/relating models	How can we understand composition, structure, properties and amounts?	How can we understand type, progression, and extent of change?	How do models of matter and energy explain the chemical and physical changes occur?
↑	Level III: Formulation Relating ideas and concepts, simple models	How can we think about interactions between atoms? (bonding)	How can we think about rearrangement of atoms? (Relating particulate, macro, conservation)	Enthalpy, entropy, heat
	Level II: Recognition Language, definitions, symbols and simple algorithms	How do chemists describe matter? (Atomic symbols, octet rule)	How do chemists describe change? (Chemical symbols, conservation of mass)	Chemical reactions associated with heat transfer Systems surroundings
Novice	Level I: Notions Everyday experience, logical reasoning, Lacks correct chemistry models	What do you know about matter? (solids, liquids, gases)	What do you know about change? (Stuff happens)	Hot/cold

[a]The proposed Energy variable is still under development and not as well articulated at this time.

over the course of instruction. Like learning progressions, the variables of the Perspectives framework represent an instantiation both of understanding the "big" or enduring ideas of a discipline of science[14] and a tracing out and expanding on such "key ideas" to describe how students grasp increasing levels of complexity over time for core concepts.[15]

The pathway of understanding that is emerging from research using the ChemQuery assessment system starts with students' Notions, which includes their everyday experiences and values their logical reasoning. The next step is Recognition. At this level students are using normative definitions, terms, symbols, and algorithms in their explanations at a unirelational level; that is, they have a simple idea for a single evolving chemistry concept. The next step in students' understanding is Formulation: at this level, students can reason and integrate multiple ideas, concepts, or topics in chemistry. Construction is hypothetical. We have not gathered enough data to confirm our conjecture at this time. Nevertheless, the design has implications for both how students' responses are scored and how the scores can be interpreted. The Perspectives framework is then coupled with construct-referenced IRT analysis to map student progress.

ChemQuery: Measuring Student Understanding

With the use of the Perspectives framework, patterns in student responses are analyzed with item-response theory Rasch psychometric models. Item-response models are statistical models that express the probability of an occurrence, such as the correct response on an assessment question or task, in terms of estimates of a person's ability and the difficulty of the question or task. Specifically, the scores for a set of student responses and the questions are calibrated relative to one another on the same scale (a "logit" or log of the odds scale) and their fit, validity, and reliability estimated.[13] These scores are matched against the ChemQuery Perspectives framework describing levels of success in chemistry along each variable from novice student up through the expert. A key aspect of this type of assessment is the location of the question difficulty and student ability to correctly answer the question item.

Figure 1 shows a diagram of the output produced from this type of analysis using ACER Conquest 3.0 software to generate a Wright map of student scores. The Conquest output includes both student-ability levels and item difficulties on the

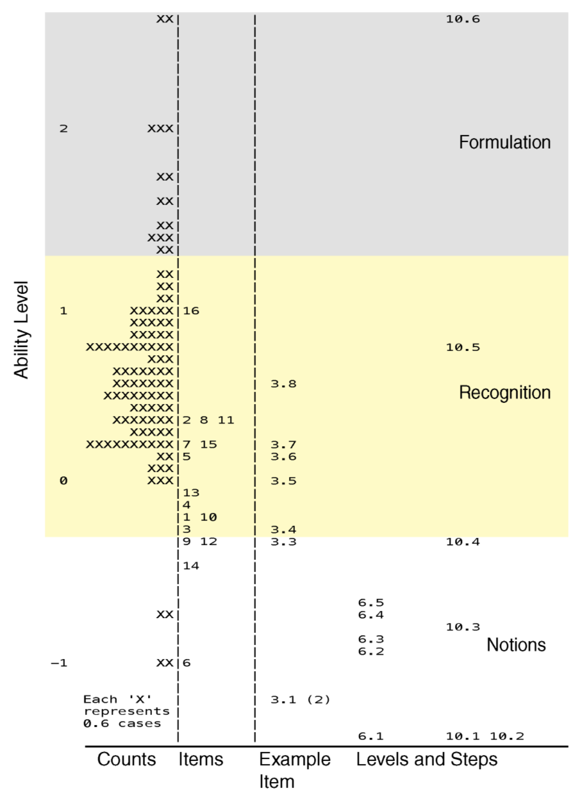
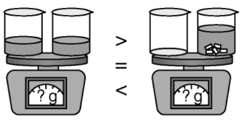
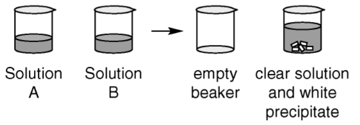


