JOURNAL OF CHEMICAL EDUCATION

Article

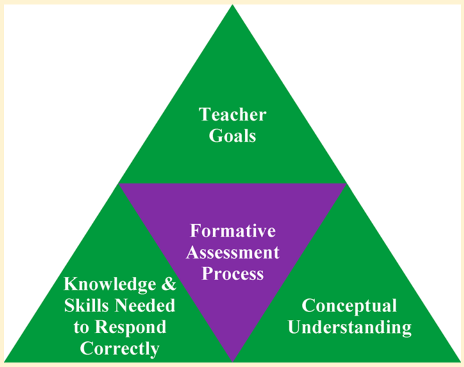
pubs.acs.org/jchemeduc

Formative Assessment in High School Chemistry Teaching: Investigating the Alignment of Teachers' Goals with Their Items

Benjamin Sandlin, Jordan Harshman, and Ellen Yezierski*

Department of Chemistry & Biochemistry, Miami University, 363 Hughes Laboratories, 651 East High Street, Oxford, Ohio 45056, United States

Supporting Information



ABSTRACT: A 2011 report by the Department of Education states that understanding how teachers use results from formative assessments to guide their practice is necessary to improve instruction. Chemistry teachers have goals for items in their formative assessments, but the degree of alignment between what is assessed by these items and the teachers' goals has not previously been investigated. This understanding of teachers' goal-setting will identify strengths and limitations in their formative assessment processes. In this qualitative project, we have characterized this alignment of assessment items with learning objectives with data collected from 19 high school chemistry teachers from 10 states. These teachers participated in semi-structured interviews describing their goals and use of a teacher-developed formative assessment they had administered to their classes. The teachers provided 41 items which were analyzed for this study. To evaluate the content knowledge and skills required for successful completion of each item, the items were evaluated by one researcher by listing statements and equations required to demonstrate an understanding and solve each individual problem. A different member of the team who conducted the interviews listed the assessment goals as stated by the teachers for each item to characterize the degree teacher's goals are emulated by their items. The author team evaluated the alignment of teachers' assessment goals to what is assessable by items. The results discussed show that teachers who address conceptual chemical phenomena in their goals, in line with the NGSS standards of using concepts to assess student understanding, are more likely to have a goal that is assessable by their assessment items than those who do not assess concepts. From findings and prior literature, we recommend that teachers use conceptual goals in conjunction with problem solving goals to assess student understanding, particularly with items that require computation.

KEYWORDS: *High School/Introductory Chemistry, Chemical Education Research, Testing/Assessment, Stoichiometry, Thermodynamics, Gases*

FEATURE: Chemical Education Research

INTRODUCTION

Formative assessments are of paramount importance to gauge learning in many chemistry classrooms. Formative assessment is defined more by what one does with the results versus a specific form of assessment.[1] In particular, formative assessments are primarily used to provide feedback to both teachers and learners.[2] Summative assessments are primarily aimed toward evaluation of students' learning, making them an assessment *of* learning as opposed to *for* learning.[1] Evaluating the alignment of assessment items with teacher goals provides useful measurements for teachers to determine students' understanding of concepts,[2−5] particularly chemical phenomena. Without this alignment, inferences that teachers make from formative assessment data cannot inform their teaching in a valid manner.[6,7]

As an example, if a teacher determines that students do not have accurate mental particulate models of gas behavior (aligned with the kinetic molecular theory) by evaluating correctness of a quantitative Charles' Law exercise, this teacher would be basing his/her conclusions on evidence (student responses to the exercise) that is not aligned with the learning goal. Furthermore, misaligned assessments lead to false positives (e.g., "students understand stoichiometry" when they do not) and false negatives (e.g., "students don't understand relationships among gas law variables" when they do) that have a detrimental effect on teaching and learning chemistry.

The need for conceptual understanding of chemistry has been recognized by the Next Generation Science Standards, NGSS,[8] which stresses the importance that students "...develop their understanding of the four core ideas in the physical sciences. These ideas include the most fundamental concepts from chemistry and physics...."[8] Without a complete understanding of material, superficial understandings are sometimes evaluated using formative assessments and demonstrate that students have a level of understanding much higher than they actually do.[9] In alignment with the NGSS, conceptual

ACS Publications

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX

understanding is different than problem solving, which may be as simple as replacing a variable with a number.

