CHEMICAL REVIEWS

Review

Cite This: Chem. Rev. 2018, 118, 6053−6087

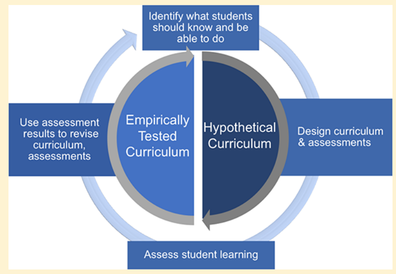
pubs.acs.org/CR

Chemistry Education Research—From Personal Empiricism to Evidence, Theory, and Informed Practice

Melanie M. Cooper* and Ryan L. Stowe

Department of Chemistry, Michigan State University, 578 S. Shaw Lane, East Lansing, Michigan 48824, United States

ABSTRACT: This Review of Chemistry Education Research (CER) provides an overview of the development of research in chemistry education from the early days, when ideas about how to teach chemistry and help students learn were guided by practitioner wisdom, to current research that is based on theories of learning and provides evidence from which to make arguments about improving teaching and learning. We introduce the dominant learning theories that have guided CER over the years and attempt to show how they have been integrated into modern research in chemistry education. We also provide examples of how this research can be used to inform the development and use of educational materials. Because CER literature is vast, we chose to limit the research we reviewed to those studies that help us answer three driving questions: (1) What should students know and be able to do with that knowledge? (2) How will we know that students have developed a coherent and useful understanding of chemistry? (3) What evidence do we have about how to help students develop a deep and robust understanding of chemistry?



CONTENTS

Received: January 10, 2018
Published: June 12, 2018

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

1. INTRODUCTION

In 1924, in the first issue of *The Journal of Chemical Education* (JCE), W. A. Patrick wrote "What Kind of Research is Essential to Good Teaching?",[1] in which he contended that "research work" in teaching is characterized by combing "the subject of chemistry from end to end for facts and for methods of exposition that will make such facts live and real to his students". Patrick offers no precise definition of what making facts "live" really means, how one might measure it, or why it matters. His sole source for the article seems to be "gut instinct". In fact, the entirety of the first issue of JCE contains no references whatsoever. Indeed, the majority of the chemistry education literature in the early 20th century was of a kind with Patrick's piece, that is, it consisted largely of opinion pieces and laboratory exercises. While the opinions and experiences of chemistry faculty are important, they are typically not systematic investigations into a topic of interest guided by data collected and analyzed via appropriate methods (i.e., they are not research). Of course, this begs the question: What is chemistry education research (CER)? Simply stated, CER is concerned with teaching and learning in chemistry, investigated through a variety of qualitative and quantitative methods.[2] Researchers in chemistry education explore a broad range of areas, including the mechanisms by which students construct understanding of chemistry principles and the barriers that impede construction; the development of instruments that measure understanding, attitudes, identity, and other affective constructs; how evidence about student learning can be incorporated into curriculum design; and how to measure the impact of curricular transformations. It is important to emphasize that CER, and discipline-based education research (DBER) more broadly, is grounded in the "priorities, worldview, knowledge, and practices" of the discipline being studied.[2] However, while a deep knowledge of chemistry principles is vital, it is not sufficient: an understanding of the methods and principles of science education, educational psychology, and cognitive science are also necessary. Chemistry education researchers are uniquely qualified to think deeply about what expertise in chemistry looks like and how it might be developed.

Given the breadth of the field, we have elected to focus this Review on the evolution of the peer-reviewed chemical education literature from a body of work that was once dominated by personal opinions to a rich corpus of studies grounded in theories of learning and supported by evidence. Because of the vast number of studies in CER over the years, we have chosen to exclude studies on precollege chemistry education, learning in the laboratory, and studies on the affective domain (such as identity, motivation, expectations, value, and interest). This is certainly not meant to imply that these studies are unimportant but rather speaks to the need to focus Reviews such as this one lest whole areas of scholarship be shortchanged. We cannot hope to do justice to all CER studies in the literature in one Review. To guide our discussion, we will focus principally on the literature surrounding three questions that have been of interest to the CER community for many years:

(1) What should students know and be able to do with that knowledge?

(2) How will we know that students have developed a coherent and useful understanding of chemistry?

(3) What evidence do we have about how to help students develop a deep and robust understanding of chemistry?

The focus of the chemistry education community has changed over time; not all of these questions have been investigated at all points in the history of CER. Additionally, as will become increasingly apparent as we progress, our three guiding questions are interrelated in significant ways. If we want evidence that students can accomplish the learning objectives of a chemistry course (question no. 1), we must figure out how to assess whether or not they have developed a coherent and useful understanding of the discipline (question no. 2). To aid students in constructing such an understanding, it would be helpful to know about, and understand, the unique challenges the chemical sciences pose for the learner and to have evidence-based strategies to combat these challenges (question no. 3). It should be stated that we aim to present both studies relevant to teaching and learning in chemistry and also current ideas about the best models for curriculum development, assessment, and practice informed by this research. For example, the question of "what students should know and be able to do" after a particular course in chemistry cannot be directly answered by a study or a set of studies—it requires reflective and analytical work that is informed by evidence but is not itself research.

1.1. Nature of Evidence in CER Studies

CER studies differ from those in chemistry because (for the most part) the systems being studied are composed of people rather than molecules and are therefore subject to the vagaries of human behavior. Over the years methodologies developed by learning scientists have informed CER studies, allowing the collection of data from which evidence-based arguments could be made. Because CER is focused on how students learn about the behavior of atoms and molecules rather than directly studying the atoms and molecule themselves, the theories that guide the research, the experimental methodologies, and the data-collection instruments must differ from those utilized in traditional chemistry research.

Most CER studies begin with a theoretical lens through which to view and explain the data that are collected. The studies themselves may be quantitative, in which large data sets are collected and analyzed, or qualitative—typically studies of smaller groups, often involving student interviews. For example, consider a study in which an intervention, that is guided by a particular learning theory, is administered to students. A quantitative study might involve a statistical analysis in which student scores on a quiz or exam after this intervention are compared to those of students who did not receive the intervention. Such a study can provide evidence about whether the intervention had an effect but not about the mechanism underpinning this effect. We may infer that the reason the intervention had an effect is because it was designed to build on a particular learning theory, but evidence for our inferred mechanism would need to arise from a separate, well-designed qualitative study. For example, students from the intervention might be interviewed and their responses coded to identify emergent themes that could provide evidence of the mechanism by which the intervention acted.

As virtually all CER involves some aspect of student behavior, there is a great deal of inherent complexity in the systems being examined that must be considered when designing studies. For example, students receiving an instructional intervention (the "treatment" group) may not be (indeed cannot be) the same as those in the control group. To make a *causal* claim that a

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

particular intervention is responsible for an outcome, the two groups of students being compared must be as similar as possible and must have been exposed to similar conditions (apart from those varied in the experiment). The "gold standard" for this type of experiment is a randomized control-treatment design,[3] that is, students are randomly assigned to either treatment or control groups. For most CER researchers this is, to say the very least, impractical. Furthermore, such studies may provide the illusion of rigor where none exists, particularly for studies on small populations without replication.[4] It is, however, possible to match groups of students by their background (sex, major, grades, etc.) to try to minimize differences between the two groups—that is, researchers can use a quasi-experimental design. The type of instrument used to detect the outcome of an experiment is also crucial to the integrity of CER studies. Such instruments are typically assessment items that are administered to the students to determine the impact of a given intervention. As we will see later, the assessments used to gather evidence may or may not be appropriate to the task at hand, that is, they may not provide valid and reliable data. That being said, appropriately designed mixed-methods studies involving both quantitative and qualitative components can provide evidence of both how an intervention may change outcomes and why it is effective.

This Review will not (and cannot) be a comprehensive listing of all CER papers ever published (of which there are thousands). Instead, we will focus on papers that embody the community's evidentiary standard for claims related to our three questions. It should be noted that standards of persuasive evidence in CER have changed over time and are now generally accepted to be those described in the National Research Council (NRC) Report on Discipline-Based Education Research (DBER)[2] outlined in Box 1.

Box 1. Characterizing the Strength of Conclusions Supported by the Evidence Base[2]

A Limited Level of Evidence

- Few peer-reviewed studies of limited scope with some convergence of findings or convergence with nonpeer-reviewed literature or with practitioner wisdom.

A Moderate Level of Evidence

- A well-designed study of appropriate scope that has been replicated by at least one other similar study. Often such evidence will include both quantitative and qualitative data; OR
- A few large-scale studies (e.g., across multiple courses, departments, or institutions) with similar results; OR
- A moderate number of smaller-scale studies (e.g., in a single course of section) with general convergence but possibly with contradictory results. If the results are contradictory, more weight must be given to studies that reflect methodological advances or a more current understanding of teaching and learning, or are conducted in more modern learning environments.

Strong Evidence

- Numerous well-designed qualitative and/or quantitative studies, with high convergence of findings.

There are a number of prior reviews[5−7] that have surveyed the CER literature from 2000 to 2013 and several other reviews in the literature that focus on particular areas of CER research including teaching and learning of thermodynamics[8] and chemical kinetics,[9] problem solving behavior in organic chemistry,[10] and spatial reasoning.[11] We direct readers to these reviews if an exhaustive listing of papers is desired.

Before we move to the research studies from the CER literature, we will provide a brief historical overview of the progression of chemistry education research from individual chemist's opinions about teaching and learning to a sophisticated research enterprise grounded in theories of learning and evidence-based arguments.

2. SHORT HISTORY OF CHEMICAL EDUCATION

2.1. Era of "Personal Empiricism", 1880−1964

Although collegiate courses in chemistry have been taught in the United States since 1795, most universities did not have multiple classes in the chemical sciences until the beginning of the 20th century.[12] In the late 1800s, recognition that an understanding of chemistry could contribute to increased productivity in industrial and agricultural arenas prompted a rapid expansion in the number of chemistry courses offered at the university level.[12,13] Promoters of early chemistry courses emphasized the practical nature of the discipline and the importance of cultivating "experimental methods for the development of mental power".[14] Because of their emphasis on practical applications, primordial classes in chemistry were almost entirely descriptive in nature and focused on processes that had obvious industrial and/or agricultural applications (e.g., fertilizer synthesis or analysis).[14]

Early courses in chemistry were not guided by a research base on how people learn. In fact, there were those in the late 1800s who doubted that teaching and learning could be informed at all by systematic investigations. Harvard philosopher Josiah Royce, in his 1891 article "Is There a Science of Education?", wrote that he was unwilling "to apply so pretentious and comforting a name as 'Science' to any exposition of the laborious and problematic art of the educator."[15] The vast majority of those who wrote on chemistry education in the period from 1880 to 1960 derived their advice solely from their own experience, a sort of "personal empiricism" if you will. For example, in his 1893 article in the *Journal of the American Chemical Society* "How Chemistry is Best Taught", Charles F. Mabery contends (with no supporting evidence) that "the guiding star to successful teaching in chemistry is the personality and enthusiasm of the instructor."[14] The prime objective of such successful teaching was "maintaining a profound interest." It is almost certain that many instructors today ascribe to this approach, although there is still little evidence to support the notion that an enthusiastic and likeable instructor is necessarily more adept at producing students who have developed a robust and useful understanding of chemistry.

It should also be mentioned that Mabery, like his contemporaries, was not concerned with students developing an understanding of atoms and molecules that could be used to predict and explain chemical phenomena; the courses he advocated for (and likely taught) were aimed at imparting technical skills and knowledge. Thus, he was very concerned with what students should "know and be able to do with their knowledge" (question no. 1), but his answers were driven by the needs of industry. This emphasis on the descriptive, rather than theoretical, aspects of the discipline continued into the 1960s.[16] Indeed, there are a vast number of other papers published in JCE and elsewhere that make claims related to our three questions of interest without any supporting evidence. We include this study

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

by Mabery primarily as a historical baseline by which to frame future work.

Although personal empiricism predominated in the early days of university chemistry education, there are a few examples of data-driven investigations. For instance, in 1923, Cornog and Colbert surveyed faculty from 27 different institutions about what they taught in general chemistry, examined the content of the eight most widely used chemistry textbooks, and analyzed final examination questions obtained from surveyed institutions. From these sources, their goal was to gain a sense of what instructors believed students should know and be able to do after a course in general chemistry (question no. 1), what materials they used, and what evidence about student learning was being elicited by the examinations.[17] According to these authors, most teachers reported that they emphasized the theoretical basis for chemistry, but upon examination, theory was neither reinforced by the textbooks used nor tested in the examinations, both of which focused principally on descriptive material. As the authors note, examinations "set the goal of student effort, to a degree teachers little dream of." Thus, they argue if dissonance exists between the stated course focus and the exam focus, the implicit messages conveyed by exams will almost certainly win out. Cornog and Colbert's advice in this was prescient and has been echoed many times since, both in the science education literature[18] and education literature[19−23] more broadly. That is, the types of assessment used in a chemistry course send a strong message to students about what is actually important. This is a subject we will return to as we consider question no. 2.

A significant shift in the curricular emphasis of college chemistry courses occurred in the late 1950s when Cornell chemistry professors Michell J. Sienko and Robert A. Plane released the book *Chemistry: Principles and Properties*.[24] Most prior texts placed significant emphasis on descriptions of how one might prepare various substances of industrial or agricultural import (such as elemental phosphorus or steel). In fact, when he began his teaching career, Alex Johnstone observed that "a set of model notes dated 1900... were identical to those I was working from in 1960!"[13] In a break with tradition, Sienko and Plane assembled their book, more-or-less from scratch, using the primary chemistry literature. The essence of their philosophy was to present accurate theory before descriptive topics in order that students "learn the most useful things".[25] When calibrating what to include in their book, Sienko and Plane made use of a "nodometer system": if students fell asleep in class when a topic was discussed, coverage of that topic was deemed too unclear or esoteric.[25] The book resultant from their efforts was extremely popular with faculty across the country, and four subsequent editions were published. A great many contemporary college chemistry texts feature roughly the same collection of topics compiled by Sienko and Plane 60 years ago.

Chemistry: Principles and Properties was assembled (and endlessly emulated) without any empirical basis for its efficacy in promoting desirable learning outcomes. Sienko and Plane, themselves trained expert chemists, assumed that an assemblage of material consistent and pleasing to them (and capable of keeping students awake) would also appear consistent and meaningful to students. This assumption was not buttressed by any literature in science education. Johnstone writes, "most of us who were in the business of curriculum development at that time fell into the pit of assuming what excited us would inevitably excite our students... few knew enough about how young people learn to avoid the pitfalls of being carried away by mature enthusiasms."[16] Curriculum developers during the "era of personal empiricism" were certainly concerned with what students should know and be able to do with that knowledge (question no. 1), but they paid little heed to the necessary coupling of this question with the second of our three: how can you determine if students actually understand what you want them to understand and can use it productively? Informed by modern literature regarding how students learn to construct and reconstruct their conceptual understanding[26−31] and the nature of novice understanding of chemical concepts,[32−36] we can now say that answering question no. 2 is exceedingly difficult (as will soon become apparent).

Systematic investigations into teaching and learning in chemistry were almost nonexistent until the midpoint of the 20th century. Many opinions were expressed and contested in the literature, and "research in education" was occasionally brought up (sometimes with negative connotations),[37,38] but "chemistry education research" was not regarded as a subfield of chemistry and there were no full-time practitioners of it. The complexities of developing a conceptual understanding of atoms and molecules that is explanatory and predictive were not studied with any depth or rigor.

2.2. Piaget, Vygotsky, Ausbel, and Novak: Constructivism, Social Constructivism, and Meaningful Learning, 1964−2000

To get at the third of our questions of interest (What evidence do we have about how to help students develop a deep and robust understanding of chemistry?), it is necessary to understand how students develop molecular-level understanding. The first empirical work on the development of childhood understanding to permeate the chemistry education literature was that of Swiss clinical psychologist Jean Piaget.[39,40] Piaget observed that children think quite differently from adults and wrote of the slow development of knowledge and intelligence over time, as new external stimuli (from coursework or simply exploring the world) are assimilated into the child's existing knowledge.[41−44] Piaget, building from the work of John Dewey,[45] laid the empirical foundations for constructivism: the notion that new ideas emerge from pre-existing ones and are constructed by the student, not transferred from the instructor. More precisely, Piaget claimed that new stimuli both provoke a response based on an individual's prior experience and also alter their knowledge structure (a process called "accommodation" in Piaget's work). Constructivism has been an extraordinarily powerful idea in chemistry education research. Piaget's proposed mechanism by which new ideas emerge from old (by a process of disturbing existing knowledge structures and subsequently re-equilibrating them)[46] was an important contribution to research on how students develop appropriate conceptual understanding and, with modifications, forms the basis for more recent learning theories that will be discussed shortly.[47]

One of the first to report Piagetian-informed work in the chemistry education literature was the theoretical physicist Robert Karplus.[40,48] In his manuscript "Chemical Phenomena in Elementary School Science",[39] Karplus proposed that the "personal natural philosophy" of young students must be carefully developed from a base of past experiences and knowledge and poses several questions for further research: "How much background must the children accumulate before an explanation based on structural unity makes sense to them, and how many impressions can the children retain without confusion before these are organized by a structural explanation?" Neither

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

of these important questions were investigated with rigor for many years.

In the early 1970s, teachers of undergraduate chemistry began to take notice of Karplus' work in the elementary arena and discuss its potential for informing college instruction. In 1974 Beryl Craig made the audacious claim that teaching might be enhanced by talking with one's students: "it would seem that the teaching and compiling of any course would be improved if the teacher of that course would spend time with students, finding out what they understood about the concepts they were studying."[49] Stated differently, figuring out whether or not students understand the concepts of a course and can use their understanding is important and merits study (question no. 2!). Craig understood that students are not "blank slates" who receive knowledge from their instructors with perfect fidelity but instead are individuals with pre-existing knowledge structures that perceive new information in light of what they already know. Further, she appreciated that a full appraisal of students' knowledge structures cannot be easily obtained from traditional tests. This sentiment represents a fairly major sea-change from the prevailing assumption that instructors could extrapolate student's understanding from their scores on a quiz or exam.

The response of the chemistry education community to Piaget's constructivism can perhaps best be summarized by George Bodner's piece in the *Journal of Chemical Education*, "Constructivism: A Theory of Knowledge".[50] Bodner succinctly summarized Piaget's ideas by saying: "knowledge is constructed in the mind of the learner."[50] That is, learning is an active process in which the learner integrates new knowledge into pre-existing structures in an attempt to assemble a working model of reality. Thus, "learning" is not creating mental copies of what is true (which cannot be done with certainty) but instead is assembling knowledge structures that "get the job done" (whatever the job happens to be).

Piaget focused attention solely on the mind of the individual (and is for that reason often known as a "cognitive constructivist"). In the 1980s, the science education community increasingly realized that influences beyond a given individual play a powerful role in shaping individual and communal values, identity, and norms. During this period, the work of Russian psychologist Lev Vygotsky (1896−1934) was "rediscovered". Vygotsky viewed knowledge as "shaped by micro- and macro-cultural influences" that "evolves through increasing participation within communities of practice."[51] Thus, his focus was not on the internal structure of concepts and the individual student but instead the social contexts within which learning occurs (Vygotsky is thus identified as a "social constructivist"). Vygotsky stressed the importance of meaningful, contextualized, "whole" activities (such as conducting scientific investigations) as fundamental units of instruction rather than rote skill building bereft of context. He also introduced the "zone of proximal development" (ZPD), which is defined as "the distance between the actual developmental level (of a student) as determined by independent learning and the level of potential development as determined through learning under the guidance of a teacher, or in collaboration with more capable peers."[52] As we will see, Vygotskys' ideas were prescient and form the basis of the collaborative and group work that has become so prevalent in the last 20 years.[53−55]

Like Piaget and Vygotsky, American psychologist David Ausubel (1918−2008) was influential in refining constructivism. He defined "meaningful learning" as learning that fosters an organization of knowledge into relational frameworks. Rote learning, by contrast, involves memorizing a fact without any notion of how that fact relates to other stored knowledge. Ausubel believed that for meaningful learning to occur three criteria must be met: (1) The student must have appropriate prior knowledge to which the new knowledge can be connected. (2) The new knowledge must be perceived as relevant to this prior knowledge. (3) The student must *choose* to make these connections (i.e., to learn meaningfully).[56−58] This third criterion is an acknowledgment that it is ultimately the choice of the student about how and what to learn.

Joseph Novak elaborated on Ausubel's notion of "meaningful learning" (and constructivism more generally) by proposing that "meaningful learning underlies the constructive integration of thinking, feeling, and acting."[59,60] Thus, according to Novak, knowledge had to be connected across three domains for meaningful learning to occur. These three domains (thinking, feeling, and acting) were more precisely defined as the cognitive domain (what is to be learned), the affective domain (the attitudes and motivations of the learner), and the psychomotor domain (the physical activities that accompany learning). Clearly this theory is most applicable to learning in the laboratory but could also apply to learning activities in lecture settings that involve student interaction with handheld models, for example.

These theories are obviously not mutually exclusive and indeed there are commonalities: they all propose that knowledge is constructed in the mind of the learner, and they emphasize that, for useful and transferrable learning to occur, connections must be constructed between what is known and what is to be learned, and that students have agency and may choose not to learn. In point of fact, modern theories of learning are typically amalgams of cognitive and social constructivist views. For those of us who wish to develop learning environments that help students construct coherent and useful knowledge of chemistry, it is important to be mindful of constructivism in all its forms.

3. DEVELOPMENT OF EXPERTISE AND THE CHARACTER OF EXPERT KNOWLEDGE

In 2001, the National Research Council published a synthesis of available research on learning entitled *How People Learn*.[61] As we are interested in student development of more coherent and useful (i.e., expert-like) knowledge of chemistry, the differing character of expert and novice knowledge structures as defined by this consensus study is worth exploring. While we are not expecting our students to become experts by the end of a course, or even by the end of a degree program, we are expecting them to develop expertise and more expert-like knowledge structures. By using and understanding the principles that characterize expert-like knowledge, we will be able to guide our discussions of theory-driven chemistry education research. Characteristics of expert knowledge, as discussed in *How People Learn*, are shown in Box 2.

As educators, we are particularly concerned with the idea that experts' knowledge is organized and contextualized. That is, experts have linked coherent knowledge structures that allow knowledge to be used in different ways and that help facilitate integration of new information (i.e., learning). The efficacy of approaches designed to foster the development of more expert-like knowledge structures will be discussed under question no. 3. Note that expertise as described by *How People Learn* is very much consistent with constructivism as described earlier. Interconnected, contextualized, expert-like knowledge structures can be thought of as the product of careful cognitive construction that occurs in and is affected by one's local context and community.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Box 2. Five key characteristics of expert knowledge as described by *How People Learn*[61]

1. Experts notice features and meaningful patterns of information that are not noticed by novices.
2. Experts have acquired a great deal of content knowledge that is organized in ways that reflect a deep understanding of their subject matter.
3. Experts' knowledge cannot be reduced to sets of isolated facts or propositions but, instead, reflects contexts of applicability: that is, the knowledge is "conditionalized" on a set of circumstances.
4. Experts are able to flexibly retrieve important aspects of their knowledge with little attentional effort.
5. Although experts know their disciplines thoroughly, this does not guarantee that they are able to teach others.

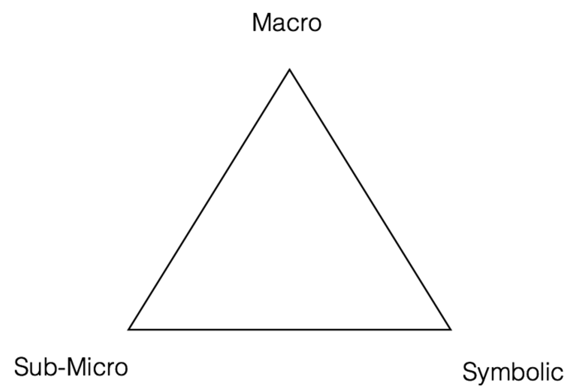


Figure 1. Johnstone's triangle, representing multilevel thought in chemistry. Adapted with permission from ref 65. Copyright 1991 John Wiley & Sons, Inc.

3.1. Development of Expertise in Chemistry: Johnstone's "Triplet"

While there is much to be learned from the domain-general learning theories described above, there are particular challenges to understanding the world at the molecular level. Various "levels of thought" are used by expert chemists to describe, explain, and communicate phenomena, some of which are more intuitively accessible than others. As macroscopic beings, we can often readily describe and measure various aspects of materials we can hold and see. Thinking through explanations at the molecular level requires us to reason using entities far beyond our experience and is thus markedly less intuitive (as Herron and colleagues observed in the 70s).[62] If we wish to communicate our molecular-level reasoning to others, we are faced with the need to learn and use whole specialized symbolic languages that describe atoms, molecules, and their transformations. These different "levels of thought" were defined and described by Alex Johnstone in his 1982 paper "Macro- and Microchemistry".[63] The "descriptive and functional level" of thought is engaged when we experience, observe, and describe macroscopic phenomena. Use of symbols and formulas to describe chemical entities and their transformations relies on "representational-level" thought. Explanations about the submicroscopic behavior of atoms and molecules leverage "atomic and molecular-level" thinking. Johnstone argued that, while it may be relatively easy for trained chemists to consider multiple levels simultaneously, it is not at all straightforward for novices to do so. That is, a trained chemist may observe a macroscopic phenomenon such as ice melting and produce a symbolic representation to describe it ($H_2O(s) \rightarrow H_2O(l)$), while maintaining a mental model of the water molecules gradually gaining enough energy to overcome their intermolecular interactions and begin to move. In contrast, a novice will almost certainly have a more fragmented understanding of the same melting process. For example, some students provide explanations and models of melting or boiling in which bonds (and not intermolecular interactions) break, to produce new chemical species,[32] often while maintaining that when bonds break energy is released.[64]

Multilevel thought in chemistry can be represented as a triangle with each "level" (e.g., macro, micro, and symbolic) inscribed on a corner (Figure 1).[65] Students often run into trouble, Johnstone argued, when instruction is focused at the middle of the triangle from day one (i.e., on all levels simultaneously), or when the instruction takes place primarily at the symbolic level.