Figure 1. Wright map and ChemQuery item difficulty levels with generalized-item thresholds.

Wright map. This output is viewed as a map in which both the students and items are located so a description of the student understanding can be described linked to the item and level of answer observed (item and step). The vertical histogram represents the number of students ranked by their ability or understanding. The "Ability" on the y-axis is a log of the odds scale. It is the 50% probability that a student will get a question right at that level. The items are placed in locations relative to levels of student responses. This location is the item parameter, which can be used to link across data sets. The step is related to the score for that item based on patterns in student responses.

dx.doi.org/10.1021/ed300796r | J. Chem. Educ. 2014, 91, 472−479

Table 2. ChemQuery Change Items and Student Response Exemplars

Notions and Recognition for Levels I–II, by Item		Notions, Recognition, and Formulation for Level III, by Item
Item 3. When a clear aqueous solution of calcium chloride ($CaCl_2$) is mixed with a clear aqueous solution of sodium carbonate (Na_2CO_3), a white solid precipitate forms. Does the mass of the solution change after the two solutions are mixed and a solid forms? Explain. 	Item 10. When Solution A is mixed with Solution B, a white solid is formed in a clear, colorless liquid.  The white solid precipitate is AgCl and the clear, colorless liquid consists of Na^+, NO_3^-, and H_2O. Identify the chemicals in Solutions A and B.	Item 6. Consider the burning of methane (CH_4): $CH_4(g) + 2O_2(g) \rightarrow CO_2(g) + 2H_2O(g)$ You are told to mix enough methane and oxygen together so that each completely reacts and is used up with nothing left over. Which of the following will result in no excess reactants of methane and oxygen? (a) 16 g of CH_4 and 32 g of O_2 (b) 16 g of CH_4 and 64 g of O_2 (c) 1 mole of CH_4 and 1 mole of O_2 (d) 1 mole of CH_4 and 4 moles of O_2 Explain your thinking.
Level III Formulation: *Correct answer reached at Level II. Student answers don't extend to Level III.*		**Level III Formulation:** "B, because Moles = Mass/molar mass we would have 2 moles oxygen and 1 mole methane 1 m = 16 g/ 16 g/m 2 m = 64g/ 32g/m"
Level II Recognition: "No, the mass is conserved (law of conservation of mass)"	**Level II Recognition:** "NaCl and $AgNO_3$"	**Level II Recognition:** [Circle A] "For every one mole of CH_4—you need double for O_2 1:2, so 16:32"
Level I Notions: "Because they mix together the precipitate is formed and it is solid so it will have more mass."	**Level I Notions:** "Silver, chlorine, sodium, water, nitrogen, and oxygen" "A: NaO_2H B: NaOHgCl"	**Level I Notions:** [Circles B] "Because when you combine them they turn into something new." [Circles C] "Because they would probably cancel each other out."

ChemQuery items 3, 6, and 10 are included as examples with the steps associated with student scores as examples. Students' responses measuring between −1 and −0.25 fall in the Notions Level, scores between −0.25 and 1.5 are in Recognition, and scores above 1.5 measure in Formulation in the ChemQuery assessment system. These cut scores were determined by looking at the item locations and associated student scores (steps).

ChemQuery: Describing Student Understanding

For this study we focused on the two variables, Matter and Change. The Matter variable, items, scoring rubrics, and analysis are described in a previous publication.[7] Example of Change items and scoring rubrics including exemplars of student understanding are described here.