Over the past years, instructional sensitivity has been highlighted as an integral criterion in determining the extent of alignment of assessments with their intended content goals.[10] Instructional sensitivity examines the degree to which student performance and quality of instruction align.[11] When teachers use items that are sensitive to their instruction, they can draw more valid conclusions about their instruction based on student performance data from the items.[10] To date, no one method for evaluating instructional sensitivity has emerged as a best practice, but several authors provide broad guidelines for appropriate methods.[12−14] Work on instructional sensitivity highlights the importance of evaluating the extent to which assessments and assessment items are able to inform instruction.

The alignment of assessment items and assessment purpose is one consideration in an existing assessment process known as data-driven inquiry (DDI).[6,15−18] DDI is a process defined by four steps (italicized): *Defining a goal* for assessments, *collecting evidence* from students, analyzing data to *make conclusions* about teaching and learning, and finally *taking pedagogical actions* that address or support the conclusions made. The ways in which these steps are carried out are described in detail in a literature review.[19] Additionally, Harshman and Yezierski argue the DDI literature does not consider disciplinary content. Coffey et al.[8] make a similar claim in previous formative assessment research, including the aspect of aligning learning objectives with assessments has "focus[ed] attention to strategies and techniques...generally presumed traditional notions of disciplinary content as a body of information...with an emphasis on terminology...."[19] Without full consideration to the rich disciplinary content of chemistry, it is not always possible to align assessments with the learning objectives.

Of notable exception to this lack of disciplinary substance, a study by Tomanek, Talanquer, and Novodvorsky researched the factors involved when secondary science teachers chose formative assessment tasks.[20] In this study, in-service and preservice teachers were given three probes, each of which asked them to choose from a variety of assessment tasks and why they chose as they did. Interestingly, they found that both types of teachers used the factor "provides evidence of understanding/misunderstanding (makes knowledge/understanding/thinking visible)" infrequently in the selection of a task in all three probes. Similarly, the factor that specifically focused on alignment between the task, curriculum, teaching, and learning objectives was only observed in one of the three probes, indicating that alignment of these features was not readily observed. While the study by Tomanek et al.[20] provides evidence that secondary science teachers do not always consider alignment using prompts that were largely biology-focused, the degree to which chemistry teachers focus on alignment is still unaddressed.

We propose to focus on teachers as the unit of analysis (as opposed to students) as we investigate the effect of alignment of assessment items on teachers' analyses of assessment data. In both cases of assessment types (formative and summative), teachers collect data that can be used to make data-driven conclusions and decisions. As such, our study is of relevance to both forms of assessment, although we place a heavy emphasis on the formative type, as it is more important to analyze data directly from classrooms and not restrict the tools to one type of assessment. The research question that guided this study was: To what level are the assessments of high school chemistry teachers instructionally sensitive (aligned with their learning goals)? We believe that this assessment process and considerations discussed will be of high interest and value to chemistry educators who wish to draw valid conclusions about student learning and instruction based on everyday assessment results.

■ THEORETICAL FRAMEWORK

The theoretical and methodological framework that guided this study is the four-step process of DDI and assessment theories therein.[21] This study began after collecting data for a larger qualitative study by Harshman and Yezierski,[19] which was based on a DDI framework.

■ METHODS

This qualitative study incorporated multiple levels of data collection including interviews, generating summaries, and determining content and skills that individual items assessed. Each of these steps, listed as subheadings, are described in detail below. First, interviews were conducted and transcribed verbatim. Second, summaries of these transcripts were produced as a means to reduce the data into a more manageable form. Lastly, propositions (statements) of requisite knowledge were generated for items that teachers provided for the interviews. These three documents made up the data corpus used in our analysis.

Interviews

The second author conducted interviews with 19 high school chemistry teachers regarding their formative assessment practice as described by a previous qualitative study.[19] This larger study and the one described here were both approved by the local IRB and compliant with all regulations. The locations, years of experience, and Science Teachers Efficacy and Beliefs Instrument[22] scores for the teachers involved can be found in the Supporting Information. The schools from which teachers were invited were drawn from a random selection stratified by scores on the science portion of the National Assessment of Educational Progress standardized exams. Each teacher participated in an interview lasting approximately 1 h. In Phase I of the interviews, teachers were asked for their general views on assessment, such as providing a definition and how they use assessment in their chemistry classes to evaluate student learning and their own teaching. In Phase II, teachers analyzed their own classroom assessments (defined as a homework, lab, classroom activity, or quiz) which had to be something that the teacher had already given to the class, collected from the students, and evaluated in whatever way they deemed fit. Questions asked by the second author aligned with the process of DDI, such as what the goal of the assessment was, if the students had met the goal, why the teacher believes they had or had not met the goal, and what action they would take as a result.