Johnstone's three levels (and the triangle depicting them) have come to be colloquially referred to as the "chemistry triplet".[66,67] Both the simplicity and power of the triplet struck a chord with the chemistry education community, and it has been used (and reimagined) a great many times. Among the more popular reimaginings of Johnstone's triangle is that in which the three vertices of the triangle characterize levels of representation rather than levels of thought.[68,69] Through this lens, the three levels became phenomenological (i.e., representations of experienced phenomena), model (i.e., representations supportive of qualitative explanations of phenomena, such as molecular-level drawings), and symbolic (i.e., representations supportive of quantitative explanations of phenomena such as equations and formulas).[69] A detailed discussion regarding the many facets of the chemistry triplet (and how one might conceptualize them into a more complex knowledge space) is provided by Talanquer.[67] Amid so many "reimaginings" of the chemistry triplet, it is easy to lose sight of Johnstone's reason for conceiving it in the first place: he sought to illustrate the unreasonable cognitive demand imposed by chemistry instruction that focuses on all three thought levels simultaneously.

Johnstone also introduced ideas about information processing to the chemical community that still resonate today. He adapted a model,[70] based on research from cognitive science, in which events are screened by a perception filter, and then information that passes the filter enters the short-term memory or working memory space, where "thinking" occurs (Figure 2). Processing tasks may be aided by information retrieved from long-term memory and/or further information admitted through the perception filter.[71] Unfortunately, an individual's working memory has very limited space to accommodate and work with information.[72−75] As Johnstone noted:[76] "It is a limited shared space in which there is a trade-off between what has to be held in conscious memory and the processing activities required to handle it, transform it, manipulate it, and get it ready for storage in long-term memory store. If there is too much to hold, there is not enough space for processing; if a lot of processing is required, we cannot store much."

In light of the limited capacity of humans' working memory,[72−75] one might well wonder why experts seem able to process much more information in their domain of specialty than novices. As it turns out, experts consider, store, and retrieve information in a different manner than novices. For example, a novice might perceive a chemical equation as a random jumble of letters and numbers and therefore process it as many disparate

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

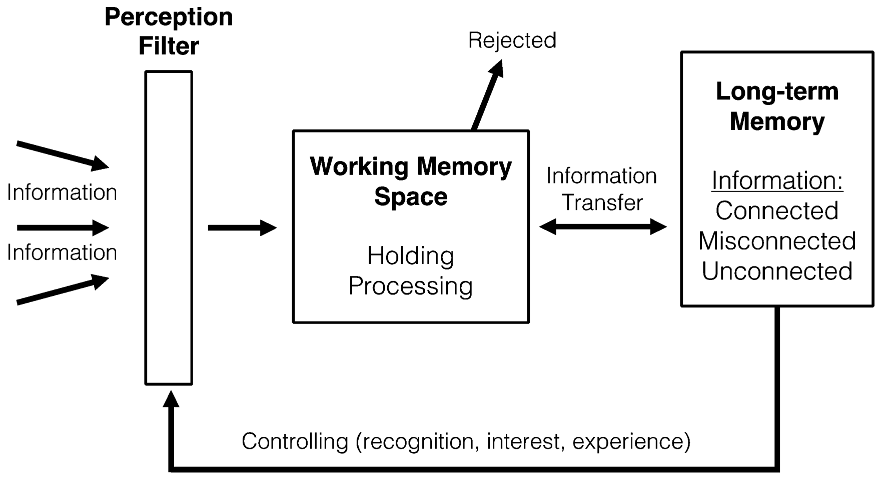


Figure 2. Johnstone's information processing model. Reproduced from ref 77 with permission from The Royal Society of Chemistry.

factoids. Considered in such a way, even a very simple equation might overwhelm one's working memory. If, instead, a student was able to ascribe meaning to the various parts (chunks) of a chemical equation, much more efficient processing becomes possible. Related knowledge can be consolidated in our minds into "chunks" and processed as a single unit (instead of multiple fragments) as we develop expertise. Speaking generally, expert knowledge structures are much more organized, connected, and contextualized than the fragmented structures characteristic of novices.[61] For this reason, experts find information integration and retrieval much less taxing than novices.

Clearly, if we want students to have any chance at understanding chemistry, we must work to promote effective consolidation of information into chunks. It will not do to have students completely overwhelmed by every chemical equation they encounter. Thus, in a general sense, many of the challenges to teaching and learning chemistry can be managed by helping students construct an expert-like knowledge structure that incorporates the three thought levels of Johnstone's triplet (question no. 3). However, as Johnstone himself noted,[78] while experts operate at the center of the triangle, beginning students cannot. Materials that provide multiple representations can overwhelm students' working memory and can be counter-productive. For example, Kozma and Russell[79] showed experts and novices multiple representations of phenomena and asked the subjects to group them in ways that made sense. While both experts and novices produced groups, the experts made large groups that were characterized by conceptual underpinnings, while novices focused on surface features (an observation that reoccurs time and again in studies on the differences between novices and experts[80,81]) and produced more groups. That is, the novices actually had a more fragmented and less meaningful outcome than the experts.

3.2. Conceptual Change

To support the development of more expert-like knowledge frameworks in students, it is important to understand something about the nature of students' existing knowledge and how it might be changed or built upon. In the early 1970s and 1980s, the predominant view was that the understanding of beginners contained a great many fully formed "false beliefs" or "misconceptions". Piaget argued that, when dissonance arises between our pre-existing ideas and our experiences, our conceptual schemas can be tweaked to fit the new sensory data. Given that utility rather than absolute truth guides the compilation and integration of knowledge under a constructivist paradigm, individuals may have very different views on what is and is not correct. The ability to assemble wildly different "working models of reality" is, Bodner argued, responsible for some students' developing misconceptions. In this early work, "misconceptions" were defined as "concepts or ideas which from the point of view of the average professional... lead to unacceptable solutions or answers to questions or problems in the context of a course."[50] They arise, it was argued, because they "fit" our experiences well enough to be integrated into our knowledge framework. Implicit in this argument is the notion that novices can develop naïve theories elaborated with all sorts of wrong ideas that work "well enough" to not be rejected. A tremendous amount of work in the CER community (continuing to the present day) has been dedicated to cataloging and characterizing students' incorrect beliefs.[82−114] Work in this vein reflected an emergent realization that students enter the classroom with prior beliefs and knowledge that must be considered and dealt with through instruction. For this Review, we are less interested in lists of misconceptions than in their character, organization, how they arise, and approaches to reconstructing more appropriate ideas. Therefore, we make no attempt to exhaustively document all of the many aspects of chemistry that students struggle with and instead refer the reader to the references cited.

In the early days of the "misconceptions movement", false beliefs were viewed as coherent entities to be overcome. The process of "overcoming" was thought to be pragmatically driven, that is, students would hold onto current ideas unless they had cause to reject them. Strike and Posner established one of the first, and most influential, models for so-called "rational change" of concepts.[30] They claimed that replacement of one "central concept" with another comes about when:

- Currently held concepts (or slight variations of these concepts) are unable to solve problems of interest;
- The student can see how "experience can be structured by a new concept sufficiently to explore the possibilities inherent in it";[30]
- The new conception seems to have the capacity to solve problems that were unsolvable using prior concepts; this is also consistent with other knowledge held by the individual;
- The new concept is capable of being built upon in useful ways.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Implicit in this model of radical restructuring was the notion that "central concepts" are stable, coherent entities that have been adopted and sustained by a desire to comprehend experience (as well as other, perhaps tangential, commitments). Further, it was suggested that productive conceptual change in the science classroom would come about through the initiation of a "cognitive conflict" in which the inadequacies of current conceptions are made apparent. That is, if students were provided with a phenomenon that could not be explained with their incorrect ideas, then they would replace the old ideas with a new, ideally more accurate, conception.

The idea of cognitive conflict as a mechanism of conceptual change is attractive to many instructors, because the wholesale replacement of one idea by another during instruction points to a simple solution. However, many misconceptions are not amenable to this "treatment", as anyone who has tried to change students' ideas about the energy changes during bond formation and bond breaking can attest.[115,116] More modern models of conceptual change represent misconceptions as concrete manifestations of a more complex cognitive structure. Some scholars adhere to a view that a great many novice ideas are organized according to "coherent underlying and organizing presuppositions",[117] while others hold that incorrect student responses to a question are generated in situ from finely grained intellectual resources.

The notion that deeper implicit conceptions underpin learners' ideas arose from examination of the naïve theories of infants. Investigators found that, well before babies could articulate their ideas, they expressed relatively adult expectations about the physical world (e.g., objects do not move on their own, solid objects cannot move through one another). From these observations it was hypothesized that infants have an implicit "framework theory" consisting of presuppositions including "that 'rest is the natural state of physical objects' and 'motion needs to be explained'".[118] Enrichment of naïve framework theories to enable the construction and use of canonical explanatory models is alleged to require a whole host of ontological, representational, and epistemological changes and is thought to occur gradually (not in a rapid paradigm shift).[28,29,119] Importantly, framework theories are implicit constructs and not the sort of rationally constructed theories referred to in Strike and Posner's work. However, like early "theory−theory" conceptual change adherents, framework theory proponents assume some coherence in students' ideas so long as those ideas draw on the same implicit presuppositions.

Other prominent conceptual change theorists view naïve ideas as consisting of simpler building blocks (pieces of knowledge, if you will), derived from experience and instruction, that are dynamically woven and rewoven in different contexts.[120,121] Much of the work that informs this "knowledge in pieces" perspective emerged in the context of physics ideas and concepts, that is, macroscopic, observable phenomena. diSessa proposed that some of these simple building blocks in physics are phenomenological primitives (p-prims),[26,27] which are intuitive ideas based on personal observation that are useful for problem solving in particular contexts. They are not reasoned out but emergent from the way we interact with the world.[122] Perhaps the most famous p-prim is the idea that more effort leads to more of a result and greater resistance causes less result.[47] This idea might be productively activated whether moving a couch or learning Ohm's law in physics class. However, this "more means more" p-prim is not so relevant to molecular-level phenomena. Students who believe that bigger molecules have higher boiling points are unlikely to activate ideas about how molecules interact based on their molecular structure.[32]

From diSessa's "Knowledge in Pieces" (KiP) perspective, incorrect responses to a particular question do not indicate a well-formed "wrong theory" (or even a coherent set of presuppositions arranged into a framework theory) but rather a suboptimal coordination of knowledge fragments in the context of the question being asked. Emphasis here is not on the base elements of the concept being incorrect (they are "resources" that are not in and of themselves right or wrong) but on activating inappropriate resources in the context of a question. The notion of a "concept" in diSessa's work is much more dynamic than that supposed in early misconceptions literature (and is more of a "conceptual ecology" than a monolithic concept). Conceptual change under the KiP paradigm is not a whole-cloth switch from a wrong idea to a right one but instead a reconstruction and reweaving of one's existing conceptual ecology (i.e., which resources are activated in certain contexts).

It should be emphasized that the notion that novices have coherent, stable, wrong theories and the idea that misconceptions are resultant from dynamic coordination of resources (like p-prims) are two poles of a spectrum of views. Taber has proposed a model that allows for both stable knowledge frameworks and in situ construction of ideas.[123] He holds that dynamic, on-the-fly, idea construction is a fall-back for individuals that do not possess (or do not recognize they possess) an appropriate conceptual framework for the problem at hand. Such frameworks are thought to emerge from "complexes of concepts and their relationships" that "may become established, 'stored' and readily activated." Taber contends that "evidence of stable, extended patterns of thinking in... an abstract area would seem to imply that the learner has constructed permanent representations of conceptual frameworks in long-term memory." By contrast, inconsistent student thinking about a particular conceptual area could be seen as evidence of on-the-fly idea construction. Taber and García-Franco found evidence that students use primitive intellectual resources when attempting to explain a range of phenomena and suggest that the p-prims identified in physics education research may themselves "reflect more general abstracted patterns and could in effect be facets reflecting a more basic reasoning principle."[124]

There remains considerable uncertainty about the character of concepts and the nature of conceptual change. Exacerbating matters is the fact that virtually no persuasive evidence exists to justify one mechanism of conceptual change over another, and it is reasonable to suggest that the mechanism of change almost certainly depends on the concept to be understood. The purpose of including a section on conceptual change in this Review is to demonstrate that misconceptions are not necessarily well-characterized monolithic entities to be readily and formulaically surmounted. There is reason to suppose that mechanisms of conceptual change may be to some degree discipline and context dependent. One might be able to assemble something resembling a correct view of Newtonian mechanics from intuitively understood p-prims,[125] but there is virtually no chance of coordinating everything needed for a coherent molecular-level understanding of chemistry through experience alone. Additionally, there is little to be gained by simply cataloging misconceptions without paying heed to the mechanisms of their emergence, their organization, and their character. A misconception elicited by a given prompt might be an aggregation of fragments coordinated in the context of the question being asked and thus far more dynamic than a word or

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

phrase on a list would imply. Finally, it is not overly useful if the community fixates solely (or largely) on what students cannot do without offering validated strategies to build productive conceptual understanding. That being said, recognizing the ideas that students bring with them is crucial for effective instruction, because constructing knowledge on problematic foundations is unlikely to produce deeper understanding. As we will discuss in section 4, these ideas can be elicited by specifically designed assessment instruments such as concept inventories.

3.3. Type 1 and 2 Thinking

Up to now we have considered how the brain processes and stores information but not how that information is retrieved or how we decide which information to use—or even whether this is a conscious decision. Research in this area has converged on ideas that were made accessible in Nobel laureate Daniel Kahneman's book *Thinking Fast and Slow*[126] and were originally developed in the 1970s.[127,128] This research has provided solid evidence that humans often make irrational decisions and are subject to all kinds of cognitive biases and heuristics. Emergent from this work is the theory that humans have (at least) two kinds of thinking, which is sometimes referred to as dual-processing. Type 1 "is the brain's fast, automatic, intuitive approach", and Type 2 is "the mind's slower, analytical mode, where reason dominates."[126,129] Type 1 is the default mode for most of us, most of the time. It allows us to make rapid responses and to perform multiple tasks simultaneously with little cognitive effort (for example, we can drive a car and talk at the same time). However, for complex tasks we (should) use type 2 thinking, which requires cognitive effort and attention (that is, we should not drive a car and attempt to text or solve a complex problem at the same time). In fact, while most researchers agree unequivocally that multitasking is a myth,[130] students who are only using type 1 thinking may well believe that using a phone or surfing the web are not taking up cognitive resources, despite emerging evidence that such behavior does impact learning and grades.[131,132] Unfortunately, because we are evolutionarily adapted to default to type 1 thinking, we have to consciously override the impulse to give the simplest most intuitive answer to assessment tasks, or use heuristics that reduce the cognitive load.[133] For example, Maeyer and Talanquer reported on a number of "fast and frugal" heuristics that students use to answer ranking tasks (such as which compound has the highest boiling point).[133,134] They categorized the heuristics that students used as "recognition, representativeness, one reason decision making, or arbitrary trend."[134] As an example, some students chose NaCl as their response to a question simply because they had seen it before (i.e., the "recognition" heuristic). In a study on structure–property relationships, we saw many students using the "more means more" heuristic (which might also be classified as a p-prim). More bonds, more hydrogens, or more oxygens were thought to correspond to higher boiling point.[32] The behavior of molecular-level species cannot often be predicted by ideas intuited from macroscopic experiences, and therefore reworking and recontexualizing p-prims may not be useful for helping students construct understanding in chemistry. The student who predicts that the heavier molecule will always have the higher boiling point will often be wrong.

That being said, most practicing chemists *do* make use of cognitive shortcuts; when we say that "like dissolves like", it is a proxy for a long sequence of concatenated ideas and skills. Experts may also use productive heuristics and routines to solve problems, for example, experts typically use disciplinary strategies to assign R and S configurations (rather than mentally rotating complex molecules). Interestingly, Treagust and co-workers found that higher-performing students were able to confidently employ algorithmic methods using mole ratios to identify limiting reagents, whereas less-able students tended to figure out the limiting reagent from first-principles, as they were taught in class, and were still less successful.[135] The difference between experts and novices is that, when prompted, an expert will be able to explain why like dissolves like, give insight into the solution process and the associated enthalpic and entropic changes, and provide a wide range of examples of different types. Students, on the other hand, may not get further than the fact that "like dissolves like". They may be able to use it as a rule of thumb without understanding the underlying concepts. Teaching such cognitive shortcuts requires explicit discussion that what is being taught *is a shortcut* and not a reason. If this is not made clear, problematic understanding can result. For example, students may think that bonds form because atoms "want" eight electrons[136] or that reactions happen because of a drive toward stability,[36] and thus be unable to provide more scientifically appropriate reasoning. Teaching such shortcuts without appropriate support often leads to didaskalogenic (teacher-induced) conceptual confusion.

Up to now we have discussed much of the background for this Review in a more-or-less chronological order beginning with the "personal empiricism" of the early 20th century and advancing to more modern theories of learning, conceptual change, and multilevel thought in chemistry. This was done to illustrate the integration of learning theory into CER over time and show the evolution of the field from the personal empiricism of the 1890s to studies that are driven by learning theories and provide evidence. We are now equipped with the theoretical fundamentals needed to explore more modern work in CER. To that end, each of our three guiding questions will now be considered, together with contemporary studies relevant to that question. Recall that significant relationships exist between each of our questions, and so it is unavoidable that some overlap in our discussion will occur. It should also be noted that the vast majority of CER studies revolve around introductory and general chemistry courses, and while there are also a substantial number of studies involving organic chemistry, upper-level courses are less well-represented.

4. WHAT STUDENTS SHOULD KNOW AND BE ABLE TO DO WITH THAT KNOWLEDGE

In this section, we consider the impact of learning theories on what chemistry knowledge students should know and, perhaps more importantly, what they should be able to do with it. We begin with a short historical discussion of the way the content of chemistry courses has changed over the years, from its highly practical origins in descriptive chemistry to the infusion of chemistry theories that were intended to support student learning. As we will see, it is only more recently that theories about how people learn have begun to inform what students should learn.

4.1. What Students Should Know

As demonstrated by Cornog and Colbert's 1924 study,[17] what students should know after a course in chemistry has been a topic of interest for as long as chemistry education literature has existed. In the era of descriptive chemistry courses, content was driven by the needs of industrial and agricultural employers. A degree in chemistry signified that one was a skilled technician

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Table 1. Comparison of "Big Ideas" over the Last 20 Years

Gillespie[147]	Atkins[148]	AP Chemistry Big Ideas[149]	ACS General Chemistry Curriculum Map Anchoring Concepts[150,151]	CLUE Core Ideas[152]
(1) atoms, molecules, and ions	(1) matter is composed of atoms	(1) atoms	(1) atoms	(1) atomic/molecular structure and properties
(2) the chemical bond	(2) elements form families	(2) chemical and physical properties	(2) bonding	(2) electrostatic and bonding interactions
(3) molecular shape and geometry	(3) bonds form by sharing electron pairs	(3) reactions: rearrangement of atoms and electrons	(3) structure/function	(3) energy
(4) kinetic theory chemical reaction	(4) shape is of the utmost importance	(4) rates/kinetics	(4) intermolecular forces	(4) change and stability in chemical systems
(5) energy and entropy	(5) molecules interact with one another	(5) thermodynamics/energy	(5) chemical reactions	
	(6) energy is conserved	(6) bonds and interactions	(6) energy and thermodynamics	
	(7) energy and matter tend to disperse		(7) kinetics	
	(8) there are barriers to reaction		(8) equilibrium	
	(9) there are only four fundamental types of reaction		(9) measurement and data	
			(10) visualization and scale	

who could apply defined procedures to the synthesis of "artificial manures" or new alloys with intriguing properties. Beginning in the late 1950s, more theoretically grounded chemistry courses were designed to focus on those aspects of the discipline deemed important by faculty. The methods used to determine importance were sometimes idiosyncratic (i.e., the nodometer used by Sienko and Plane) and were rarely informed by learning theory. In fact, one could persuasively argue that the content in most modern textbooks is an elaboration of the topics originally compiled by Sienko and Plane 60 years ago.[137,138] It is perhaps due to the community's tendency toward elaboration of existing texts rather than wholesale reinvention that modern textbooks have ballooned in size relative to Sienko and Plane's *Chemistry: Principles and Properties* (which weighs in at a svelte 618 pages, not including references; most modern texts are twice that size).[137] Interestingly, several recent textbooks have shuffled the order of their topics to put a discussion of atoms first.[139−141] There is no reported theoretical basis either for the original order of the text or this rearrangement, and the material covered remains much the same as traditional Sienko and Plane based curricula. In fact, a study by Esterling and Bartels indicates that simply rearranging the topics of a textbook does not improve student achievement as measured by course grade.[142]

Although it might appear that general chemistry curricula have changed little since the 1960s, there have been a number of attempts to identify what *should* be taught. These attempts were often driven by (1) a recognition that general chemistry in particular has become bloated—a mile wide and an inch deep—and while topics have been added, little has been removed, and (2) an understanding that the majority of students in most chemistry courses are not destined to be chemistry majors.[143−146] This is a point that deserves to be emphasized: students in our classes may become doctors, teachers, farmers, social workers, and nurses, as well as scientifically literate citizens, and their needs and interests should be addressed.

As we learn more about how experts' knowledge is structured, there has also been a tendency to try to organize what students learn around big ideas. In the late 1990s Gillespie[147] and Atkins[148] independently identified what they both called the "great ideas" that "form the basis of modern chemistry". Gillespie characterized six great ideas, and Atkins characterized nine. While there was no underlying theoretical basis for either of their lists, they are strikingly similar to each other and to some later efforts and are the product of lifetimes of thinking about how chemistry should be taught.

The notion that curriculum development should be centered around big ideas has now become the basis of a number of reform efforts. Those most relevant to chemistry include the Advanced Placement (AP) reinvention project (six big ideas)[149] and a project of the American Chemical Society's (ACS) Examinations Institute (tied to 10 anchoring concepts) to identify what should be taught in every course in chemistry.[150] A comparison of these big ideas is provided in Table 1 along with a set of core ideas generated as part of a curriculum design and subsequent course transformation at Michigan State University ("CLUE Core Ideas"). Big ideas vary in grain size and in what is included between initiatives (and even within a single initiative). However, most include ideas about the structure and properties of matter, bonding and interactions, energy changes and dissipation, and changes and stability in chemical systems.

It is worth noting that, even though many big ideas appear similar, the way they are operationalized may be quite different. For example, these big ideas can be thought of as overarching concepts that underlie all of chemistry, or alternatively they may be considered as objectives that students must work toward understanding. For example, energy could be considered as a concept that ties chemistry ideas together or as a specific objective such as "to break a bond requires an input of energy".[150,151] That is, these big ideas could be considered goals for learning or frameworks upon which learning can be built. While the learning goal aspect of these ideas is important and perhaps easier to attend to in a traditional curriculum structure, the overarching nature of big ideas is better aligned with what we know about the development of coherent expert-like knowledge frameworks.

The most elaborated and granular listing of what one should know after a course in chemistry is the Anchoring Concepts Content Map (ACCM) assembled by the American Chemical Society's Exams Institute.[150,151] The ACCM was constructed with the goal of specifying content that would be used to draft the nationally normed ACS exams across various subdisciplines of chemistry. Thus, the ACCM is meant to specify anchoring concepts for an entire undergraduate curriculum. Accordingly, "anchoring concepts" are defined as "concepts that are important in chemistry and are generally covered in some fashion in virtually every course."[150]

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

The ACCM specifies a hierarchical structure that places anchoring concepts at the highest level. Each anchoring concept is further subdivided into "enduring understandings", which are further split into "subdisciplinary articulations" that specify how higher-level concepts should play out in particular chemistry courses. The most finely grained level of the ACCM are "content details" that specify the individual pieces of knowledge that are typically tested on ACS exams. There are currently 263 individual content details within the general chemistry anchoring concept content map published by the ACS, and over 1100 details for the undergraduate chemistry curriculum as a whole. The construction of this map is a significant accomplishment, and an exhaustive listing of each content detail is a useful service to the chemistry community. However, while the ACCM was constructed by volunteers who were required to have experience teaching the particular course for which a given content map was being developed, these volunteers were not required to have any background in research on teaching and learning. As such, the ACCM represents the conventional wisdom about course content rather than an attempt to construct a coherent foundational framework for chemistry. It might be argued (and in fact we have argued[152]) that the construction of such a fragmented map lies in direct opposition to the original intent of the big ideas. That is, it is difficult for students to link such fragments together to form a coherent and consistent framework for understanding. This problem is exacerbated if assessment is focused (as recommended) at the content details level. A student can certainly balance a chemical equation (a skill) without understanding that conservation of matter provides the rationale for their endeavor. We have recently contended that one must have evidence that knowledge is coherent and linked and that to assume an expert-like knowledge structure from assessment of fragments is unwarranted (Figure 3).[152]

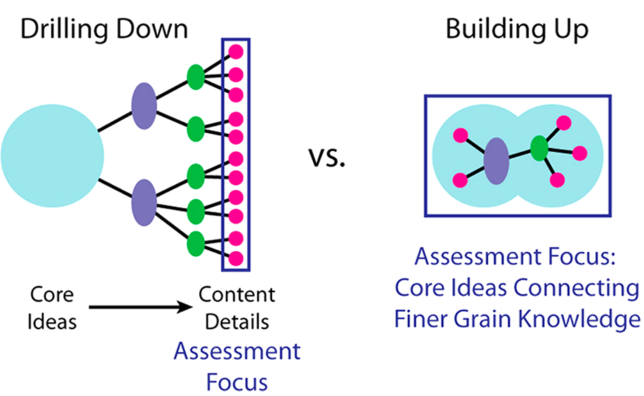


Figure 3. Comparison between the evidence for knowledge coherence gained from assessing content details versus that obtained from assessing core ideas. Reprinted with permission from ref 152. Copyright 2017 American Chemical Society.