In ChemQuery Level I Notions, we find that students can articulate their ideas about change and matter; they use prior experiences, observations, logical reasoning, and knowledge to provide evidence for their ideas that *stuff happens*. For example, students responding to a limiting reactant question (Item 6) often respond with logical reasoning that the substances "cancel out" or with ideas of *stuff happens* with responses such as "when you combine...you make something new." This type of Notions-level thinking is also illustrated in ChemQuery Item 3, in which students use their experience that solids are heavier than liquids; for Item 10, many students list all of the ingredients either by name or symbol (Table 2). The student responses at this level focus largely on macroscopic descriptions of change or will use information from the question to construct a reasonable response. It seems that students measuring at Level 1 Notions rarely hold models that include ideas of conservation at the particulate level and stay more focused at the macroscopic level.

In comparison, students scoring in Level II Recognition begin to explore the language (terms and concepts) and specific symbols used by chemists to describe change more fluently with early ideas of conservation. Specifically in Item 3 students correctly apply the term "conservation of mass" to explain that the mass does not change when two solutions are mixed, and in Item 10 students correctly combine ions into the associated solutions. ChemQuery Items 3 and 10 do not measure beyond Level II Recognition. However, for items that do, student responses at Level II Recognition include accurate definitions, although students' understanding is not fully developed so that student reasoning often is limited to causal instead of explanatory mechanisms.

Student responses can be correct and incomplete with limited understanding of chemistry in Level II Recognition. For example, for Item 6 on limiting reactants, students at Level II Recognition correctly recognize limiting reactants and can interpret a chemical equation but are not making the link of moles to molecules at the particulate level. Overall, at Level II Recognition, student responses tend to focus on one correct link of understanding without consideration or understanding of the other links necessary for a more robust model of explanation to the question posed. Formulation is the next level as described in the ChemQuery assessment system. At this level we are observing students correctly linking conceptual and quantitative, and macroscopic with a particulate model of understanding. Specifically, in Item 6, students are making the link between symbols, grams, and moles. Additionally, these levels are meant to be cumulative; a student cannot reach Level III Formulation without moving through Levels II and I.

Comparing the Assessments

Overall, the AP exam and ChemQuery have different formats, different item questions, and different statistical measures, but both the AP exam and ChemQuery are designed to measure what students know about a common set of ideas in chemistry. Yet, each exam design has implications for both how students' responses are scored and how the scores can be interpreted. This research determines how the two exams correlate, and what each tells us about what students "know".

The AP exam is norm-referenced with a focus on problem-solving skills, meaning that the scales are aligned most directly to what a range of students achieve in high school chemistry rather than to actual concepts of college chemistry. Student scores on the AP chemistry exam are summed then normally distributed. From there the scores are divided into five groups with approximately 15−20% of the students in each group, as shown in Table 3. In 2006 when the scores were collected for the study comparison, 6150 schools offered the AP chemistry exam to a total of 78,453 students, 54% male, 45% female. (On a side note, nearly one-quarter of students are awarded a score of 1.) Overall, the AP score is determined by how students' scores compare to each other, for a norm-referenced sample, with psychometric measures of the items based on how scores were distributed the last three years, and how well college students do on the item questions.

Table 3. 2006 AP Chemistry Exam Score Distribution[a]

Exam Grade	Number of Examinees	AT, %
Score of 5	11,796	15.0
Score of 4	14,340	18.3
Score of 3	17,775	22.7
Score of 2	15,493	19.7
Score of 1	19,049	24.3
Total	78,453	100.0

[a]Data from ref 1.

In contrast, the resulting IRT analysis using the Perspectives construct tracks student learning by making people-to-discipline measurements that assess links between student understanding and chemistry domain concepts, rather than the people−to−people comparisons, which is common with norm-reference tests. In other words, students are placed on a scale of understanding of chemistry along each variable rather than focusing on referencing how they did in comparison to other students who took the exam.