Finally, the second author chose 2−4 individual items from the provided assessment and inquired about those questions specifically in Phase III. For each individual question chosen, teachers were asked about their goals, conclusions, evidence, and actions, similar to Phase II. In addition, teachers were asked specific questions about the chemistry content embedded in each assessment item (e.g., "Why did you choose a 3:1 molar ratio in this stoichiometry problem?"). After all interviews were transcribed verbatim, a total of 41 individual assessment items

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX

from the 19 teachers were gathered for analysis. The complete interview protocol and an example assessment are provided in the Supporting Information.

Summaries

For each of the teacher-generated assessment items that teachers were asked about specifically in the interviews, a detailed summary of the corresponding teacher's goals, evidence, conclusions, and actions was created from their responses. Great care was taken to ensure that the words used in the summaries were the same words used by the teachers, minimizing misinterpretations due to different connotations of words. In general, goals were summarized first by looking at the responses after the interview question, "What is/are your goal(s) for the item?" as well as other places in the transcript about that particular item. The same occurred for conclusions, evidence, and actions by looking at responses to their respective questions in the interview. Once the second author generated this summary for all of the teachers' items, approximately 5% of the items were summarized by the corresponding author. Comparison of these items yielded minimal and often grammatical differences. The first author then reviewed all other items a second time, adjusting mostly for how much information to provide based on the comparison of the two authors' work. See the Supporting Information for an example transcript, assessment item, and its corresponding summary.

Propositions of Requisite Knowledge

The first author answered/responded to the chemistry content of each of the 41 items that were obtained from the interviews. This was done to gain an understanding of what was required for a student to (a) understand, and/or (b) solve the item. In many cases, these two sets of requirements were different. As a result, two lists of propositions (statements) were generated for each of the 41 items: One set of propositions described what is *required to simply solve the item correctly* ("*propositions for solving*") and one set described what is *required to understand the concepts behind the item* ("*propositions for understanding*"). See Figure 1 for examples of the two types of propositions that were generated for each item.

The propositions for understanding described the chemical theories and principles required to gain a full understanding of the phenomenon, while the propositions for solving did not involve any more detail than what was absolutely necessary to respond to the item correctly.

These propositions and equations were compiled and sent to the second and third authors for content validation. Propositions were negotiated until all authors agreed that the propositions listed represented the appropriate knowledge requirements for each teacher-generated assessment item. The first author (who was not previously familiar with the interviews) was responsible for the original development of the propositions so that the other authors would not skew the propositions of the items by knowing what the teachers' goals were, as they were familiar with the previous interviews conducted. Propositions for all items can be found in the Supporting Information.

Comparisons

Using the propositions and summaries, we were able to determine the alignment of teachers' goals and conclusions (data in the summaries) with what was assessed by the items (propositions). For each of the 41 items, we compared the summaries (capturing teachers' descriptions viewed through a DDI lens) to the propositions with the aim of generating descriptors that effectively characterized the features of the data corpus. These descriptors were revised and tested as we applied them to more and more items. Through multiple iterations and discussion among the research team, categories became a list of codes that we systematically applied to each item. Originally, the authors set out to determine if the propositions that the first author created aligned or did not align with the teachers' ideas. However, reducing this to a dichotomy severely limited our understanding of data and the conclusions we could make. Alternatively, we present the features that we examined in Table 1, which represent our coding scheme.

E. Essay - Explain both quantitatively and qualitatively. If 5.0 mole of B is allowed to react with 2.0 moles of A, in the reaction $3A + 2B \rightarrow A_3B_2$, explain the role of the limiting reactant in determining the amount of A_3B_2 that can be produced, and the amount of excess reactant left over. (10 pts)

Propositions Required for Understanding:

1. A limiting reactant will stop a reaction from progressing in the energetically favored direction.
2. An excess reactant will not affect a reaction from progressing.
3. An excess reactant will be left over after a reaction.
4. Stoichiometry is used to take known amounts of a compound or element and, through coefficient conversion factors, determine the amount of an unknown compound or element.