Focusing chemistry coursework on big ideas that are central to a discipline and provide support for a range of connected concepts, rather than learning objectives, offers the potential to foster more coherent and useful student knowledge. Big ideas in this vein are not thought of as separate topics but instead as large conceptual areas that are necessary to understanding a discipline. For example, "structure–property relationships" is typically not taught as a chapter in any general chemistry course but underlies almost every topic. That is, phase changes can (and should) be linked to the core ideas of "structure–property relationships", "energy", and "bonding and interactions". Kinetics can be linked to the core ideas of "structure–property relationships", "energy", "bonding and interactions", and "change and stability in chemical systems". In this way, with appropriately designed assessments, students will have a better opportunity to build more expert-like connected frameworks of knowledge.

4.2. What Students Should Be Able To Do with Their Knowledge

Having a vast repertoire of facts and skills means little if the student cannot take that knowledge and use it in some context. The chemistry education community has expended substantial effort attempting to define what productive use of knowledge looks like and in what context such "use" should occur.

4.2.1. Problem Solving. Historically, problem solving has been a major goal of chemistry education; after all, as George Bodner has noted, "problem solving is what chemists do".[153] But chemists do many things; if we are to understand how to support student learning, it is perhaps not optimal to collect all these activities together as if problem solving were some monolithic activity. Solving numerical problems using a provided equation, proposing organic syntheses of target compounds, constructing mechanisms of reactions, identifying patterns in data and making deductions from them, modeling chemical phenomena by computation, and identifying an unknown compound from its spectroscopic properties could all be (and have been) described as problem solving. However, these activities require different patterns of thought, background knowledge, skills, and different types of evidence of student mastery. Nevertheless, over the years there have been many research studies on problem solving, and several literature reviews on this aspect of chemistry education alone.[154,155] In fact, much of the early work in CER revolved around a wide range of different tasks loosely grouped under the umbrella of "problem solving".

That being said, many would agree with the definition of problem solving as "what you do when you don't know what to do."[156,157] This deceptively simple definition is more complex than it may seem at first glance. For instance, a novice might well not "know what to do" in a scenario that appears routine to an expert. That is, the characteristics of the problem solver impact what we might classify as a problem. The type of problem may also affect whether or not a student has an opportunity to figure out a strategy when they do not know what to do. In open-ended problems (i.e., the design of a synthesis or a case-based recommendation for environmental cleanup), students must chart a path forward while choosing among many possible actions. Very often, open-ended problems have several reasonable answers. By contrast, closed problems (such as most of the end-of-chapter exercises in commercially available texts or the online items that come with many homework systems) have one answer and can be solved by using an algorithmic approach or pattern recognition.[135,155,158,159] They bear little resemblance to what it is that chemists do, and yet in many instances, these exercises are the sum-total of students' experiences in chemistry.

Much of the research on problem solving has focused on characterizing the difficulties that students have with solving both open- and close- ended numerical problems. One thread that runs through all of it—which may seem obvious to the practitioner, is that students are often able to answer well-defined (i.e., close-ended) problems, without making use of conceptual understanding.[135,155,158,159] That is, students may be

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

able to carry out calculations, but often have little understanding of what it is they are calculating or why it matters.

Some researchers have used the information processing model,[77] introduced to the CER community by Johnstone, as a lens to study problem solving (see Figure 2). In this model, problem solving takes place in the working memory, which can only hold a finite number of ideas (e.g., concepts and/or operations)—typically between 5 and 9.[72] The more complex the problem, the more operations must be held in working memory and the higher the "cognitive load". Recall that the coherent, connected knowledge framework characteristic of domain experts enables ready access to relevant concepts and skills as well as an ability to "chunk" large fragments. Consolidation of information into related chunks enables experts to make use of the 5−9 "slots" in their working memory much more efficiently than novices and thereby lower their cognitive load. Researchers have investigated the impact of cognitive load on problem solving in chemistry[160] and the associated idea of cognitive complexity[161,162] and have found that, indeed, the cognitive load of a given problem is inversely correlated with success on that problem. However, as noted before, problem solving has a wide range of meanings, and therefore the type of problem almost certainly has an impact on the resources that students must call on. Overton and co-workers investigated the relationships between the cognitive resources required to solve both algorithmic and open-ended problems (in which students were required to analyze data and where there was no one correct answer).[163] These researchers found that the best predictor of algorithmic problem solving success was student performance on a counting recall test that required simultaneous processing and storage of information within working memory. By contrast, student performance on open-ended problems was better predicted by instruments involving backward digit recall and a figural intersection test. As the authors note, there is "a dissociation between the cognitive resources underlying algorithmic and open-ended problem solving." In other words, problem solving is not a monolithic ability.

As the problem-solving literature has been extensively reviewed elsewhere,[154,155] we will not delve deeply into work on problem solving per se. Instead we will point readers to potentially more useful and tractable approaches to characterizing what it is that chemists do.

4.2.2. Scientific and Engineering Practices. There is a significant body of work in chemistry education research focused on what students should be able to do. Collectively, these activities have become known as scientific practices.[164] For example, the construction and use of models have long been recognized as central to the understanding of chemistry. As Justi and Gilbert state:[165] "Chemists model both the phenomena they observe and the ideas with which they try to explain such phenomena—that is, at both the macroscopic and the submicroscopic levels[16]—by the use of analogy with what they already know. The outcomes of such a process are expressed in a concrete, visual, mathematical and/or verbal mode of representation, sometimes by using special symbols that constitute chemical language (e.g., formulae of compounds)." Models are so important in chemistry because "the explanations of the natures of substances and of their transformations are essentially abstract. Thinking with models enables chemists to visualise the entities or processes that are being enquired into, to more readily plan experimental activities on them,[166] and to support the processes of reasoning and constructing knowledge[167]".[168] Most of the earlier research on models and modeling focuses on how precollege students, understand the nature of models and much of it pertains to how students understand the nature of models,[169] for example, whether the model is a copy of reality, or a representation, or is created to test ideas and may be changed by the modeler to accommodate new observations. As we will discuss later, investigating student ability to construct and use models can be a fruitful approach to assessing how students understand chemical phenomena.

There is a similar body of work around the practice of argumentation, which Erduran defines as "the coordination of evidence and theory to support or refute an explanatory conclusion, model or prediction".[170] Much of this work has focused on the Toulmin argumentation patterns that students use as they discuss experimental observations or data.[170,171] These patterns generally begin with data (either collected by students or provided), followed by supporting statements (or warrants and backing) emergent from student discussions on data analysis that eventually lead to a conclusion or claim consistent with the data. Toulmin argumentation patterns for students enrolled in a variety of chemistry courses (including general chemistry, physical chemistry, and organic chemistry) have been analyzed and reported. For example, de Arellano and Towns[35] reported that lack of background information and incorrect warrants were a major source of error for undergraduate organic chemistry students interviewed about reactions involving alkyl halides. Pabuccu and Erduran, upon studying how preservice teachers engaged in argumentation about conformational analysis, found that higher performers tended to have more rebuttals in their discourse and that lower performers had difficulty evaluating the credibility of evidence.[172] Studies such as these provide us with evidence about how students construct arguments that may be used (see sections 5 and 8) to design appropriate materials to support students.

In contrast to argumentation, where the goal is justification of an uncertain conclusion, explanation is a practice in which students use scientific principles to explain a known phenomenon.[173] When students engage in argument they are using evidence to support a claim, whereas constructing an explanation requires students to use what they know to explain how or why a given phenomenon occurs. Again, there is an emerging body of work on the role of explanation in chemistry education and the ways in which students construct explanations. For example, Talanquer has reported that students often rely on implicit knowledge and heuristic reasoning strategies to simplify the task,[174,175] while Cooper and co-workers have focused on how to support students to produce causal mechanistic explanations of phenomena.[176,177] We will return to these ideas in the assessment section.

In 2012, a consensus study from the National Academies of Science entitled *The Framework for K-12 Science Education* (hereafter referred to as *The Framework*) was published.[164] The *Framework* provides a vision for how science education should be structured based on a synthesis of current research. Learning in science, as described by *The Framework*, should involve developing *and using* interconnected, contextualized (i.e., expert-like) knowledge structures. Most of the primary literature base underpinning *The Framework* is just as applicable to collegiate study as it is to K-12 science education. The Framework characterizes eight scientific and engineering practices (SEPs) (Box 3), each of which connects to and overlaps with others; they characterize the different things that chemists (and scientists more generally) do.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Box 3. Scientific and Engineering Practices[164]

1. Asking questions (for science) and defining problems (for engineering)
2. Developing and using models
3. Planning and carrying out investigations
4. Analyzing and interpreting data
5. Using mathematics and computational thinking
6. Constructing explanations (for science) and designing solutions (for engineering)
7. Engaging in argument from evidence
8. Obtaining, evaluating, and communicating information

While some of these are perhaps unfamiliar in a teaching situation, many of the SEPs are fairly self-explanatory, and taken together they might be thought of as the disaggregated components of inquiry. For example, while there are several SEPs that are clearly associated with laboratory learning (asking questions, planning and carrying out investigations, analyzing and interpreting data, and engaging in argument from evidence), all of the SEPs can and should be incorporated into all aspects of science learning. It should be noted that many traditional chemistry teaching laboratories do not in fact require students to engage with SEPs, because the topic of the investigation, the experimental procedure, and the appropriate data-analysis methods are often provided.[178,179]

There is no implicit or explicit hierarchy of scientific and engineering practices. That is, explanation is not to be taken as necessarily more sophisticated, or higher order, than, for example, asking questions. This stands in marked contrast to Bloom's taxonomy,[180] which (for example) describes explanation as a low level competency. There is ample evidence in the literature that molecular-level explanations of macroscopic properties[32] require students to reason in sophisticated ways with abstract concepts and are far from low-level thought. Although Bloom's taxonomy continues to be occasionally referenced in the chemistry education literature,[181,182] it has generally fallen out of favor, in part because of the speculative nature of its hierarchy and the lack of a theoretical basis for its approach.

Zoller and co-workers have proposed a hierarchical approach to what students should be able to do that has elements of both the scientific practices and Bloom's taxonomy.[183] They divide tasks into those that require higher-order cognitive skills (HOCS) and those that rely on lower-order cognitive skills (LOCS). HOCS involve such activities as question asking,[184] problem solving, decision making, and critical thinking, all of which require "evaluative thinking",[185] while LOCS involve more traditional types of tasks involving algorithmic numerical problem solving, recall of facts, and simple procedures. Interestingly, Zoller reports that students tend to prefer examinations that emphasize understanding, analysis, and synthesis rather than rote learning, while at the same time instructors continue to construct examinations largely composed of items that only require LOCS.[186]

We believe that this approach is entirely consistent with the use of SEPs, all of which would presumably be classified as requiring HOCS. However, unlike outcomes such as critical thinking or problem solving, which may have very different instantiations depending upon the context, SEPs are better characterized and more easily operationalized in the form or assessments. If you do not know what you are aiming for (as is often the case when trying to assess vague constructs), it is difficult to get there.

Importantly, SEPs are not intended to be abstracted from context but rather as opportunities for students to extend and adapt their knowledge. As such, *The Framework* describes learning in science as a slow buildup of interconnected, contextualized, and useful knowledge anchored to core domain ideas, extended through engagement in scientific and engineering practices, and linked to other disciplines through crosscutting concepts. *The Framework* characterizes this as three-dimensional (3D) learning.[187] Equal emphasis is placed both on what students should know and also what they should be able to do with that knowledge (as the two are synergistic). Stated differently, SEPs allow students to use their knowledge to design, reason, model, and explain, which in turn leads to more connected knowledge.

Sevian and Talanquer[188] have defined a set of practices different from those in *The Framework* as part of their "Chemical Thinking" learning progression. Chemical Thinking focuses explicitly on the work of professional chemists and how one might enable students to engage in that work.[188] Course content is built around promoting an understanding of practices deemed central to modern chemists: analysis ("the development and application of strategies for detecting, identifying, separating, and quantifying chemical substances"), synthesis ("the design of new substances from synthetic routes"), and transformation ("controlling chemical process for non-synthetic purposes, such as harnessing chemical energy").[188] "Crosscutting disciplinary concepts", which serve as the highest level of organization for course content in Chemical Thinking, are very different from the big ideas specified by other content maps in that they are defined by a core question rather than a coherent conceptual area. For example, "chemical causality" is driven by the question "why do chemical processes occur?", which incorporates ideas related to both energy and atomic/molecular structure. Like *The Framework*, Chemical Thinking emphasizes both knowledge and use of that knowledge in productive, authentic practices.

Regardless of the ideas and practices around which a curriculum is developed, it is important for students to receive and respond to the message that both knowledge and the ways that knowledge is used are crucial aspects of learning chemistry. This will involve developing explicit descriptions of what students are expected to learn and be able to do (i.e., learning objectives, learning performances, or performance expectations) and ensuring these descriptions are the focus of assessment.

4.3. Summary

There have been a number of attempts to identify what it is that students should know and be able to do. In the early days, these approaches were developed by practicing expert chemists' beliefs about the needs of industry rather than being guided by theory and evidence. Our current understanding of how people learn indicates that, to develop expert-like knowledge frameworks, ideas should be connected and contextualized. The identification of big ideas in chemistry helps facilitate this process if the knowledge that students learn is connected to these big ideas rather than being learned in isolation. We also know that helping students use this knowledge by explicitly incorporating scientific and engineering practices into the learning environment is more likely to support the development of expert-like thinking.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

5. HOW TO KNOW THAT STUDENTS HAVE DEVELOPED A COHERENT AND USEFUL UNDERSTANDING OF CHEMISTRY

Unfortunately, we can never really know the full extent of what students know and can do. Instead, we must rely on tasks that can produce evidence, from which we can make inferences about student competence. The evidence we view as convincing and the inferences we deem appropriate are all guided by the assumptions we make about the nature of learning in a domain.[189] The National Research Council's (NRC) report *Knowing What Students Know* provides a model of assessment that involves reasoning from evidence. That is, assessments must elicit evidence of student understanding, from which we can make an argument about what a student knows and can do. The process of reasoning from assessment-derived evidence can be portrayed via the "assessment triangle" (Figure 4). Each corner of the triangle represents a foundational, underlying element of the assessment process: (1) "a model of student *cognition* and learning in the domain, (2) the kinds of *observations* that will provide evidence of students' competencies, and (3) an *interpretation* process for making sense of the evidence."[189] For an assessment to be effective, all three elements should be both dependent on and connected to the other two.

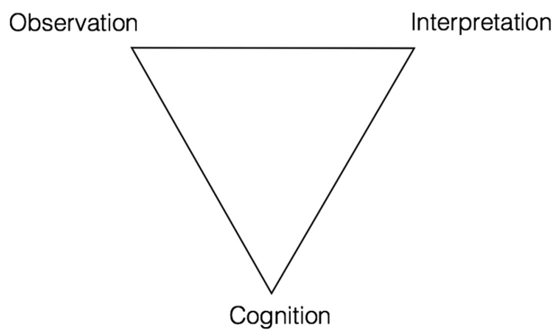


Figure 4. Assessment triangle as described by *Knowing what Students Know*.[189]

The *cognition* corner of the triangle represents a theory about how students develop expertise in a domain. A theory of learning is required so that the set of knowledge and skills needed for a task can be defined. For example, a theory might model a typical progression from novice to expert and suggest assessments that are capable of differentiating between more and less expert-like understanding.[189] Note that here a "theory of cognition" is not referring to a generalizable theory about how people learn (such as constructivism) but rather a finely grained, evidence-based account of how groups of individuals tend to progress from novice to expert-like in a domain. For example, if we accept that students will develop more expert-like knowledge if their learning is organized around core ideas in the context of SEPs and crosscutting concepts, then a well-defined account of 3D learning in chemistry can serve as our cognitive framework for developing assessments.

The *observations* (or assessment items and their potential student responses) suggested by one's theory of cognition are not arbitrary but instead are determined by beliefs about the sort of prompts or situations likely to elicit useful evidence of students' knowledge and skills. The nature of useful evidence varies depending on the purpose of a given assessment. As a general rule, there is no all-purpose assessment. Experienced assessment designers are mindful of the purpose for which they are creating assessment tasks and attempt to design tasks that gather the strongest evidence possible within the constraints of the context being assessed. The nature of the evidence should be established by research in teaching and learning (in chemistry, in our case).

To *interpret* assessment-derived data, one must have a well-developed sense of what inferences these data support. That is, we must be able to take the data from an assessment and use it as evidence to support the argument that students are developing some aspect of the desired cognition. This interpretation step may take different forms. For large-scale assessments of the type developed by the ACS Examinations Institute, the method of interpretation is typically a statistical model that characterizes patterns of responses. On the other hand, formative assessments designed to provide feedback to inform future instruction may be viewed through a qualitative interpretive lens. For example, coding schemes designed to infer various types of understanding from student responses to a prompt are interpretive tools sometimes used in chemistry education research. Specific examples of such schemes will be discussed later.

Deployment and use of assessments depends a great deal on course curriculum and the instructional methods used to enact it. By curriculum, we mean both the learning objectives that are typically developed to guide instruction and the scope and sequence of material presented. Instruction refers principally to the teaching methods and learning activities used to realize the vision set forth by the curriculum in the classroom. Optimally, the goals of assessments, curricula, and instruction should be aligned (i.e., directed toward the same or complementary ends). Thus, we should assess those things that students should know and be able to do (laid out in learning objectives) rather than solely assessing fragments of knowledge and skills that are not evidence of a coherent understanding.

Because a discussion of "how we know students have developed a coherent and useful understanding of chemistry" is essentially a discussion of assessments, we will focus our attention on three broad assessment categories: exams developed by the ACS Exams Institute, concept inventories, and other alternate assessments reported in the literature. Our discussion of each of these categories will be guided by the assumptions, either implicit or explicit, made by assessment designers concerning the three elements of the assessment triangle.

5.1. Large-Scale Assessments Developed by the ACS Examinations Institute

Chemistry is unique among the STEM disciplines in that it has a long history of community-developed assessments at the national level.[190] The ACS Examinations Institute began in 1930[191] and since then has provided a wide range of chemistry examinations that cover the whole undergraduate chemistry curriculum. As noted earlier, these exams are designed by practitioner consensus about what topics should be assessed and in what ways. Because these exams are given to large numbers of students across many different situations, it is essential that they are valid (that is, they assess what they are intended to assess) and reliable (that is, the exam should give similar results with similar students). The exams should be appropriate for all students, and particular groups of students should not be disenfranchised (or privileged).

Over the years, the ACS Examinations Institute exams have changed considerably reflecting changes in enacted curricula, the needs of instructors, the resources available for grading, and the use of modern test theory to help produce valid and reliable

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

exams. Interestingly, early general chemistry examinations used a wider range of question formats than they do now, including tiered multiple choice, multiple True/False questions, and questions that required students to judge the plausibility of a particular claim from evidence.[192] The content of these early exams often included chemistry that might be unfamiliar to a more modern audience (remember that, in the early years, chemistry students were often being trained to be chemical technicians). ACS Examinations Institute tests have changed over time to reflect how chemistry teaching itself has changed, incorporating more theory and different kinds of visualizations. Test questions have settled into a multiple-choice format that makes grading easier, and current exams are aligned with the chemistry curriculum maps.[151] Changes in the organic chemistry exam showed a movement from simple recall to more conceptual questions from the 1940s to 2012, with an increase in spectroscopy questions and a decrease in questions involving qualitative analysis.[193]

A major appeal of these exams is that they allow individual instructors to compare their students to national norms. That is, each exam is normed according to data submitted by instructors, so that it is possible to see where a particular class average falls relative to the national sample of students who took the exam. Over the years the data from exams has been mined to identify whether item order matters or whether items behave differently with different groups.[194] Modern test theory,[195] which allows the test developer to identify item difficulty, has also been used to construct tests that span a range of student abilities. This kind of information results in fairer tests that are psychometrically reliable and valid for a wider range of students. However, as noted earlier, typical ACS exam items are focused on assessing finely grained content details. In part, this is because the tests must be reliable and valid: the more complex the question, the more difficult it is to produce large-scale assessments that have appropriate psychometric qualities. Importantly, knowledge of content detail fragments does not guarantee deep and robust understanding. This issue, coupled with the concern that students can answer numerical problems without understanding their underlying concepts,[196] led to the development of various tests that were intended to assess conceptual understanding.

In the mid-1990s a "conceptual" ACS general chemistry exam was developed that involved no overt calculations and used a wider range of representations (particulate and graphical) in response to the growing unease in the community that students who could answer numerical problems did not understand their conceptual basis. Perhaps not surprisingly, after surveying approximately 1400 instructors, Holme and co-workers found that "conceptual understanding" means different things to different people.[197] Five "definition fragments" were most common in survey respondents' notions of conceptual understanding: transfer (to a new setting), depth (of understanding), predict/explain, problem solving, and translation (among representations and scales). While this is an emergent set of components generated by practitioners, there is no underlying theory base for this set, and many of the terms are themselves not well-defined (problem solving) or are extraordinarily difficult to achieve ("transfer" to a new setting).[198]

Although the ACS Examinations Institute has never discussed their tests using the language of the assessment triangle (Figure 4), we propose that the anchoring concept content maps (ACCMs) serve as the theory of cognition upon which the exams are based. Taken together, they represent an expression of the knowledge one should master over the course of a baccalaureate degree in chemistry. As content maps exist for several individual courses, one may extrapolate a progression from novice to expert content knowledge by examining a sequence of ACCMs. To gather evidence that students can integrate knowledge from multiple chemistry courses into expert-like understanding, the ACS Exams Institute designed the Diagnostic of Undergraduate Chemistry Knowledge (DUCK) Exam.[199] The DUCK nests item prompts within scenarios and requires students to integrate knowledge from several traditional chemistry courses (as expressed by the ACCM for those courses). However, there is as yet no research on how effectively the DUCK measures student expertise in chemistry. Crucially, it is not the aim of the ACS Exams Institute to provide finely grained, nuanced appraisals of what students know and can do with their knowledge. Instead, Institute products are designed to provide a valid and reliable means of comparing cohorts of students against national norms. The ACCMs provide a theory of cognition that is suitable to this task.

5.2. Concept Inventories

Beginning in the late 1980s, researchers began to develop and validate assessment items using a different approach than the ACS Examinations Institute. Rather than disciplinary experts in the community constructing assessment items based on their expert knowledge, researchers interviewed students to identify problematic understandings and, based on the results of these interviews, constructed assessment items. Among the first assessments developed using this philosophy were "diagnostic assessments" published by Treagust.[87,200] These were typically two-tier multiple-choice tests (i.e., a yes/no question followed by a set of statements meant to enable justification of the yes/no). Diagnostic assessments were precursors to the well-known class of assessments called concept inventories (CIs), the first of which was published by Hestenes et al.[201] The first CI designed for use in chemistry, known as the Chemistry Concepts Inventory, was published by Mulford and Robinson in 2002.[97] At present, a large number of such inventories exist for various concepts and subdisciplines of chemistry.[34,34,202−208] Generally, CIs are developed based on student interview data and/or the misconceptions literature, and newer ones often target a particular idea, for example, conceptions of acid strength[34] or enzyme−substrate binding.[204]

Concept inventories are designed to elicit whether students have scientifically appropriate ideas, that is, just as Treagust implied in the earliest of these, they are diagnostic instruments. As Ausubel famously wrote,[57] "The most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly." If a student's prior knowledge is faulty, it is unlikely that related new knowledge will be robust and useful. Recall, the theory of meaningful learning:[59,60,209,210] the idea that what is learned must be integrated with existing knowledge and must also be perceived by the learner as being useful. As discussed earlier, we now know the idea that student misconceptions can be identified early, and then remedied by overwriting with the correct information, is problematic in many instances.[26,47] Despite this, instructors who know and understand the prior knowledge that students arrive with are more likely to productively address these ideas during instruction. This is part of what is known as pedagogical content knowledge (PCK):[211] that is, the knowledge associated with how particular ideas may be problematic for students and how to help students construct coherent understanding.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Question	Rubric
Dimethyl ether (CH_3OCH_3) and ethanol (CH_3CH_2OH) have the same molecular formula, but one of these compounds is a liquid at room temperature and the other is a gas. Draw Lewis structures for each substance and use them to help you determine which substance is a liquid. Provide a molecular level explanation for your choice, being sure to include a discussion of the interactions and energy changes involved.	• Correct Lewis structures **Claim:** • Ethanol is a liquid at room temperature, while dimethyl ether is not **Evidence**: • Ethanol is capable of hydrogen bonding, while dimethyl ether is not. **Reasoning:** • The interactions between ethanol molecules are stronger than the interactions between dimethyl ether molecules • Therefore, more energy is needed to overcome the attractions between ethanol molecules compared to dimethyl ether molecules, • More energy needed to overcome attractive forces between molecules corresponds to a higher temperature, meaning ethanol will boil at a higher temperature.

Figure 5. Prompt designed to engage students in argumentation from evidence. Adapted with permission from *J. Chem. Educ.* **2015**, *92* (8), 1273−1279.

A significant problem inherent in the format of almost all CIs is the tendency of multiple-choice assessment items to overstate students' understanding.[212,213] The constraints imposed on the student by a selected response prompt can help guide them to a correct answer, even if their underlying reasoning is inappropriate.[32] Further, as has been discussed by diSessa,[26,27,47,122] Elby,[214] and Hammer[215] (among others), many misconceptions are almost certainly not coherent or stable and thus attempts to discern their presence or absence may be oversimplifying the situation. Stated differently, prompts designed with the express purpose of ferreting out a misconception might cause a coalescence of fragments that may give the appearance that the student has a robust misconception. If the context of the prompt is removed, the same student who appeared (by CI) to have a particular misconception might instead reveal a set of fragmentary ideas that, while not expert-like, is also not a rationally constructed wrong theory.[32] As we will see, the nature of the prompt has a significant impact on the evidence that is evinced in the students' responses.