■ STUDY DESIGN

This study compared the scores of 64 students on the ChemQuery assessment to their AP chemistry exam scores. All of the students in this study were enrolled in an AP chemistry class taught by the same instructor at a public high school in the northeast. The participating students had completed one year of high school chemistry previously. Sixteen ChemQuery items were selected from the Matter and Change item pool for the ChemQuery assessment. The item parameters were previously calibrated from thousands of high school and college chemistry from previous studies, so only a subset of items were necessary to determine student ability levels in this study. A ChemQuery pretest was administered in the fall of 2005 between the first and second year of chemistry instruction, with the posttest administered in the spring of 2006 immediately following the advanced placement exam. The ChemQuery assessment was administered as an ungraded quiz to highly motivated students. All students completed the quiz during the class period. Student posttest scores on the ChemQuery assessment were compared to self-reported AP chemistry scores.

Overall, the students in this study performed better than average on the AP chemistry exam with 25%, 20%, and 23% of students scoring 3, 4, and 5 on the exam (Table 4), respectively, compared to the 2006 national averages reported by the College Board as 15%, 18%, and 23%, respectively (Table 3). Note that only 9% of students scored a 1, which is below the national average of 25%.[1] These high scores are to be expected from the sample selection of the students, who were considered highly motivated by the teacher and who attended a high school that is recognized as a high-achieving school.

Table 4. Distribution of Student AP Chemistry Exam Scores[a]

Exam Grade	Number of Examinees	AT, %
Score of 5	15	23.44
Score of 4	13	20.31
Score of 3	16	25.00
Score of 2	14	21.88
Score of 1	6	9.38
	64	100.00

[a]Data are self-reported by students.

To determine students' score on ChemQuery, scoring rubrics with exemplars for each item question were used. Scoring involved assigning a raw score for each item a student answered based on the item exemplars, which resemble the examples from Table 2 with scores of 0−8 assigned for the various levels and sublevels (steps). Student scores were then compiled to produce a raw score for each student, with higher raw scores indicating that a student is at the higher levels of the construct and lower raw scores indicating that a student is at the lower level of the construct.

With the use of ACER Conquest 3.0 software, a Wright map of student posttest scores was generated (Figure 1). Inter-rater reliability was determined using Krippendorff's α coefficient. Responses from 10−14 students on 15 of the items were scored by all of the 7 raters, with responses from 10 students on one item (item 88a) scored by 6 raters. Krippendorff's α coefficient provides the summary of agreement among two or more independent raters on polytomous items.[16] The value of unity indicates perfect inter-rater reliability and the value of zero indicates the lack of reliability. The average Krippendorff's α value over the entire set of items was estimated to be 0.64.

The relationship between student AP scores and their scores on the ChemQuery assessment system was found to have a Pearson correlation of 0.67, with $p < 0.01$ on a two-tailed significance test. This indicates a clear relationship between students' self-reported AP exam scores and their measured ability level scores using the ChemQuery assessment system (Figure 2). In other words, students scoring high on the ChemQuery exam also scored high on the AP exam, for this sample. Therefore, in this study, conceptual understanding as measured by the ChemQuery exam is correlated to the problem-solving skills measured by the AP chemistry exam. We can now discuss the correlation between the conceptual understanding that develops from Notions to Recognition in the ChemQuery assessment system linked to more traditional measures, such as the AP chemistry exam. This study shows that the two measures are correlated but that they are quite different interpretations of students' ability or understanding of

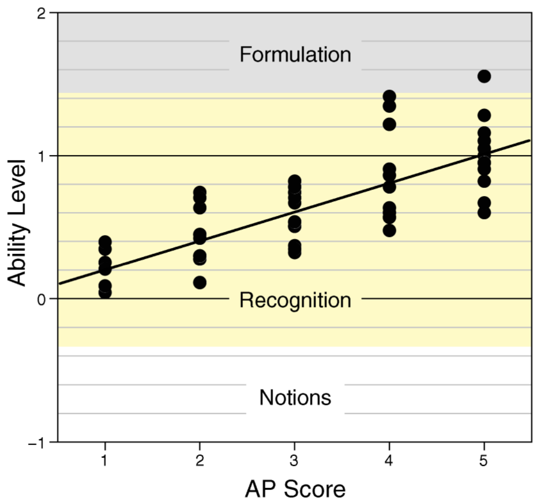


Figure 2. ChemQuery ability levels versus AP scores.

chemistry. A significant finding of this study shows that students who are passing the AP exam with scores of 3, 4, and 5 are all falling in Level II Recognition as described in the Perspectives of Chemists on the ChemQuery exam, which allows us to discuss what the AP score means in terms of their conceptual understanding of chemistry.