Propositions Required for Solving the Problem:

1. In stoichiometry, one must convert given moles of a compound by the mole ratio of unknown reagent/known reagent coefficients to result in moles of the known reagent.
2. By comparing the stoichiometric yield of both reactants to the product, one can determine that the limiting reagent is the one that yields the least amount of product.
3. One can determine the amount of excess reactant remaining by completing a second stoichiometric step of converting the limited amount of product to the excess reactant, then subtracting out the amount that is needed to react with the limiting reactant to ascertain the final product.

Figure 1. Example item provided by a teacher and the corresponding propositions of understanding and solving.

Some codes originated as themes from the interviews collected, others arose out of the propositions that were created, and some were generated to apply to all items with the aim of capturing broader themes that could emerge. Each item was coded according to the binary outcome in Table 1. An example of this process is given in Figure 2, which shows the outcome based on the summary displayed in Figure 1.

■ RESULTS AND DISCUSSION

The initial data retrieval resulted in summarizing the frequency of the codes used on all 41 assessment items. A frequency table of all items coded can be found in Table 2.

This table demonstrates several key properties of the data set. First, more than double the items presented *multiple goals* as did a *single goal*. Second, the number of items that did not align *conceptual propositions* outweighs those that did. This would suggest that, in general, our sample of teachers have many learning goals that often do not align with conceptual underpinnings of chemical phenomena. Furthermore, only

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 1. Coding Criteria for Comparisons

Code	Description
Single vs Multiple Concepts Assessed	This was the determination of whether or not an individual item assessed multiple concepts/skills or just one. This consideration arose as a theme from the interviews conducted. The following criteria were used to determine if an item assessed multiple concepts: If multiple equations and/or unit conversions were required to get the correct answer (and if they were not the target of the item), if the goal(s) for the item could not be described by one "idea", or if teachers used the word "solve", we assumed the teacher was seeking to determine multiple concepts or skills in the item (unless the teacher has elaborated on what solving the problem entails). As an example, a stoichiometry problem that requires students to write/balance an equation assesses that as well as mole-to-mole ratios (multiple concepts) whereas giving information in moles of reactant, providing a balanced equation, and asking to convert to moles of product just assesses mole-to-mole ratios (single concept). Those that do meet the above criteria were categorized as assessing a single concept. *Outcome: Single or Multiple.*
Specific vs Ambiguous Goals	This was the determination of whether or not a teacher's goal for a specific item was specific or ambiguous, and was a reoccurring theme that came from the interviews conducted. To be labeled a specific goal, teachers must propose a mechanism to solve or answer the question by identifying each single concept/skill needed. An ambiguous goal would provide no mechanism for how to solve or answer the item and was often indicated by words such as "understand", "know", or "recognize" being used in the goal. *Outcome: Specific or Ambiguous.*
Goals Aligned with Conceptual Proposition	This was the determination of whether or not a teacher's goals were aligned with the propositions for understanding. This code came from the propositions created. To be coded as being aligned with the proposition for understanding, teachers had to express the same idea, not necessarily in the same words, that represent the chemical phenomena in the propositions. *Outcome: Aligned or Not Aligned.*
Goals Aligned with Problem Solving Propositions	This was the determination of whether or not the teacher's goals aligned with the propositions for solving. Creation of the propositions generated this code. Similar to the previous code, its focus was more on the essence of the teacher's ideas as opposed to the exact wording used. *Outcome: Aligned or Not Aligned.*
Algorithmic	In order to determine if an item involved "plugging and chugging", a colloquial term used to describe the process where students follow a mathematical algorithm to arrive at a solution (sometimes with little thought given to the concepts behind the algorithm), we used the following criteria: If the item called for use of an equation, numbers and units were present in the item, and/or the item was solvable without chemical knowledge, then the item itself was coded as algorithmic. *Outcome: Yes or No.*
Algorithmic Recognition	This was coded when a teacher identified, either directly or indirectly, that an item was an algorithmic exercise. Often this was determined by the teachers' reliance on the words "calculate", "determine amount", "how much", or "solve" in describing their goals. Additionally, if the teacher made it appear as though the students could successfully respond to the item without chemical knowledge, they were coded as recognizing an algorithmic exercise. *Outcome: Yes or No.*
Assessable Goals	This determined whether or not the goals listed by the teacher were assessable by the item in a valid manner. The criteria for whether or not an item could assess a teacher's goal was largely context-oriented and is exemplified in Figure 2. This term arose out of the comparisons from the propositions and the teacher's goals. *Outcome: Assessable and Not Assessable.*