Despite limitations inherent in the formatting of CIs, prompting students to consider why a given answer is incorrect may be a useful exercise. Talanquer[216] has recently reported that having students explain why the wrong answer is wrong can activate student knowledge about a particular concept and override heuristic reasoning (type 1 cognition) that is often prompted by such questions. Thus, reflection on incorrect answers may induce a more coherent coordination of resources than simply answering the selected response prompt.

Although little explicit mention of the assessment triangle occurs in the concept inventory literature,[217] the assumptions made by CIs may be mapped onto the triangle's three vertices. For instance, CI designers assume, either implicitly or explicitly, that student understanding of concepts is more-or-less coherent and stable and that the character of a student's conceptual understanding can be gleaned from their answer selection on multiple choice items. As we have just discussed, there is little evidence that these assumptions are warranted. Thus, CIs should be viewed as diagnostic of particular knowledge fragments coordinated in the context of a prompt, which may or may not reveal stable conceptual difficulties. CIs have also been used to measure learning gains,[218] using prepost analyses. However, for all the reasons cited above, results from such studies must be considered with caution.

5.3. Alternative Assessments

While there has been general acceptance (at least among those who think about such things) that students construct their own knowledge, there has not been a concomitant shift in the way this knowledge is assessed. As Shepard notes, "One hundred years ago, various recall, completion, matching, and multiple-choice test types, along with some essay questions, fit closely with what was deemed important to learn."[219] This type of test is still common today, despite the fact that we now have a fair amount of evidence that such tests may not provide strong evidence about what students know and can do.

Most recent scholarly work dedicated to eliciting evidence of what students know and can do has focused on attempts to capture knowledge-in-use. That is, assessment items that require students to draw on knowledge anchored to core ideas to predict, explain, and/or model phenomena. In the language of *The Framework*, such assessments would be 3-dimensional because they require disciplinary core ideas to be extended through engagement with scientific practices and crosscutting concepts.[220,221]

Prompts that can provide evidence of facility with practices and core ideas are more likely to demonstrate knowledge coherence by illumining the chain of inferences students use when, for instance, crafting a molecular-level explanation. By contrast, it is difficult (and perhaps impossible) to gain strong evidence of coherent student understanding in the absence of knowledge use. Consider, for example, a question from the Chemistry Concept Inventory[97] that asks students to select which substance (water or alcohol) takes more energy to heat to a specific temperature. The correct answer choice is "water takes more heat energy". The corresponding correct response to the "justification" tier is that "water has a higher specific heat so it heats slower." However, this justification does not address the reasons why water has a higher specific heat than alcohol (i.e., the

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

relative strength of intermolecular interactions and the fact that there are many more water molecules per gram).

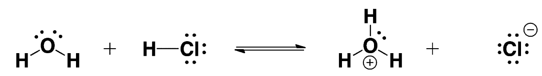
The path to development of assessment tasks (observations in the assessment triangle) that can elicit evidence of student engagement in contextualized practices has been the focus of several publications.[220−222] Laverty et al. have published a protocol that is useful in setting some basic criteria that assessment items must meet to have the potential to engage students in each of the three dimensions (SEP, core ideas, and crosscutting concepts).[223] According to these authors, tasks designed to produce evidence of students using their knowledge in the context of a SEP typically require explicit prompts for student reasoning. In the absence of asking students to justify their claims, it is argued, we cannot know whether they have thought through their answer using appropriate logic and principles or used a readily recalled rule of thumb.[224] Further, by requiring students to think deeply about a topic, we are making it clear that reliance on heuristics (i.e., type 1 thinking) is not what we are looking for. As we discussed earlier, students tend to value what is assessed. If we do not change the nature of the assessments in a transformed course, students are unlikely to change their approach to learning.

A prompt designed to engage students in argumentation from evidence is shown in Figure 5. This item was carefully scaffolded to elicit evidence that students understand the relationship between molecular-level structure and macroscopic properties. Accordingly, students are encouraged to make use of molecular representations as part of an argument grounded in forces and energy. The prompt shown in Figure 5 provides much stronger evidence that students can infer molecular-level causes from macroscopic observations than all-too-typical prompts that simply require one to choose the compound with the higher boiling point. In point of fact, ranking tasks are very often answered using simple heuristics rather than first-principles reasoning.[134] Questions with the potential to engage students in a scientific practice can be scored by rubrics such as that shown in Figure 5 and also coded to describe the level of sophistication evident in student responses.

It is crucial to note that the criteria published by Laverty et al. are the requirements for a task to have the *potential* to engage students in an SEP but do not guarantee that such engagement will occur. Students must receive and respond to the message that type 2 thinking is required in a particular context for a prompt to *actually* engage students in a scientific and engineering practice. There is evidence that the way the prompt for an assessment item is structured can send powerful messages to students about how they should respond.[225] An overly structured task can provide students with too much information and thereby overestimate student understanding (as multiple-choice items tend to do).[226] By contrast, if too little structure is offered, students might not understand what constitutes an acceptable answer, particularly if they are unused to responding to such items. The "goldilocks zone" of prompt development is that middle ground where there is enough structure to convey the importance of supporting claims with evidence and reasoning.

The importance of calibrating prompt structure was shown in a study on reasoning about acid−base reactions.[176] Initial prompts (prompt 1 Figure 6) elicited mainly descriptions of the reaction. The second prompt (prompt 2, Figure 6) activated cognitive resources that allowed more students to provide a fuller causal mechanistic explanation. By asking separately about how and why a reaction occurred, more students understood that a descriptive account was not a sufficient response.

For this reaction:



Prompt 1 (Spring 12)

a. How would you classify this reaction? Please explain your reasoning.
b. Please explain your reasoning for what you think is happening at the molecular level for this reaction.

Prompt 2 (Spring 15)

a. How would you classify this reaction? Please explain why you chose that classification.
b. Describe in full detail what you think is happening on the molecular level for this reaction. Specifically, discuss the role of each reactant.
c. Using a molecular level explanation, please explain why this reaction occurs. Specifically, why the reactants form the products shown.
d. Please draw arrows to indicate how this reaction occurs.

Figure 6. Two prompts designed to elicit student explanations about an acid−base reaction. Adapted with permission from ref 176. Copyright 2016 American Chemical Society.

The question now arises: what are the attributes of student responses that signify appropriate use of contextualized scientific ways of thinking (i.e., practices)? In contrast to many assessment items, responses to 3-dimensional tasks can vary substantially in character. The spectrum of possible responses to a prompt varies according to the SEP being assessed *and also* the particular context in which it is used. The character of the evidence and reasoning that might be used to justify claims about, say, phase changes differs substantially from what may be useful in explaining atomic emission spectra. "Coding schemes" by which student SEP engagement might be categorized have been published for a variety of practices (e.g., explanation,[176,177,227] analysis, and use of data[228]) in several contexts (intermolecular forces,[33,177,229] reaction rate,[228] and reaction cause[176]). These schemes are useful in tracking the evolution in student thinking over time as well as the response of student cohorts to a particular curricular or pedagogical intervention. Coding schemes may either emerge from student data (i.e., prompt responses or interviews) or from expert views on solving particular problems.[227] The vast majority of such schemes published in the literature were derived from student answers to specifically designed questions.

Although the bulk of our discussion regarding alternative assessments has focused on knowledge use in the context of scientific practices, there are other potentially productive ways students might demonstrate what they know and can do. For example, Zoller[183] has advocated for the development of examinations that target higher-order cognitive skills (HOCS). The multipart examination prompt in Figure 7 has been presented as an example of exemplary HOCS questions.[181] Clearly the HOCS questions could also be "tagged" as requiring scientific practices to answer them.

5.4. Summary

As should be apparent by now, there is no easy answer to "how will we know when students have a coherent and useful understanding of chemistry?" We can gather only indirect evidence of understanding from assessment tasks and from there make inferences about the nature of students' knowledge structures. Assessments are the means by which the chemistry education community calibrates and evaluates its experiments, and so the assumptions underpinning them are tremendously important. There are a range of assessment types that can be used

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

In a battery factory, workers are exposed to ZnS and $CdCl_2$ (in the manufacture of electrodes); HCl (in the preparation of the electrolytic bridge): oily grease (from the oily metal parts); CH_2Cl_2 (a solvent for cleaning the grease); and H_2S. A suggestion was made to replace the water by petroleum for washing the workers working clothes.

1. Do you think that the idea of replacing the water with petroleum is good from the point of view of cleaning the cloth? Explain (Question level HOCS) (Alternative designation: Engaging in argument from evidence)
2. What is the possible source of the poisonous H2S in the battery factory? Explain and write the relevant chemical equation (Question level LOCS) (Alternative designation: Constructing explanations – but would require an additional prompt to elicit appropriate reasoning about the mechanism of the reaction)
3. Based on the chemistry you know propose a simple practical method to overcome the H2S problem in the factory (Question level LOCs+) (Alternative designation: with additional prompting this could address the SEP Designing solutions – which is formally an engineering practice).
4. Do you think that the idea of replacing the water with petroleum is good from the point of view of the environment outside the factory? Explain (Question level HOCS) (Alternative designation: Engaging in argument from evidence)

Figure 7. HOCS and LOCS questions and their corresponding SEPs. Adapted by permission from Springer Nature: From Teaching to KNOW to Learning to THINK in Science Education by Zoller and Nahum. Copyright 2012.[183]

for a variety of purposes. The ACS nationally normed examinations provide evidence about student knowledge of content details and can be compared to others on a national scale.[190] Concept inventories provide diagnostic information about student knowledge of particular concepts but cannot provide evidence about how robust that understanding is. Inferences grounded in student use of knowledge typically provide stronger evidence of knowledge structure coherence than those grounded solely in content mastery,[19,230,231] but few other generalizations can be made. That being said, as we move on to question 3, the evidence we seek must come from the types of assessments that are available to the researchers.

6. EVIDENCE ABOUT HOW TO HELP STUDENTS DEVELOP A DEEP AND ROBUST UNDERSTANDING OF CHEMISTRY

Our focus for the preceding sections has been on learning objectives (what should students know and be able to do) and assessments (how we will know students have molecular-level understanding). In this section, we will turn our attention to studies that suggest ways in which we might help students develop a deep and robust understanding of chemistry. In particular, we will examine the strength of the evidence underpinning major pedagogical and curricular approaches published in the CER literature. Three major areas of research on improving learning will be discussed: (1) student engagement, which involves changes in pedagogical approaches; (2) research on visualization and representation of the molecular level; and (3) instructional technology and the evidence about online and computer-supported education; in section 7 we will consider curriculum design, which ties much of the preceding research on teaching and learning together.

We should note that the majority of studies outlined here support their conclusions with only limited evidence (as defined by the NRC DBER consensus study[2]). That is, there are few examples of large-scale, longitudinal, or multi-institutional research findings. Research is expensive, and CER studies typically take years, not weeks or months. Funding for such longitudinal studies is rare, and therefore many published studies are small-scale studies of a single cohort. Aggregation of the findings from smaller studies must be done with great care. There is an emerging understanding that studies that are underpowered (that is, quantitative studies that do not have a large enough group of participants) can produce findings that appear to be statistically significant but are not reproducible with larger cohorts.[4,232]

Work on improving student molecular-level understanding is, in essence, focused on improving measurable student outcomes and using those outcomes as a proxy for understanding. There are, of course, many possible outcomes and a great many ways they might be measured. It falls to the CER community to determine which outcomes most persuasively indicate coherent and useful understanding and how those might be best measured. Content knowledge as measured by the ACS examinations Institute exams or Concept Inventories is one of the most commonly reported outcomes in the CER literature. Other outcomes that may act to mediate learning, such as cognitive and affective expectations,[233] attitudes,[234,235] identity,[236] meaningful learning (in the laboratory),[237] self-efficacy, and metacognition,[238] have also been measured and reported in the literature.

Because this Review cannot hope to be comprehensive of every aspect of chemistry education research, we will leave a longer discussion of noncontent-related measures to others. However, it should be noted that, as predicted by the learning theories of Ausubel and Novak, many factors beyond the material to be learned can affect student learning in powerful ways that we are only beginning to understand.[239]

6.1. Research on Student Engagement

Perhaps the most profound changes in higher education over the past 25 years stem from the recognition that students actively construct their understanding[50,51,240] during the course of their learning activities. Therefore, learning environments, experiences, and curricula that explicitly promote construction of knowledge have been proposed and enacted as ways to improve student understanding. Work that aims to promote "active learning"[241−243] or "student-centered learning" employs a wide variety of approaches, ranging from the use of clickers in large enrollment classes[244] to specifically designed lecture−lab studio classes where students work in collaborative groups.[243,245] There is strong evidence that providing students with opportunities to discuss, engage, and interact with each other and the instructor

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

improves outcomes as measured by course grades and/or concept inventories. The first large study on the effects of interactive engagement was published by Hake in 1998,[218] using a prepost Concept Inventory score as the measure of change for over 60 physics courses. Hake showed that courses using engagement techniques tended to show significant gains on the CI. In the largest to date study of active learning effects,[242] Freeman and co-workers published a meta-analysis of 225 studies that showed students in traditional lecture-based (STEM) classes were 1.5 times more likely to fail than those in active learning classes, with the most marked improvements occurring for students in the lower half of the typical grade distribution. While the Hake and Freeman studies are powerful and persuasive findings, there are several limitations to this meta-analysis approach, as the authors necessarily relied on the methods used in the original studies surveyed. The assessments used in the analyzed studies, which were either course grades or concept inventories, may not provide strong evidence of coherent and useful disciplinary knowledge (as discussed earlier). Further, as the authors did not disaggregate types of active learning in their analysis, it is impossible to infer anything about the relative effectiveness of particular methods. Viewed in the most reductionist sense, these studies can be interpreted as indicating that the umbrella category of active learning is universally superior to traditional lecture. Such a simplistic interpretation might lead some to condemn lecture as useless at best and harmful at worst. However, there is evidence that the active learning versus lecture dichotomy may be too simplistic and that particular sorts of engagement might, in fact, prime students to consolidate their understanding from engaging with a lecture. For example, Schwartz and Bransford have shown that having students analyze contrasting cases, followed by a lecture in which the conceptual underpinnings of those cases is developed and explained, is an effective approach to instruction.[246]

In light of the potentially powerful impact of active learning on student understanding, it is natural to wonder *how* active engagement improves outcomes. Essentially all active learning methods facilitate, in some manner, student interaction with each other or the teacher. Recall that Vygotsky (and many social constructivists since Vygotsky[51]) believed that knowledge is coconstructed via interactions with other people—either peers or knowledgeable others. Methods that promote construction of knowledge in a community, from a sociocultural perspective, would thus be expected to improve student understanding. As we will see, evidence supporting the efficacy of particular coconstruction strategies exists but is fairly limited. Our discussion will focus on the two most widely known strategies for knowledge coconstruction in chemistry: peer-led team learning and process-oriented guided inquiry learning.

Peer-led team learning (PLTL) uses experienced undergraduates (Vygotsky's knowledgeable other) to lead teams of students as they interactively engage with tasks specifically designed for teams. Typically, PLTL does not require changes in curriculum or assessment, and the teams meet out of the normal class time. Wilson and Varma-Nelson have published a review of the numerous ways PLTL has been implemented.[54] They note that PLTL has been shown to improve course grade distributions in 13 of the 39 surveyed studies.[247−252] Other notable findings were that affective outcomes such as motivation or attitudes were sometimes (but not always) improved. Studies on the effects of PLTL on the peer leaders themselves indicate that the peer leaders generally self-report increased levels of confidence, self-efficacy, and content knowledge. In contrast, Chan and Bauer reported no difference in exam scores for students who were randomized into either PLTL groups or a control group (balanced for time on task and study).[253] He proposed that, because many of the earlier studies were quasi-experimental (that is students self-selected into the PLTL sections), perhaps activities such as PLTL attract students who are motivated to take advantage of them.

Process-oriented guided inquiry learning (POGIL) is another approach specifically designed around a social-constructivist approach to learning. POGIL is characterized by students working in groups to construct concepts emergent from data by discussion and explanation. A recent meta-analysis[254] of studies on the effectiveness of POGIL reported similar results to the Freeman study: that is, the effects on achievement were marginal, but the odds of passing the course were substantially improved for students in POGIL courses versus traditional courses. Because POGIL courses typically provide students with experiences that involve scientific practices, it might be expected that this approach would support the development of student expertise. However, there are, as yet, no reports of assessments used to evaluate POGIL that compare the ability to predict, explain, or model with students in traditional courses. For this reason, it is currently not possible to say whether or not POGIL promotes the development of knowledge useful in making sense of the world.

Although it is widely recognized in the CER literature that opportunities for knowledge coconstruction with peers can improve learning outcomes, the exact parameters of collaborative learning environments that are most beneficial are less well-understood. For example, there are varying recommendations on how to constitute student groups for collaborative problem-solving, specifying, among other things, appropriate group size and heterogeneity. Much of the original research in this area was with much younger students,[255] and it is not clear that (for example) a group of senior chemistry majors working on a complex semester capstone project needs the same kind of group structure as a group of first-semester students in an unfamiliar laboratory setting. That being said, most practitioners recommend that groups be as heterogeneous as possible to tap into the different strategies that a diverse group can provide,[256] with the caveat that students who may be underrepresented (such as females or minorities) should not be isolated within a group.

In one of the few studies aimed at identifying the effect of group composition on problem-solving ability at the undergraduate level, students were assigned to pairs on the basis of their score on the Group Assessment of Logical Thinking (GALT),[257] to work on online open-ended chemistry problems intended to simulate qualitative analysis problems.[258] On average, students improved by working in a collaborative group; this improvement was maintained in subsequent problem sets and, more importantly, when subsequently working alone. The composition of the group did not seem to matter because each possible combination of students improved their problem solving, *except* for groups that contained only students who had scored in the lowest category on the GALT. This finding supports Vygotsky's notion of the knowledgeable other: that interaction with more-expert peers can improve one's understanding. Groups with students who were all low performers were unlikely to succeed, possibly because they did not know how to get started on the problems. In contrast, low-scoring students paired with students who had stronger logical-thinking skills improved just as much as other students in their ability to solve qualitative analysis problems and maintained their

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

improvement in subsequent problems. In this study, the female students improved more than males. In fact, working in groups appears to level the playing field for females. It has been proposed that instructional approaches that place value on social interactions improve female motivation and achievement.[259] For example, a study of undergraduate engineers showed that female students tend to use more collaborative strategies when learning and that collaborative strategies improved their course grades.[260]

While there is little doubt that providing students with opportunities to collaborate improves outcomes for lower-performing students, the mechanism by which this occurs must be elicited by studies that probe how groups interact and function. To that end, researchers have studied the interactions of students working in collaborative groups to identify the kinds of productive interactions that may lead to improved learning outcomes. Most often, qualitative research methods have been used to analyze student discourse in order to search for productive patterns of interaction. For example, students interactions in a physical chemistry setting were analyzed using Toulmin argumentation patterns, which enable characterization of students' use of evidence to support claims.[261] The authors propose the emergence of *sociochemical norms* as students discuss and analyze data to make arguments about the meaning of the data.[262] In a related study, Lewis and co-workers report that general chemistry students are able to coconstruct arguments and that they are able to resolve claims that are in error without intervention by a peer leader.[263] In general, if learning opportunities that allow students in groups to engage in scientific practices are appropriately structured, then the kinds of dialogue that allow cognitive growth, reflection, and metacognition are probably more likely to occur.

One thing that appears to be quite consistent among all these studies is that they all seem to promote improvement for lower-performing students. There are few reports of improvements for higher-achieving students. There are a number of possible reasons for this observation. It may be that the assessments used in these instances are not appropriate to measure improvements in higher-achieving students. That is, the students may be "maxed out" and there is little room for improvement. If students were assessed with more challenging items, it is possible that the higher-performing students would also increase their scores.

Promotion of student metacognition is a prime goal of many teaching strategies that fall under the active learning umbrella. Metacognition has been described as *thinking about thinking*,[264] or the capacity to reflect about one's own thinking.[265] Theoretical models of metacognition generally propose two components: knowledge of cognition, which includes knowing about things or how things should be done, and regulation of cognition, which involves knowing when and why things should be done. There are a number of researchers who have addressed the role of metacognition in learning. Clearly the idea of reflecting on how and why a particular task should be done should lead to more thoughtful approaches to a particular task. McGuire and co-workers[266] have reported that a 1 h seminar designed to teach students metacognitive learning strategies improves outcomes for those students who attend. Sandi-Urena et al. have developed and validated an instrument (the metacognitive activities instrument MCA-I) in which students self-report the degree to which they engage in metacognitive behaviors while solving problems. Interestingly, if this instrument is used in prepost studies, where the intervention is designed to promote metacognition, the scores on the MCA-I go *down*.[267,268] This is almost certainly attributable to students in the pre condition lacking understanding of what is involved in problem-solving tasks, as well as a heightened awareness of what is necessary in the post condition.

Active or student-centered learning is often closely associated with a particular classroom format. For example, the "flipped classroom" approach, in which students read the text or watch out-of-class videos and then come to class prepared for class activities has been enthusiastically adopted by many who desire greater student engagement.[269−274] Unfortunately for classroom flippers, it is quite difficult to provide strong evidence for the efficacy of flipped course formats. To persuasively suggest improvements in learning are due to flipping would require control-treatment studies in which students are randomly assigned to each condition, and some way of ensuring that the instructor of each section is equally invested in both approaches. Most studies of flipped classroom outcomes suffer from one or more experimental design issues including small sample sizes, comparison of nonequivalent groups, difficulty in controlling instructor bias, and a tendency to compare the flipped classroom to a traditional classroom, rather than another classroom in which evidence-based practices are being used. Because the flipped classroom is based on the same theoretical constructs as other engaged learning approaches, there is no reason to suppose that it is more or less effective than those other approaches until appropriate studies are performed.

The use of classroom response systems has also been promoted as a way to promote student engagement, particularly in large enrollment classes where such systems can be added with relative ease. Remote response systems (clickers) can enable a class to answer a wide variety of questions—from multiple-choice prompts to open-ended, group responses. The review by MacArthur and Jones[275] on the use of clickers in chemistry classrooms reported that clicker use is accompanied by disparate results (some studies showed improvement, and some did not) but that all the reports that produced improvement in learning required some kind of collaboration. It is thus reasonable to infer that it is not the clickers themselves that improve learning but rather the collaboration that they can mediate.[250] As we will see is the case generally with technology in learning, it is not the tool but rather how it is used that can promote learning.

6.2. Research on Representation of the Molecular Level

Every discipline has particular difficulties associated with it. Physics may seem problematic because students arrive with preconceived notions about how the macroscopic world behaves. Biology deals with systems that are so complex it is difficult to simplify them in a way that makes sense while still maintaining a connection to core ideas of the discipline. However, chemistry deals with the unseen and (to the naked eye) unseeable world of atoms and molecules. To understand chemistry, students must learn different characterizations of the molecular level, including different types of structural representations and different approaches to visualizing molecular-level phenomena.

Because the entities that are responsible for the macroscopic behavior of substances are too small to see and their behavior is governed by the vagaries of quantum mechanics, chemistry educators have long sought ways to make the molecular level more accessible to students. Much of this work has been gathered under the general umbrella of "visualization". Johnstone's triangle explicitly calls out representations of the molecular level as being crucial to an understanding of chemistry, and the ACS curriculum map includes visualization as an anchoring

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

concept (although one might classify visualization as a skill rather than a concept). In 1948, long before Johnstone published his triangle, Garrett discussed how the use of physical molecular models can serve as resources to the lecturer. He noted that "the mental picture of molecules in motion, the simple structural facts, and a picture of size and configuration of atoms and molecules make it possible to build most of the core of our course in general chemistry in a consistent pattern which is a solid foundation for further work, and which gives a very simple treatment of most of the topics. In this way, general chemistry is no longer a collection of seemingly unrelated facts—it has a golden thread of simple, sound logic on which to relate the necessary descriptive material." It may seem that there is nothing new under the sun; Garrett's optimism about the use of diagrams and models to represent the atomic/molecular scale has resonated through thousands of chemistry classrooms and lecture halls. Students are often advised to purchase molecular models,[276] and increasingly sophisticated diagrams, movies, and simulations are included with modern chemistry books and online materials. The question is to what degree is Garrett's optimism supported by evidence? Answering this question is complex, as you might expect, because of the many types of representations/models and the many ways they might be used.

Much of the research on the use of handheld models has been carried out in the context of organic chemistry. Stull and co-workers[277] propose three ways in which using 3D models (both physical and virtual) might help students develop what is called representational competence in organic chemistry: (1) 3D models allow students to off-load cognition. That is, models reduce the load on the working memory by reducing what students must hold in their memory as they manipulate the model, thus freeing up more cognitive resources to work on the problem at hand. (2) Using multiple representations (i.e., models in conjunction with structure drawing) helps students develop more complete mental models. (3) 3D models provide ways for students to connect and integrate new knowledge with existing knowledge by "offering multiple perceptual modalities to support encoding and recall of the represented molecules". That is, physical enactment and nonvisual information such as touch and sensing of self-motion should lead to better information encoding and recall.

Stull and co-workers' proposed mechanisms by which models improve representational competence are grounded in evidence to varying degrees. For example, it is reasonable to infer that the use of models significantly improves students' ability to translate between different representations of structures by off-loading cognition. However, Stull and co-workers' proposal that physical manipulation of molecular models would lead to improved ability to translate among visualizations has not held up to experimental scrutiny thus far. No difference in performance was observed between students who used handheld models and those who used virtual models (manipulated by either gesture control or traditional mouse/trackpad inputs). This suggests that the use of virtual models is as effective as handheld physical models, although this finding clearly warrants further study.[278] Interestingly, students who learned to use models to guide their drawing of different representations performed significantly better on a delayed post-test, even when models were not available. This suggests that the use of models scaffolds learning, rather than acting as a crutch.