■ ANALYSIS

It is reasonable to assume that students scoring a 5 on the AP exam could correctly answer the Item 88 ChemQuery limiting reactant problem posed in the introduction (Box 1). However, this is not necessarily the case in light of the statistical analysis. The correlation between the exams indicates that students who received a score of 3 or above on the AP exam tend to answer the conceptual questions such as ChemQuery Item 88 in the Level II Recognition on the Perspectives framework. The benefit of using criterion-referenced measurement in the ChemQuery assessment system compared to norm-referenced measurement of the AP chemistry exam is that it allows one to match scores from a large-scale assessment to a construct of student understanding. In this case, we can provide a general description of the types of answers we have from students at different levels of understanding across two different assessments. Specifically here, we can describe the conceptual understanding of students who received a score of 5 on the AP chemistry exam.

For example, ChemQuery Item 88 regarding the combustion reaction of hydrogen and oxygen is an example that reaches to Level III Formulation. Students at Level II Recognition correctly answer, “How many moles of water vapor are produced?” but are unlikely to correctly answer, “How many molecules of water vapor are produced?” Using the Wright map output, students in Recognition have only a 50% probability of answering the follow-up question on molecules correctly. Student responses tend to focus on one correct link of understanding without consideration or understanding of the other links necessary for a more robust model of explanation to the question posed. In other words, student answers on the ChemQuery exam typically responded correctly to the question about mole number, but statistically most students in this study were not able to answer the corresponding question on the number of molecules correctly.

Measures of student understanding using the ChemQuery assessment system indicate a limited understanding of chemistry by not fully connecting terms, symbols, and algorithms to achieve a more sophisticated and intellectually flexible model of understanding. The link between the macroscopic, symbolic, and submicroscopic (particulate) as described by Johnstone seems to explain the difficulties we are observing in the data as students are developing their understanding toward Formulation.[17] Specifically, item 88 supports Lythcott’s research that describes the disconnect between students’ problem solving and conceptual understanding.[2] Overall our results are consistent with the literature on problem solving versus conceptual understanding in both chemistry and physics, which acknowledges that students can problem solve but maintain misconceptions that limit their conceptual understanding.[2−5] Therefore, it is not surprising then that the high-scoring AP students are not answering the correlating conceptual ChemQuery question correctly. Yet, are these measures of understanding equivalent to a first-semester college-level chemistry course?

The AP exam is validated by comparing student knowledge of high school examinees with that of students in introductory-level college chemistry courses.[6] Our evidence to date using the ChemQuery assessment system shows that students typically enter high school chemistry in the lower areas of Level 1 Notions. After one year of college-level chemistry, students in our data sets are in the upper region of Level II and lower region of Level III, between Recognition and Formulation. The sample group consisted of 973 students in their first postsecondary course in general chemistry at a single university.[19] Percentage distribution of student scores on the Perspectives framework over the course of instruction are shown in Figure 3. Note that

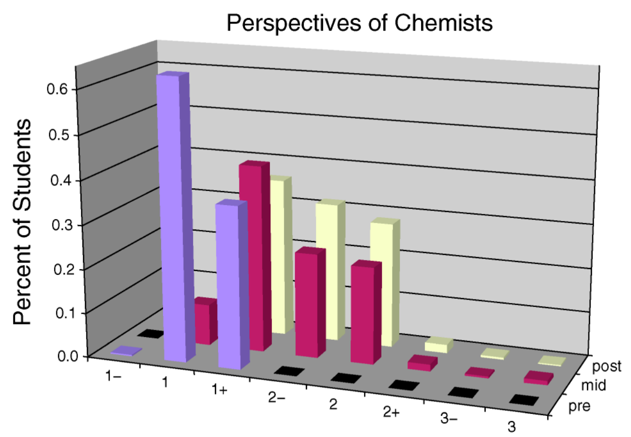


Figure 3. Distribution of general chemistry college students on Perspectives in first-year chemistry.

within each Perspective level, there are 3 sublevels of −1, 1, and 1+ in Level I Notions, for example.