Amy #23
If 5.0 mole of B is allowed to react with 2.0 moles of A, in the reaction $3A + 2B \rightarrow A_3B_2$, explain the role of the limiting reactant in determining the amount of A_3B_2 that can be produced, and the amount of excess reactant left over. (10 pts)

Code	Outcome
Single vs. Multiple	Multiple
Specific vs. Ambiguous	Ambiguous
Aligned Conceptual Goals	Not Aligned
Aligned Solving Goals	Not Aligned
Plug and Chug	Not Plug and Chug
Aligned Plug and Chug	Aligned
Assessability	Yes

Figure 2. An example of the coding process from the summaries of teachers, as compared to the data of the items.

Table 2. Number of Items Coded under Each Criterion

Code	Outcome	Frequency
Goals	*Single*	13
	Multiple	28
Goal Specificity	*Specific*	23
	Ambiguous	18
Conceptual Propositions	*Aligned*	15
	Not Aligned	26
Solving Propositions	*Aligned*	20
	Not Aligned	21
Algorithmic	*Yes*	5
	No	36
Identified as Algorithmic	*Yes*	5
	No	36
Assessability	*Assessable*	30
	Not Assessable	11

half of the items were aligned with the propositions for understanding, and 11 goals for the items were not assessable. This further suggests that, aside from not aligning conceptually, teachers did not necessarily account for all content considerations and/or included features not validly assessable by the item, both of which are discussed further below.

In addition to the findings from examining overall frequencies of codes, we also found several themes that address our research question, particularly how holding *single* or *multiple goals* affected the assessment, the ambiguity of goals related to assessability, what goals were associated with *algorithmic* items, the importance of conceptual understanding in goal-setting, and the implications of setting nonassessable goals.

Single versus Multiple Goals

While having *multiple goals* for a question does not intrinsically mark the question as being "bad," *multiple goals* can lead to disparities in teachers' abilities to interpret information. Teachers that ask questions with *multiple goals* need to ensure the question addresses each individual goal. Of all items that were analyzed, 28 were identified as having *multiple goals*; out of these 28, 15 were identified to *have ambiguous goals* with nine of those being classified as *not assessable*. This result is further highlighted by the fact that, in the entire study, only 11 goals were *not assessable*, indicating that the vast majority of nonassessable goals were present with teachers who set many goals for an item. Teachers who address *multiple goals* from a question, therefore, exhibited a proclivity to assess non-assessable goals.

Ambiguity of Goals versus Assessability

Since *ambiguous goals* by nature deter teachers from making specific conclusions based on assessment results, we investigated relationships between those goals coded as *ambiguous* and those coded as *nonassessable*. We hypothesized that goals that were *nonassessable* would be as such because of their lack of specificity, as this was found with this sample of teachers[19] and suggested in other research.[6] Our study found that an *ambiguous goal* stated by a teacher does not necessarily lead

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX

to a nonassessable item. In our study, out of the 11 *nonassessable* goals, 5 had *specific goals*.

The teacher in Figure 3 set the *ambiguous goal* of "reviewing atomic structure" (we presume she meant molecular structure)

Bond	Average Bond Dissociation Energy (kJ/mol)
C–H	415
O=O	495
C=O	799
O–H	463

$$CH_4(g) + 2\ O_2(g) \rightarrow CO_2(g) + 2\ H_2O(g)$$

The standard free energy change, $\Delta G°$, for the reaction above is -801 kJ at 298 K.

a) Use the table of bond dissociation energies to find $\Delta H°$ for the reaction above.

Stated Goal: "To see if they understood the atomic structure concept from Chem I, cause they would've had to remember the number of bonds and the energy for each bond... I'm seeing how much they remember Lewis Structures back in Chem I... I also wanted to see, it kind of divided my students into those that really did the homework... The only ones that really worked in depth on the homework and practice got A's."

Figure 3. An example of an item with a nonassessable goal.

and the *specific goal* of seeing who did the homework and who did not. These goals do not address the specific concept (energy as a state function) or skill (using bond dissociation energies in proper molar ratios to calculate enthalpy of formation) that students must use to solve the item. Seeing which students are doing their homework relies on the assumption that students who perform well on one assignment did so because they have been keeping up with the homework, which while possibly correlated, is not validly determined by examining student scores. The incorporation of these considerations is required for teachers to validly draw inferences about their students' learning, as is outlined in DDI[23] and instructional sensitivity.[10] This further exposes the necessity to consider the content of items in goal-setting; a problem like the one in Figure 3 assesses much more content than just molecular structure.