Unfortunately, fewer than half of students with ready access to molecular models make use of them.[279] This remains unchanged even if instructors strongly encourage student model use. Requiring students to use models leads to significantly improved performance on representational translation tasks relative to students who have access to but do not choose to use models.[280] The reluctance of students to make use of available models is reflected in other studies in which students are provided with the choice of representations to use. Not surprisingly, they typically use the most familiar representations rather than the best representation for the job at hand. Using eye-tracking software, Williamson and co-workers found that students tended to fixate on familiar ball-and-stick representations rather than electrostatic potential maps when asked to determine charge delocalization.[281] In a similar vein, Cox[282] found that students who were asked to assign configurations to chiral centers tended to use familiar representations (i.e., those used by the instructor) such as sawhorse or wedge-dash structures rather than 3D manipulable structures. At a minimum, this suggests that instructors who want students to use either physical models or virtual models must ensure that students are familiar with how to manipulate them and what they represent.

A significant proportion of the chemistry education research literature, including a number of reviews, has been focused on computer visualization of the molecular level.[283−289] There is some overlap between computer visualization, virtual molecular models, and simulations of molecular-level phenomena (for instance, they each involve computer-generated imagery). In general, computer visualizations depict some aspect of a molecular-level phenomenon in a dynamic way. Unlike simulations, computer visualizations do not allow the use to change the system variables and see those changes borne out in the visualization. Thus, a computer-generated video of molecular behavior as water changes phase would fall under the heading of visualization rather than simulation.

Much of the research on computer visualizations in chemistry has focused on elucidating those aspects of molecular-level visualizations that are helpful, what inappropriate ideas might be induced by visualizations, and how design of visualizations can impact learning. Increased understanding of the particulate nature of matter is a major focus of this corpus of work.[208,285] In their early paper in this area, Williamson and Abraham propose that the increases in conceptual understanding seen in their studies "may be due to the superiority of the formation of more expert-like, dynamic mental models of particle behavior in these chemical processes."[285] Other work has focused on the design of visualizations: for example, should the visualizations be (relatively) realistic, such as those developed by Tasker[290] in which molecules and ions are solvated and constantly in motion, or should they be less visually rich, so that students can attend to important ideas? Do visualizations that contain extraneous details and large numbers of molecules in motion overwhelm the working memory of beginning students? Rosenthal and Sanger proposed that, when given a choice, students should see the simplified visualization first, followed by the richer examples.[288] However, there is some debate about this approach, because simpler visualizations may also increase incorrect ideas.[287,291,292] Similarly, there is no evidence-based consensus on the ways in which visualizations should be introduced into learning activities. However, Tasker has proposed that a scaffolded introduction to rich visualizations over time is more likely to produce improved understanding and ability to reason at the molecular level.[293]

As technology has improved, computer visualizations have proliferated. Most commercial texts now provide a full suite of such materials, and while it might seem obvious that such visualizations should help students connect the three "levels" of

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

Johnstone's triad (macroscopic, symbolic, and particulate), the evidence for their efficacy is somewhat sparse. Providing students with molecular-level visualizations should improve the way they think about molecular-level structures and phenomena; there is little agreement about how and when such visualizations are effective. As Stieff has noted,[294] "There is increasing evidence that dynamic visualizations can help improve science learning; however, the observed improvements in learning outcomes are often marginal, which challenges the wisdom of their widespread adoption."

Proponents of molecular-level visualizations and models often mention the potential of representational tools to aid students with limited spatial reasoning skills. It has been proposed that, because chemistry requires one to visualize at the molecular level, students with poor spatial reasoning skills may find the subject especially challenging. While this may seem a reasonable inference, there is little evidence for a causal relationship between spatial reasoning ability and success in chemistry.[11,289] The typical tests used in such studies are not highly correlated with most chemistry tasks,[295] and while there are a few studies that do support the notion that higher spatial skills are correlated with chemistry achievement in some areas, the reported effects are small.[289,296] More recently, Lopez and co-workers reported that spatial visualization skills are not correlated with problem solving in organic chemistry.[297] In fact, Stieff has shown that, for tasks such as assigning configuration, experts tend to use discipline-specific strategies rather than mental rotations and that students who are taught such strategies are significantly more accurate at assigning configuration than those who were taught to use mental rotations.[298] Further, while mental rotation is one of the few areas of cognition where females tend to have lower abilities than males, use of discipline-specific strategies removes any differences.[298]

6.2.1. Student-Constructed Visualizations: Drawing and Modeling. While much of the research related to visualization has focused on computer-supported approaches, there is a growing recognition that student-constructed visualizations can serve as a way to make student thinking visible[299] and also to help students construct appropriate mental models of molecular-level processes. Research in this area spans a wide range of studies from those that investigate how students draw representations of molecular-level phenomena (e.g., dissolution,[300] phase changes,[32] or intermolecular interactions[33]) to studies that investigate how students construct and use chemistry-specific symbols and iconography to convey both chemical structure[301] and reaction mechanisms.[302,303]

It has been proposed that drawing can be a powerful support for learning science,[299] and chemistry in particular.[294,304] Studies on the efficacy of drawing as an instructional tool have shown mixed results, and just as with other instructional approaches, the way that drawing is integrated into the learning activity affects outcomes. A drawing activity must be accompanied by appropriate scaffolds and supports that guide student attention and constrain the features to be depicted.[292] Ainsworth et al. have suggested[299] that it is not the act of drawing itself that brings about understanding but rather the process by which students must determine what aspects of a particular phenomenon to represent. That is, drawing is akin to self-explanation,[305] or perhaps even metacognition; the act of drawing requires students to plan, monitor, and evaluate what they are putting forth as a representation.

Drawing may also provide instructors with valuable information about student understanding. As previously mentioned under the heading of question 2, strong evidence of coherent and actionable understanding can only be obtained by tasks that require students to use their knowledge to predict, explain, or model. Construction of an appropriately elaborated molecular visualization can serve as part of an explanation, prediction, or model and may thus help make student thinking visible. For example, in a multimodal study on how students understand and represent intermolecular forces, students wrote apparently coherent descriptions of what they meant by intermolecular forces (IMFs). However, when they were asked to draw three molecules of ethanol and indicate the location of IMFs, the majority of students chose to represent each type of IMF (hydrogen bonding, dipole−dipole interactions, and London dispersion forces) as interactions (often bonds) within an ethanol molecule.[33,229] Of 97 students in one trial, only one student consistently drew IMFs as operating *between* molecules of ethanol. In this study students' drawings provided evidence about what they understood about IMFs that contrasted sharply with their written definitions. Clearly, students were able to memorize textbook IMF definitions without understanding the molecular basis for those definitions.

Several authors have incorporated drawing as part of instruction augmented by dynamic visualizations and simulations.[208,291,292,306] For example, Kelly and Jones found that many students were able to construct an appropriate representation of sodium chloride dissolving in water with the aid of relevant visualizations. Unfortunately, these visualizations also appeared to promote incorrect ideas among some of the tested student population. In addition, Kelly and Jones found that students are often unable to transfer the model they construct with the aid of a visualization to a similar process. For example, although students were able to construct a molecular-level representation of sodium chloride dissolving in water with assistance from a visualization, they could not transfer their understanding to the reverse process of salt precipitating from solution.[291]

6.2.2. Research on Student-Constructed Representations of Structures and Reaction Mechanisms. Drawing chemical structures is a different skill than drawing a representation of a molecular-level phenomenon. One might imagine that this skill is easier to acquire, as ostensibly students must follow relatively simple rules that do not necessarily require an understanding of the chemistry itself. Unfortunately, there is ample evidence that students often struggle with generating chemical structures. While there are numerous reports on improved approaches to structure drawing, few are supported by data. One exception to this was informed by a report that students in organic chemistry courses often struggled with drawing simple Lewis structures without being given structural cues.[301] On further investigation, it was determined that many students did not understand the purpose of drawing such structures.[301,307] Even some chemistry graduate students did not understand that structural representations such as these can (and should) be used to predict chemical and physical properties. For these students, learning to draw structures did not meet the criteria for meaningful learning, that is, it was not perceived as relevant or useful. To test whether or not emphasizing the utility of Lewis structures might improve students' ability to draw such structures, students in a transformed curriculum where the sequence of ideas and skills that students must use to determine properties from structures was explicitly linked were compared to matched cohorts of students from traditional curricula. This study showed that students from the transformed curriculum

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

were significantly better at drawing Lewis structures than those in the traditional cohort and that the difference in ability increased over the course of two semesters.[33,229] Moreover, students were also more aware that such structures could be used to predict properties,[307] and again, the difference in improvement increased over the span of two years of chemistry—through organic chemistry.

Perhaps the fact that many students are unaware of the connection between structure and properties contributes to the findings from a range of researchers on the difficulties that students have in drawing reaction mechanisms in organic chemistry.[303,308−311] If students do not emerge from general chemistry with the ideas and skills necessary to succeed in organic chemistry, they will have little recourse other than to memorize, rather than understand, how molecules interact. Ideally, we might want students asked to draw a particular mechanism to identify the electrophile and nucleophile from an inspection of reactant structure, use curved arrows from the nucleophilic center to the electrophilic center, draw the result of that electron movement, and carry on until a recognizable product is produced. In practice, there are numerous research studies showing that many students do not follow this approach. Bhattacharyya and Bodner described graduate student struggles with mechanism tasks,[308] where graduate students said "it gets me to the product" but showed little understanding of how and why arrows fly the way that they do.

After conducting a meta-analysis of the research on mechanistic reasoning in organic chemistry, Bhattacharyya[310] proposed that graduate students solve mechanism problems by (1) mapping the reactant onto the product and (2) determining whether the process is a one- or multi-step reaction, which leads to either recall of canonical reactions (e.g., S_N2) or functional group transformations. Often, adding the arrows is the last step. That is, even graduate students do not tend to use the predictive power of arrow pushing. It should be noted that most studies investigating mechanistic reasoning are conducted using prompts that give the product of a particular reaction. There are fewer studies in which students are asked to predict the products by using mechanistic arrows, but the results of those that exist are not encouraging. Grove and co-workers found that less than half of students in a typical organic course tended to use mechanisms at all, even when prompted.[303] Indeed, around 20% of students actually added mechanistic arrows after drawing the (memorized) product. Interestingly there was no difference between mechanism users and nonmechanism users in the likelihood of correct product prediction or in overall course grades. However, those students who used mechanisms appropriately were more likely than the nonmechanism users to predict the correct product for a reaction that they had not seen before.[309]

Most studies related to student mechanism use provide suggestions for how to improve the teaching of curved-arrow formalism, but as yet there are no studies providing strong evidence that a particular approach develops improved use of arrows to predict mechanisms. However, it is important to note that typical organic examination questions do not tend to promote the kind of predictive mechanistic thinking described here. Analysis of prompts from examinations given at elite institutions as well as questions on an ACS organic chemistry examination shows that there are few items that are designed to elicit explicit reasoning.[222] We cannot be surprised if students struggle with predictive mechanistic reasoning if they have been able to get by throughout their chemical careers without it—as evidenced by Bhattacharyya's review.[310]

There are a number of ongoing studies aimed at improving students' mechanistic reasoning. For example, Flynn and Featherstone[302] have designed a curriculum in which students learn to use mechanistic arrows before they learn why substances interact. Assessments during the first part of this courses consist of problems in which students are provided with chemical structures and mechanistic arrows for reactions that they have not learned yet. Students are asked to (1) predict the structure of the product from given starting materials and mechanistic arrows or (2) fill in the mechanistic arrows on the starting materials that yield the given product. Flynn's "formalism first" approach to curriculum transformation was designed to reduce the demand on student working memory. She proposes that learning to use mechanistic arrows fluidly should lower the cognitive load of any mechanism task and thereby enable students to attend to complex predictive mechanisms rather than be bogged down by the rules governing arrow pushing. However, it remains to be seen whether teaching this formalism in the absence of reasoning will improve the situation.

6.3. Research on the Efficacy of Instructional Technology

Among the most enthusiastically accepted innovations in chemistry teaching and learning are those involving the use of instructional technology. At national meetings in chemistry education, the technology sessions are invariably crowded, and most publishers provide technology-support materials ranging from simulations of phenomena, to visualizations, to computerized testing materials. Indeed, whole courses have been transformed into online learning experiences.

6.3.1. Online Learning. While the landscape of online learning is changing rapidly, as more flexible and responsive systems are being developed, there is little evidence that localizing all or part of a course on the web has any significant (positive) impact on learning. A review of the research on online education conducted in 2013[312] concluded that there is "little evidence to support broad claims that online or hybrid learning is significantly more effective or significantly less effective than courses taught in a face-to-face format" while also highlighting the need for further studies on this topic. Just as with studies on flipped classrooms, it is extraordinarily difficult to conduct rigorous studies that compare randomly assigned students in treatment and control groups. The majority of studies show that students who *complete* online courses fare just as well as those who attend face-to-face classes. However, low-income and underprepared students drop out of online classes at a higher rate than observed in face-to-face settings.[313] This observation complicates the stated desire of many to realize greater access to education through the use of online coursework.[314] In a study of 18 896 students enrolled in community college courses, Xu and Jaggars[315] found that "males, younger students, and students with lower GPAs each faced a significantly greater penalty for taking online courses." Thus, the hype[316] that surrounded the inception of MOOCs (massive open online courses) has not (yet) resulted in the open-access education for all that some predicted. A review of the status of online learning in postsecondary education cautions that "skepticism is warranted concerning any analysis of learning outcomes that has been done by self-interested stakeholders."[312] Just as drug companies outsource the clinical trials of their products, perhaps so should vendors of online learning materials.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

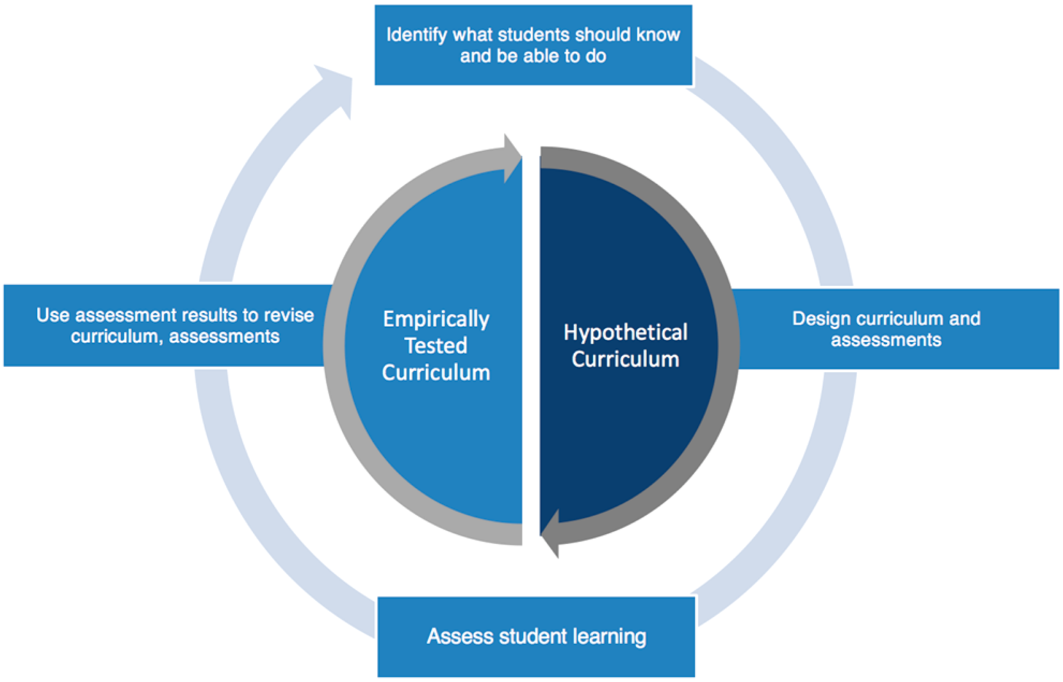


Figure 8. Curriculum design cycle.

There are a number of potential reasons why underprepared students do not fare well with online learning: lack of access, lack of support, and lack of structure are all thought to contribute to the higher dropout rate.[312] Online courses are typically taught in a traditional lecture mode, and while some may provide opportunities for collaboration, it may not be enough to offset the other barriers to completion that face underprepared students. There are some attempts to design collaborative environments that might provide the kind of structure and support that face-to-face learning can deliver. Varma-Nelson and co-workers studied how students who were assigned to online PLTL study groups fared, compared to face-to-face groups. Although the online groups earned somewhat lower grades and higher DFW percentages than their face-to-face peers, the effect size was small, meaning that the online PLTL was basically as effective as face-to-face meetings.[317]

Beginning with a single solid-state chemistry course offered by MIT in 2012, a fairly large number of chemistry MOOCs have been developed.[318] Unfortunately, the low pass rate (7%) of that first MIT MOOC is all too typical of later courses. Until these courses take advantage of the affordances of technology to support student learning with collaborative, supportive interactive materials, rather than simply moving a traditional face-to-face lecture onto the web, it seems unlikely that the most at-risk students will benefit from such accessibility.

6.3.2. Simulations and Games. Simulations and games are both approaches to instructional technology that rely on interactive computer models. Simulations allow students to explore a given phenomenon by manipulating variables and monitoring changes. They are often designed to allow students to design and run virtual experiments from which data may be collected and analyzed. For example, there are simulations for experiments to determine rate laws or to discover the effect of perturbations on equilibrium.[319] Many of these simulations provide molecular-level visualizations, but they are different from simple visualizations because students can manipulate the conditions. In contrast, games are usually played in a more informal environment and often are organized around a skill (e.g., nomenclature or balancing equations). Just as with other aspects of instructional technology, the claims of game and simulation developers are not strongly justified by evidence. According to the NRC consensus study report *Learning Science with Computer Games and Simulations*,[320] "There is moderate evidence that simulations motivate students' interest in science and science learning, and less evidence about whether they support other science learning goals." Similarly, the authors of the NRC report conclude that the evidence to support the use of games in learning is very limited.

While the NRC report was published in 2011, the recommendations and research agenda they propose is still valid. Just as with other areas of learning we have discussed, the tools we use, whether simulations, games, visualizations, or clickers, must be used in a way that aligns with what we know about how people learn, and the research on their efficacy must be well-designed.

7. CURRICULUM DESIGN: TYING EVIDENCE TO PRACTICE

In this section we briefly discuss how the research on student learning in chemistry has been implemented in approaches to curriculum design. By this we mean the sequence of topics, the activities and tasks that students engage with, and the mechanisms for determining the efficacy of one curriculum design over another. Indeed, one might look upon this section as putting together evidence-based approaches to supporting learning into a coherent course. Our discussion here will mostly focus on research about general chemistry curriculum design because there is as yet very little work on effective strategies for upper-level course and curriculum design.

As noted earlier, the sequence and scope of curricula have historically been driven by practitioner wisdom, and while there have been periodic calls for reform,[132−135] most of which center on streamlining the curriculum, there is little evidence that mainstream texts have heeded these cries. Current general

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

chemistry textbooks show no signs of diminution, even though it is practically impossible to learn all the material they contain in two semesters (although it is certainly possible to *present* all the material). It should also be noted that these texts reflect the desires of the reviewers and adopters of the texts, and as such are not necessarily driven by any evidence of efficacy.

One well-regarded approach to curriculum design is known as backward design.[321] In this process, the curriculum developer specifies the goals for what students should know and be able to do by the end of a learning period (course, major, etc.) in the form of learning objectives or performance expectations. Then, the curriculum is designed to support these performance expectations. Assessments that are intended to measure students' ability to complete stated learning objectives would be developed concurrently with other instructional materials. Ideally, the curriculum developers would then investigate student performance on the assessments and use this data to revise the curriculum. Analysis and refinement of a curriculum in light of data from assessments should occur iteratively in a design research cycle[322,323] (Figure 8). In this way, we might begin with a new hypothetical curriculum, look for evidence supporting its efficacy, and identify weaknesses suggested by student performances. Then, the curriculum would be revised so as to shore up these weaknesses and evaluated again by student performances. Curricula emergent from several design research cycles can rightly be described as empirically tested. Backward design provides us with a framework for curriculum reform. If we know what we want students to know and do at the end, we can plot a course that would lead us to that goal.

The factors that can guide curriculum design are essentially those we have already highlighted in this Review: identifying what students should know and be able to do, designing appropriate assessments, and implementing techniques shown to improve learning. One other factor that should be considered is the order of topics and the ways that they are connected. A constant theme that emerges when we are looking at the development of more expert-like conceptual frameworks is that of connections among what is already known, what is to be learned, and what context the knowledge will be used in. There is an emerging body of evidence that indicates the progression of ideas in a particular course or curriculum is crucial to development of understanding.[324,325] The interactions of atoms and molecules are often counterintuitive, and the resources necessary for predicting, explaining, and modeling their behavior must be derived in large part from instruction. To not overwhelm the processing capacity of students, careful scaffolding and sequencing of concepts is necessary.

The general chemistry curriculum Chemistry, Life, the Universe, and Everything (CLUE) provides an example of the potential of curricula to foster improved learning outcomes across a range of competencies.[138] CLUE is built upon intertwined progressions of core ideas (structure−property relationships, bonding and interactions, energy, and change and stability). Several replicated studies demonstrate that students enrolled in CLUE are significantly better able to draw structures,[326] understand the nature of intermolecular forces,[33,229] understand the nature of structure−property relationships,[307] and reason mechanistically about acid−base reactions[176] than a matched cohort of traditionally taught students. These improvements were maintained over the course of the next two semesters of organic chemistry. The DFW rate for general chemistry after CLUE implementation decreased from 35% to 17%, while scores on the ACS conceptual general chemistry exam were above the national average.

Another general chemistry curriculum based on the Chemical Thinking learning progression has also been adopted in large enrollment courses.[327] This curriculum revolves around eight essential questions, such as the following: How do we distinguish substances? How do we determine structure? How do we predict properties? Students enrolled in this course scored higher on the ACS conceptual exam than students from a traditional course at the same institution, and there was a small but significant increase in the distribution of letter grades in the subsequent organic course. Unfortunately, as Talanquer and Pollard note,[327] there are several major challenges to implementing such transformed curricula, including fidelity of implementation, student buy-in, faculty buy-in, and appropriate assessment instruments.

7.1. Context and Relevance

Perceptions of irrelevance have long plagued chemistry education. Atoms and molecules are far removed from experience, and their behavior can seem unimportant to the day-to-day routine of most. In the late 1990s Osborne and Collins echoed this notion, saying, "school science engages when it makes connections to pupils' everyday lives... physics and chemistry... have less points of contact with pupil's experiences."[328] Around the same time, the Dutch ministry of education noted that students in The Netherlands "do not see... a connection between what they learn about chemistry at school and the chemical reality of the world around them."[329] From our prior discussion of meaningful learning, it should be evident that students are more likely to choose to learn meaningfully if they see the knowledge they are gaining as relevant. Numerous researchers[330−333] have discussed how relevance and context can and should guide curriculum development, and while most of this discussion has taken place in the literature pertaining to precollege science, the ideas and frameworks developed can provide a guide for curriculum developers in higher education. However, Eilks and co-workers indicate that the idea of relevance has not been precisely defined, and many authors use the term to refer to different kinds of relevance, for example:

(1) "Relevance for preparing students for potential careers in science and engineering;
(2) Relevance for understanding scientific phenomena and coping with the challenges in a learner's life; and
(3) Relevance for students becoming effective future citizens in the society in which they live."

Complicating matters further, it is often extremely challenging for students to construct molecular-level explanations, predictions, and models for relatable macroscopic phenomena. The world we can see and touch is determined by interactions and energy changes at the molecular level that are far more complex than the simple systems often presented as exemplars in school science. There is thus enormous potential for assessments grounded in so-called rich contexts to overwhelm students' processing capacity and force them to resort to regurgitating memorized factoids. Very little work has carefully examined how intellectual resources might be developed and coordinated to enable students to make sense of more complex, potentially meaningful systems.

Currently, in higher-education chemistry courses, attempts to focus on relevance to society (and the concomitant preparation of informed citizens) are often limited to nonscience majors.[334]

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

For example, the ACS text *Chemistry in Context* uses a case-based approach in which current societal issues are used as a framework to teach chemical principles. While there are a few general chemistry curricula targeted at particular disciplinary majors (for example, engineering), most commercially available texts still address what chemistry faculty believe that chemistry majors should know rather than the needs of the students in the courses.

8. CONCLUSIONS

In this Review, we have attempted to chronicle how chemistry education research has evolved over the years from an individual's ideas about how to improve student learning to a sophisticated research enterprise driven by data and evidence. In contrast to the era of personal empiricism, modern CER studies are grounded in an understanding of how people learn that has emerged from contemporary learning theories. It is now generally accepted in the CER community that knowledge is not transferred intact from instructor to student but rather is constructed in the mind of the learner.[50,51,240] This construction does not take place on a blank slate, as students arrive in our courses with background ideas and preconceptions that may or may not be appropriate. As a consequence, these pieces of knowledge may contain errors and unscientific ideas. We also know that many of these ideas cannot be "overwritten" merely by telling students the correct version.[47] Although we do not have strong evidence for the mechanisms involved in construction and reconstruction of ideas, there is evidence that many learners do not have stable mental models about chemical concepts.[32] Unlike experts, students' knowledge tends to be fragmentary and not well-contextualized;[26,122] most researchers therefore believe that students must be helped to connect and, if necessary, reweave their ideas to construct more coherent, robust understanding of atoms and molecules.

One way to facilitate the construction of more expert-like knowledge frameworks is to develop learning experiences that allow students to make connections to core ideas in chemistry.[152,164,188] Relationships between concepts that may be obvious to experts often go unnoticed by novices. Thus, explicit linkage of course content back to foundational disciplinary ideas can aid students in appropriately organizing their knowledge. Course content can be extended, adapted, and connected to core ideas in the context of scientific practices such as constructing and using models, analyzing and evaluating data, using evidence to construct and support claims, and constructing explanations.[152,177] Curricula that support students with explicit connections to core ideas in the context of scientific practices are more likely to build understanding. Gains in understanding are even more pronounced if these connections are built over time, scaffolded, and revisited as students' understanding deepens.[325,335,336]

Tasks (both formative and summative) that require students to make connections should support student construction of more expert-like frameworks.[222] In this context, it is useful to remember that assessments send a strong message to students about what matters.[18−23] When everything is assessed, everything appears to be equally important, and students can be quickly overwhelmed by what appears to be a vast swathe of material. In such a scenario, there is no recourse for students but to memorize everything and regurgitate fragments of knowledge. Focusing on assessing fewer important ideas helps students develop deeper knowledge by encouraging intentional, Type 2 thinking about those concepts most fundamental to a discipline.