Additionally, it should be recognized that scores on many of the ChemQuery questions can range to a much higher level on the Perspectives framework with such results achievable by at least some university-level general chemistry students, as has been demonstrated independently by first-year college students in this study.[18] So even though we chose some of the easier items for illustrative purposes in the Perspectives description for this paper, it was the high school students’ ability, or understanding, of chemistry that was limiting their scores.

■ IMPLICATIONS FOR ASSESSMENT

This analysis sheds some light on the level of conceptual understanding of high-scoring students on the AP chemistry exam and raises a call for uniting conceptual and problem solving to develop deep understanding and complex reasoning. Specifically, this study finds that problem solving correlates to conceptual understanding but that conceptual understanding lags behind algorithmic problem solving. The implication is that as a community we need to look at student learning more coherently by integrating what is known about how students learn with the organization of the domain of chemistry.

There is a large body of research that describes how students learn, but it has limited application to chemistry and does not seem well incorporated into chemistry assessments such as the AP chemistry exam as described here.[20−22] Additionally, much of the research on student learning has been deconstructed into various knowledge types such as problem solving, conceptual change, transfer, representations, and reasoning, rather than looking at the integration of knowledge such that conceptual reasoning informs problem solving and can be measured with students.[23] Talanquer speaks to these dichotomies and describes 10 in chemistry education, including: "misconceptions versus scientific conceptions", "submicroscopic versus symbolic", and "algorithmic versus conceptual".[24] Research in chemistry education not only falls into these distinct categories but tends to be topic-specific to the domain of chemistry, such as describing student understanding of the mole,[25−28] thermodynamics,[29−31] or chemical change, for example.[32−34] Our concern is that when the research looks at just one knowledge type or one specific topic, an overall understanding of how students learn chemistry is missing, which is where we seem to find ourselves today in chemistry education.

Yet as stated in the NRC Report, the primary goal of advanced study in any discipline should be for students to achieve a deep conceptual understanding of the discipline's content and unifying concepts,[6] which seems to imply thinking more deeply about what the big ideas or unifying concepts are. The current hypothesis in science education is that organizing subject domains such as chemistry around "big ideas" will improve student learning.[15,35] In the AP chemistry redesign, six "Big Ideas" have been described:[36]

1. Structure of matter
2. Bonding and intermolecular forces
3. Chemical reactions
4. Kinetics
5. Thermodynamics
6. Chemical equilibrium

In ChemQuery we have posited three big ideas at the high school level: Matter, Change, and Energy. Talanquer has a different approach that emphasizes chemistry as an applied science organized around these essential questions: What is it? (analysis); How do I make it? (synthesis); How do I change it? (transformation); and How do I explain it? (modeling).[37] Depending on how one interprets the science framework and Next Generation Science Standards, there can be two−four "big ideas" in chemistry. From physical science core ideas there is PS 1, Matter and its implications; PS 3, Energy; and possibly PS 2, Motion and Stability: Forces and interactions with matter and energy—these are crosscutting concepts that are to be taught from K−12 and across the disciplinary core ideas.[15,35] However, this is all hypothetical. No one really knows whether the "big ideas" will help students develop explanatory power, or which "big ideas" are the most fruitful; however, the movement to integrate and unify ideas has gained momentum in K−12 science education and may prove fruitful to help bridge the gap between problem solving and conceptual understanding.

One challenge then is to create good questions that capture the void between problem solving and the correlated conceptual understanding of the unifying concepts to better measure how student understanding of chemistry develops. One suggestion from our research would be to use simple numbers such as ChemQuery Item 6 and Item 88 that emphasize conceptual understanding of chemistry with simple mathematical reasoning. Another suggestion is to measure and value the development of correct explanations found in student responses to chemistry problems, which may include scoring answers beyond right or wrong to value incorrect answers that are shown to lead to deeper understanding over the course of instruction. For example, student responses in Level II Recognition contain correct chemistry that may not be fully developed to produce a right answer.