Algorithmic and Recognition

Out of the 41 items, only five were identified as *algorithmic* by the authors. From those five, every teacher that was interviewed concerning the *algorithmic* item was able to identify it as an *algorithmic* item, meaning each teacher recognized it solely assessed calculations. Conversely, 37 items were identified as not being *algorithmic*. Out of these 37 items, 32 were also identified by the teachers as not being *algorithmic*, suggesting that five items that were not identified as *algorithmic* by the authors were identified as *algorithmic* by the teachers that formulated them. This demonstrates that there are a number of teachers that failed to recognize that their items assessed more than just algorithmic calculations.

For example, Mandisa brought a question that we did not identify as not an *algorithmic* problem, but she did. This suggests that the teachers sometimes disregarded the need for conceptual understanding regardless of the fact that the problem necessitated some of a conceptual understanding of chemical phenomena. This is further supported by evidence provided in the next section.

Conceptual Understanding versus Solving the Problem

Table 3 shows a contingency table for items which were associated with propositions for *problem solving* and *conceptual understanding* from the teachers compared to the propositions generated by the authors. This shows that, generally, if a teacher can identify *conceptual understanding* goals, s/he could also identify *problem solving* goals. However, the opposite is not supported by our results. This again provides evidence that teachers sometimes view understanding of chemistry in terms of ability to solve problems correctly, an issue that has been raised from very early research in chemistry education.[24,25]

Table 3. Alignment of Teachers' Goals with Propositions for Problem Solving and Conceptual Understanding

		Problem Solving: Aligned	Problem Solving: Not Aligned	
Concepts	Aligned	14	1	*15*
	Not Aligned	6	20	*26*
		20	*21*	**41**

As an example, Figure 4 shows an item provided by Jess. To solve the exercise, Jess claimed that a student must "use stoichiometry," but she did not identify the conceptual issues that stoichiometry represents, which is the meaning of mole-to-mole ratios considering the law of conservation of matter and particulate-level modeling of substances. This lack of attention to the conceptual domain of chemistry can limit teachers in making instructional decisions based solely on ability to solve exercises and follow algorithms, which ignores a wealth of other valuable information contained within student solutions to these items.

9. Hydrogen gas can be prepared in the following manner: $Fe(s) + H_2SO_4 \rightarrow FeSO_4 + H_2$. The volume of a balloon filled with this gas is 2.0×10^4 L. What mass of iron is needed to react with an excess of acid? Assume the reaction takes place at room temperature, at sea level, and has a 100% yield.

Figure 4. Goal listed by the teacher was to use stoichiometry to solve for the mass of iron needed to react.

Nonassessable Goals

Of the 11 goals the authors determined to be not assessable, none were identified as *algorithmic*. This indicated that *algorithmic* items, for the most part, are assessable from teachers' perspectives. However, the occurrence of approximately one-quarter of the items having nonassessable goals is alarming. In Table 4, we have examined the types of nonassessable goals exhibited in this data and supplied the implication to teaching and learning.

■ LIMITATIONS

The authors recognize that there are limitations to the above study. One such limitation is the fact that the first author did

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX

Table 4. Categories of Nonassesable Goals, Implications, and Examples of Each Category

Category of Nonassessable Goal	Implication of the Category	Iterations (Teacher, Item)
Reviewing an equation, conversion, common value (boiling point of water), or concept.	This category of nonassessable goals is not to suggest that review is a poor use of class time. Rather, it is to indicate that the goal of "giving a review" is not assessable. Teachers who use items with a similar goal in mind neglect to address concepts or solutions to problems in their goals—topics that are assessable.	Britt, 2 Mandisa, 20 Mandisa, 18 Natalie, 2a Kari, 2b
Determining if students can recognize an equation, common value or concept.	Using an item to determine if a student can merely recognize a topic does not examine if the student *understands* the topic; this specific distinction demonstrates the danger of using items to test student recognition.	Mandisa, 20 Francesca, 8 Francesca, 12 Michael, 1 Michael, 2 Matthew, 14
Goals that do not require student responses.	Addressing a nonchemical topic, such as determining if the student did homework or not, is detrimental to the item. These nonassessable goals are set apart from the previous two because they lack chemical merit.	Bart, 6 Matthew,14 Natalie, 2a

not design/conduct the interviews and therefore could not probe specifically about alignment issues outside of what the second author probed. However, sufficient evidence was gathered in order to make conclusions. Any knowledge of the interviews could have biased the first author, so this was also a means of reducing threats to validity. A second limitation is that the propositions for understanding and solving created by the first author and validated by the second and third were not standardized across several institutions. However, given the chemistry expertise of the three authors, it is likely that other content experts would agree with these propositions.