Weaving together resources into coherent and useful disciplinary knowledge is a difficult process that requires considerable effort on the part of the student.[126,128,129] Stated differently, development of expertise is inherently an active process. Indeed, there is strong evidence that learning environments designed to promote student engagement and interactions both among peers and with the instructor improve outcomes, particularly for lower-performing students.[218,242] The mechanisms of improvement may be explained by social constructivist learning theory:[52] as students interact with each other and with more knowledgeable others, they have to consider their ideas, refine, reflect, and explain. There are many ways to bring about this kind of active engagement, and most of them have been shown to improve overall average course grades or performance on concept inventories.

While much of the aforementioned research is relevant to learning in general, there are particular difficulties with learning chemistry because all of the causal mechanistic drivers for phenomena of interest occur at the atomic and molecular level. Students can learn to make the connection between macroscopic properties of materials and the molecular level, but it requires a long chain of causal inferences—almost none of which can be intuited by experience. Students must make sense of representations and diagrams that attempt to convey ideas about things that cannot be seen, and so visualizations, structures, and other tools can impart not only meaning but also incorrect ideas that may hinder learning.

8.1. Where We Go from Here

While we have come a long way, the research discussed in this Review points to a number of potentially fruitful areas for further research. While evidence-based strategies for supporting students in constructing a coherent and useful molecular-level understanding have been an area of significant interest in the CER community and there are several promising studies, very little of what has been done until now meets the criteria for strong evidence (Box 1). In this last section we offer some suggestions for next steps that could build on our current understanding of how students learn chemistry.

8.1.1. Shifting Focus to a More Constructivist Perspective. Much of the prior work in CER has emerged from a deficit model about what students do *not* know. Student ideas have been characterized from the standpoint of misconceptions and difficulties rather than inappropriate activation of mental resources (i.e., fragments of knowledge) or naïve ideas about the molecular world. Although it is certainly important to understand what knowledge students are building on, it is equally important to realize that students' molecular-level understanding is more likely to be dynamic and fragmented rather than a collection of stable wrong ideas. Ultimately, our goal is to determine how to help students weave their fragmentary knowledge bases into a tapestry that is more expert-like. Current evidence about conceptual change indicates that different mechanisms may be operational depending on the nature of the knowledge. As Taber has written, "it is important to be able to advise teachers when their best response to students' ideas is to ignore them, to actively challenge them, and to consider them as suitable starting points for developing toward scientific models." We do not have a good sense of when each of these options is most appropriate, nor do we have much information about how to reweave fragments into accurate and appropriately sophisticated models of phenomena. Reimagining whole courses to focus on integration and use of knowledge

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

anchored to core ideas may provide a promising avenue toward facilitating student resource development and coordination. Data from analysis of the transformed general chemistry courses illustrate the potential of de novo curricula, designed according to modern learning theories, to help students make sense of the molecular world.[229,326,327] Unfortunately, there is very limited research on how students learning can be supported in upper-level courses.

Although Schunn and co-workers have provided powerful evidence that whole transformed curricula are more effective at improving student outcomes than slotted-in reformed modules,[337] wholesale course transformation may not be an option for many. For this reason, we need a better sense of what ideas and skills might be developed through less-ambitious interventions. For example, we know that students in traditional general chemistry courses often have trouble constructing Lewis structures and do not have a coherent understanding of IMFs. Is it possible to help students master these ideas without restructuring large amounts of the curriculum?

8.1.2. Understanding How To Effectively Make Use of Technology. While we have some evidence to support the use of visualizations and simulations, there is no consensus on how they should be designed and used. Under what circumstances are visualizations or simulations useful? What design criteria are important? How important is it that students draw molecular-level representations, or is it sufficient to observe them on a computer? Along these same lines, while instructional technology in general has great potential, there is currently little evidence about improved learning resultant from any particular implementation of that technology. We need to understand more about online learning and how to construct materials that build on the significant promise of technology and take advantage of our current understanding of how people learn. This means going beyond putting lectures on the Internet and taking advantage of the affordances of technology to provide evidence-based learning environments.

8.1.3. Designing Appropriate Assessments. Assessments are the metrics by which we judge success both in the classroom and in CER studies. It is therefore vitally important that the tools used to gather evidence of student understanding are appropriately designed and validated and that the context in which they are used is appropriate. In the broadest sense, assessments should gather evidence about what students know and can do. Assessments (tests, quizzes, workshop activities, and homework assignments) must go beyond recall and algorithms. A few targeted assessments have been developed that give evidence of expertise development, but much more is needed to examine student adaptation and extension of knowledge to new contexts (both within a course and related courses).

8.1.4. Integrating Relevant Contexts into Instruction. Although there is broad consensus that chemistry coursework should ideally be relevant to students' lives and priorities, scant evidence exists about how this might be best accomplished. As we noted earlier, it is uniquely challenging to tie molecular-level understanding to compelling real-life scenarios due to the long chain of inferences needed to relate observations to their molecular-level cause. Even relatively simple macroscopic phenomena require coordination of a great many intellectual resources. For example, modeling dissolution of a salt in a solvent requires ready command of ideas associated with atomic and molecular structure, bonding, intermolecular forces, energy transfer and transformation, and perhaps entropy. If students cannot activate the necessary resources (or do not know to activate them), they will be forced to memorize rather than understand, and this may set the pattern for the whole course. This means that there is a tension between providing the background material needed as a basis for student understanding and making this material relevant and contextual. Developing curricular materials and pedagogies that support development and coordination of intellectual resources while providing opportunities for students to figure out the molecular-level cause for relatable phenomena will be a major challenge.

8.1.5. Examining How Transformed Teaching and Learning Might Be Disseminated and Sustained. To take advantage of research findings, we need to know more about how faculty can be supported to adopt evidence-based transformations. Research on how and why transformation efforts succeed or fail is vital to the success of such enterprises. Most current efforts focus on what have become known as evidence-based instructional practices (EBIPS),[338] which typically consist of ways to engage students. Studies on the implementation of EBIPs tend to show that their uptake is quite low, as is the fidelity with which they are implemented.[339,340] That being said, as we have seen in this Review, transforming instruction by incorporating research on teaching and learning involves more than incorporating engagement techniques. Further research is needed on implementation of evidence-based curriculum materials and assessments.

8.2. Closing Thoughts

We should note that many questions also remain in the areas not covered by this Review. For example, we do not know how identity, mindset, motivation, or interest interact to help or hinder student success in chemistry. There are questions to be answered about the efficacy of laboratory learning and how it should be assessed. We have very few studies in which data are disaggregated by gender or ethnicity, as well as a paucity of longitudinal or replication studies.

The CER community has made a great deal of progress over the last century, and while there is much work to be done, the evidence gathered so far can point us in productive directions. While it is tempting to go with gut instinct and personal experience, as scientists we must go where the evidence points. As we have noted earlier, "personal experience is not appropriate for scientists, yet when it comes to education, personal experience seems to be an acceptable substitute for evidence. Unfortunately, most scientists' beliefs about education are rarely based on objective evidence but rather on what they imagine to be true. While personal experience in the classroom can give valuable insights, it is not data."[341]

AUTHOR INFORMATION

Corresponding Author

*mmc@msu.edu.

ORCID

Melanie M. Cooper: 0000-0002-7050-8649

Ryan L. Stowe: 0000-0002-5548-495X

Notes

The authors declare no competing financial interest.

Biographies

Melanie Cooper is the Lappan-Phillips Professor of Science Education and Professor of Chemistry at Michigan State University. Her research has focused on improving teaching and learning in large enrollment general and organic chemistry courses at the college level, and she is a

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

proponent of evidence-based curriculum reform and assessment. She is a Fellow of the American Chemical Society, the American Association for the Advancement of Science, and the Royal Society of Chemistry and is a member of the National Research Council Advisory Board on Science Education (BOSE). She was a member of the Leadership team for the Next Generation Science Standards (NGSS) and has also served on the NRC committees that produced the DBER report and the new report on Strengthening Undergraduate Research Experiences. She has received a number of awards including the ACS Award for Achievement in Research on Teaching and Learning in Chemistry (2014), the Norris Award for Outstanding Achievement in Teaching of Chemistry (2013), and the Outstanding Undergraduate Science Teacher Award from the Society for College Science Teaching (2011). She earned her B.S., M.S., and Ph.D. in chemistry from the University of Manchester, England.

Ryan Stowe received his Ph.D. from the Scripps Research Institute in 2016. His doctoral studies, conducted under the guidance of Prof. William Roush, focused on the design and synthesis of histone deacetylase inhibitors and the development of methods useful in natural products synthesis. In 2015, he served as a Christine Mirzayan Science and Technology Policy Fellow with the Board on Science Education at the National Academy of Sciences. His current work, as a research associate mentored by Prof. Melanie Cooper, is focused on supporting molecular-level sense-making in high school chemistry students and engaging undergraduates in scientific practices in the context of organic chemistry.

REFERENCES

(1) Patrick, W. A. What Kind of Research Is Essential to Good Teaching? *J. Chem. Educ.* **1924**, *1*, 16.

(2) The National Research Council. *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering*; National Academies Press: Washington, DC, 2012.

(3) NIH, National Cancer Institute. Randomized Control Trial—National Library of Medicine; https://www.ncbi.nlm.nih.gov/pubmedhealth/PMHT0025811/ (accessed Dec 8, 2017).

(4) Cooper, M. M. The Replication Crisis and Chemistry Education Research. *J. Chem. Educ.* **2018**, *95*, 1−2.

(5) Towns, M. H. New Guidelines for Chemistry Education Research Manuscripts and Future Directions of the Field. *J. Chem. Educ.* **2013**, *90*, 1107−1108.

(6) Towns, M. H.; Kraft, A. Review and Synthesis of Research in Chemical Education from 2000−2010. *A white paper for the National Academies National Research Council Board on Science Education, Committee on Status, Contributions, and Future Directions of Discipline Based Education Research*; **2011**.

(7) Teo, T. W.; Goh, M. T.; Yeo, L. W. Chemistry Education Research Trends: 2004−2013. *Chem. Educ. Res. Pract.* **2014**, *15*, 470−487.

(8) Bain, K.; Moon, A.; Mack, M. R.; Towns, M. H. A Review of Research on the Teaching and Learning of Thermodynamics at the University Level. *Chem. Educ. Res. Pract.* **2014**, *15*, 320−335.

(9) Bain, K.; Towns, M. H. A Review of Research on the Teaching and Learning of Chemical Kinetics. *Chem. Educ. Res. Pract.* **2016**, *17*, 246−262.

(10) Graulich, N. The Tip of the Iceberg in Organic Chemistry Classes: How Do Students Deal with the Invisible? *Chem. Educ. Res. Pract.* **2015**, *16*, 9−21.

(11) Harle, M.; Towns, M. A Review of Spatial Ability Literature, Its Connection to Chemistry, and Implications for Instruction. *J. Chem. Educ.* **2011**, *88*, 351−360.

(12) Powers, S. Chapter XIII. Chemical Education. *J. Am. Chem. Soc.* **1926**, *48*, 237−254.

(13) Hale, H. The History of Chemical Education in the United States from 1870 to 1914. *J. Chem. Educ.* **1932**, *9*, 729.

(14) Mabery, C. F. How Chemistry Is Best Taught. *J. Am. Chem. Soc.* **1893**, *15*, 463−475.

(15) Royce, J. Is There a Science of Education? *Educ. Rev.* **1891**, *1*, 15−25.

(16) Johnstone, A. H. The Development of Chemistry Teaching: A Changing Response to Changing Demand. *J. Chem. Educ.* **1993**, *70*, 701.

(17) Cornog, J.; Colbert, J. C. What We Teach Our Freshmen in Chemistry. *J. Chem. Educ.* **1924**, *1*, 5.

(18) Momsen, J.; Offerdahl, E.; Kryjevskaia, M.; Montplaisir, L.; Anderson, E.; Grosz, N. Using Assessments to Investigate and Compare the Nature of Learning in Undergraduate Science Courses. *CBE-Life Sci. Educ.* **2013**, *12*, 239−249.

(19) Scouller, K. The Influence of Assessment Method on Students' Learning Approaches: Multiple Choice Question Examination versus Assignment Essay. *High. Educ.* **1998**, *35*, 453−472.

(20) Scouller, K. M.; Prosser, M. Students' Experiences in Studying for Multiple Choice Question Examinations. *Stud. High. Educ.* **1994**, *19*, 267−279.

(21) Snyder, M. D. B. R. *The Hidden Curriculum*, 1st ed.; Knopf: New York, 1971.

(22) Entwistle, N. J. Approaches to Learning and Perceptions of the Learning Environment. *High. Educ.* **1991**, *22*, 201−204.

(23) Crooks, T. J. The Impact of Classroom Evaluation Practices on Students. *Rev. Educ. Res.* **1988**, *58*, 438−481.

(24) Plane, R. A.; Sienko, M. J. *Chemistry: Principles and Properties*, 1st ed.; Mcgraw-Hill: New York, 1957.

(25) Borman, S. A Tale of Two Textbooks—History and Influence of Sienko & Plane and Morrison and Boyd Detailed at Symposium. *Chem. Eng. News* **2005**, *83*, 48−51.

(26) diSessa, A. A. Knowledge in Pieces. In *Constructivism in the computer age*; Forman, G. E., Pufall, P. B., Eds.; Lawrence Erlbaum Associates: Hillsdale, NJ, 1988; pp 49−70.

(27) diSessa, A. A. Toward an Epistemology of Physics. *Cogn. Instr.* **1993**, *10*, 105−225.

(28) Vosniadou, S. Examining Cognitive Development from a Conceptual Change Point of View: The Framework Theory Approach. *Eur. J. Dev. Psychol.* **2014**, *11*, 645−661.

(29) Vosniadou, S.; Skopeliti, I. Conceptual Change from the Framework Theory Side of the Fence. *Sci. Educ.* **2014**, *23*, 1427−1445.

(30) Posner, G. J.; Strike, K. A.; Hewson, P. W.; Gertzog, W. A. Accommodation of a Scientific Conception: Toward a Theory of Conceptual Change. *Sci. Educ.* **1982**, *66*, 211−227.

(31) Chi, M. T. H. Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *J. Learn. Sci.* **2005**, *14*, 161−199.

(32) Cooper, M. M.; Corley, L. M.; Underwood, S. M. An Investigation of College Chemistry Students' Understanding of Structure−Property Relationships. *J. Res. Sci. Teach.* **2013**, *50*, 699−721.

(33) Cooper, M. M.; Williams, L. C.; Underwood, S. M. Student Understanding of Intermolecular Forces: A Multimodal Study. *J. Chem. Educ.* **2015**, *92*, 1288−1298.

(34) Bretz, S. L.; McClary, L. Students' Understandings of Acid Strength: How Meaningful Is Reliability When Measuring Alternative Conceptions? *J. Chem. Educ.* **2015**, *92*, 212−219.

(35) Cruz-Ramírez de Arellano, D.; Towns, M. H. Students' Understanding of Alkyl Halide Reactions in Undergraduate Organic Chemistry. *Chem. Educ. Res. Pract.* **2014**, *15*, 501−515.

(36) Yan, F.; Talanquer, V. Students' Ideas about How and Why Chemical Reactions Happen: Mapping the Conceptual Landscape. *Int. J. Sci. Educ.* **2015**, *37*, 3066−3092.

(37) Bodner, G. M. Status, Contributions, and Future Directions of Discipline-Based Education Research: The Development of Research in Chemical Education as a Field of Study. *A white paper for the National Academies National Research Council Board on Science Education, Committee on Status, Contributions, and Future Directions of Discipline Based Education Research*; **2011**.

(38) Holmes, H. N. College Research. *J. Chem. Educ.* **1924**, *1*, 81.