Additionally, if unifying or big ideas are proposed to be useful, then assessment questions within an organizing framework like the Perspectives of Chemists could test whether, indeed, students are developing understanding with explanatory models of chemistry implied by each proposed organizing framework.

Some of the goals as stated in the NRC report for the redesign of AP courses that are pertinent in light of this research are for students to achieve a deep conceptual understanding of the discipline's content and unifying concepts; to use assessment to certify mastery; and to ensure that the assessments measure conceptual understanding and complex reasoning.[6] Furthermore, as recognized by the NRC and the AP chemistry redesign efforts, the AP chemistry exam has wide-ranging effects on teachers, curriculum, sequencing, and students,[6,19] based on the common tenet in education that "what you test is what you get". The assumption is that assessment drives instruction, which implies that the change in the AP redesign will drive changes in instruction in order to align curriculum, instruction, and assessment to facilitate "deep learning".

■ SUMMARY

This research highlights the need to better unite conceptual and problem-solving abilities to describe student understanding. Currently, we find that traditional problem-solving questions are easier for students to answer correctly with conceptual understanding lagging a step behind, even though the conceptual questions seem easier to us in chemistry education. Specifically, conceptual understanding as measured by the ChemQuery assessment system, and scores on the AP chemistry exam are correlated yet reveal different interpretations of students' ability or understanding of chemistry. Our data analysis finds that the level of conceptual understanding is only at Level II Recognition, even for those students scoring 5 on the AP exam, which suggests that many of even the higher-performing students are much more competent at problem solving in chemistry than they are in answering the "simpler" conceptual questions.

It seems that we are measuring two distinct yet highly correlated types of understanding that will behoove us as a community to more effectively address. This is not new, but the correlation between the two types of understanding is. Deep understanding that links problem solving to conceptual understanding may develop later and may require instruction and assessments that measure and value how correct explanations to chemistry problems develop. By recognizing the power of the AP chemistry exam to drive changes to improve student understanding of chemistry, our hope is that this research can contribute to the mission of the AP chemistry exam redesign efforts with the ultimate goal of helping *all* students learn chemistry.

AUTHOR INFORMATION

Corresponding Author

*E-mail: jennifer.claesgens@nau.edu.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This material is based upon work supported by the U.S. National Science Foundation under Grant Nos. ESI-9730634, DUE-0125651, and 0737056 (CCLI).

REFERENCES

(1) College Board. Setting a Policy for AP Chemistry. http://www.collegeboard.com/prod_downloads/ipeAPC/apchempg060309_51087.pdf (accessed Spring 2006). (Note: Because the data used in comparison was specific to the 2006 AP Exam, the access date used during research is provided. Unfortunately, the data set is no longer available on the current Web site.)

(2) Lythcott, J. *J. Chem. Educ.* **1990**, *67*, 248−252.

(3) Bodner, G. *J. Chem. Educ.* **1991**, *68*, 385−388.

(4) Hestenes, D. *Phys. Teach.* **1993** *30*, 141−158.

(5) McCloskey, M. Naïve Theories of Motion. In *Mental Models*; Gentner, D., Stevens, A. L., Eds.; Erlbaum Associates: Hillsdale, NJ, 1983; pp 299−324.

(6) National Research Council. *Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools*. Center for Education, Division of Social Sciences and Education, National Academies Press: Washington, DC, 2002.

(7) Claesgens, J.; Scalise, K.; Wilson, M.; Stacy, A. *Sci. Educ.* **2009**, *93*, 56−85.

(8) College Board. Learn about AP. http://www.collegeboard.com/student/testing/ap/sub_chem.html (accessed Jan 2014).

(9) Claesgens, J.; Scalise, K.; Wilson, M.; Stacy, A. *Assess. Update* **2008**, *20*, 6−9.