CONCLUSIONS

This study indicates that the teachers examined fell into various categories indicating instructional sensitivity of their items. There were four types of goal-to-item relationships:

Multiple Goals: Teachers commonly have *multiple goals* in mind when they ask questions. The data indicate that teachers can use multiple goals to asses multiple topics in one item.

Ambiguous Goals: Teachers who use *ambiguous goals* cannot specifically determine what students do and do not understand. While the *ambiguous goal* may be assessable, for example, if the student knows how to do stoichiometry, the goal cannot define the performance standard. As such, the goal is not useful in designing/selecting assessment items and inhibits making valid conclusions about student learning or instruction from assessment data.

Identification of *Problem Solving*: The goals that fall into this group have underlying concepts that can be addressed by the item, but the teacher does not address them in their goals.

Well-aligned: The data show that teachers who craft well-aligned items are likely guided by conceptual and computational goals, which will lead to more effective assessment items. Considering these four groupings identified, we concluded that several of our sample of chemistry teachers described learning goals that were not detailed enough to provide appropriate sensitivity to instruction, and more so misaligned conceptually based learning goals and assessment items.

IMPLICATIONS

Teachers who have *multiple goals* must be cautioned that each of these goals needs to embrace an assessable topic to be effective. Moreover, to optimize student understanding, the goals should specifically assess a concept to ascertain the students' understanding of the chemical concepts. Special care must be taken to assess these *conceptual goals* rather than simple problem solving. Teachers who use ambiguous goals should also use caution, in that a simple skimming over of topics in a chemistry course can lead to student confusion and lack of understanding. Well-designed items address conceptual goals that, in the process, also encompass mathematical concepts. The items that should be crafted are items that conform to this well-designed definition.

In accordance with the NGSS, conceptual goals are required to be assessed to reliably and accurately assess student understanding on most items. This study highlights the importance of setting these goals and verifies the NGSS in their standard. The goals that are assessable, more often than not, are those that address conceptual issues first. Because of this, teachers should use conceptual goals to gauge student understanding and not rely solely on mathematical algorithms to determine if a student truly grasps a topic. Looking forward, teachers should create items that assess conceptual goals in conjunction with mathematical or problem solving goals to determine student understanding. Such a shift in assessment practices strongly aligns with recommendations in the NGSS.

ASSOCIATED CONTENT

Supporting Information

Information about the demographics, example transcript and summary, propositions for all items, and interview protocol. The Supporting Information is available on the ACS Publications website at DOI: 10.1021/acs.jchemed.5b00163.

AUTHOR INFORMATION

Corresponding Author

*E-mail: yeziere@miamioh.edu:.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors acknowledge the participating teachers for providing items and interviews for this project, as well as the Yezierski and Bretz CER Groups at Miami University.

■ REFERENCES

(1) William, D. *Formative assessment and contingency in the regulation of learning processes*. Paper presented at the Annual Meeting of American Educational Research Association. Philadelphia, PA, April 2014.

(2) Pellegrino, J. Assessment of science learning: Living in interesting times. *J. Res. Sci. Teach.* **2012**, *49* (6), 831−841.

(3) Irons, A. *Enhancing Learning through Formative Assessment and Feedback*; Routledge: New York, NY, 2008.

(4) Anderson, L. W. *Classroom Assessment: Enhancing the Quality of Teacher Decision Making*; L. Erlbaum Associates: Mahwah, NJ, 2003.

(5) Taylor, G. R. *Informal Classroom Assessment Strategies for Teachers*; Scarecrow Press, Inc.: Lanham, MD, 2003.

(6) Hamilton, L.; Halverson, R.; Jackson, S.; Mandinach, E.; Supovitz, J.; Wayman, J. *Using Student Achievement Data to Support Instructional Decision Making (NCEE 2009−4067)*; National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education: Washington, DC. 2009.