(39) Karplus, R. Part III. Curriculum Project Reports: The Science Curriculum Improvement Study-Report to the Piaget Conference. *J. Res. Sci. Teach.* **1964**, *2*, 236−240.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(40) Karplus, R. Chemical Phenomena in Elementary School Science. *J. Chem. Educ.* **1966**, *43*, 267−269.
(41) Furth, H. G. *Piaget and Knowledge: Theoretical Foundations*; Prentice Hall: Englewood Cliffs, NJ, 1969.
(42) Piaget, J.; Duckworth, E. Genetic Epistemology. *Am. Behav. Sci.* **1970**, *13*, 459−480.
(43) Flavell, J. H. *The Developmental Psychology of Jean Piaget*; The University Series in Psychology; Van Nostrand and Co., Inc.: Princeton, NJ, 1964; Vol. *2*.
(44) Piaget, J. *Structuralism*; Psychology Press: New York, 2015.
(45) Dewey, J. *Democracy and Education: An Introduction to the Philosophy of Education*; Macmillan: New York, 1916.
(46) Piaget, J.; Brown, T.; Thampy, K. J. *Equilibration of Cognitive Structures: The Central Problem of Intellectual Development*; University of Chicago Press: Chicago, 1985.
(47) diSessa, A. A. A History of Conceptual Change Research: Threads and Fault Lines. In *The Cambridge Handbook of the Learning Sciences*; Cambridge University Press: New York, 2014; pp 88−108.
(48) Karplus, R. The Science Curriculum Improvement Study. *J. Res. Sci. Teach.* **1964**, *2*, 293−303.
(49) Craig, B. S. The Philosophy of Jean Piaget and Its Usefulness to Teachers of Chemistry. *J. Chem. Educ.* **1972**, *49*, 807.
(50) Bodner, G. M. Constructivism: A Theory of Knowledge. *J. Chem. Educ.* **1986**, *63*, 873.
(51) Windschitl, M. Framing Constructivism in Practice as the Negotiation of Dilemmas: An Analysis of the Conceptual, Pedagogical, Cultural, and Political Challenges Facing Teachers. *Rev. Educ. Res.* **2002**, *72*, 131−175.
(52) Vygotsky, L. S. *Mind in Society: The Development of Higher Psychological Processes*; Cole, M., John-Steiner, V., Scribner, S., Souberman, E., Eds.; Harvard University Press: Cambridge, MA, 1978.
(53) Gosser, D. K.; Roth, V. The Workshop Chemistry Project: Peer-Led Team-Learning. *J. Chem. Educ.* **1998**, *75*, 185−187.
(54) Wilson, S. B.; Varma-Nelson, P. Small Groups, Significant Impact: A Review of Peer-Led Team Learning Research with Implications for STEM Education Researchers and Faculty. *J. Chem. Educ.* **2016**, *93*, 1686−1702.
(55) Gosser, D. K.; Cracolice, M. S.; Kampmeier, J. A.; Roth, V.; Strozak, V. S.; Varma-Nelson, P. *Peer-Led Team Learning: A Guidebook*; Prentice Hall: Upper Saddle River, NJ, 2000.
(56) Ebenezer, J. V. Making Chemistry Learning More Meaningful. *J. Chem. Educ.* **1992**, *69*, 464−467.
(57) Ausubel, D. P. *Educational Psychology: A Cognitive View*; Holt, Rinehart and Winston: New York, 1968.
(58) Ausubel, D. P. *The Psychology of Meaningful Verbal Learning*; Grune & Stratton: New York, 1963.
(59) Bretz, S. L. Novak's Theory of Education: Human Constructivism and Meaningful Learning. *J. Chem. Educ.* **2001**, *78*, 1107−1115.
(60) Novak, J. D.; Gowin, D. B.; Kahle, J. B. *Learning How to Learn*, 1st ed.; Cambridge University Press: Cambridge, U.K., 1984.
(61) The National Research Council. *How People Learn: Brain, Mind, Experience, and School: Expanded ed.*, 2nd ed.; National Academies Press: Washington, DC, 2000.
(62) Herron, J. D. Piaget for Chemists. Explaining What "Good" Students Cannot Understand. *J. Chem. Educ.* **1975**, *52*, 146−150.
(63) Johnstone, A. H. Macro- and Microchemistry. *Sch. Sci. Rev.* **1982**, *64*, 377−379.
(64) Teichert, M. A.; Stacy, A. M. Promoting Understanding of Chemical Bonding and Spontaneity through Student Explanation and Integration of Ideas. *J. Res. Sci. Teach.* **2002**, *39*, 464−496.
(65) Johnstone, A. H. Why Is Science Difficult to Learn? Things Are Seldom What They Seem. *J. Comput. Assist. Learn.* **1991**, *7*, 75−83.
(66) Taber, K. S. Revisiting the Chemistry Triplet: Drawing upon the Nature of Chemical Knowledge and the Psychology of Learning to Inform Chemistry Education. *Chem. Educ. Res. Pract.* **2013**, *14*, 156−168.
(67) Talanquer, V. Macro, Submicro, and Symbolic: The Many Faces of the Chemistry "Triplet". *Int. J. Sci. Educ.* **2011**, *33*, 179−195.
(68) Gabel, D. Improving Teaching and Learning through Chemistry Education Research: A Look to the Future. *J. Chem. Educ.* **1999**, *76*, 548−554.
(69) *Multiple Representations in Chemical Education*, 1st ed.; Gilbert, J. K., Treagust, D. F., Eds.; Models and Modeling in Science Education; Springer: Dordrecht, 2009; Vol. *4*.
(70) Greene, J.; Hicks, C. *Basic Cognitive Processes*; Milton Keynes, Open University Press: 1984.
(71) Baddeley, A. Working Memory: Looking Back and Looking Forward. *Nat. Rev. Neurosci.* **2003**, *4*, 829−839.
(72) Miller, G. A. The Magical Number Seven, plus or Minus Two: Some Limits on Our Capacity for Processing Information. *Psychol. Rev.* **1956**, *63*, 81−97.
(73) De Ribaupierre, A.; Fagot, D.; Lecerf, T. Working Memory Capacity and Its Role in Cognitive Development: Are Age Differences Driven by the Same Processes Across the Lifespan? In *Cognitive Development and Working Memory: A Dialogue between Neo-Piagetian Theories and Cognitive Approaches*;Barrouillet, P., Gaillard, V., Eds.; Psychology Press: Hove, East Sussex, 2011; pp 105−133.
(74) Cowan, N.; Chen, Z.; Rouder, J. N. Constant Capacity in an Immediate Serial-Recall Task: A Logical Sequel to Miller (1956). *Psychol. Sci.* **2004**, *15*, 634−640.
(75) Mathy, F.; Feldman, J. What's Magic about Magic Numbers? Chunking and Data Compression in Short-Term Memory. *Cognition* **2012**, *122*, 346−362.
(76) Johnstone, A. H. Chemistry Teaching - Science or Alchemy? *J. Chem. Educ.* **1997**, *74*, 262−268.
(77) Johnstone, A. H. Chemical Education Research in Glasgow in Perspective. *Chem. Educ. Res. Pract.* **2006**, *7*, 49−63.
(78) Johnstone, A. H. You Can't Get There from Here. *J. Chem. Educ.* **2010**, *87*, 22−29.
(79) Kozma, R.; Russell, J. Multimedia and Understanding: Expert and Novice Responses to Different Representations of Chemical Phenomena. *J. Res. Sci. Teach.* **1997**, *34*, 949−968.
(80) de Jong, T.; Ferguson-Hessler, M. G. M. Cognitive Structures of Good and Poor Novice Problem Solvers in Physics. *J. Educ. Psychol.* **1986**, *78*, 279−288.
(81) Smith, J. I.; Combs, E. D.; Nagami, P. H.; Alto, V. M.; Goh, H. G.; Gourdet, M. A. A.; Hough, C. M.; Nickell, A. E.; Peer, A. G.; Coley, J. D.; et al. Development of the Biology Card Sorting Task to Measure Conceptual Expertise in Biology. *CBE-Life Sci. Educ.* **2013**, *12*, 628−644.
(82) Ben-Zvi, R.; Eylon, B.-S.; Silberstein, J. Is an Atom of Copper Malleable? *J. Chem. Educ.* **1986**, *63*, 64−66.
(83) Novick, S.; Nussbaum, J. Junior High School Pupils' Understanding of the Particulate Nature of Matter: An Interview Study. *Sci. Educ.* **1978**, *62*, 273−281.
(84) Novick, S.; Nussbaum, J. Pupils' Understanding of the Particulate Nature of Matter: A Cross-Age Study. *Sci. Educ.* **1981**, *65*, 187−196.
(85) Cros, D.; Maurin, M.; Amouroux, R.; Chastrette, M.; Leber, J.; Fayol, M. Conceptions of First-year University Students of the Constituents of Matter and the Notions of Acids and Bases. *Eur. J. Sci. Educ.* **1986**, *8*, 305−313.
(86) Cros, D.; Chastrette, M.; Fayol, M. Conceptions of Second Year University Students of Some Fundamental Notions in Chemistry. *Int. J. Sci. Educ.* **1988**, *10*, 331−336.
(87) Treagust, D. F. Development and Use of Diagnostic Tests to Evaluate Students' Misconceptions in Science. *Int. J. Sci. Educ.* **1988**, *10*, 159−169.
(88) Mas, C. J. F.; Perez, J. H.; Harris, H. H. Parallels between Adolescents' Conception of Gases and the History of Chemistry. *J. Chem. Educ.* **1987**, *64*, 616−618.
(89) Stavy, R. Children's Conception of Gas. *Int. J. Sci. Educ.* **1988**, *10*, 553−560.
(90) Yarroch, W. L. Student Understanding of Chemical Equation Balancing. *J. Res. Sci. Teach.* **1985**, *22*, 449−459.
(91) Andersson, B. Pupils' Explanations of Some Aspects of Chemical Reactions. *Sci. Educ.* **1986**, *70*, 549−563.
(92) de Vos, W.; Verdonk, A. H. A New Road to Reactions. Part 4. The Substance and Its Molecules. *J. Chem. Educ.* **1987**, *64*, 692−694.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(93) Stavridou, H.; Solomonidou, C. Physical Phenomena–chemical Phenomena: Do Pupils Make the Distinction? *Int. J. Sci. Educ.* **1989**, *11*, 83−92.
(94) de Vos, W.; Verdonk, A. H. A New Road to Reactions: Part III. Teaching the Heat Effect of Reactions. *J. Chem. Educ.* **1986**, *63*, 972−974.
(95) Gussarsky, E.; Gorodetsky, M. On the Concept "Chemical Equilibrium": The Associative Framework. *J. Res. Sci. Teach.* **1990**, *27*, 197−204.
(96) Garnett, P. J.; Garnett, P. J.; Hackling, M. W. Students' Alternative Conceptions in Chemistry: A Review of Research and Implications for Teaching and Learning. *Stud. Sci. Educ.* **1995**, *25*, 69−96.
(97) Mulford, D. R.; Robinson, W. R. An Inventory for Alternate Conceptions among First-Semester General Chemistry Students. *J. Chem. Educ.* **2002**, *79*, 739−744.
(98) Pedrosa, M. A.; Dias, M. H. Chemistry Textbook Approaches to Chemical Equilibrium and Student Alternative Conceptions. *Chem. Educ. Res. Pract.* **2000**, *1*, 227−236.
(99) Pozo, R. M. D. Prospective Teachers' Ideas about the Relationships between Concepts Describing the Composition of Matter. *Int. J. Sci. Educ.* **2001**, *23*, 353−371.
(100) Nicoll, G. A Report of Undergraduates' Bonding Misconceptions. *Int. J. Sci. Educ.* **2001**, *23*, 707−730.
(101) Özkaya, A. R. Conceptual Difficulties Experienced by Prospective Teachers in Electrochemistry: Half-Cell Potential, Cell Potential, and Chemical and Electrochemical Equilibrium in Galvanic Cells. *J. Chem. Educ.* **2002**, *79*, 735−738.
(102) Niaz, M.; Aguilera, D.; Maza, A.; Liendo, G. Arguments, Contradictions, Resistances, and Conceptual Change in Students' Understanding of Atomic Structure. *Sci. Educ.* **2002**, *86*, 505−525.
(103) Özmen, H.; Ayas, A. Students Difficulties in Understanding of the Conservation of Matter in Open and Closed-System Chemical Reactions. *Chem. Educ. Res. Pract.* **2003**, *4*, 279−290.
(104) Treagust, D. F.; Tan, K. C. D.; Goh, N. K.; Chia, L. S. Major Sources of Difficulty in Students' Understanding of Basic Inorganic Qualitative Analysis. *J. Chem. Educ.* **2004**, *81*, 725−732.
(105) Gopal, H.; Kleinsmidt, J.; Case, J.; Musonge, P. An Investigation of Tertiary Students' Understanding of Evaporation, Condensation and Vapour Pressure. *Int. J. Sci. Educ.* **2004**, *26*, 1597−1620.
(106) Canpolat, N.; Pinarbasi, T.; Sözbilir, M. Prospective Teachers' Misconceptions of Vaporization and Vapor Pressure. *J. Chem. Educ.* **2006**, *83*, 1237−1242.
(107) Canpolat, N. Turkish Undergraduates' Misconceptions of Evaporation, Evaporation Rate, and Vapour Pressure. *Int. J. Sci. Educ.* **2006**, *28*, 1757−1770.
(108) Chiu, M.-H. A National Survey of Students' Conceptions of Chemistry in Taiwan. *Int. J. Sci. Educ.* **2007**, *29*, 421−452.
(109) Özmen, H. Determination of Students' Alternative Conceptions about Chemical Equilibrium: A Review of Research and the Case of Turkey. *Chem. Educ. Res. Pract.* **2008**, *9*, 225−233.
(110) Cakmakci, G. Identifying Alternative Conceptions of Chemical Kinetics among Secondary School and Undergraduate Students in Turkey. *J. Chem. Educ.* **2010**, *87*, 449−455.
(111) Villafañe, S. M.; Loertscher, J.; Minderhout, V.; Lewis, J. E. Uncovering Students' Incorrect Ideas about Foundational Concepts for Biochemistry. *Chem. Educ. Res. Pract.* **2011**, *12*, 210−218.
(112) Naah, B. M.; Sanger, M. J. Student Misconceptions in Writing Balanced Equations for Dissolving Ionic Compounds in Water. *Chem. Educ. Res. Pract.* **2012**, *13*, 186−194.
(113) Turányi, T.; Tóth, Z. Hungarian University Students' Misunderstandings in Thermodynamics and Chemical Kinetics. *Chem. Educ. Res. Pract.* **2013**, *14*, 105−116.
(114) Eymur, G.; Çetin, P.; Geban, Ö. Analysis of the Alternative Conceptions of Preservice Teachers and High School Students Concerning Atomic Size. *J. Chem. Educ.* **2013**, *90*, 976−980.
(115) Cooper, M. M.; Klymkowsky, M. W. The Trouble with Chemical Energy: Why Understanding Bond Energies Requires an Interdisciplinary Systems Approach. *CBE Life Sci. Educ* **2013**, *12*, 306−312.
(116) Boo, H. K. Students' Understandings of Chemical Bonds and the Energetics of Chemical Reactions. *J. Res. Sci. Teach.* **1998**, *35*, 569−581.
(117) Brown, D.; Hammer, D. Conceptual Change in Physics. In *International handbook of research on conceptual change*; Vosniadou, S., Ed.; Routledge: New York, 2008; pp 127−154.
(118) Ioannides, C.; Vosniadou, S. The Changing Meaning of Force. *Cogn. Sci. Q.* **2002**, *2*, 5−61.
(119) Chi, M. T. H.; Slotta, J. D.; De Leeuw, N. From Things to Processes: A Theory of Conceptual Change for Learning Science Concepts. *Learn. Instr.* **1994**, *4*, 27−43.
(120) Minstrell, J. Explaining the "at Rest" Condition of an Object. *Phys. Teach.* **1982**, *20*, 10−14.
(121) Minstrell, J. Teaching Science for Understanding. In *Toward the thinking curriculum: current cognitive research*; Resnick, L. B., Klopfer, L., Eds.; Association for Supervision and Curriculum Development: Alexandria, VA, 1989; pp 129−149.
(122) diSessa, A. Alternative Conceptions and P-Prims. In *Encyclopedia of Science Education*; Gunstone, R., Ed.; Springer: Dordrecht, 2015; pp 34−37.
(123) Taber, K. S. Conceptual Resources for Learning Science: Issues of Transience and Grain-size in Cognition and Cognitive Structure. *Int. J. Sci. Educ.* **2008**, *30*, 1027−1053.
(124) Taber, K. S.; García-Franco, A. Learning Processes in Chemistry: Drawing Upon Cognitive Resources to Learn About the Particulate Structure of Matter. *J. Learn. Sci.* **2010**, *19*, 99−142.
(125) McCloskey, M. Naive Theories of Motion. In *Mental Models*; Genter, D., Stevens, A. L., Eds.; Psychology Press: New York, 2014; pp 299−322.
(126) Kahneman, D. *Thinking, Fast and Slow*; Farrar, Straus and Giroux: New York, 2011.
(127) Evans, J. *Bias in Human Reasoning: Causes and Consequences*, 1st ed.; Psychology Press: Hove, 1990.
(128) Wason, P. C.; Evans, J. S. B. T. Dual Processes in Reasoning? *Cognition* **1974**, *3*, 141−154.
(129) Evans, J. S. B. T.; Stanovich, K. E. Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspect. Psychol. Sci.* **2013**, *8*, 223−241.
(130) Sana, F.; Weston, T.; Cepeda, N. J. Laptop Multitasking Hinders Classroom Learning for Both Users and Nearby Peers. *Comput. Educ.* **2013**, *62*, 24−31.
(131) Junco, R. In-Class Multitasking and Academic Performance. *Comput. Hum. Behav.* **2012**, *28*, 2236−2243.
(132) Dietz, S.; Henrich, C. Texting as a Distraction to Learning in College Students. *Comput. Hum. Behav.* **2014**, *36*, 163−167.
(133) Talanquer, V. Chemistry Education: Ten Heuristics To Tame. *J. Chem. Educ.* **2014**, *91*, 1091−1097.
(134) Maeyer, J.; Talanquer, V. Making Predictions about Chemical Reactivity: Assumptions and Heuristics. *J. Res. Sci. Teach.* **2013**, *50*, 748−767.
(135) Chandrasegaran, A. L.; Treagust, D. F.; Waldrip, B. G.; Chandrasegaran, A. Students' Dilemmas in Reaction Stoichiometry Problem Solving: Deducing the Limiting Reagent in Chemical Reactions. *Chem. Educ. Res. Pract.* **2009**, *10*, 14−23.
(136) Talanquer, V. When Atoms Want. *J. Chem. Educ.* **2013**, *90*, 1419−1424.
(137) Spencer, J. N. General Chemistry Course Content. *J. Chem. Educ.* **1992**, *69*, 182−186.
(138) Cooper, M.; Klymkowsky, M. Chemistry, Life, the Universe, and Everything: A New Approach to General Chemistry, and a Model for Curriculum Reform. *J. Chem. Educ.* **2013**, *90*, 1116−1122.
(139) Burdge, J.; Overby, J. *Chemistry: Atoms First*; McGraw Hill: Boston, MA, 2012.
(140) Zumdahl, S. S.; Zumdahl, S. A. *Chemistry: An Atoms First Approach*; Cengage Learning: Boston, MA, 2011.
(141) McMurrey, J.; Fay, R. C. *General Chemistry Atoms First*; Pearson Prentice Hall: Upper Saddle River, NJ, 2010.
(142) Esterling, K. M.; Bartels, L. Atoms-First Curriculum: A Comparison of Student Success in General Chemistry. *J. Chem. Educ.* **2013**, *90*, 1433−1436.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(143) Cooper, M. The Case for Reform of the Undergraduate General Chemistry Curriculum. *J. Chem. Educ.* **2010**, *87*, 231−232.
(144) Lloyd, B. W.; Spencer, J. N. The Forum: New Directions for General Chemistry: Recommendations of the Task Force on the General Chemistry Curriculum. *J. Chem. Educ.* **1994**, *71*, 206−209.
(145) Havighurst, R. J. Reform in the Chemistry Curriculum. *J. Chem. Educ.* **1929**, *6*, 1126−1129.
(146) Nameroff, T. J.; Busch, D. H. Exploring the Molecular Vision: Report from a SOCED Invitational Conference. *J. Chem. Educ.* **2004**, *81*, 177−179.
(147) Gillespie, R. J. The Great Ideas of Chemistry. *J. Chem. Educ.* **1997**, *74*, 862−864.
(148) Atkins, P. Chemistry: The Great Ideas. *Pure Appl. Chem.* **1999**, *71*, 927−929.
(149) The College Board. *AP Chemistry: Course and Exam Description*; The College Board: New York, 2014.
(150) Murphy, K.; Holme, T.; Zenisky, A.; Caruthers, H.; Knaus, K. Building the ACS Exams Anchoring Concept Content Map for Undergraduate Chemistry. *J. Chem. Educ.* **2012**, *89*, 715−720.
(151) Holme, T.; Murphy, K. The ACS Exams Institute Undergraduate Chemistry Anchoring Concepts Content Map I: General Chemistry. *J. Chem. Educ.* **2012**, *89*, 721−723.
(152) Cooper, M. M.; Posey, L. A.; Underwood, S. M. Core Ideas and Topics: Building Up or Drilling Down? *J. Chem. Educ.* **2017**, *94*, 541−548.
(153) Bodner, G. M. Research on Problem Solving in Chemistry. In *Chemistry Education*; García-Martínez, J., Serrano-Torregrosa, E., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2015; pp 181−202.
(154) Bodner, G. M.; Herron, J. D. Problem-Solving in Chemistry. In *Chemical Education: Towards Research-based Practice*; Gilbert, J. K., de Jong, O., Justi, R., Treagust, D. F., van Driel, J. H., Eds.; Science & Technology Education Library; Springer: Dordrecht, 2002; pp 235−266.
(155) Gabel, D.; Bunce, D. M. Research on Problem Solving: Chemistry. In *Handbook of Research on Science Teaching and Learning*; Gabel, D. L., Ed.; MacMillan: New York, 1994; pp 301−326.
(156) Martinez, M. E. *Learning and Cognition: The Design of the Mind*, 1st ed.; Pearson: Upper Saddle River, NJ, 2010.
(157) Bassok, M.; Novick, L. R. Problem Solving. In *The Oxford Handbook of Thinking and Reasoning*; Holyoak, K. J., Morrison, R. G., Eds.; Oxford University Press: Oxford, U.K., 2012.
(158) Camacho, M.; Good, R. Problem Solving and Chemical Equilibrium: Successful versus Unsuccessful Performance. *J. Res. Sci. Teach.* **1989**, *26*, 251−272.
(159) Tingle, J. B.; Good, R. Effects of Cooperative Grouping on Stoichiometric Problem Solving in High School Chemistry. *J. Res. Sci. Teach.* **1990**, *27*, 671−683.
(160) Tang, H.; Kirk, J.; Pienta, N. J. Investigating the Effect of Complexity Factors in Stoichiometry Problems Using Logistic Regression and Eye Tracking. *J. Chem. Educ.* **2014**, *91*, 969−975.
(161) Knaus, K.; Murphy, K.; Blecking, A.; Holme, T. A Valid and Reliable Instrument for Cognitive Complexity Rating Assignment of Chemistry Exam Items. *J. Chem. Educ.* **2011**, *88*, 554−560.
(162) Raker, J. R.; Trate, J. M.; Holme, T. A.; Murphy, K. Adaptation of an Instrument for Measuring the Cognitive Complexity of Organic Chemistry Exam Items. *J. Chem. Educ.* **2013**, *90*, 1290−1295.
(163) Clair-Thompson, H. S.; Overton, T.; Bugler, M. Mental Capacity and Working Memory in Chemistry: Algorithmic versus Open-Ended Problem Solving. *Chem. Educ. Res. Pract.* **2012**, *13*, 484−489.
(164) The National Research Council. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*; National Academies Press: Washington, DC, 2012.
(165) Justi, R.; Gilbert, J. Models and Modeling in Chemical Education. In *Chemical education: Towards research-based practice*; Gilbert, J., De Jong, O., Justi, R., Treagust, D. F., Van Driel, J. H., Eds.; Kluwer Academic Publisher: Dortrecht, Netherlands, 2003; pp 47−68.
(166) Tomasi, J. Models and Modeling in Theoretical Chemistry. *J. Mol. Struct.: THEOCHEM* **1988**, *179*, 273−292.
(167) Nersessian, N. J. Model-Based Reasoning in Conceptual Change. In *Model-Based Reasoning in Scientific Discovery*; Magnani, L., Nersessian, N. J., Thagard, P., Eds.; Kluwer Academic Publishers: New York, 1999; pp 5−22.
(168) Gilbert, J. K.; Justi, R. Models of Modeling. In *Modelling-based Teaching in Science Education*; Gilbert, J. K., Justi, R., Eds.; Models and Modeling in Science Education; Springer: Dordrecht, 2016; pp 17−40.
(169) Grosslight, L.; Unger, C.; Jay, E.; Smith, C. L. Understanding Models and Their Use in Science: Conceptions of Middle and High School Students and Experts. *J. Res. Sci. Teach.* **1991**, *28*, 799−822.
(170) Jimenez-Aleixandre, M.; Erduran, S. Argumentation in Science Education: An Overview. In *Argumentation in science education: perspectives from classroom-based research*; Jimenez-Aleixandre, M., Erduran, S., Eds.; Springer: Dordrecht, 2008; pp 3−27.
(171) Erduran, S.; Simon, S.; Osborne, J. TAPping into Argumentation: Developments in the Application of Toulmin's Argument Pattern for Studying Science Discourse. *Sci. Educ.* **2004**, *88*, 915−933.
(172) Pabuccu, A.; Erduran, S. Beyond Rote Learning in Organic Chemistry: The Infusion and Impact of Argumentation in Tertiary Education. *Int. J. Sci. Educ.* **2017**, *39*, 1154−1172.
(173) Osborne, J. F.; Patterson, A. Scientific Argument and Explanation: A Necessary Distinction? *Sci. Educ.* **2011**, *95*, 627−638.
(174) Talanquer, V. Exploring Dominant Types of Explanations Built by General Chemistry Students. *Int. J. Sci. Educ.* **2010**, *32*, 2393−2412.
(175) Talanquer, V. Explanations and Teleology in Chemistry Education. *Int. J. Sci. Educ.* **2007**, *29*, 853−870.
(176) Cooper, M. M.; Kouyoumdjian, H.; Underwood, S. M. Investigating Students' Reasoning about Acid−Base Reactions. *J. Chem. Educ.* **2016**, *93*, 1703−1712.
(177) Becker, N.; Noyes, K.; Cooper, M. Characterizing Students' Mechanistic Reasoning about London Dispersion Forces. *J. Chem. Educ.* **2016**, *93*, 1713−1724.
(178) Domin, D. S. A Review of Laboratory Instruction Styles. *J. Chem. Educ.* **1999**, *76*, 543−547.
(179) Buck, L. B.; Bretz, S. L.; Towns, M. H. Characterizing the Level of Inquiry in the Undergraduate Laboratory. *J. Coll. Sci. Teach.* **2008**, *38*, 52−58.
(180) Bloom, B. S. *Taxonomy of Educational Objectives, Handbook 1: Cognitive Domain*, 1st ed.; Longman Group: London, 1956.
(181) Cook, E.; Kennedy, E.; McGuire, S. Y. Effect of Teaching Metacognitive Learning Strategies on Performance in General Chemistry Courses. *J. Chem. Educ.* **2013**, *90*, 961−967.
(182) 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.
(183) Zoller, U.; Nahum, T. L. From Teaching to KNOW to Learning to THINK in Science Education. In *Second International Handbook of Science Education*; Springer International Handbooks of Education; Springer: Dordrecht, 2012; pp 209−229.
(184) Zoller, U. The Fostering of Question-Asking Capability: A Meaningful Aspect of Problem-Solving in Chemistry. *J. Chem. Educ.* **1987**, *64*, 510−512.
(185) Zoller, U. Are Lecture and Learning Compatible? Maybe for LOCS: Unlikely for HOCS. *J. Chem. Educ.* **1993**, *70*, 195−197.
(186) Zoller, U. Algorithmic, LOCS and HOCS (Chemistry) Exam Questions: Performance and Attitudes of College Students. *Int. J. Sci. Educ.* **2002**, *24*, 185−203.
(187) Cooper, M. M.; Caballero, M. D.; Ebert-May, D.; Fata-Hartley, C. L.; Jardeleza, S. E.; Krajcik, J. S.; Laverty, J. T.; Matz, R. L.; Posey, L. A.; Underwood, S. M. Challenge Faculty to Transform STEM Learning. *Science* **2015**, *350*, 281−282.
(188) Sevian, H.; Talanquer, V. Rethinking Chemistry: A Learning Progression on Chemical Thinking. *Chem. Educ. Res. Pract.* **2014**, *15*, 10−23.
(189) The National Research Council. *Knowing What Students Know: The Science and Design of Educational Assessment*; National Academies Press: Washington, DC, 2001.
(190) Holme, T. Assessment and Quality Control in Chemistry Education. *J. Chem. Educ.* **2003**, *80*, 594−596.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(191) ACS Examinations Institute. A Short History, ACS Exams; https://uwm.edu/acs-exams/about-us/history/ (accessed Aug 17, 2017).
(192) Brandriet, A.; Reed, J. J.; Holme, T. A Historical Investigation into Item Formats of ACS Exams and Their Relationships to Science Practices. *J. Chem. Educ.* **2015**, *92*, 1798−1806.
(193) Raker, J. R.; Holme, T. A. A Historical Analysis of the Curriculum of Organic Chemistry Using ACS Exams as Artifacts. *J. Chem. Educ.* **2013**, *90*, 1437−1442.
(194) Kendhammer, L.; Holme, T.; Murphy, K. Identifying Differential Performance in General Chemistry: Differential Item Functioning Analysis of ACS General Chemistry Trial Tests. *J. Chem. Educ.* **2013**, *90*, 846−853.
(195) Embretson, S. E.; Reise, S. P. *Item Response Theory for Psychologists*, 1st ed.; Psychology Press: Mahwah, NJ, 2000.
(196) Nurrenbern, S. C.; Pickering, M. Concept Learning versus Problem Solving: Is There a Difference? *J. Chem. Educ.* **1987**, *64*, 508−510.
(197) Holme, T. A.; Luxford, C. J.; Brandriet, A. Defining Conceptual Understanding in General Chemistry. *J. Chem. Educ.* **2015**, *92*, 1477−1483.
(198) Lobato, J. Alternative Perspectives on the Transfer of Learning: History, Issues, and Challenges for Future Research. *J. Learn. Sci.* **2006**, *15*, 431−449.
(199) ACS Division of Chemical Education Examinations Institute. *Diagnostic of Undergraduate Chemistry Knowledge Exam*; 2013.
(200) Treagust, D. F. Diagnostic Assessment of Students' Science Knowledge. *Learn. Sci. Sch. Res. Reforming Pract.* **1995**, *1*, 327−436.
(201) Hestenes, D.; Wells, M.; Swackhamer, G. Force Concept Inventory. *Phys. Teach.* **1992**, *30*, 141−158.
(202) Brandriet, A. R.; Bretz, S. L. The Development of the Redox Concept Inventory as a Measure of Students' Symbolic and Particulate Redox Understandings and Confidence. *J. Chem. Educ.* **2014**, *91*, 1132−1144.
(203) Luxford, C. J.; Bretz, S. L. Development of the Bonding Representations Inventory To Identify Student Misconceptions about Covalent and Ionic Bonding Representations. *J. Chem. Educ.* **2014**, *91*, 312−320.
(204) Bretz, S. L.; Linenberger, K. J. Development of the Enzyme−substrate Interactions Concept Inventory. *Biochem. Mol. Biol. Educ.* **2012**, *40*, 229−233.
(205) Dick-Perez, M.; Luxford, C. J.; Windus, T. L.; Holme, T. A Quantum Chemistry Concept Inventory for Physical Chemistry Classes. *J. Chem. Educ.* **2016**, *93*, 605−612.
(206) Wren, D.; Barbera, J. Psychometric Analysis of the Thermochemistry Concept Inventory. *Chem. Educ. Res. Pract.* **2014**, *15*, 380−390.
(207) Wren, D.; Barbera, J. Gathering Evidence for Validity during the Design, Development, and Qualitative Evaluation of Thermochemistry Concept Inventory Items. *J. Chem. Educ.* **2013**, *90*, 1590−1601.
(208) Yezierski, E. J.; Birk, J. P. Misconceptions about the Particulate Nature of Matter. Using Animations To Close the Gender Gap. *J. Chem. Educ.* **2006**, *83*, 954−960.
(209) Novak, J. D. *A Theory of Education*; Cornell University Press: Ithaca, NY, 1977.
(210) Ausubel, D. P.; Novak, J. D.; Hanesian, H. *Educational Psychology: A Cognitive View*; Holt, Rinehart and Winston: New York, 1978.
(211) Magnusson, S.; Krajcik, J.; Borko, H. Nature, Sources, and Development of Pedagogical Content Knowledge for Science Teaching. In *Examining Pedagogical Content Knowledge*; Gess-Newsome, J., Lederman, N. G., Eds.; Science & Technology Education Library; Springer: Dordrecht, 1999; pp 95−132.
(212) Linn, R. L. *Educational Measurement*, 3rd ed.; Macmillan: New York, 1989.
(213) Hubbard, J. K.; Potts, M. A.; Couch, B. A. How Question Types Reveal Student Thinking: An Experimental Comparison of Multiple-True-False and Free-Response Formats. *CBE-Life Sci. Educ.* **2017**, *16*, ar26.
(214) Hammer, D.; Elby, A. On the Form of a Personal Epistemology. In *Personal Epistemology: The Psychology of Beliefs about Knowledge and Knowing*; Hofer, B. K., Pintrich, P. R., Eds.; Lawrence Erlbaum Associates: Mahwah, NJ, 2002; pp 169−190.
(215) Hammer, D. Misconceptions or P-Prims: How May Alternative Perspectives of Cognitive Structure Influence Instructional Perceptions and Intentions? *J. Learn. Sci.* **1996**, *5*, 97−127.
(216) Talanquer, V. Concept Inventories: Predicting the Wrong Answer May Boost Performance. *J. Chem. Educ.* **2017**, *94*, 1805−1810.
(217) Jorion, N.; Gane, B. D.; James, K.; Schroeder, L.; DiBello, L. V.; Pellegrino, J. W. An Analytic Framework for Evaluating the Validity of Concept Inventory Claims. *J. Eng. Educ.* **2015**, *104*, 454−496.
(218) Hake, R. R. Interactive-Engagement versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses. *Am. J. Phys.* **1998**, *66*, 64−74.
(219) Shepard, L. A. The Role of Assessment in a Learning Culture. *Educ. Res.* **2000**, *29*, 4−14.
(220) The National Research Council. *Developing Assessments for the Next Generation Science Standards*; National Academies Press: Washington, DC, 2013.
(221) Underwood, S. M.; Posey, L. A.; Herrington, D. G.; Carmel, J. H.; Cooper, M. M. Adapting Assessment Tasks To Support Three-Dimensional Learning. *J. Chem. Educ.* **2018**, *95*, 207−217.
(222) Harris, C. J.; Krajcik, J. S.; Pellegrino, J. W.; McElhaney, K. W. *Constructing Assessment Tasks That Blend Disciplinary Core Ideas, Crosscutting Concepts, and Science Practices for Classroom Formative Applications*; SRI International: Menlo Park, CA, 2016.
(223) Laverty, J. T.; Underwood, S. M.; Matz, R. L.; Posey, L. A.; Carmel, J. H.; Caballero, M. D.; Fata-Hartley, C. L.; Ebert-May, D.; Jardeleza, S. E.; Cooper, M. M. Characterizing College Science Assessments: The Three-Dimensional Learning Assessment Protocol. *PLoS One* **2016**, *11*, e0162333.
(224) Stowe, R.; Cooper, M. M. Practice What We Preach: Assessing "Critical Thinking" in Organic Chemistry. *J. Chem. Educ.* **2017**, *94*, 1852−1859.
(225) Jin, H.; Anderson, C. W. Developing Assessments For A Learning Progression on Carbon-Transforming Processes in Socio-Ecological Systems. In *Learning Progressions in Science*; SensePublishers: Rotterdam, 2012; pp 151−181.
(226) Lee, H.-S.; Liu, O. L.; Linn, M. C. Validating Measurement of Knowledge Integration in Science Using Multiple-Choice and Explanation Items. *Appl. Meas. Educ.* **2011**, *24*, 115−136.
(227) Claesgens, J.; Scalise, K.; Wilson, M.; Stacy, A. Mapping Student Understanding in Chemistry: The Perspectives of Chemists. *Sci. Educ.* **2009**, *93*, 56−85.
(228) Becker, N. M.; Rupp, C. A.; Brandriet, A. Engaging Students in Analyzing and Interpreting Data to Construct Mathematical Models: An Analysis of Students' Reasoning in a Method of Initial Rates Task. *Chem. Educ. Res. Pract.* **2017**, *18*, 798−810.
(229) Williams, L. C.; Underwood, S. M.; Klymkowsky, M. W.; Cooper, M. M. Are Noncovalent Interactions an Achilles Heel in Chemistry Education? A Comparison of Instructional Approaches. *J. Chem. Educ.* **2015**, *92*, 1979−1987.
(230) Stanger-Hall, K. F. Multiple-Choice Exams: An Obstacle for Higher-Level Thinking in Introductory Science Classes. *CBE-Life Sci. Educ.* **2012**, *11*, 294−306.
(231) Zeidner, M. Essay versus Multiple-Choice Type Classroom Exams: The Student's Perspective. *J. Educ. Res.* **1987**, *80*, 352−358.
(232) Ioannidis, J. P. A. Why Most Published Research Findings Are False. *PLOS Med.* **2005**, *2*, e124.
(233) Grove, N.; Bretz, S. L. CHEMX: An Instrument To Assess Students' Cognitive Expectations for Learning Chemistry. *J. Chem. Educ.* **2007**, *84*, 1524−1529.
(234) Bauer, C. F. Attitude toward Chemistry: A Semantic Differential Instrument for Assessing Curriculum Impacts. *J. Chem. Educ.* **2008**, *85*, 1440−1445.
(235) Xu, X.; Lewis, J. E. Refinement of a Chemistry Attitude Measure for College Students. *J. Chem. Educ.* **2011**, *88*, 561−568.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(236) Hazari, Z.; Sadler, P. M.; Sonnert, G. The Science Identity of College Students: Exploring the Intersection of Gender, Race, and Ethnicity. *J. Coll. Sci. Teach.* **2013**, *42*, 82−91.
(237) 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.
(238) Cooper, M. M.; Sandi-Urena, S. Design and Validation of an Instrument To Assess Metacognitive Skillfulness in Chemistry Problem Solving. *J. Chem. Educ.* **2009**, *86*, 240−245.
(239) Linnenbrink-Garcia, L.; Patall, E. A.; Pekrun, R. Adaptive Motivation and Emotion in Education: Research and Principles for Instructional Design. *Policy Insights Behav. Brain Sci.* **2016**, *3*, 228−236.
(240) Bodner, G.; Klobuchar, M.; Geelan, D. The Many Forms of Constructivism. *J. Chem. Educ.* **2001**, *78*, 1107−1133.
(241) Bowen, C. W. A Quantitative Literature Review of Cooperative Learning Effects on High School and College Chemistry Achievement. *J. Chem. Educ.* **2000**, *77*, 116−119.
(242) Freeman, S.; Eddy, S. L.; McDonough, M.; Smith, M. K.; Okoroafor, N.; Jordt, H.; Wenderoth, M. P. Active Learning Increases Student Performance in Science, Engineering, and Mathematics. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 8410−8415.
(243) Springer, L.; Stanne, M. E.; Donovan, S. S. Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis. *Rev. Educ. Res.* **1999**, *69*, 21−51.
(244) Gibbons, R. E.; Laga, E. E.; Leon, J.; Villafañe, S. M.; Stains, M.; Murphy, K.; Raker, J. R. Chasm Crossed? Clicker Use in Postsecondary Chemistry Education. *J. Chem. Educ.* **2017**, *94*, 549−557.
(245) Apple, T.; Cutler, A. The Rensselaer Studio General Chemistry Course. *J. Chem. Educ.* **1999**, *76*, 462−463.
(246) Schwartz, D. L.; Bransford, J. D. A Time For Telling. *Cogn. Instr.* **1998**, *16*, 475−5223.
(247) Hockings, S. C.; DeAngelis, K. J.; Frey, R. F. Peer-Led Team Learning in General Chemistry: Implementation and Evaluation. *J. Chem. Educ.* **2008**, *85*, 990−996.
(248) Lewis, S. E. Retention and Reform: An Evaluation of Peer-Led Team Learning. *J. Chem. Educ.* **2011**, *88*, 703−707.
(249) Shields, S. P.; Hogrebe, M. C.; Spees, W. M.; Handlin, L. B.; Noelken, G. P.; Riley, J. M.; Frey, R. F. A Transition Program for Underprepared Students in General Chemistry: Diagnosis, Implementation, and Evaluation. *J. Chem. Educ.* **2012**, *89*, 995−1000.
(250) Smith, M. K.; Wood, W. B.; Adams, W. K.; Wieman, C.; Knight, J. K.; Guild, N.; Su, T. T. Why Peer Discussion Improves Student Performance on In-Class Concept Questions. *Science* **2009**, *323*, 122−124.
(251) Tien, L. T.; Roth, V.; Kampmeier, J. a. Implementation of a Peer-Led Team Learning Instructional Approach in an Undergraduate Organic Chemistry Course. *J. Res. Sci. Teach.* **2002**, *39*, 606−632.
(252) Varma-Nelson, P.; Coppola, B. P. Team Learning. *Chemists' Guide to Effective Teaching*; Pienta, N. J., Cooper, M. M., Greenbowe, T. J., Eds; Prentice Hall: Upper Saddle River, NJ, 2005; pp 155−169.
(253) Chan, J. Y. K.; Bauer, C. F. Effect of Peer-Led Team Learning (PLTL) on Student Achievement, Attitude, and Self-Concept in College General Chemistry in Randomized and Quasi Experimental Designs. *J. Res. Sci. Teach.* **2015**, *52*, 319−346.
(254) Walker, L.; Warfa, A.-R. M. Process Oriented Guided Inquiry Learning (POGIL®) Marginally Effects Student Achievement Measures but Substantially Increases the Odds of Passing a Course. *PLoS One* **2017**, *12*, e0186203.
(255) Slavin, R. E. Synthesis of Research of Cooperative Learning. *Educ. Leadership* **1991**, *48*, 71−82.
(256) Johnson, D. W.; Johnson, R. T. An Educational Psychology Success Story: Social Interdependence Theory and Cooperative Learning. *Educ. Res.* **2009**, *38*, 365−379.
(257) Roadrangka, V. *The Construction and Validation of the Group Assessment of Logical Thinking (GALT)*; University of Georgia: Athens, GA, 1985.
(258) Cooper, M. M.; Cox, C. T.; Nammouz, M.; Case, E.; Stevens, R. An Assessment of the Effect of Collaborative Groups on Students' Problem-Solving Strategies and Abilities. *J. Chem. Educ.* **2008**, *85*, 866−872.
(259) Cen, L.; Ruta, D.; Powell, L.; Ng, J. Does Gender Matter for Collaborative Learning? In *2014 IEEE International Conference on Teaching, Assessment and Learning for Engineering (TALE)*; 2014; pp 433−440.
(260) Stump, G. S.; Hilpert, J. C.; Husman, J.; Chung, W.; Kim, W. Collaborative Learning in Engineering Students: Gender and Achievement. *J. Eng. Educ.* **2011**, *100*, 475−497.
(261) Cole, R.; Becker, N.; Towns, M.; Sweeney, G.; Wawro, M.; Rasmussen, C. Adapting a Methodology from Mathematics Education Research to Chemistry Education Research: Documenting Collective Activity. *Int. J. Sci. Math. Educ.* **2012**, *10*, 193−211.
(262) Becker, N.; Rasmussen, C.; Sweeney, G.; Wawro, M.; Towns, M.; Cole, R. Reasoning Using Particulate Nature of Matter: An Example of a Sociochemical Norm in a University-Level Physical Chemistry Class. *Chem. Educ. Res. Pract.* **2013**, *14*, 81−94.
(263) Kulatunga, U.; Moog, R. S.; Lewis, J. E. Argumentation and Participation Patterns in General Chemistry Peer-Led Sessions. *J. Res. Sci. Teach.* **2013**, *50*, 1207−1231.
(264) Rickey, D.; Stacy, A. M. The Role of Metacognition in Learning Chemistry. *J. Chem. Educ.* **2000**, *77*, 915−920.
(265) Schraw, G. Promoting General Metacognitive Awareness. In *Metacognition in Learning and Instruction; Neuropsychology and Cognition*; Hartman, H. J., Ed; Springer: Dordrecht, 2001; pp 3−16.
(266) Cook, E.; Kennedy, E.; McGuire, S. Y. Effect of Teaching Metacognitive Learning Strategies on Performance in General Chemistry Courses. *J. Chem. Educ.* **2013**, *90*, 961−967.
(267) Sandi-Urena, S.; Cooper, M. M.; Stevens, R. H. Enhancement of Metacognition Use and Awareness by Means of a Collaborative Intervention. *Int. J. Sci. Educ.* **2011**, *33*, 323−340.
(268) Sandi-Urena, S.; Cooper, M.; Stevens, R. Effect of Cooperative Problem-Based Lab Instruction on Metacognition and Problem-Solving Skills. *J. Chem. Educ.* **2012**, *89*, 700−706.
(269) Christiansen, M. A.; Lambert, A. M.; Nadelson, L. S.; Dupree, K. M.; Kingsford, T. A. In-Class Versus At-Home Quizzes: Which Is Better? A Flipped Learning Study in a Two-Site Synchronously Broadcast Organic Chemistry Course. *J. Chem. Educ.* **2017**, *94*, 157−163.
(270) Eichler, J. F.; Peeples, J. Flipped Classroom Modules for Large Enrollment General Chemistry Courses: A Low Barrier Approach to Increase Active Learning and Improve Student Grades. *Chem. Educ. Res. Pract.* **2016**, *17*, 197−208.
(271) Fautch, J. M. The Flipped Classroom for Teaching Organic Chemistry in Small Classes: Is It Effective? *Chem. Educ. Res. Pract.* **2015**, *16*, 179−186.
(272) Flynn, A. B. Structure and Evaluation of Flipped Chemistry Courses: Organic & Spectroscopy, Large and Small, First to Third Year, English and French. *Chem. Educ. Res. Pract.* **2015**, *16*, 198−211.
(273) Hartman, J. R.; Dahm, D. J.; Nelson, E. A. ConfChem Conference on Flipped Classroom: Time-Saving Resources Aligned with Cognitive Science To Help Instructors. *J. Chem. Educ.* **2015**, *92*, 1568−1569.
(274) Hibbard, L.; Sung, S.; Wells, B. Examining the Effectiveness of a Semi-Self-Paced Flipped Learning Format in a College General Chemistry Sequence. *J. Chem. Educ.* **2016**, *93*, 24−30.
(275) MacArthur, J. R.; Jones, L. L. A Review of Literature Reports of Clickers Applicable to College Chemistry Classrooms. *Chem. Educ. Res. Pract.* **2008**, *9*, 187−195.
(276) It is important to emphasize that, throughout this section, "models" refers to physical or virtual representations of atoms and molecules, not simplifications based on scientific principles that serve to predict or explain phenomena.
(277) Stull, A. T.; Gainer, M.; Padalkar, S.; Hegarty, M. Promoting Representational Competence with Molecular Models in Organic Chemistry. *J. Chem. Educ.* **2016**, *93*, 994−1001.
(278) Stull, A. T.; Barrett, T.; Hegarty, M. Usability of Concrete and Virtual Models in Chemistry Instruction. *Comput. Hum. Behav.* **2013**, *29*, 2546−2556.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(279) Stull, A. T.; Hegarty, M.; Dixon, B.; Stieff, M. Representational Translation with Concrete Models in Organic Chemistry. *Cogn. Instr.* **2012**, *30*, 404−434.
(280) Padalkar, S.; Hegarty, M. Models as Feedback: Developing Representational Competence in Chemistry. *J. Educ. Psychol.* **2015**, *107*, 451−467.
(281) Williamson, V. M.; Hegarty, M.; Deslongchamps, G.; Williamson, K. C.; Shultz, M. J. Identifying Student Use of Ball-and-Stick Images versus Electrostatic Potential Map Images via Eye Tracking. *J. Chem. Educ.* **2013**, *90*, 159−164.
(282) Cox, C. T. An Investigation of the Effects of Interventions on Problem Solving Strategies and Abilities. Ph.D. Dissertation, Clemson University: Clemson, SC, 2006.
(283) Jones, L. L. How Multimedia-Based Learning and Molecular Visualization Change the Landscape of Chemical Education Research. *J. Chem. Educ.* **2013**, *90*, 1571−1576.
(284) Jones, L. L.; Kelly, R. M. Visualization: The Key to Understanding Chemistry Concepts. In *Sputnik to Smartphones: A Half-Century of Chemistry Education*; Orna, M. V., Ed.; Oxford University Press: Oxford, U.K., 2015; Vol. *1208*, pp 121−140.
(285) Williamson, V. M.; Abraham, M. R. The Effects of Computer Animation on the Particulate Mental Models of College Chemistry Students. *J. Res. Sci. Teach.* **1995**, *32*, 521−534.
(286) Velázquez-Marcano, A.; Williamson, V. M.; Ashkenazi, G.; Tasker, R.; Williamson, K. C. The Use of Video Demonstrations and Particulate Animation in General Chemistry. *J. Sci. Educ. Technol.* **2004**, *13*, 315−323.
(287) Tasker, R. Using Multimedia to Visualize the Molecular World: Educational Theory into Practice. *Chemists' Guide to Effective Teaching*; Pienta, N. J., Cooper, M. M., Greenbowe, T. J., Eds; Prentice Hall: Upper Saddle River, NJ, 2005; pp 195−211.
(288) Rosenthal, D. P.; Sanger, M. J. How Does Viewing One Computer Animation Affect Students' Interpretations of Another Animation Depicting the Same Oxidation−reduction Reaction? *Chem. Educ. Res. Pract.* **2013**, *14*, 286−296.
(289) Wu, H.-K.; Shah, P. Exploring Visuospatial Thinking in Chemistry Learning. *Sci. Educ.* **2004**, *88*, 465−492.
(290) VisChem. Visualising the Chemical world; http://vischem.com.au/ (accessed Aug 18, 2017).
(291) Kelly, R. M.; Jones, L. L. Investigating Students' Ability To Transfer Ideas Learned from Molecular Animations of the Dissolution Process. *J. Chem. Educ.* **2008**, *85*, 303−309.
(292) Kelly, R. M.; Jones, L. L. Exploring How Different Features of Animations of Sodium Chloride Dissolution Affect Students' Explanations. *J. Sci. Educ. Technol.* **2007**, *16*, 413−429.
(293) Tasker, R. Molecular-Level Visualisation for Educational Purposes. In *Chemical Modelling*; Springborg, M., Joswig, J.-O., Eds.; The Royal Society of Chemistry: Cambridge, 2017; Vol. *17*, pp 261−282.
(294) Stieff, M. Drawing for Promoting Learning and Engagement with Dynamic Visualizations. In *Learning from Dynamic Visualization*; Lowe, R., Ploetzner, R., Eds.; Springer: Cham, 2017; pp 333−356.
(295) Bodner, G. M.; Guay, R. B. The Purdue Visualization of Rotations Test. *Chem. Educ.* **1997**, *2*, 1−17.
(296) Pribyl, J. R.; Bodner, G. M. Spatial Ability and Its Role in Organic Chemistry: A Study of Four Organic Courses. *J. Res. Sci. Teach.* **1987**, *24*, 229−240.
(297) Lopez, E. J.; Shavelson, R. J.; Nandagopal, K.; Szu, E.; Penn, J. Factors Contributing to Problem-Solving Performance in First-Semester Organic Chemistry. *J. Chem. Educ.* **2014**, *91*, 976−981.
(298) Stieff, M. Sex Differences in the Mental Rotation of Chemistry Representations. *J. Chem. Educ.* **2013**, *90*, 165−170.
(299) Ainsworth, S.; Prain, V.; Tytler, R. Drawing to Learn in Science. *Science* **2011**, *333*, 1096−1097.
(300) Tien, L. T.; Teichert, M. A.; Rickey, D. Effectiveness of a MORE Laboratory Module in Prompting Students To Revise Their Molecular-Level Ideas about Solutions. *J. Chem. Educ.* **2007**, *84*, 175−181.
(301) Cooper, M. M.; Grove, N.; Underwood, S. M.; Klymkowsky, M. W. Lost in Lewis Structures: An Investigation of Student Difficulties in Developing Representational Competence. *J. Chem. Educ.* **2010**, *87*, 869−874.
(302) Flynn, A. B.; Featherstone, R. B. Language of Mechanisms: Exam Analysis Reveals Students' Strengths, Strategies, and Errors When Using the Electron-Pushing Formalism (Curved Arrows) in New Reactions. *Chem. Educ. Res. Pract.* **2017**, *18*, 64−77.
(303) Grove, N. P.; Cooper, M. M.; Rush, K. M. Decorating with Arrows: Toward the Development of Representational Competence in Organic Chemistry. *J. Chem. Educ.* **2012**, *89*, 844−849.
(304) Cooper, M. M.; Stieff, M.; DeSutter, D. Sketching the Invisible to Predict the Visible: From Drawing to Modeling in Chemistry. *Top. Cogn. Sci.* **2017**, *9*, 902−920.
(305) Chi, M. T. H. Active-Constructive-Interactive: A Conceptual Framework for Differentiating Learning Activities. *Top. Cogn. Sci.* **2009**, *1*, 73−105.
(306) Teichert, M. A.; Tien, L. T.; Anthony, S.; Rickey, D. Effects of Context on Students' Molecular-Level Ideas. *Int. J. Sci. Educ.* **2008**, *30*, 1095−1114.
(307) Underwood, S. M.; Reyes-Gastelum, D.; Cooper, M. M. When Do Students Recognize Relationships between Molecular Structure and Properties? A Longitudinal Comparison of the Impact of Traditional and Transformed Curricula. *Chem. Educ. Res. Pract.* **2016**, *17*, 365−380.
(308) Bhattacharyya, G.; Bodner, G. M. It Gets Me to the Product": How Students Propose Organic Mechanisms. *J. Chem. Educ.* **2005**, *82*, 1402−1407.
(309) Grove, N. P.; Cooper, M. M.; Cox, E. L. Does Mechanistic Thinking Improve Student Success in Organic Chemistry? *J. Chem. Educ.* **2012**, *89*, 850−853.
(310) Bhattacharyya, G. Trials and Tribulations: Student Approaches and Difficulties with Proposing Mechanisms Using the Electron-Pushing Formalism. *Chem. Educ. Res. Pract.* **2014**, *15*, 594−609.
(311) Bhattacharyya, G. From Source to Sink: Mechanistic Reasoning Using the Electron-Pushing Formalism. *J. Chem. Educ.* **2013**, *90*, 1282−1289.
(312) Lack, K. A. *Current Status of Research on Online Learning in Postsecondary Education*; Ithaka S+R: 2013; p 72.
(313) *Online Learning: Does It Help Low-Income and Underprepared Students?*; Assessment of Evidence Series; http://ccrc.tc.columbia.edu/publications/online-learning-low-income-underprepared.html (accessed Sep 14, 2017).
(314) *MOOCs: Expectations and Reality*; https://library.educause.edu/resources/2014/5/moocs-expectations-and-reality (accessed Sep 14, 2017).
(315) Xu, D.; Jaggars, S. S. The Effectiveness of Distance Education across Virginia's Community Colleges: Evidence from Introductory College-Level Math and English Courses. *Educ. Eval. Policy Anal.* **2011**, *33*, 360−377.
(316) Mintz, B. S. *Lifting All Boats: How MOOCs Can Bring Higher Ed Together*; https://campustechnology.com/articles/2013/07/24/lifting-all-boats-how-moocs-can-bring-higher-ed-together.aspx; 2013 (accessed Jan 2, 2018).
(317) Smith, J.; Wilson, S. B.; Banks, J.; Zhu, L.; Varma-Nelson, P. Replicating Peer-Led Team Learning in Cyberspace: Research, Opportunities, and Challenges. *J. Res. Sci. Teach.* **2014**, *51*, 714−740.
(318) Leontyev, A.; Baranov, D. Massive Open Online Courses in Chemistry: A Comparative Overview of Platforms and Features. *J. Chem. Educ.* **2013**, *90*, 1533−1539.
(319) *PhET Chemistry Simulations*; https://phet.colorado.edu/en/simulations/category/chemistry.
(320) The National Research Council. *Learning Science through Computer Games and Simulations*; Honey, M. A., Hilton, M. L., Eds.; National Academies Press: Washington, DC, 2011.
(321) Wiggins, G. P.; McTighe, J. *Understanding by Design*, 2nd ed.; Association for Supervision and Curriculum Development: Alexandria, VA, 2005.
(322) Brown, A. L. Design Experiments: Theoretical and Methodological Challenges in Creating Complex Interventions in Classroom Settings. *J. Learn. Sci.* **1992**, *2*, 141−178.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087