(10) Scalise, K.; Claesgens, J.; Wilson, M.; Stacy, A. *Chem. Educ. Res. Pract.* **2006**, *7*, 170−184.

(11) Watanabe, M.; Nunes, N.; Mebane, S.; Scalise, K.; Claesgens. *J. Sci. Educ.* **2007**, *91*, 683−709.

(12) Strauss, A.; Corbin, J. *Basics of Qualitative Research: Grounded Theory Procedures and Techniques*; Sage Publications: Newbury Park, CA, 1990.

(13) Wilson, M. *Constructing Measures: An Item Response Modeling Approach*; Erlbaum: Mahwah, NJ, 2005.

(14) Smith, C. L.; Wiser, M.; Anderson, C. W.; Krajcik, J. Implications of Research on Children's Learning for Standards and Assessment: A Proposed Learning Progression for Matter and the Atomic-Molecular Theory. *Measurement* **2006**, *4*, 1−98.

(15) National Research Council. *A Framework for K−12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, DC, 2011.

(16) Krippendorff, K. *Content Analysis: An Introduction to Its Methodology*; Sage: Beverly Hills, CA, 1980.

(17) Johnstone, A. *J. Comp. Assisted Learn.* **1991**, *7*, 75−83.

(18) Scalise, K.; Stacy, A.; Douskey, M.; Daubenmire, P.; Lim, T.; Sinapuelas, M.; Rao, S. ChemQuery as a Formative Assessment Instructional Strategy for Chem 1. Paper presented at the 240th ACS National Meeting, Boston, MA, 2008.

(19) Spencer, J.; Hnatow, J. AP Chemistry: Course and Exam Review. In *Chemistry in the National Science Education Standards*; Lowery-Bretz, S., Ed.; American Chemical Society: Washington, DC, 2008; pp 109−117.

(20) National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academy Press: Washington, DC, 2001.

(21) Bransford, J. D., Brown, A. L.; Cocking, R. R. *How People Learn: Brain, Mind, Experience, and School: Expanded Edition.*; National Academy Press: Washington, DC, 1999.

(22) Driver, R.; Squires, A.; Rushworth, P.; Wood-Robinson, V. *Making Sense of Secondary Science: Research into Children's Ideas*; Routledge: London, 1994.

(23) Glaser, R. *Am. Psychol.* **1984**, *39*, 93−104.

(24) Talanquer, V. *J. Chem. Educ.* **2012**, *89*, 1340−1344.

(25) Astudillo, L.; Niaz, M. *J. Sci. Educ. Tech.* **1996**, *5*, 131−140.

(26) Furio, C.; Guisasola, J. *Int. J. Sci. Educ.* **2000**, *22*, 1285−1304.

(27) Gabel, D.; Sherwood, R. *J. Res. Sci. Teach.* **1984**, *21*, 843−852.

(28) Staver, J.; Lumpe, A. *J. Res. Sci. Teach.* **1995**, *32*, 177−193.

(29) Boo, H. *J. Res. Sci. Teach.* **1998**, *35*, 569−581.

(30) Greenbowe, T.; Meltzer, D. *Int. J. Sci. Educ.* **2003**, *25*, 779−800.

(31) Teichert, M.; Stacy, A. *J. Res. Sci. Teach.* **2002**, *39*, 464−496.

(32) Hesse, J.; Anderson, C. *J. Res. Sci. Teach* **1992**, *29*, 277−299.

(33) Johnson, P. *Int. J. Sci. Educ.* **2002**, *24*, 1037−1054.

(34) Yarroch, W. *J. Res. Sci. Teach.* **1985**, *22*, 449−459.

(35) Achieve. Next Generation Science Standards; National Research Council: Washington, DC, 2013. http://www.nextgenscience.org/ (accessed Jan 2014).

(36) College Board. AP Chemistry Course Details. https://apstudent.collegeboard.org/apcourse/ap-chemistry/course-details (accessed Jan 2014).

(37) Talanquer, V. Chemistry Education: Ten Facets To Shape Us. *J. Chem. Educ.* **2013**, *90* (7), 832−838.