(7) Datnow, A.; Park, V.; Wohlstetter, P. *Achieving with Data: How High-Performing School Systems Use Data to Improve Instruction for Elementary Students*; University of Southern California, Center on Educational Governance: Los Angeles, CA, 2007.

(8) NGSS Lead States. *Next Generation Science Standards: For States, By States*; The National Academies Press: Washington, DC, 2013.

(9) Coffey, J. E.; Hammer, D.; Levin, D. M.; Grant, T. The missing disciplinary substance of formative assessment. *J. Res. Sci. Teach.* **2011**, *48* (10), 1109−1136.

(10) Ruiz-Primo, M. A.; Li, M.; Wills, K.; Giamellaro, M.; Lan, M. C.; Mason, H.; Sands, D. Developing and evaluating instructionally sensitive assessments in science. *J. Res. Sci. Teach.* **2012**, *49* (6), 691−712.

(11) Kosecoff, J. B.; Klein, S. P. *Instructional sensitivity statistics appropriate for objectives-based test items*. Paper presented at the Annual Conference of the National Council on Measurement in Education, Chicago, IL, 1974.

(12) Popham, W. J. Instructional insensitivity of tests: Accountability's dire drawback. *Phi Delta Kappan.* **2007**, *89* (2), 146−155.

(13) Burstein, L. *Conceptual Considerations in Instructionally Sensitive Assessment*; Technical Report 333; Center for the Study of Evaluation, National Center for Research on Evaluation, Standards, and Student Testing, Graduate School of Education & Information Studies, University of California: Los Angeles, CA, 1989.

(14) Polikoff, M. S. Instructional sensitivity as a psychometric property of assessments. *Educ. Meas.: Issues and Practice* **2010**, *29* (4), 3−14.

(15) Calfee, R. C.; Masuda, W. V. Classroom assessment as inquiry. In *Handbook of Classroom Assessment. Learning, Adjustment, And Achievement*; Phye, G. D., Ed.; Academic Press: San Diego, CA, 1997.

(16) Brunner, C.; Fasca, C.; Heinze, J.; Honey, M.; Light, D.; Mandinach, E.; Wexler, D. Linking data and learning: The Grow Network study. *J. Educ. Stud. Placed Risk* **2005**, *10* (3), 241−267.

(17) Halverson, R.; Prichett, R. B.; Watson, J. G. *Formative Feedback Systems and the New Instructional Leadership*; University of Wisconsin: Madison, WI, 2001.

(18) Gallagher, L.; Means, B.; Padilla, C. *Teachers' Use of Student Data Systems to Improve Instruction, 2005 to 2007*; U.S. Department of Education, Office of Planning, Evaluation and Policy Development, Policy and Program Studies Service: Jessup, MD, 2008.

(19) Harshman, J.; Yezierski, E. J. Guiding teaching with assessments: High school chemistry teachers' use of data-driven inquiry. *Chem. Educ. Res. Pract.* **2015**, *16*, 93−103.

(20) Tomanek, D.; Talanquer, V.; Novodvorsky, I. What do science teachers consider when selecting formative assessment tasks? *J. Res. Sci. Teach.* **2008**, *45* (10), 1113−1130.

(21) Harshman, J.; Yezierski, E. J. Assessment data-driven inquiry: A review of how to use assessment results to inform chemistry teaching. *Sci. Educator*, submitted for publicarion, **2015**.

(22) Bleicher, R. E. Revisiting the STEBI-B: Measuring self-efficacy in preservice elementary teachers. *Sch. Sci. Math.* **2004**, *104* (8), 383−391.

(23) Means, B.; Chen, E.; DeBarger, A.; Padilla, C. *Teachers' Ability to Use Data to Inform Instruction: Challenges and Supports*; Office of Planning, Evaluation and Policy Development, U.S. Department of Education: Washington, DC, 2011.

(24) Nurrenbern, S. C.; Pickering, M. Concept learning versus problem solving: Is there a difference? *J. Chem. Educ.* **1987**, *64* (6), 508−10.

(25) Bodner, G. M.; Herron, J. D. Problem solving in chemistry. In *Chemical Education: Research-Based Practice*; Gilbert, J. K., de Jong, O., Justi, R., Treagust, D. F., van Driel, J. H., Eds.; Kluwer Academic Publishers: Dordecht, The Netherlands, 2002.

DOI: 10.1021/acs.jchemed.5b00163
J. Chem. Educ. XXXX, XXX, XXX−XXX