(323) Barab, S.; Squire, K. Design-Based Research: Putting a Stake in the Ground. *J. Learn. Sci.* **2004**, *13*, 1−14.
(324) Corcoran, T.; Mosher, F. A.; Rogat, A. *Learning Progressions in Science: An Evidence Based Approach to Reform*; RR-63; Consortium for Policy Research in Education, Teachers College, Columbia University: 2009.
(325) Duncan, R. G.; Hmelo-Silver, C. E. Learning Progressions: Aligning Curriculum, Instruction, and Assessment. *J. Res. Sci. Teach.* **2009**, *46*, 606−609.
(326) Cooper, M. M.; Underwood, S. M.; Hilley, C. Z.; Klymkowsky, M. W. Development and Assessment of a Molecular Structure and Properties Learning Progression. *J. Chem. Educ.* **2012**, *89*, 1351−1357.
(327) Talanquer, V.; Pollard, J. Reforming a Large Foundational Course: Successes and Challenges. *J. Chem. Educ.* **2017**, *94*, 1844−1851.
(328) Osborne, J.; Collins, S. Pupil's and Parents' Views of the School Science Curriculum. *Sch. Sci. Rev.* **2000**, *82*, 23−31.
(329) Van Aalsvoort, J. Logical Positivism as a Tool to Analyse the Problem of Chemistry's Lack of Relevance in Secondary School Chemical Education. *Int. J. Sci. Educ.* **2004**, *26*, 1151−1168.
(330) Marks, R.; Eilks, I. Promoting Scientific Literacy Using a Sociocritical and Problem-Oriented Approach to Chemistry Teaching: Concept, Examples, Experiences. *Int. J. Environ. Sci. Educ.* **2009**, *4*, 231−245.
(331) Stuckey, M.; Hofstein, A.; Mamlok-Naaman, R.; Eilks, I. The Meaning of 'Relevance'in Science Education and Its Implications for the Science Curriculum. *Stud. Sci. Educ.* **2013**, *49*, 1−34.
(332) *Relevant Chemistry Education: From Theory to Practice*; Eilks, I., Hofstein, A., Eds.; Sense Publishers: Rotterdam, 2015.
(333) Gilbert, J. K.; Bulte, A. M. W.; Pilot, A. Concept Development and Transfer in Context-Based Science Education. *Int. J. Sci. Educ.* **2011**, *33*, 817−837.
(334) Fahlman, B. D.; Purvis-Roberts, K. L.; Kirk, J. S.; Bentley, A. K.; Daubenmire, P. L.; Ellis, J. P.; Mury, M. T. *Chemistry in Context: Applying Chemistry to Society*, 9th ed.; McGraw-Hill Education: New York, 2017.
(335) 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. *Meas. Interdiscip. Res. Perspect.* **2006**, *4*, 1−98.
(336) Krajcik, J. S. The Importance, Cautions and Future of Learning Progression Research. In *Learning Progressions in Science*; Alonzo, A. C., Gotwals, A. W., Eds.; Sense Publishers: Rotterdam, 2012; pp 27−36.
(337) Roblin, N. P.; Schunn, C.; McKenney, S. What Are Critical Features of Science Curriculum Materials That Impact Student and Teacher Outcomes? *Sci. Educ.* **2018**, *102*, 260−282.
(338) Baker, L. A.; Chakraverty, D.; Columbus, L.; Feig, A. L.; Jenks, W. S.; Pilarz, M.; Stains, M.; Waterman, R.; Wesemann, J. L. Cottrell Scholars Collaborative New Faculty Workshop: Professional Development for New Chemistry Faculty and Initial Assessment of Its Efficacy. *J. Chem. Educ.* **2014**, *91*, 1874−1881.
(339) Dancy, M.; Henderson, C. Pedagogical Practices and Instructional Change of Physics Faculty. *Am. J. Phys.* **2010**, *78*, 1056−1063.
(340) Ebert-May, D.; Derting, T. L.; Hodder, J.; Momsen, J. L.; Long, T. M.; Jardeleza, S. E. What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *BioScience* **2011**, *61*, 550−558.
(341) Cooper, M. M. Data-Driven Education Research. *Science* **2007**, *317*, 1171.

DOI: 10.1021/acs.chemrev.8b00020
Chem. Rev. 2018, 118, 6053−6087